International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-2, Issue-2, May 2012

Performance Analysis of Patient Specific Epilepsy Risk Level Classifications from EEG Signals Using Two Tier Hybrid (Fuzzy, Soft Decision Trees Models and MLP Neural Networks) Classifiers

R.Harikumar , M.Balasubramani, T.Vijayakumar

*Abstract— **This paper compares the performance analysis of a two tier hybrid Fuzzy, Soft Decision Tree (SDT) models and Multi layer Perceptron (MLP) neural networks in optimization of patient specific epilepsy risk levels classifications from EEG (Electroencephalogram) signals. The fuzzy classifier (level one) is used to classify the risk levels of epilepsy based on extracted parameters like energy, variance, peaks, sharp and spike waves, duration, events and covariance from the EEG signals of the patient. Soft Decision Tree (post classifier with max-min and min-max criteria) of three models and MLP neural networks are applied on the classified data to identify the optimized risk level (singleton) which characterizes the patient's state. The efficacies of these methods are compared with the bench mark parameters such as Performance Index (PI), Sensitivity, Specificity and Quality Value (QV). A group of twenty patients with known epilepsy findings are analyzed. High PI such as 95.88 % was obtained at QV's of 22.43 in the SDT model of (16-4-2-1) with Method-II (min-max criteria) and for MLP (4-4-1) 99.9%and 24.43 when compared to the value of 40% and 6.25 through fuzzy classifier respectively. It was identified that the SDT models and MLP (4-4-1) are good post classifier in the optimization of epilepsy risk levels. SDT models were well accounted for low training cost over heads. A part from the training cost MLP neural networks outperformed SDT classifiers in classifying the epilepsy risk levels.***

Index Terms— EEG Signals, Epilepsy, Fuzzy Logic, Soft Decision Trees, Multi Layer Perceptron (MLP) neural networks, Risk Levels.

I. INTRODUCTION

Medical expert systems are a challenging field, requiring the synergy of different scientific areas. The representation of medical knowledge and expertise, the decision making in the presence of uncertainty and imprecision, the choice and adaptation of suitable model, are some issues that a medical expert system should take under consideration [1].

Manuscript received on April 20, 2012

Dr.R.Harikumar,Professor, ECE Department,.Bannari amman Institute of Technology, Sathyamangalam,India 9942932239 (e-mail:harikumarrajaguru@gamil.com).

M.Balasubramani Assistant Professor, ECE Department, .Info Institute of Engineering, Coimbatore, India 9486770500 (e-mail: rishiveda@yahoo.com.)

T.Vijayakumar Assistant Professor,(Sr.Gr) Department of IT, Bannari Amman Institute of Technology, Sathyamangalam9965214777(e-mail:vishal_16278@gamil.com)

Uncertainty is traditionally treated in probabilistic manner; recently, however, methods based on fuzzy techniques have gained ground [2]. The model's parameter adaptation (training) amounts to optimizing a properly constructed "error" function. There is a variety of methods with diverse features that may properly understand the subtleties of the optimization procedures and is a key to choose effective training approach [3].Now, the subject of neural networks has become very popular in many areas such as signal processing and pattern recognition [4],[5]. Actually, the applications of neural networks in pattern recognition problems go back to the days of simple perceptrons in the 1950's [6]. Many advantages for neural networks have been cited in the literature [7],[8]. Most important of them, for the pattern recognition problems, seems to be the fact that neural network- based approaches are usually non parametric even statistical information could also possible to incorporated for improving their performance[9]. Also neural networks can extract nonlinear combinations of features, and the resulting discriminating surfaces can be very complex. These characteristics of neural networks can be very attractive in a decision tree classifier where one has to determine the appropriate feature subsets and the decision rules at each internal node [10]. There are various neural network models such as Hopfield nets, the Boltzmann machine and Kohonen self- organizing feature maps, to name a few; the most popular network by far, however, is the Multi Layer Feed Forward Network [11]. Even though Multi Layer Perceptron (MLP) Networks and SDT's are two very different techniques for classification, the similarities in the distributed nature of decision- making in both processes has motivated some researchers [12],[13],[14].

A. General Techniques and Motivation

EEG is an important clinical tool for diagnosing, monitoring and managing neurological disorders related to epilepsy. This disorder is characterized by sudden recurrent and transient disturbances of mental function and/or movements of body that results in excessive discharge group of brain cells [15]. The presence of epileptiform activity in the EEG confirms the diagnosis of epilepsy, which sometimes confused with other disorders producing similar seizure like activity [16]. Between seizures, the EEG of a patient with epilepsy may be characterized by occasional epileptic form transients-spikes and sharp waves. The different types of epileptic seizures are characterized by different EEG waveform patterns [17]. With real-time monitoring to detect epileptic seizures gaining widespread recognition, the advent of computers has made it possible to effectively apply a host of methods to quantify the changes occurring based on the EEG signals [2].One of them is a classification of risk level of epilepsy using Fuzzy techniques [4].This paper addresses the application and comparison of SDT models and MLP neural networks towards optimization of fuzzy outputs in the classification of epilepsy risk levels.

Weber etal, [18] have proposed a three stage design of an EEG seizure detection system. The first stage of the seizure detector compress the raw data stream and transforms the data into variables that represent the state of the subject's EEG. These state measures are referred to as context parameters. The second stage of the system is a neural network that transforms the state measures into a smaller number of parameters that are intended to represent measures of recognized phenomena such as small seizure in the EEG [19]. The third stage consists of a few simple rules that confirm the existence of the phenomena under consideration. Similarly, this paper also presents a three stage designs for epilepsy risk level classification. The first stage extracts the required seven distinct features from raw EEG data stream of the patient in time domain. The next stage transforms these features into a code word through a Fuzzy system with six alphabets which represents the patient's state in a five distinct risk levels for a two second epoch of EEG signal per channel. The last stage is a MLP neural network or SDT which optimizes the epilepsy risk level of the patient.

II. MATERIALS AND METHODS

The EEG data used in the study were acquired from twenty epileptic patients who had been under the evaluation and treatment in the Neurology Department of Sri Ramakrishna Hospital, Coimbatore, India. A paper record of 16 channel EEG data is acquired from a clinical EEG monitoring system through 10-20 international electrode placing method.

