International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 05 Special Issue 01 | 2020

Neural Network Classifier Based on Wavelet Based Feature of EEG Signals for Diagnosis of Epilepsy

PRAVIN A. KHARAT[1], ASHWINI D. BHOPLE[2] MANJIRI KARANDE[3]
[1]Associate Professor, Department of Computer Science & Engineering.
Padmashri Dr. V. B. Kolte College of Engineering Malkapur ((Maharashtra)
[2]Assistant Professor, Department of Computer Science & Engineering.
Sharadchandra Pawar College of Engineering, Otur (Maharashtra)
[3]Assistant Professor, Department of Computer Science & Engineering
Padmashri Dr. V. B. Kolte College of Engineering Malkapur ((Maharashtra)
[1]pravinakharat82@gmail.com, [2]ashwinibhople@gmail.com,[3]manjiri.karande@gmail.com

Abstract: - Epilepsy is one of the major fields of application of EEG. Now a days, identification of epilepsy is accomplished manually by skilled neurologists. Who are very small in number. In this work, we propose a methodology for automatic detection of normal, interictal and ictal conditions from recorded of EEG signals. We used the wavelet transform for the feature extraction and obtained statistical parameters from the decomposed wavelet coefficients. The Generalized Feed Forward Neural Network (GFFNN), Multilayer Perceptron (MLP), Elman Neural Network (ENN) and Support vector Machine (SVM) are used for the classification. The performance of the proposed system was evaluated in terms of classification accuracy, Sensitivity, Specificity and overall accuracy. For the generalization and true learning the results are also obtained on reverse tagging. The sensitivity analysis is used to remove the insignificant input channel and reduce the complexity of the network. Robustness of classifier to noise is also verified by introducing controlled Gaussian and Uniform noise in input.

Key-words: - Artificial Neural Network, Classification Accuracy,Daubechies,EEG, Waveletand Entropy

1 Introduction

Epilepsy is one of the most common neurological disorders second only to stroke; it affects about 0.8% of the world's total human population. The most important treatment is pharmacological; however in 25% of patients, seizures are drug resistant. Epilepsy is manifested by a sudden and recurrent brain malfunction which has its origin in excessive and hyper synchronous activity of neurons. The seizures occur at random and impair the normal function of the brain. During the seizure, the electroencephalogram (EEG) changes dramatically, its amplitude increases by an order of magnitude and characteristic patterns varying in time and frequency appear. Electroencephalography is the most useful and cost effective modality for the diagnosis of epilepsy. The detection of these abnormalities by the visual inspection of EEG signals is complex and time consuming process and it requires highly skilled doctors. In most of the cases, epilepsy is controlled by the proper medical treatment. For that purpose, the proper and earlier diagnosis of epilepsy is required. In some cases, surgical treatment for removal of the epileptic part is also available. Recently, a new method is introduced, in which, part of brain iselectrically stimulated to avoid the arrival of seizure. Automatic detection of seizure is very important part of such a treatment.

Several automated diagnostic systems for epilepsy diagnosis have been suggested in the literature [1-7]. In 2001 M. Akin et al. [8] reported artificial neural network for the diagnosis of epilepsy and they designed multi layer feed forward NN for it. For the input to the neural network, the sub frequencies like α, ϴ, β and ɗ were extracted from the EEG signal by using wavelet transform. In 2007 V. Shrinivasan *et al.* [9] proposed automated epileptic EEG detection system, in which two different neural networks, namely, Elman network and probabilistic neural network are used. Time-domain feature of EEG signal called approximate entropy reflecting the nonlinear dynamics of the brain activity was used as an input to the NN. Samanwoy Ghosh *et al.* [10] proposed two stage classifier for the accurate and robust EEG classification based on the nine-parameter mixed-band features, also the principal component analysis is used for the feature enhancement technique. For the classification, Sheng-Fu-Liang *et al*. [11] used two linear methods: linear least squares and linear discriminate analysis and two nonlinear models: back propagation neural network and support vector machine with radial basis function kernel. For the EEG feature extraction; approximate entropy, EEG power spectrum and the principal component analysis techniques are used. Pravin Kumar S. *et al*. [12] used three entropies, namely, wavelet entropy, spectral entropy and sample entropy to exploit the important diagnostic information from EEG signal. Two neural network, namely, recurrent Elman network and radial basis network are used in his proposed model. A similar entropy based automated system is proposed by Rajendra Acharya *et al*. [13], who used approximate entropy, sample entropy, phase entropy (S1) and phase entropy (S2) for the feature extraction. He used seven different classifiers: Fuzzy Sugeno Classifier, Support Vector Machine, K-Nearest Neighbor, Probabilistic Neural Network, Decision Tree, and Gaussian mixture model and Naive Bayes Classifier. Recently, Abdulhamit Subasi [14] proposed the EEG signal classification using wavelet feature extraction technique. He used Daubechies wavelet of order 4 (Db4) for decomposing the EEG signal in to D1-D5 and one final approximation A5. For further dimensionality reduction he extracted the four statistical parameters, namely, Mean, Average Power, Standard Deviation and Ratio of the absolute mean values of adjacent sub-band from the approximate and detail coefficients. Patnaik L. M. *et al*. [15] proposed similar wavelet based system for the epileptic EEG detection. In his proposed system, he used wavelet transform for feature extraction and obtained the statistical parameter from the decomposed wavelet coefficients. A feed-forward back propagation artificial neural network was used for classification purpose. We recently proposed the statistical parameter and principal component analysis based technique for the diagnosis of epilepsy [17-19].

This paper explores the similar type of method for thediagnosis of epilepsy. The epilepsy diagnosis problem is modeled as three group classification problem. The three groups are: 1) Healthy subject (Normal EEG) 2) Epileptic subject during seizure free interval (Interictal EEG) and 3) Epileptic subject during seizure activity (Ictal EEG).We have also performed the sensitivity analysis for reducing the complexity of neural networks.