A. Artifact Rejection and Acquisition of EEG Data

Effective elimination of artifact (head and body movement, perspiration, and low frequency instrument artifacts, under 1 Hz; high frequency artifact including gross muscle potentials, 30-50 Hz; and eye movements, under 3 Hz, in the frontal channels) from the collected data is an essential step in preparation of data for analysis. It is neither sensible nor correct to apply elaborate computer analysis to data contaminated with artifacts [20]. The traditional practice has been to select visually a representative segment of artifact free data for computer analysis. These procedures obviously introduce an element of subjective bias in data selection. With an EEG signal free of artifacts, a reasonably accurate detection of epilepsy is possible; however, difficulties arise with artifacts. This problem increases the number of false detection that commonly plagues all classification systems. With the help of neurologist, artifact free EEG records with distinct features were selected. These records were scanned by Umax 6696 scanner with a resolution of 600dpi.

Since the EEG records are over a continuous duration of about thirty seconds, they are divided into epochs of two second duration each by scanning into a bitmap image of size 400x100 pixels. A two second epoch is long enough to detect any significant changes in activity and presence of artifacts and also short enough to avoid any repetition or redundancy in the signal [21]. The EEG signal has a maximum frequency of 50Hz and so, each epoch is sampled at a frequency of 200Hz. Each sample corresponds to the instantaneous amplitude values of the signal, totaling 400 values for an epoch. The different parameters used for quantification of the EEG are computed using these amplitude values by suitable programming codes. The parameters are obtained for three different continuous epochs at discrete times in order to locate variations and differences in the epileptic activity [22]. Twenty EEG records were used for both training and testing. These EEG records had an average length of six seconds and total length of 120 seconds. The patients had an average age of 31 years. A total of 960 epochs of 2 seconds duration are used.

B .Fuzzy System as a Level One Classifier

Figure 1 enumerates the overall epilepsy risk level (Hybrid Fuzzy-SDT and Neural network) classifier system. The motto of this research is to classify the epilepsy risk level of a patient from EEG signal parameters [23]. This is accomplished as:

1) Fuzzy classification for epilepsy risk level at each channel from EEG signals and its parameters.
2) Each channel results from fuzzy classifier are optimized using soft decision trees six types, since they are at different risk levels and highly nonlinear.
3) Comparison of performance of fuzzy classification with the Soft decision Tree models and MLP neural networks optimization methods are analyzed.

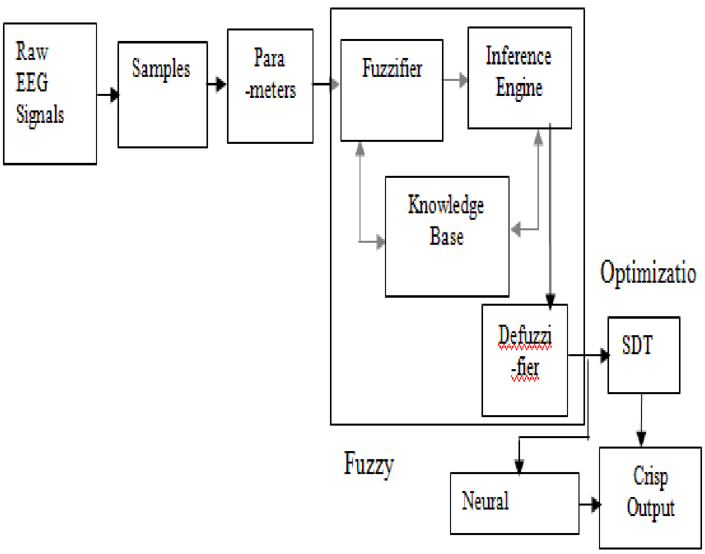

Figure 1. Hybrid Fuzzy – SDTand Neural Network Classification System

1. The energy in each two-second epoch is given by

$$E = \sum_{i=1}^{n} x_i^2 \qquad (1)$$

Where x_i is signal sample value and n is number of samples. The scaled energy is taken by dividing the energy term by 1000.

2. The total number of positive and negative peaks exceeding a threshold is found.

3. Spikes are detected when the zero crossing duration of predominantly high amplitude peaks in the EEG waveform lies between 20 and 70 ms and sharp waves are detected when the duration lies between 70 and 200ms.

4 The total numbers of spike and sharp waves in an epoch are recorded as events.

5. The variance is computed as σ given by

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

Where $\mu = \frac{\sum_{i=1}^{n} x_i}{n}$ is the average amplitude of the epoch.

6 .The average duration is given by

$$D = \frac{\sum_{i=1}^{p} t_i}{p} \qquad (3)$$

Where t_i is one peak to peak duration and p is the number of such durations.

7. Covariance of Duration. The variation of the average duration is defined by

$$CD = \frac{\sum_{i=1}^{p}(D - t_i)^2}{pD^2} \qquad (4)$$

C. Fuzzy Membership functions

The energy is compared with the other six input features to give six outputs. Each input feature is classified into five fuzzy linguistic levels viz., *very low, low, medium, high* and *very high* [24]. The triangular membership functions are used for the linguistic levels of energy, peaks, variance events, spike and sharp waves, average duration and covariance of duration. The output risk level is classified into five linguistic levels namely *normal, low, medium, high* and *very high.*

D. Fuzzy Rule Set

Rules are framed in the format

IF Energy is low AND Variance is low THEN Output Risk Level is low

In this fuzzy system we have five linguistic levels of energy and five linguistic levels of other six features such as variance, peaks, events, spike and sharp waves, average duration and covariance of duration. Theoretically there may be 5^6 (that is 15625) rules are possible but we had considered the fuzzy pre -classifier as a combination of six two inputs and one output (2×1) system. With energy being a constant one input the other input is selected in sequential manner. This two inputs one output (2×1) fuzzy system works with 25 rules. We obtain a total rule base of 150 rules based on six sets of 25 rules each. This is a type of exhaustive fuzzy rule based system [25].