2 Discrete Wavelet Transform

The transform of a signal is just another way of representing the signal. It does not change the information content present in the signal. The Wavelet Transform provides a time-frequency representation of the signal.The Continuous Wavelet Transform (CWT) is provided by equation 1, where *x(t)* is the signal to be analyzed. *ψ(t)* is the mother wavelet or the basis function. All the wavelet functions used in the transformation are derived from the mother wavelet through translation (shifting) and scaling (dilation or compression).

$$X_{WT}(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t).\psi\left(\frac{t-s}{s}\right) dt \qquad (1)$$

The mother wavelet used to generate all the basis functions is designed based on some desired characteristics associated with that function. The translation parameter τ relates to the location of the wavelet function as it is shifted through the signal.Thus, it corresponds to the time information in the Wavelet Transform. The scale parameter s is defined as |1/frequency| and corresponds to frequency information. Scaling either dilates (expands) or compresses a signal. Large scales (low frequencies) dilate the signal and provide detailed information

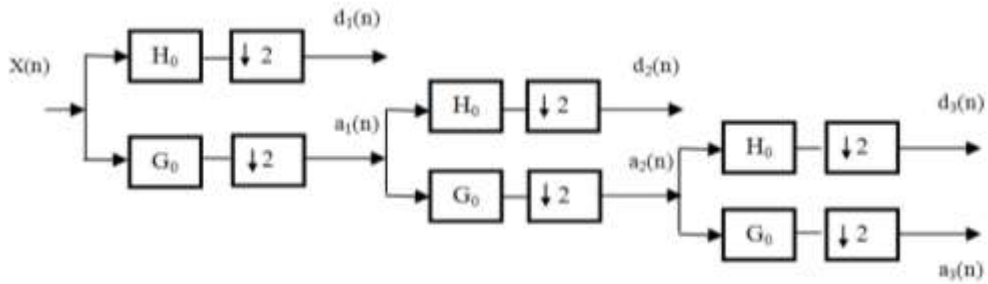

Fig.1: Approximation and detail decomposition of three level DWT

hidden in the signal, while small scales (high frequencies) compress the signal and provide global information about the signal.

The Wavelet Series is just a sampled version of CWTand its computation may consume significant amount of time and resources, depending on the resolution required. The Discrete Wavelet Transform (DWT), which is based on sub-band coding is found to yield a fast computation of Wavelet Transform. It is easy to implement and reduce the computation time and resources required.

The DWT is computed by successive lowpass and highpass filtering of the discrete time-domain signal as shown in figure 1. This is called the Mallat algorithm or Mallat-tree decomposition. Its significance is in the manner it connects the continuous-time multi resolution to discrete-time filters. In the figure, the signal is denoted by the sequence $x[n]$, where n is an integer. The low pass filter is denoted by G_0 while the high pass filter is denoted by H_0. At each level, the high pass filter produces detail information$d[n]$, while the low pass filter associated with scaling function produces coarse approximations $a[n]$.

At each decomposition level, the half band filters produce signals spanning only half the frequency band. This doubles the frequency resolution as the uncertainty in frequency is reduced by half. In accordance with Nyquist's rule, if the original signal has a highest frequency of ω, which requires a sampling frequency of 2ω radians, then it now has the highest frequency of $\omega/2$ radians. It can now be sampled at a frequency of ω radians thus discarding half the samples with no loss of information. This decimation by 2 halves the time resolution as the entire signal is now represented by only half the number of samples. Thus, while the half band low pass filtering removes half of the frequencies and thus halves the resolution, the decimation by 2 doubles the scale. Selection of suitable wavelet and number of levels of decomposition is very important in the analysis of signals using DWT. The wavelet can be chosen depending on how smooth the signal is and also on the basis of the amount of computation involved. The number of levels of decomposition is chosen based on the dominant frequency components of the signals. The levels are chosen such that those part of the signal that correlate well with the frequencies required for classification of the signal are retained in the wavelet coefficients.

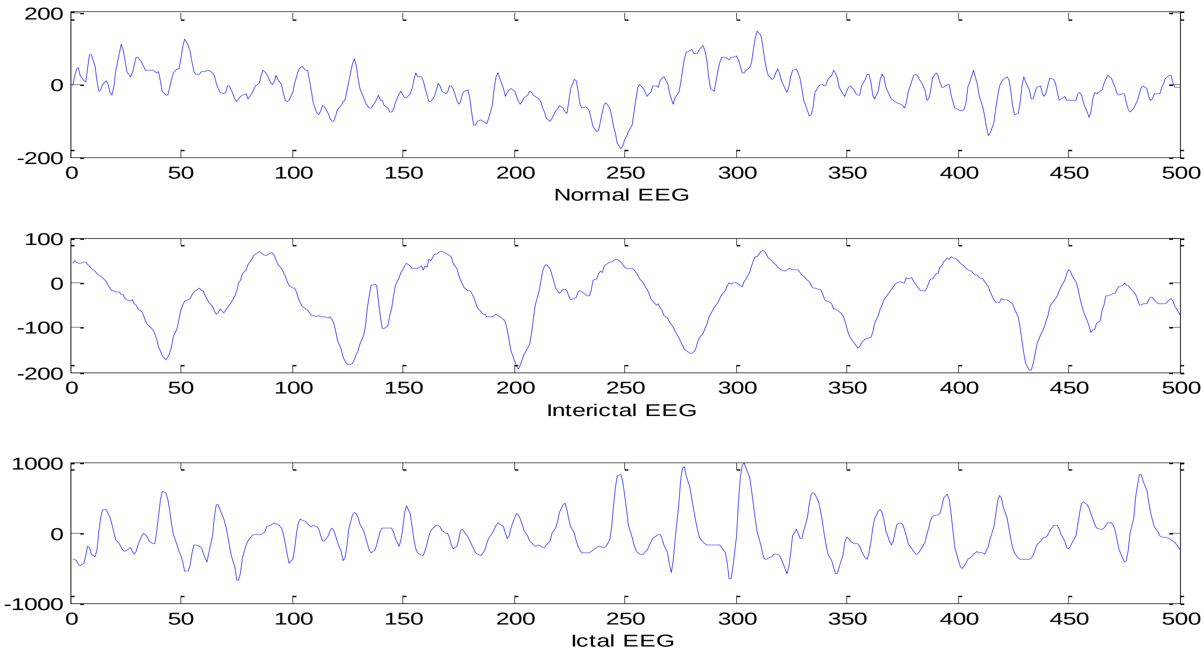

Fig.2: Sample EEG signals from set A, D and E(top to bottom)

3 Dataset

The EEG data considered for this work is extracted from University of Bonn EEG database which is available in public domain [16]. The complete database is comprised of five sets of dataset referred to as A-E. Each dataset contains 100 single channel EEG segment without any artifacts with 23.6-sec. Set A and B contain recording obtained from surface EEG recording that were carried out on five healthy volunteers using a standardized electrode placement scheme. Set C and D contained only activity measured during seizure free interval, segments in set D were recorded within the epileptognic zone and those in the set C from the hippocampal formation of apposite hemisphere of the brain. Set E only contains the seizure activity.