E. Estimation of Risk Level in Fuzzy Outputs

The output of a fuzzy logic represents a wide space of risk levels. This is because there are sixteen different channels for input to the system at three epochs. This gives a total of forty-eight input output pairs. Since we deal with known cases of epileptic patients, it is necessary to find the exact level of risk the patient. This will also aid in the development of automated systems that can precisely classify the risk level of the epileptic patient under observation. Hence an optimization of the outputs of the fuzzy system is necessary. This will improve the classification of the patient and can provide the EEGer with a clear picture. A specific coding method processes the output fuzzy values as individual code. Since working on definite alphabets is easier than processing numbers with large decimal accuracy, we encode the outputs as a string of alphabets. The alphabetical representation of the five classifications of the outputs is shown in table.1

Table.1Representation of Risk level Classifications

Risk Level	Representatio
Normal	U
Low	W
Medium	X
High	Y
Very High	Z

A sample output of the fuzzy system with actual patient readings is shown in fig. 2 for eight channels over three epochs. It can be seen that the Channel 1 shows medium risk levels while channel 8 shows very high risk levels. Also, the risk level classification varies between adjacent epochs.

Epoch 1	Epoch 2	Epoch 3
YYYYXX	ZYYWYY	YYYXYZ
YYYXYY	ZZYZZZ	YYYXYZ
YYYYYY	ZZYZZZ	ZYYYZZ
ZYYYZZ	ZZYZYY	YYYXXZ
YYYYYY	YYYXYY	YYYYYZ
YYYYYY	YYYXYY	YYYXYY
YYYYYY	YYYYYY	YYYYYY
ZZYZYZ	ZZYZZZ	ZZYZZZ

Figure 2. Fuzzy Logic Output

The Performance of the Fuzzy method is defined as follows [21]

$$PI = \frac{PC - MC - FA}{PC} \times 100 \qquad (5)$$

Where PC – Perfect Classification, MC – Missed

Classification, FA – False Alarm,

PI= [(0.5-0.2-0.1)/0.5] *100 =40%.

The perfect classification represents when the physicians and fuzzy classifier agrees with the epilepsy risk level. Missed classification represents a true negative of fuzzy classifier in reference to the physician and shows High level as Low level. False alarm represents a false positive of fuzzy classifier in reference to the physician and shows Low level as High level. The performance for Fuzzy classifier is as low as 40%.The limitations of Fuzzy techniques are analyzed below.

F. Limitations of Fuzzy Techniques

Now, the nonlinearities and limitations associated with fuzzy outputs in describing the epilepsy risk levels are to be identified. Let the fuzzy outputs as shown in figure 2 is coded with appropriate numerical values. These numerical values are associated with the probability of each coded epilepsy risk level patterns. The five risk levels are encoded as Z>Y>X>W>U in binary strings of length five bits using weighted positional representation as shown in table 2. Encoding each output risk level of the fuzzy output gives us a string of six chromosomes, the value of which is calculated as the sum of probabilities of the individual genes. For example, if the output of an epoch is encoded as ZZYXWZ, its value would be 0.333331,[26]. Now the each input patterns are encoded in the numerical form of the range 0-1.

Table 2. Binary Representation of Risk levels

Risk Level	Code	Binary String	Weight	Probability
Very high	Z	10000	16/31=0.51612	0.086021
High	Y	01000	8/31=0.25806	0.043011
Medium	X	00100	4/31=0.12903	0.021505
Low	W	00010	2/31=0.06451	0.010752
Nomal	U	00001	1/31=0.03225	0.005376
		11111=31	Σ=1	

To illustrate the non linearity a statistical measure of cross correlation between the two adjacent epoch patterns was chosen. Thus the cross correlation function $\mathbf{r_{xy}(m)}$ of the epochs x(n) and y(n) is defined [27] by the equation (6) and assuming that both sequence have been measured from n=0 to n=N-1, in our case n=1to 16.

$$r_{xy}(m)=\begin{cases}\dfrac{1}{N}\displaystyle\sum_{n=0}^{N-m-1}x(n+m)y(n),\ for,0\le m\le N-1\\ \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-|M|-1}x(n)y(n+M),\ for,-(N-1)\le m\le 0\end{cases} \qquad (6)$$

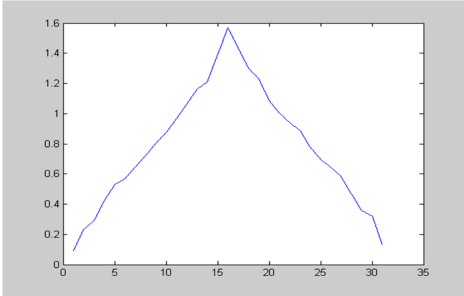

Figure.3 Cross Correlation Function plot for the Adjacent Epochs in fuzzy based Epilepsy Risk Level Outputs

As shown in the figure.3 the cross correlation $r_{xy}(m)$ plot emulates the occurrence of highly non periodic patterns in the fuzzy outputs. Therefore any closed form of solution will be failed in this purpose of optimization. Hence, it will be advisable to prefer non linear techniques instead of a linear one, such a one type is SDT. Since, SDT is a common way to solve a wide variety of ill-posed problems which is not necessarily treated as hard constraint one. A pertinent explanation for the SDT optimization is given below.

III.SOFT DECISION TREES FOR OPTIMIZATION OF FUZZY OUTPUTS AS HYBIRD CLASSIFIER

The EEG signals are inherently complicated due to their non-Gaussian, non-stationary, and often non linear nature. On the top of that, the small amplitude of these signals reinforces their sensitivity to various artifact removal and noise sources. Our objective is to merge the epilepsy risk level representation, with approximate reasoning capabilities, and symbolic decision trees while preserving advantages of both: uncertainty handling and gradual processing of the former with the comprehensibility, popularity, and ease of application of the later. Decision trees are attractive because they can approximate global complex decision regions by the union of simple local decision regions at various levels of the trees. In contrast to conventional single stage classifiers where each data sample is tested against all classes, thereby reducing efficiency, in a soft decision tree a sample is tested against only certain subsets of classes, therefore unnecessary computations are eliminated. The high dimensionality problem associated with multi criteria decision and minimum training samples are curtailed by the use of SDT.

Apart from several advantages there are some pertinent drawbacks associated with decision trees which areas follow i) Errors may accumulate from level to level in a large tree. Therefore one cannot simultaneously optimize both accuracy and efficiency; for any given accuracy a bound on efficiency must be satisfied. ii) Increased in number of terminals when number of classes is large and this lead to increase the search time and memory space requirements. iii) Finally, there may be difficulties involved in designing optimal SDT. The performance of SDT strongly depends on how well the tree is designed. The main objectives of SDT are, to classify correctly as much of training samples as possible, generalized beyond the training sample so that unseen samples could be classified with high accuracy (which is also a characteristics gleam of neural networks), easy for updating as more training samples are available, and a simpler structure is also possible. Only one feature is examined at each node. The algorithms are feasible only for a small number of features, else the size of the lattice becomes large and storage space requirements become a problem. This is because during the optimization process intermediate results must be fully accessible. This perhaps is the main limitation. The problem of designing a truly optimal SDT is a very difficult one. In fact it has been shown by Hyafil and Rivest [28] that the problem of constructing optimal binary trees, optimal in the sense of minimizing the expected numbers of tests required to classify an unknown sample is an NP- complete problem and thus very unlikely of non-polynomial time complexity. It is conjectured that more general problems, i.e. the problem with a general cost function or minimizing the maximum number of tests (instead of average) to classify an unknown sample would also be NP- complete. They also conjecture that no sufficient algorithm exists and thereby supply motivation for finding efficient heuristics for constructing near-optimal decision trees.

A. Algorithm for SDT Optimization

The design of SDT can be decomposed into the following tasks, i)the appropriate choice of tree structure, ii) the choice of feature subsets to be used at each internal node and iii) the choice of the decision rule or strategy to be used at each internal node. The various heuristic methods for construction of SDT can roughly be divided into four categories: Bottom-up approaches, Top-Down approaches the hybrid approach and tree Growing – pruning approaches [29]. A decision tree using bottom-up approach was constructed and studied. Using max-min soft decision measures, pair wise distances between a priori defined classes are computed and at each step the two classes with the node decision are merged to form a new group, and this process is repeated until one is left with one group at the root which will be the optimized epilepsy risk level patterns. This has some of the characteristics of an unsupervised hierarchical clustering approach. The tree constructed this way, the more obvious discriminations are done first, and more subtle ones at later stages of the tree. From a processing point of view, these types of trees are highly recommended. The generic representation of SDT optimization is explained, let W= [P_{ij}] be the co –occurrence matrix with (i,j) elements which represents fuzzy based epilepsy risk level patterns of single epoch. There are 48 (16x3) epochs are available. Three models of SDT such as (16-8-4-2-1), (16-4-2-1), and (16-2-1)were selected for optimization of fuzzy patterns. A decision strategy of Method –I (Max-min) or Method –II (Min-max) were applied at each nodes in the above three SDT models. Therefore six types of SDT models were obtained.

In the case of (16-8-4-2-1) model an epoch of (16x1) elements were considered as the leaf nodes of the tree. The next level of tree was named as B with eight decision nodes, which was followed by C level with four soft decision nodes. Further level was designated as D level with two nodes and the final level was the E level with single node which was the root of the tree. The following decisions were performed at the each node of the tree.

Max-Min Method I.

Let R_I , R_{I+1} be the ith and i+1 th leaf to be decided at next level B_I , as

1. B_i=max (R_I,R_{I+1}),& B_{i+1}=min (R_{I+2},R_{I+3}) and at next C_J level
2. C_i=min (B_I,B_{I+1}) & C_{i+1}=max (B_{I+2},B_{I+3}) and at next D_I level
3. D_i=max (C_I,C_{I+1}) & D_{i+1}=min (C_{I+2},C_{I+3}) and at next E level
4. E_I =min (D_I ,D_{I+1}).

In the case of **Min –Max Method II** procedure the following decisions are taken at the nodes of B, C, D, and E levels

1. B_i=min (R_I,R_{I+1}),& B_{i+1}=max (R_{I+2},R_{I+3}) and at next C_J level
2. C_i=max (B_I,B_{I+1}) & C_{i+1}=min (B_{I+2},B_{I+3}) and at next D_I level
3. D_i=min (C_I,C_{I+1}) & D_{i+1}=max (C_{I+2},C_{I+3}) and at next E level
4. E_I =max (D_I ,D_{I+1}).

The above algorithm is depicted in the figure 4.Each SDT model is trained and tested by means of MSE Function. Since our model is patient specific in nature, we are applying 48 (3x16) patterns for each SDT model. As the number of patterns in each database for training is limited, each model is trained with one set of patterns (16) for zero mean square error condition and tested with other two sets of patterns (2x16). After network is trained using these, the classification performance of test set is recorded. The testing process is monitored by the Mean Square Error (MSE) which is defined as [9]

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(O_i - T_j)^2 \qquad (7)$$

Where Oi is the observed value at time i, T_j is the target value at model j; j=1-10, and N is the total number of observations per epoch and in our case, it is 16. Table.3 depicts the training and testing performance of six types of SDT models for twenty patients. The squared error (e_i^2) from equation (7) between the input and the output of the SDT is converted into the confidence score using relation **C_i=exp (-λe_i^2)** where refers to the SDT index [30]. In this paper λ=1was chosen.

Table .3 Estimation of MSE in SDT models

SDT Models	Mean Square Error (MSE) Index		Confidence score
	Training	Testing	C_i=exp(-λe_i^2)
Method-I			
16-8-4-2-1 (AR1)	5.2E-03	5.9E-03	0.9941
16-4-2-1 (AR2)	8.9E-03	8.4E-03	0.9916
16-2-1(AR3)	9.1E-03	9.32E-03	0.9961
Method-II			
16-8-4-2-1 (AR1)	1.66E-02	7.24E-03	0.9927
16-4-2-1 (AR2)	**9.6E-04**	**2.04E-03**	**0.998**
16-2-1 (AR3)	1.66E-02	0.101	0.989

The average confidence score for each SDT model is also tabulated in the table 3. SDT (16-4-2-1) model with method –II (Min-max criteria) provided better training and testing MSE. Hence, SDT (16-4-2-1) model with method-II was selected as an appropriate post classifier for optimization of fuzzy outputs in epilepsy risk level classification. The following section explains about the role of neural networks as a post classifier for epilepsy risk levels.