All signals were recorded with 128-channel amplifier system, using an average common reference. After 12 bit analog-to-digital conversion, the data were written continuously onto the disk of a data acquisition computer system at sampling rate of 173.61 Hz. Band pass filter setting were 0.53-40 Hz.

Three sets of EEG data have been selected for the training and testing of neural network: set A for healthy subject, set D for epileptic subjects during a seizure free interval that indicates interictal activity and set E contains seizure activity which indicates ictal activity. An example of first 500 sampling point of three EEGs for normal, interictal and ictal activity are magnified and displayed in fig. 2.

4 Choice of Mother Wavelet and Number of Levels

There are a number of basis functions that can be employed as the mother wavelet for Wavelet Transformation. Since the mother wavelet produces all wavelet functions used in the transformation through translation and scaling, it determines the characteristics of the resulting wavelet transform. Therefore, the details of the particular application should be taken into account and the appropriate mother wavelet should be chosen in order to use the Wavelet Transform, effectively.

The numbers of tests are performed with different types of wavelet and one which gives maximum efficiency is selected for the decomposition. As Daubechies wavelet is known for detecting the change in frequency, the wavelet coefficients are extracted by using DB2. The number of levels of decomposition is chosen based on the dominant frequency components of the signals. The levels are chosen such that those part of the signal that correlatewell with the frequencies required for classification of the signal are retained in the wavelet coefficients.Daubechies order 2 wavelet transform was applied to the normal, interictal and ictal EEG signal. The number of levels chosen for the decomposition is determined as three. Fig.3 to fig. 5 show the three different levels of approximation coefficients (A1-A4) and details coefficients (D1-D4) of normal, interictal and ictal EEG signal, respectively.

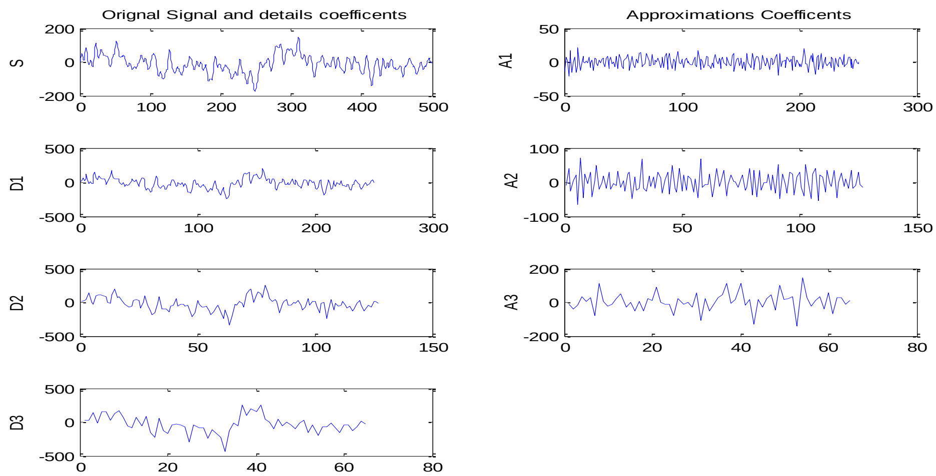

Fig. 3: Db2 level 3 decomposition of normal EEG signal

5 Feature Extractions

The extracted wavelet coefficients of EEG signal provide the time-frequency representation of the signal. Some statistical features were also extracted from the wavelet coefficients for matching length of feature vector. Following statistical features are used to represent the time-frequency distribution of the EEG signal.

1. Mean of the absolute value of both the approximation and detail coefficients.
2. Median of the approximation and detail vector.
3. Mode that is most frequently occurring value in both the sub-bands.
4. Maximum and minimum values from the approximation and detail vector.
5. Range that is the difference between maximum and minimum value in each sub bands.
6. Standard deviation of the coefficient in each sub bands.
7. Median and mean absolute deviation from both the sub bands.

In this way, for three levels decomposition, total 54 statistical parameters are extracted from three approximations and three details coefficients. In addition to these 54 statistical parameters extracted from basebands, five wavelet entropies namely Shannon, log energy, threshold, user and norm entropy are extracted from the EEG signal.The percentage energy corresponding to approximate and detail coefficients was also extracted, which is the vector of length four. Each of dataset A, D and E contains 100 EEG segments. Therefore the size of final feature vector matrix is 300 x 63. According to the statisticalparameter, a typical scatter plot for a few sample is shown in fig. 6.from this plot it is seen that the decision boundaries discriminating between different class are quite complex, nonlinear and overlapping.

6 Classification Using ANN

Artificial Neural Network evolved as a powerful tool for pattern recognition, classification, prediction and pattern completion [20-23]. ANN is an inspiration from biological neurons. The artificial neuron is the most basic computational unit of information processing in ANNs. The knowledge about the problem is distributed among them and between the connection weights which link the neurons. The training algorithm attempts to adjust the various weights (and biases) and set them to a value such that the ANN performs better at the applied input. Thus, the entiretraining process is a means of evaluating the rightcombination of weights and biases for which the ANN performs at its best.This performance and trainingdepend on the number of hidden layers, as well as on the neurons in these hidden layers.A good ANN architecture gives the best performance in the least number of layers and the least number of neurons. This performance is measured using the testing data set. At the training stage, the feature vector is applied as input to neural network and network adjusts its variable parameter, the weights and biases, to capture the relationship between input pattern and outputs.

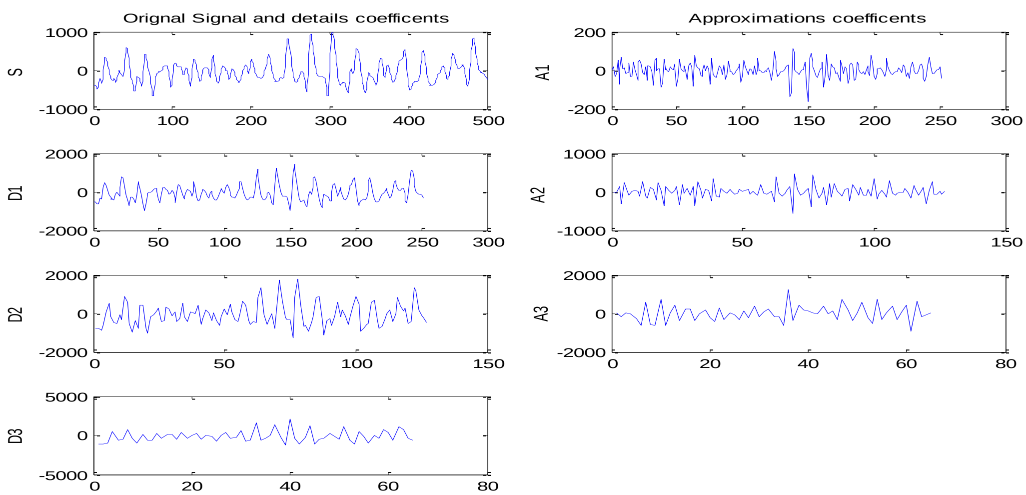

Fig. 4: Db2 level 3 decomposition of ictal EEG signal

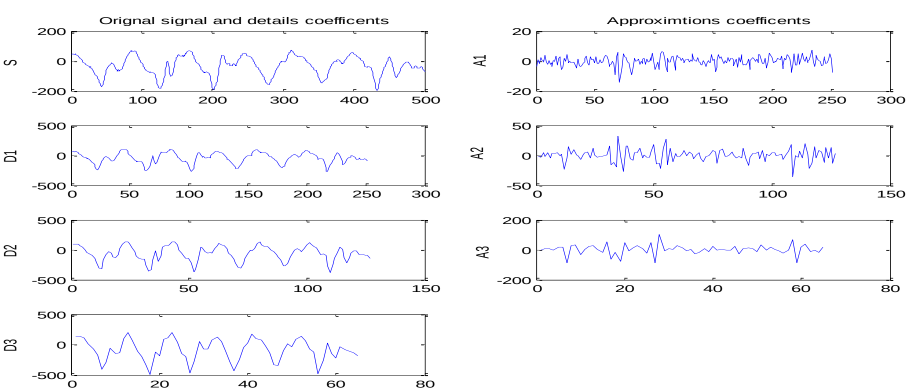

Fig. 5: Db2 level 3 decomposition of interictal EEG signal

7 Selections of Neural Network parameters

The neural networks were developed by systematic parameter optimization in view of the reasonable performance of the classifier. The training parameters were selected to obtain the best performance. The performance is decided on the basis of average classification accuracy. After several different experiments, such as number of hidden layers, size of hidden layers, value of the moment constant andlearning rate, and type of activation functions the optimal parameters are decided. Data tagged for training,testing and cross validation in each experiment were selected by trial and error method.For the generalization and randomization the results are also obtained on reverse tagging.

Table 1 presents the classification results of GFFNN, MLP, ENN and SVM for normal tagging and reverse tagging. Two different data partitions are used with different tagging orders. In the first case (forward tagging), the first 60 % samples (1:180) are used for training, 10 % samples (181:210) are used for CV and the last 30% samples (211:300) are for testing of classifier. In the second case (reverse tagging), the last 60 % sample (121:300) are used for training, 10% samples (91:120) are used for CV and the first30% samples (1:90) for testing of classifier.In both the cases, testing data set is divided in to three group namely group 1, group 2 and group3.Each group contains30 number of samples.Results for all these four neural networks are satisfactory in both the cases, but still classification accuracy with MLP and SVM are slightly higher as compared to the other two.

7.1 Multi-layer Perceptron neural network (MLPNN)

MLP is designed by systematic parameter optimizationas discussed earlier to give the optimal performance on the basis of best classification accuracy. The main featurevector was divided into three parts, the first 60% was used for training purpose, second is of 10%which wasused for cross validation and the last remaining 30%was used for testing purpose. It is observed that MLP with single hidden layer yields better performance. To decide the number of processing elements (PEs) in the hidden layer, the number of PEs are varied from 1 to 20 and the average minimum MSE is examined. Fig.7 shows that with 18 PEs in hidden layer, we have optimum result.

Table 1: Classification of epilepsy using ANNs

Tagging	NN	Testing			Average	CV	Training
		Group 1	Group 2	Group 3			
Normal/Forward	MLP	96.97	100	94.87	97.28	100	98.90
	SVM	100	100	100	100	100	100
	GFF	96.67	100	96.97	97.88	96.30	97.81
	ENN	95.83	89.81	93.33	92.99	89.74	87.78
Reverse	MLP	96.67	100	87.46	94.71	100	96.74
	SVM	96.30	98.04	93.94	96.09	100	100
	GFF	96.67	96.97	89.34	94.32	95.24	98.94
	ENN	87.91	100	96.67	94.86	97.78	96.30