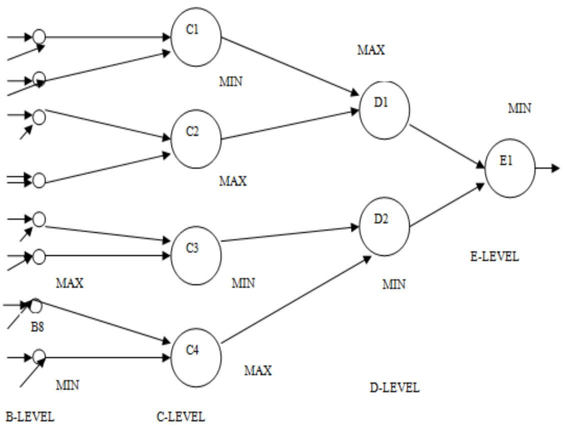

Figure. 4 Optimization of Epilepsy Risk Levels through STD (16-8-4-2-1) model with (Max-min) Method I

IV. ROLE OF NEURAL NETWORKS AS A POST CLASSIFIERS

Neural networks have been touted as having excellent potential for improving classification accuracy in patient specific diagnostic data. However, there have been few studies which have demonstrated these potential using real data sets [7]. The appeal of neural networks as pattern recognition systems is based upon several considerations. First, neural networks appear to perform as well or better than other techniques, and require no assumptions about the explicit parametric nature of distributions of the pattern data to be classified. In this regard they are similar to K-nearest neighbor algorithms. However, neural networks, once trained, are computationally more efficient.

A. Multi layer Perceptrons (MLP) Neural Network for Risk Level Optimization

Multilayer perceptrons (MLPs) are feed forward neural networks trained with the standard back propagation algorithm. They are supervised networks so they required a desired response to be trained [31]. They learn how to transform input data into a desired response, so they are widely used for pattern classification [5]. Most NN applications involve MLPs. They are very powerful pattern classifiers. With one or two hidden layers they can approximate virtually any input-output map. They have been shown to approximate the performance of optimal statistical classifiers in difficult problems. They can efficiently use the information contained in the input data. The advantage of using this network resides in its simplicity and the fact that it is well suited for online implementation [32].

The Levenberg-Marquardt (LM) algorithm is the standard training method for minimization of MSE (Mean Square Error) criteria, due to its rapid convergence properties and robustness. It provides a fast convergence, it is robust and simple to implement, and it is not necessary for the user to initialize any strange design parameters. It out performs simple gradient descent and other conjugate gradient methods in a wide variety of problems. The LM algorithm is first shown to be a blend of vanilla gradient descent and Gaussian Newton iteration. This error back propagation algorithm is used to calculate the weights updates in each layer of the network. The LM update rule is given as[32]

$$\Delta W = \left(J^T J + \mu\right)^{-1} J^T e \qquad (8)$$

Where J is Jacobian matrix of derivatives of each error to each weighted μ is a scalar, and e is error vector. If scalar μ is very large, the above method approximates gradient –descent. While if it is small the above expression becomes Gauss-Newton method. Because the GN method is faster but tends to less accurate near an error minima. The scalar μ is adjusted just like adaptive learning rate used by *trainbpx*. As long as the error gets smaller, μ is made smaller. Training continues until the error goal is met, the minimum error gradient occurs, the maximum value of μ occurs or the maximum number of epochs has finished.

B. Training and Testing Procedures for the Selection of Optimal Architecture

The primary aim of developing an ANN is to generalize the features (epilepsy risk level) of the processed fuzzy outputs. We have applied different architectures of MLP networks for optimization. The weights between input layer, the hidden layer and output layer network are trained with error back propagation algorithm to minimize the square output error to zero. The simulations were realized by Neural Simulator 4.0 of Matlab v.7.0 [33]. As the number of hidden units is gradually increased from its initial value, the minimum MSE on the testing set begins to decrease. The optimal number of hidden units is that number for which the lowest MSE is achieved. If the number of hidden units is increased beyond this performance does not improve and soon begins to deteriorate as the complexity of the neural network model is increased beyond that which is required for the problem. The training procedure for MLP Neural network is shown in the figure.5.

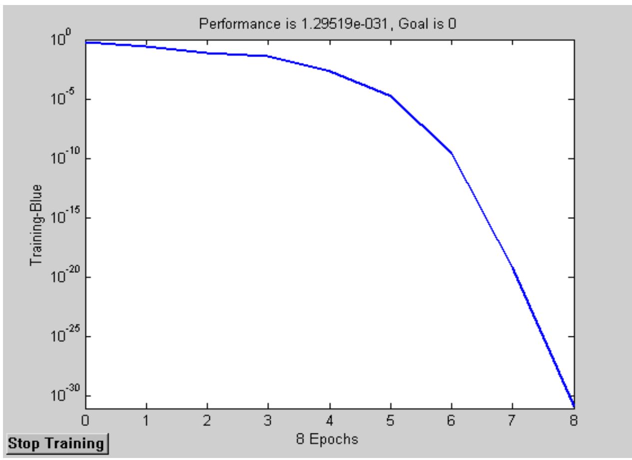

figure. 5.Training of MLP Feed forward Neural Network (2-4-2)

Performance of various training algorithms in 4-4-1 MLP Architecture is shown in table.4. LM method required less amount of training time to achieve the target than its counterpart. Therefore LM training algorithm was selected.