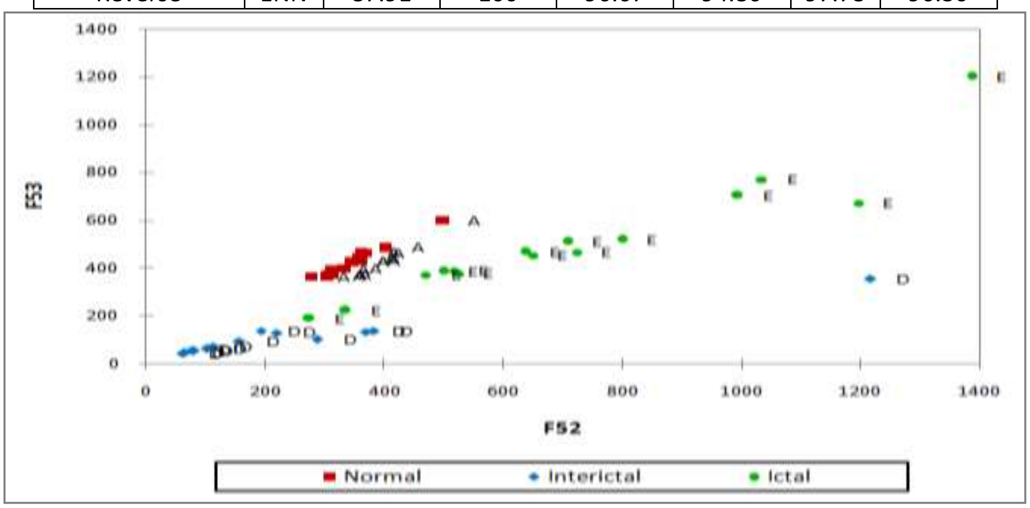

Fig. 6: Typical scatter plot of Normal, Interictal and Ictal EEG signal

Various Transfer functions, namely, Tanh, Sigmoid, Linertanh, Linear-sigmoid, Softmax, and learning rules, namely, Momentum, Conjugate-Gradient, Quick Propagation, Delta Bar Delta, and Step are verified for training and cross validation. Minimum MSE and average classification accuracy on training, testing and CV data set are compared. It is found that Tanh transfer function and momentum learning rule give the optimum results. The average minimum MSE and average classification accuracy with different transfer function and learning rule are plotted in fig. 8. Step size and momentum of hidden layer and output layer is also varied for optimum average minimum MSE and average classification accuracy. MLP is designed with the following optimal parameters.

Number of Inputs: 63; Number of Hidden Layers: 01;
Number of PEs in Hidden Layer: 18;
Hidden Layer:
Transfer function: tanh   Learning Rule: Momentum
Step size: 0.1   Momentum: 0.7
Output Layer:
Transfer function: tanh   Learning Rule: Momentum
Step size: 0.1   Momentum: 0.7

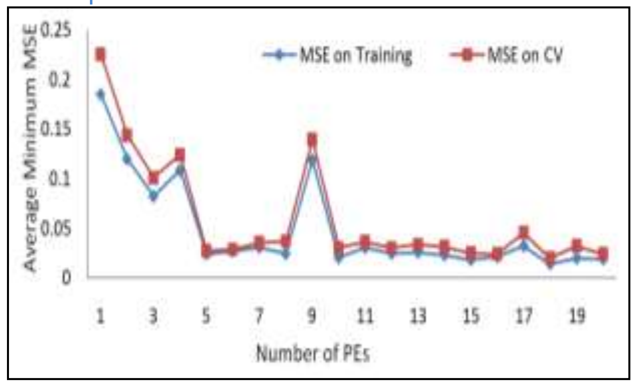

Fig. 7: Variation of average MSE with number of PEs in hidden layer

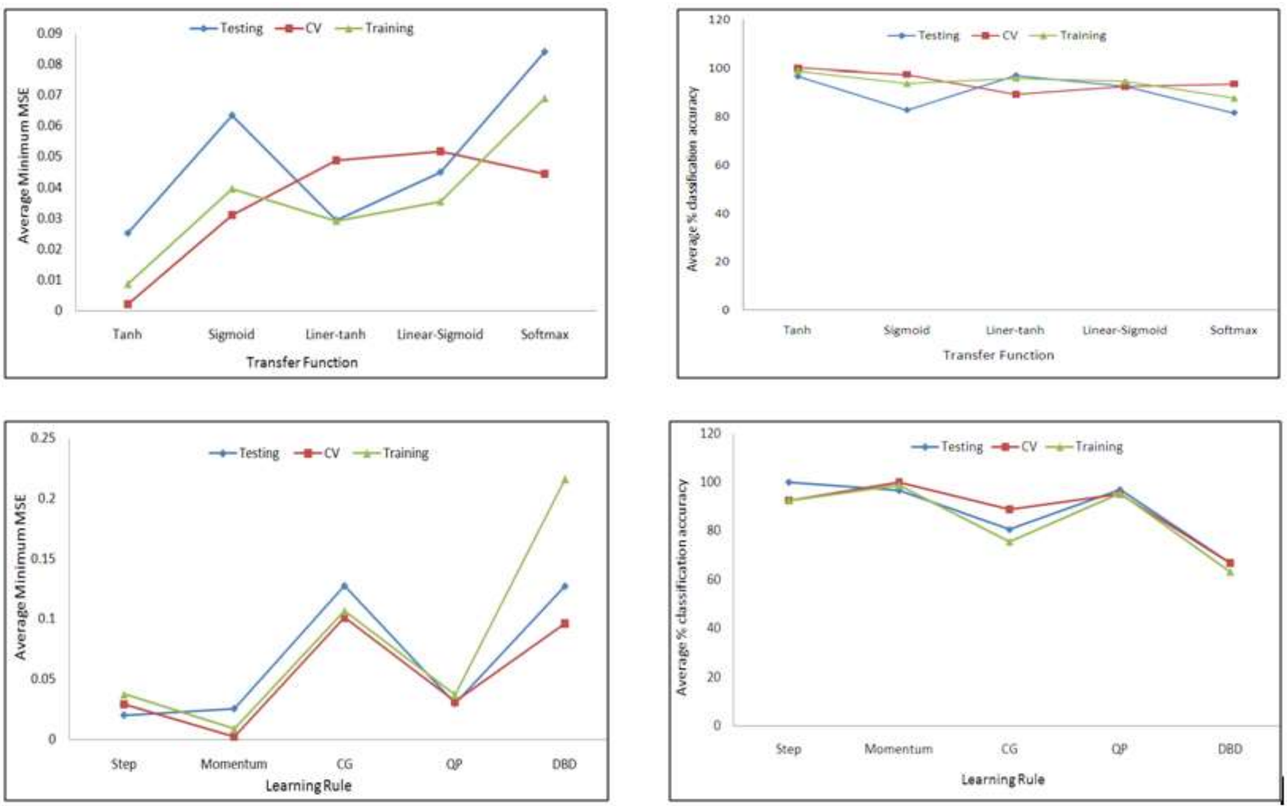

Fig.8: Graphs showing variation of average minimum MSE and average % classification accuracy with transfer function and learning rule

7.2 Support Vector Machine (SVM)

The support vector machine (SVM) is a new kind of classifier that is motivated by two concepts. First, transforming data into a high-dimensional space can transform complex problems (with complex decision surfaces) into simpler problems that can use linear discriminant functions. Second, SVMs are motivated by the concept of training and using only those inputs that are near the decision surface since they provide the most important information about the classification.It is a kind of learning machine based on statistical learning theory. The basic idea of applying SVM to pattern classification can be stated as follows: first the input vectors are mapped into one feature space, possibly in higher dimensional space, either linearly or nonlinearly, which is relevant with the kernel function. We have used the Kernel Adatron algorithm for this purpose. Then, within the feature space from the first step, optimized linear division is sought, that is a hyper plane is constructed which separates two classes. It can be extended to multi-class. SVM's training always seeks a global optimized solution and avoid over-fitting, so it has an ability to deal with a large number of features.

Kernel Adatron algorithm for the classifier:

For N dimensional space data $x_i(i = 1 \ldots N)$ this algorithm can be easily extended to network by substituting the inner product of patterns in the input space by the kernel function, leading to the following quadratic optimization problem:

$$J(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j G(x_i - x_j, 2\sigma^2) \quad (2)$$

Subject to

$$\sum_{i=1}^{N} d_i \alpha_i = 0 \; \alpha_i \geq 0, \forall_i \in \{i \ldots . N\} (3)$$

Where $G(x, \alpha\ ^2)$represents a Gaussian function, N denotes the number of samples, α_i*are* a set of multipliers (one for each sample),

$$J(x_i) = \mathrm{d_i}\left(\sum_{i=1}^{n} d_j \alpha_j \; \mathrm{G}(\mathrm{x_i} - \mathrm{x_j}, 2\sigma^2) + \mathrm{b}\right) (4)$$

And

$$M = \min \mathrm{g}(x_i) \quad (5)$$

and a common starting multiplier α_i, learning rate η, and a small threshold are chosen. Then, while M > t, we choose a pattern $x_i$and an update $\Delta\alpha_i = \eta(1 - g(x_i))$ is calculated and the update is performed if,

$\alpha_i(n) + \Delta\alpha_i > 0$

$$\alpha_i(n+1) = \alpha_i(n) + \Delta\alpha_i(n)$$
$$b(n+1) = b(n) + d_i\Delta\alpha_i \quad (6)$$

And if

$\alpha_i(n) + \Delta\alpha_i \leq 0$

$$\alpha_i(n+1) = \alpha_i(n)$$
$$b(n+1) = b(n) \quad (7)$$

After adaptation only some of the α_i are different from zero(called the support vectors). They correspond to the samplesthat are closest to the boundary between classes. This algorithmcan be considered the "on-line" version of the quadraticoptimization approach utilized for SVMs, and it can find thesame solutions as Vapnik's original algorithm for SVMs. It iseasy to implement the kernel Adatron algorithm since$g(x_i)$ can be computed locally to each multiplier, provided thatthe desired response is available in the input file. In fact, theexpression for $g(x_i)$ resembles the multiplication of an errorwith an activation, so it can be included in the framework ofneural network learning. The Adatron algorithm essentiallyprunes the RBF network so that its output for testing is given by,

$$f(\mathrm{x}) = \mathrm{sgn}\left(\sum_{\substack{i \in support \\ vector}}^{N} d_j \alpha_j \; \mathrm{G}(\mathrm{x_i} - \mathrm{x_j}, 2\sigma^2) - \mathrm{b}\right) \quad (8)$$

And cost function in error criterion is

$$\mathrm{J(t)} = \frac{1}{2}\sum_{i=1}\left(\mathrm{d_i(t)} - \left(\tanh\left(\mathrm{y_i(t)}\right)\right)\right)^2 \quad (9)$$

The feature vector was again split into three parts. First part of 70% data was used for training purpose, second part of 20% used for cross validation and remaining 10% used for testing purpose. The SVM was retrained three times to avoid any kind of biasing and to ensure true learning. Finally, the SVM based classifier is designed with following specifications,

Number of Inputs: 59; Step Size: 0.7
Kernel algorithm: Adatron

8 Result

The performance of the proposed system is measured by using the parameters like sensitivity, Specificity and overall accuracy. In medical diagnosis, sensitivity denotes the percentage of correctly classified disease individuals and specificity denotes the percentage of correctly classified individuals without the disease. For our epilepsy diagnosis system selectivity, sensitivity and overall accuracy is defined in equations(10) to (12).

$$\text{Sensitivity} = \frac{\text{Positive correctly classifies EEG segments}}{\text{Total positive EEG segments}} \quad (10)$$

$$\text{Specificity} = \frac{\text{Negative correctly classifies EEG segments}}{\text{Total Negative EEG segments}} \quad (11)$$

$$\text{Overall accuracy} = \frac{\text{correctly classifies EEG segmants}}{\text{Total EEG segments}} \quad (12)$$

Table 2: Confusion matrix for testing data set using MLP

Output / Desired	Normal	Epilipetic(Ictal)	Epileptic (Interictal)
Normal	28	0	0
Epilipetic(Ictal)	1	27	0
Epileptic (Interictal)	1	1	32

Table 3: Confusion matrix for testing data set using SVM

Output / Desired	Normal	Epilipetic(Ictal)	Epileptic (Interictal)
Normal	25	0	0
Epilipetic(Ictal)	0	32	0
Epileptic (Interictal)	0	0	33

The confusion matrices for testing dataset using MLP and SVM with the testing dataset are shown in table 2 and table 3 respectively. Table 4 shows the performance measures for MLP and SVM obtained with the testing dataset. With MLP the percentage of average classification accuracy for training, testing and CV dataset is 98.90%, 97.28% and 100% respectively. For SVM, it is 100%, 100% and 100% respectively. The overall percentage accuracy for MLP and SVM is 98.50% and 100%, respectively.