Table 4 Performance of Various Training Algorithms in 4-4-1 MLP Architecture

Sl.no	Training Algorithm	Training Epochs	Mean Square Error (MSE) Index
1	Levenberg-Marquardt (LM)	8	1.157 E-06
2	Gradient Descent Back Propagation (GD)	179	1.262 E-06
3	Gradient Descent with Momentum Back Propagation (GDM)	951	3.6E-06
4	Gradient Descent with Adaptive learning Rate Back Propagation (GDA)	200	2.53E-06

The results of the MLP back propagation neural models trained with the Levenberg-Marquardt (LM) learning algorithm are shown in table 5. The gain or learning rate η (0.3), momentum α (0.5), and training epochs are tabulated for each architecture [30].

Table 5 Estimation of MSE in Various MLP Network Architectures

Architecture	Training Epochs	Mean Square Error (MSE) Index		Confidence score
		Training	Testing	$C_i=\exp(-\lambda e_i^2)$
16-16-1	**38**	**0**	**7.31E-03**	**0.9927**
16-3-1	6	0	2.19E-02	0.9783
8-8-1	283	0	9.13E-03	0.9909
8-4-1	6	0	5.1E-02	0.9503
4-4-1	**9**	**0**	**2.83E-08**	**0.9999**
4-4-4	12	0	7.74E-03	0.992
2-2-2	3820	3.0E-08	3.7 E-08	0.9999
2-4-2	**8**	**0**	**0**	**1**
1-1-1	4538	1.08E-08	1.2E-08	0.9999

In the MLP networks testing procedures MSE index and number of epochs used for training are inversely proportional to each other. Therefore a compromise between them was achieved by taking into the consideration of larger training cost will ruin the system even though considerable accuracy is achieved in the targets (epilepsy risk levels). Therefore we had selected (16-16-1),(4-4-1) and (2-4-2) MLP network architectures due to their lesser number of training epochs.

IV. RESULTS AND DISCUSSION

To study the relative performance of these Fuzzy techniques and STD models (six types) and MLP(three models), we measure two parameters, the Performance Index and the Quality Value. These parameters are calculated for each set of twenty patients and compared.

A. Performance Index

A sample of Performance Index for a known epilepsy data set at average value is shown in table 6. It is evident that the STD optimization model (AR2) (16-4-2-1) with method II and MLP (2-4-2) charts a better performance than the fuzzy techniques because of its lower false alarms and missed classifications. Terminology is also important issue when we compare performance of methods. We submit that it is important to differentiate between the two terms of risk level prediction and risk level predictability.

Table 6. Performance Index

Methods	Perfect Classification	Missed Classification	False Alarm	Performance Index
Fuzzy Technique	50	20	10	40
SDT Method -I				
AR1	96.03	2.91	1.04	95.88
AR2	94.59	2.08	3.33	94.28
AR3	97.29	0	2.71	97.21
Method-II				
AR1	96.04	2.29	1.66	95.88
AR2	98.125	0.625	1.25	98.74
AR3	95.42	1.458	3.125	95.2
MLP Neural Networks				
16-16-1	96.24	0	3.536	96.32
4-4-1	99.37	0	0.624	99.37
2-4-2	100	-	-	100

The predictability is a necessary but not a sufficient condition for risk level prediction. Risk level predictability has to do with the sensitivity, whereas risk level prediction with both the sensitivity and specificity of a proposed and prospective methods. Hence, it is necessary to present the sensitivity and specificity of epilepsy risks levels classifier with fuzzy STD and MLP methods. These two precursors are defined as [34],

$$Sensitivity = \frac{PC}{PC + FA} \times 100 \quad (9)$$

$$Specificity = \frac{PC}{PC + MC} \times 100 \quad (10)$$

$$AverageDetection = \frac{Sensitivity + Specificity}{2} \times 100 \quad (11)$$

$$Relative - risk = \frac{Sensitivity}{Specificity} \quad (12)$$

The Average sensitivity and specificity parameters for twenty epilepsy patients in classification of epilepsy risk levels through fuzzy STD and MLP methods are shown in Table 7. It narrates that poor specificity leads to under performance and low sensitivity measures severe false alarms of the system. Cumulative detection level of classifier is represented by average detection. The relative risk indicates how far the classifier deviates from the threshold for an identified specificity.

Table 7. Results of Sensitivity, Specificity, Average

Detection and Relative risk of Classifiers for Average of all Twenty Patients.

Methods	Sensitivity	Specificity	Average Detection	Relative Risk
Fuzzy Technique	78.52	73.57	76.045	1.06
SDT Method -I				
AR1	98.95	97.28	98.115	1.017
AR2	97.85	97.8	97.825	1.001
AR3	97.29	100	98.645	0.9729
Method-II				
AR1	98.33	97.71	98.02	1.01
AR2	98.74	99.36	99.05	0.9937
AR3	96.82	98.49	97.655	0.983
MLP Neural Networks				
16-16-1	96.4	100	98.2	0.964
4-4-1	99.37	100	99.685	0.9937
2-4-2	100	100	100	1

It is noted from the table 7 that SDT AR2 (16-4-2-1) Method-II settled at the values of higher sensitivity, specificity, average detection and relative risk. The MLP neural network (2-4-2) attained the most ideal condition for all precursors of hybrid classifiers.

B. Quality Value

In Order to compare different classifier a measure that reflects the overall quality of the classifier was needed . The quality value was determined by three factors. Classification rate, Classification delay, and False Alarm rate. The quality value Q_V was defined as [21],

$$Q_V = \frac{C}{(R_{fa} + 0.2)*(T_{dly} * P_{dct} + 6 * P_{msd})} \qquad (13)$$

Where, C is the scaling constant, R_{fa} is the number of false alarm per set, T_{dly} is the average delay of the on set classification in seconds, P_{dct} is the percentage of perfect classification and P_{msd} is the percentage of perfect risk level missed. A constant C is empirically set to 10 because this scale is the value of Q_V to an easy reading range. The higher value of Q_V, the better the classifier among the different classifier, the classifier with the highest Q_V should be the best. Table 8 shows the Comparison of the fuzzy STD and MLP Neural Networks optimization techniques.