9 Sensitivity Analyses

Sensitivity analysis is used to find effect, that each of the network inputs is having on the network output. This provides feedback as to which input features are the most significant. From there, we may decide to prune the input space by removing the insignificant features. This will reduce the size of the network, which in turn reduces the complexity and the training times.

Sensitivity analysis is a method for extracting the cause and effect relationship between the inputs and outputs of the network. The network learning is disabled during this operation such that the network weights are not affected. The basic idea is that the inputs to the network are shifted slightly and the corresponding change in the output is reported either as a percentage or a raw difference.

After removing the insignificant input channel with the help of sensitivity analysis, new feature vectors are formed of the size 19*300, 16*300,17*300 and 20*300 for MLP, GFF, ENN and SVM respectively. These neural networks are again retrained with the new feature vector for the optimized result. Table 5 shows the reduced number of connection weights after the sensitivity analysis for the MLP, GFF, ENN and SVM. It is observed that the complexity of all these neural networks is considerably reduced without affecting the average classification accuracy significantly. However time complexity and space complexity of these networks have been reduced substantially.

10 Noise Sustainability of Classifier

Since the proposed classifier is to be used in real time, where measurement noise is anticipated, it is necessary to check the robustness of the classifier to noise. The Noisecomponent is used to inject random noise sources into an input or output. Since the noise is unwanted and unknown disturbance, it is difficult to check the noise sustainability of the network. To know the true performance of the network with noisy input and output, it is not common to test a network's sensitivity to uniform and Gaussian noises. The noise signals were specified by their mean and variance. Noise functions are as follows:

Noise Function

Uniform:$y = \sqrt{3\sigma^2}(x - 0.5) + \mu$

Gaussian: $y = \sigma^2\sqrt{-2logx}\cos(2\pi x) + \mu$

Wherexdenotes a pseudorandom floating point value between 0 and 1; σ denotes the square root of the variance; and μ denotes the mean.

Fig.9 and fig.10 show the performance of neural networks in presence of controlled Uniform noise and Gaussian noise. The MLP and SVM are the most robust classifier among these four classifiers. It is observed that the overall classification accuracy is decreased by less than 5% for MLP and SVM for the both Uniform and Gaussian noise.It is also observed that MLP and SVM can sustain Uniform noise up to 6% variance and Gaussian noise up to 8% variance.

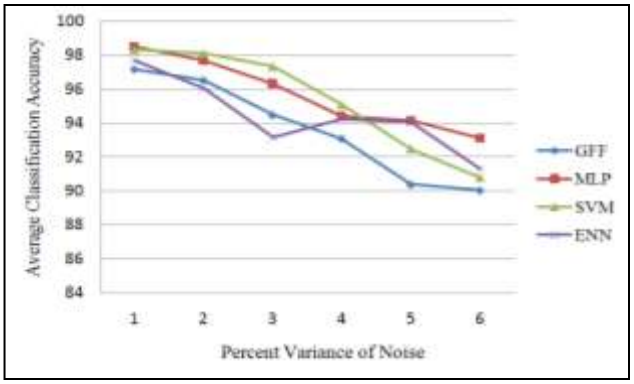

Fig9: Effect on the overall classification accuracy as variance of Uniform noise changes.

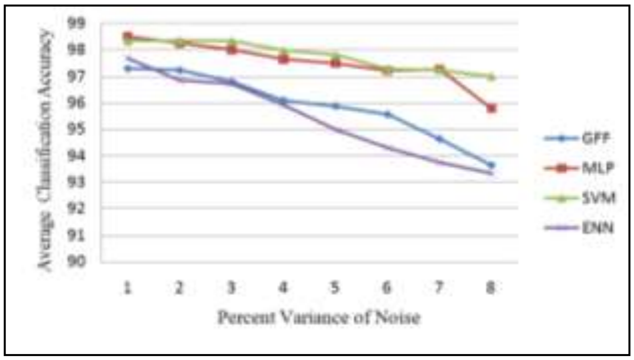

Fig10: Effect on the overall classification accuracy as variance of Gaussian noise changes.

11 Conclusion

we have proposed the wavelet based feature extraction technique for the classification of epileptic EEG signal using the artificial neural network. The problem of classification is modeled as three class classification problem. The three groups are: 1) Healthy subjects (Normal EEG) 2) Epileptic subjects during seizure free interval (Interictal EEG) and 3) Epileptic subjects during seizure activity (Ictal EEG). MLP, GFFNN, ENN and SVM are designed for this classification problem. Each neural network was retrained three times with different random initialization of connection weights so as to ensure the true learning. To avoid any kind of biasing the reverse tagging order is also used.The performance of these artificial neural networks is measured and compared in terms of percentage average classification accuracy. The percentage classification accuracy for MLP and SVM was found to be the highest amongst these four neural networks. Using MLP, sensitivity of 96.42% and 94.11% are obtained for interictal EEG and ictal EEG, respectively. For SVM, it is 100% for both interictal EEG and ictal EEG. The 100% specificity is obtained for both MLP and SVM. The complexity of neural network is considerably reduced with the help of sensitivity analysis after removing the insignificant inputs. Also the reliability and robustness of neural network is tested in presence of Uniform and Gaussian noise. It has been observed that MLP and SVM can sustain noise up to 6% and 8% variance in case of uniform and Gaussian noise respectively. For both types of noise the overall classification accuracy of MLP and SVM is reduced by only less than 5%.