Table. 8 Results of Classifiers taken as Average of all Twenty Patients

Methods	Weighted delay (s)	False-alarm rate/set	Performance Index %	Quality value
Fuzzy logic	4	0.2	40	6.25
SDT Method -I				
AR1	2.095	1.04	95.88	22.68
AR2	2.0166	3.33	94.28	21.25
AR3	1.945	2.71	97.21	22.63
Method-II				
AR1	2.0582	1.66	95.88	22.43
AR2	2	1.25	98.1	23.52
AR3	1.908	3.125	95.2	22.65
MLP Neural Networks				
16-16-1	1.92	3.536	96.32	22.05
4-4-1	1.98	0.624	99.34	24.43
2-4-2	2	0	100	25

It was observed from table 8, that STD AR2 (16-4-2-1) Method II and MLP neural networks (2-4-2) are performing well with the higher performance index and quality values. As such SDT AR3 in Method I and Method II are empowered with high false alarm rate and also low weighted delay. This indicates the lower threshold value of the classifiers. On the other hand STD AR1 in Methods I&II and AR2 Method II are suffered by high missed classification and long weighted delays. Higher delay is the mark of high threshold value of the classifiers. Hence it was a compromise to select STD (AR2) method II. The MLP (2-4-2) neural network is a slow response method with inheriting weighted delay of 2 seconds. The other two MLP neural networks (16-16-1) and (4-4-1) are exhibiting quick responses with weighted delay of 1.92 seconds and 1.98 seconds respectively. These MLPs are placed with low performance indices and quality values.

V. CONCLUSION

In this paper, we consider generic classification of the epilepsy risk level of epileptic patients from EEG signals. The parameters derived from the EEG signal are complied as data sets. Then the fuzzy logic is used to the risk level from each epoch at every EEG channel. The target was to classify perfect risk levels with high rate of classification, a short delay from onset, and a low false alarm rate. Though it is impossible to obtain a perfect performance in all these conditions, some compromises have been made. As a high false alarm rate ruins the effectiveness of the system, a low false-alarm rate is most important. Since, the fuzzy outputs are highly nonlinear in nature with dynamic probability functions. STD based optimization technique to optimize the risk level by incorporating the above goals was chosen. Off the six types of STD methods AR2 method II performs well with high PI, Quality value and with moderate time delay. MLP (2-4-2) attained ideal conditions of the epilepsy risk classifications. Further research is in the direction to compare these hybrid models with Fuzzy Support Vector Machine (SVM) model to solve this open end problem.

ACKNOWLEDGEMENTS

The authors thank the Management and the Principal of Bannari Amman Institute of Technology, Sathyamangalam and the Management of Info Institute of Engineering Coimbatore for providing excellent computing facilities and encouragement. The authors are grateful to Dr.Asokan, Neurologist, Ramakrishna Hospital Coimbatore and Dr. B.Rajalakshmi, Diabetologist, Govt. Hospital Dindigul for providing the EEG signals.

This research is also funded by AICTE RPS.:FNo 8023/BOR/RID/Rps-41/2009-10 , dated 10th Dec 2010.

REFERENCES

[1] Li Gang etal, “An Artificial –Intelligence Approach to ECG Analysis,” IEEE EMB Magazine, vol 20, April 2000, pp 95-100.

[2] K. P.Adlassnig, “Fuzzy Set Theory in Medical diagnosis”, IEEE Transactions on Systems Man Cybernetics, vol,16 , March 1986, pp 260-265.

[3] Donna L Hudson, “Fuzzy logic in Medical Expert Systems”, IEEE EMB Magazine, vol, 13,no.6, November/December 1994, pp 693-698.

[4] R.Harikumar, Dr.(Mrs). R.Sukanesh, P.A. Bharathi, “Genetic Algorithm Optimization of Fuzzy outputs for Classification of Epilepsy Risk Levels from EEG signals,” Journal of Interdisciplinary panels I.E. (India),vol.86, no.1, May 2005,pp1-10.

[5] S.Watanabe, “Pattern Recognition as a quest for Minimum Entropy”, Pattern Recognition, vol 13, 1981, pp 381-387.

[6] R.P.Lippmann, “An Introduction to Computing With Neural Nets,” IEEE ASP Magazine, April 1989, pp 4-22.

[7] Guoqiang Peter Zhang,"Neural Networks for Classification: A Survey," IEEE Transactions on Systems, Man, And Cybernetics-Part C, vol,30,no.4, November 2000, pp 451-462.

[8] S.Gelfand and H.Guo, "Tree Classifier With Multilayer Perceptron Feature Extraction," M.S. Thesis Purdue University, 1990.

[9] Nurettin Acir etal., " Automatic Detection of Epileptiform Events In EEG by A Three-Stage Procedure Based on Artificial Neural Networks," IEEE Transactions on Bio Medical Engineering, vol,52,no.1,January 2005, pp30-40.

[10] Moreno.L.etal.,"Brain Maturation Estimation Using Neural Classifier," IEEE Transactions on Bio Medical Engineering , vol,42,no.2, April 1995, pp-428-432.

[11] Dr.(Mrs.).R.Sukanesh, R.Harikumar,"A Patient Specific Neural Networks (MLP) for Optimization of Fuzzy Outputs in Classification of Epilepsy Risk Levels from EEG Signals", Journal of Engineering Letters, vol 13, No.2, September 2006, pp 50-56.

[12] I.K.Sethi, "Entropy Nets: from Decision Trees to Neural networks," Proceedings of the IEEE vol 78, no 10 (Special Issue on Neural Networks), October 1990, pp 1605-1613.

[13] I.K.Sethi, "Layered Neural net Design through Decision Trees," In Proceedings of International Symposium on Circuits and Systems, New Orleans, LA, May 1-3, 1990.

[14] I.K.Sethi, and M.Otten, "Comparison Between Entropy net and Decision Tree Classifiers," In Proceedings of International Joint Conference on Neural Networks, (IJCNN) Sandiego, CA 1989

[15] Manuca, R.,Casdagil, M.C.Savit, "Non Stationarity in Epileptic EEG and Implications for Neural Dynamics," Mathematical Bio Sciences,147, 1998, pp 1-22.

[16] D.Zumsteg and H.G.Wieser,"Presurgical Evaluation: Current Role of Invasive EEG," Epilepsia,vol.41, no. suppl 3, 2000,pp, S55-60.