Table 4: Performance measures of MLP and SVM based classifier

Neural Network	Average MSE			%Average Classification Accuracy			% Sensitivity	% Sensitivity	% of Specificity	Overall % Accuracy
	Training	Testing	CV	Training	Testing	CV	(Interictal)	(Ictal)		
MLP	0.008	0.025	0.002	98.90	96.59	100	94.11	96.42	93.33	96.66
SVM	0.005	0.028	0.038	100	100	95	100	100	100	100

Table 5: Effect of sensitivity analysis on the complexity of Neural Network

NN	Before Sensitivity Analysis			Connection Weights	Average Classification Accuracy	After Sensitivity Analysis			Connection Weights	Average Classificat ion Accuracy
	PEs in I/P	PEs in Hidden	PEs in O/P			PEs in I/P	PEs in Hidden	PEs in O/P		

	layer	layer	layer			layer	layer	layer		
MLP	63	18	03	1209	98.49	19	13	03	302	98.11
GFF	63	16	03	1065	97.29	16	11	03	223	97.65
ENN	63	12	03	807	97.66	17	15	03	363	96.59
SVM	63	210	03	14073	98.33	20	210	03	5043	96.11

List of Abbreviations:

ANN Artificial Neural Network
CV Cross Validation
DB Daubechies
DWT Discrete wavelet transform
EEG Electroencephalogram
ENN Elman Neural Network
GFF Generalized Feed Forward
I/O Input
MLP Multilayer Perceptron
MSE Mean Square Error
NN Neural Network
O/P Output
PE Processing Elements
SVM Support Vector Machine

References:

[1] Yinxia Liu,Weidong Zhou, Qi Yuan and Shuangshuang Chen (2012). Automatic Seizure Detection Using Wavelet Transform and SVM in Long-Term Intracranial EEG, Neural Systems and Rehabilitation Engineering, vol 20(6). Pp-749-755

[2] L. Szilagyi Z. Benyo and M. Szilagyi (2002).A new method for epileptic waveform recognition using wavelet decomposition and artificial neural networks, Proceeding of second joint EMBS/BMES conference, vol. 3, pp. 2025-2026.

[3] Gotman J. (1982).Automatic recognition of epileptic seizure in the EEG, Electroencephalography and Clinical Neurophysiology, vol. 54, pp. 530-540.

[4] Arulampalam G. and Bouzerdoum A. (2003). Generalized Feedforward Neura Network Classifier, Proceedings of the international joint conference on Neural Network, vol.2, pp. 1429-1434.

[5] Pravin Kumar S., Sriram N. and Benakop P. G. (2008).Automated detection of epileptic seizure using wavelet entropy feature with recurrent neural network classifier, IEEE Region 10 conference, TENCON 2008, pp. 1-5

[6] Pradhan N., Sadasivan P. K. and Arunodaya G. R. (1996). Detection of seizure activity in EEG by an artificial neural network: A preliminary study, Computer Biomedical Research, vol. 29. pp. 303-313.

[7] Anupama Shukla, Ritu Tiwari, and Prabhdeep Kaur (2009). Intelligent system for the diagnosis of epilepsy, World congress on computer science and information engineering, IEEE, vol. 5, pp.755-758.

[8] M. Akin, M. A. Arserim, M. K. Kiymik, and I. Turkoglu (2001). A new approach for diagnosing epilepsy by using wavelet transform and neural network, 23rd annual EMBS international conference, vol. 2, pp.1596-1599.

[9] Vairavan Srinivasan, Chikkannan eswaran, and Natarajan Sriraam (2007). Approximate entropy-based epileptic EEG detection using artificial neural network, IEEE transaction on information technology in biomedicine, vol, 11(3), pp.288-295.

[10] S.Gosh-Dastidar and Hojjat Adeli (2007). Principal component analysis-enhanced cosine redial basic function neural network for robust epilepsy and seizure detection, IEEE Transaction on Biomedical Engineering, vol. 54(9), pp. 1545-1551.

[11] Sheng-Fu Liang, Wan-Lin Chand and Herming Chiueh (2010). EEG-based absence seizure detection methods, Neural network (IJCNN), the 2010 international conference, pp.1 – 4.

[12] S. Pravin Kumar et. al. (2010) "Entropy based detection of epileptic seizure with artificial neural network classifiers", Expert system with applications, vol. 37(4), pp. 3284-3291.

[13] U Rajendra Acharya et. al. (2012). Automated diagnosis of epileptic EEG using entropies, Biomedical signal processing and control, Vol. 7, pp. 401-408.

[14] Abdulhamit Subasi (2007). EEG signal classification using wavelet feature extraction and mixture of expert model, Expert system with application vol. 32, pp. 1084-1093.

[15] L. M. Patnaik and Ohil K. Manyam (2008). Epileptic EEG detection using neural networks and post-classification, Computer methods and program in biomedicine, vol. 91, pp. 100-109.

[16] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, christoph Rieke, Peter David, and Christian E. Elger (2001). Indication of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, phys. Rev. E,Sate. Nonliner Soft Mattr Phys.,vol. 64, pp.061907-1-061907-8.

[17] P. A. Kharat and S. V. Dudul (2011).Effects of Input Dimensionality Reduction on the Performance of Epilepsy Diagnosis Based on Neural Network, International Journal of Machine Intelligence, vol. 3(5), pp-396-402.

[18] P. A. Kharat and S. V. Dudul. (2012).Epilepsy diagnosis based on generalized feed forward neural network, Interdisciplinary Sciences-Computational Life Science,vol. 4 (3), pp 209-214.

[19] P. A. Kharat and S. V. Dudul (2011). Clinical Decision Support System Based on Jordan/Elman Neural Networks, Recent Advances in Intelligent Computational SystemsRAICS-2011, pp. 255-259.

[20] V. N. Ghate and S. V. Dudul (2009). Induction machine fault detection using support vector machine based classifier, WSEAS transaction on system, Issue 5(8), pp. 591-603.

[21] Shiqing Zhang, Xiaoming Zhao and Bicheng Lei (2012). Facial expression recognition using sparse representation, WSEAS transaction on system, vol. 11(8), pp. 440-452.

[22] Shatnawi, Y. and Al-khassaweneh, M.(2014). Fault Diagnosis in Internal Combustion Engines Using ExtensionNeural Network,IEEE Transactions onIndustrial Electronics, vol. 61(3),pp- 1434 – 1443.

[23] Jain, S.K. and Singh, S.N.(2014). Low-Order Dominant Harmonic Estimation Using Adaptive Wavelet Neural Network,IEEE Transactions on Industrial Electronics, vol. 61(1), pp- 428-435