[17] Joel.J etal, "Detection of Seizure Precursors from Depth EEG Using a Sign Periodogram Transform," IEEE Transactions on Bio Medical Engineering, vol 51, no,4,April 2004, pp449-458,

[18] W.R.S.Webber, R.P.Lesser, R.T.Richardson and K.Wilson, "An Approach to Seizure Detection Using an Artificial Neural Network (ANN)",Electroenceph. Clin. Neurophysio., vol 98, 1996,pp 250-274.

[19] Arthur C Guyton, "Text Book of Medical Physiology", Prism Books Pvt. Ltd., Bangalore, 9th Edition, 1996.

[20] S. Balanco, H.Garcia, R.Quian Quiroja, L.Romanelli, and O.A.Rossa, "Stationarity of EEG Series," IEEE EMBS Magazine, July/August 1995, pp 395-399.

[21] Haoqu and Jean Gotman, "A Patient Specific Algorithm for Detection Onset in Long-Term EEG Monitoring-Possible Use as Warning Device", IEEE Transactions on Biomedical Engineering, vol, 44, no.2, February 1997, pp 115-122.

[22] J. Seunghan Park et al, "TDAT Domain Analysis Tool for EEG Analysis", IEEE Transactions on Biomedical Engineering ,vol, 37,no.8,August 1990, pp 803-811.

[23] Alison A Dingle et al, "A Multistage System to Detect Epileptic Form Activity in the EEG", IEEE_Transactions on Biomedical Engineering ,vol, 40, no.12, December 1993, pp 1260-1268.

[24] Pamela McCauley-Bell and Adedeji B.Badiru, " Fuzzy Modeling and Analytic Hierarchy Processing to Quantify Risk levels Associated with Occupational Injuries- Part I: The Development of Fuzzy- Linguistic Risk Levels," IEEE Transactions on Fuzzy Systems, vol,4,no.2, 1996, pp 124-31.

[25] Dr .(Mrs.).R.Sukanesh, R.Harikumar, " A Comparison of Elman and MLP Feed Forward Neural Networks for Classification of Epilepsy Risk Level Using EEG Signals,"AMSE Journal ,Modeling C, vol.67 no.1, September 2006, pp 43-60.

[26] Dr .(Mrs.).R.Sukanesh, R.Harikumar, "Diagnosis and Classification of Epilepsy Risk Levels From EEG Signals Using Fuzzy Aggregation Techniques," Journal of Engineering Letters, vol 14, no.1, March 2007, pp 90-95.

[27] C B Gupta and Vijay Gupta, "An Introduction to Statistical Methods", 22nd Ed., Vikas Publishing House Lt., 2001

[28] L.Hyafil and Rivest, "Constructing Optimal Binary Decision Trees is NP- Complete", Information Processing Letters, vol 5, no, 1 1976, pp15-17.

[29] Ronald.R.Yager, "Hierarchical Aggregation Functions Generated From Belief Structures," IEEE Transactions on Fuzzy Systems,vol.8, no.5, May 2005,pp 481-490.

[30] K.Srirama murty and B.Yegnannarayana, "Combining Evidence from Residual Phase and MFCC Features for Speaker Recognition", IEEE Signal Processing Letters, vol,13,no.1, 2006, pp 52-55.

[31] Drazen. S.etal., "Estimation of Difficult –to- Measure Process Variables Using Neural Networks", Proceedings of IEEE MELECON 2004,May 12-15,2004,Dubrovnik, Croatia, pp 387-390.

[32] Tarassenko.L, Y.U.Khan, M.R.G.Holt, "Identification of Inter-Ictal Spikes in The EEG Using Neural Network Analysis," IEE Proceedings –Science Measurement Technology, vol, 145, no.6,November 1998, pp270-278.

[33] H.Demuth and M..Beale,"Neural Network Tool Box: User's guide, Version 3.0," the Math works, Inc., Natick, MA, 1998.

[34] Dr.(Mrs.).R.Sukanesh, R.Harikumar,"Fuzzy Techniques and Statistical Tests for Epilepsy Risk Level Classification Using EEG Signals," Journal of Interdisciplinary panels I.E(.India)vol.86 ,no.1, May 2005, pp 29-36.

Dr. R. Harikumar received his B.E (ECE) Degree from REC (NIT) Trichy1988..He obtained his M.E (Applied Electronics) degree from College of Engineering, Guindy, Anna University Chennai in 1990. He was awarded Ph.D. in I &C Engg from Anna university Chennai in 2009. He has 22 years of teaching experience at college level. He worked as faculty at Department of ECE , PSNA College of Engineering & Technology, Dindigul. He was Assistant Professor in IT at PSG College of Technology, Coimbatore. He also worked as Assistant Professor in ECE at Amrita Institute of Technology, Coimbatore. Currently he is working as Professor ECE at Bannari Amman Institute of Technology, Sathyamangalam. He is guiding eight PhD theses in these areas. He has published Forty papers in International and National Journals and also published around seventy six papers in International and National Conferences conducted both in India and abroad. His area of interest is Bio signal Processing, Soft computing, VLSI Design and Communication Engineering. He is life member of IETE, IEEE and ISTE.

M.Balasubramani received his B.E (ECE) degree from BIT Sathy, Anna university Chennai in 2008.He obtained his M.E (VLSI DESIGN) degree from BIT Sathy ,Anna University Coimbatore in 2010. He has one year five months of teaching experience at college level. He is pursuing his doctoral programme as a research scholar in Bio Medical Engineering under guidance of **Dr. R.Harikumar** at Department of ECE, Bannari Amman Institute of Technology Sathyamangalam. Currently he is working as Assistant Professor in Department of ECE at Info Institute Engineering, Coimbatore. His area of interest is Bio signal Processing, Soft computing, Embedded systems and Communication Engineering. He is member of ISTE.

T.Vijayakumar received his B.E (ECE) degree from Bharathiyar University in 2003 .He obtained his M.E (Computer Communication)degree from Anna university Chennai in 2007. He has five years of teaching experience at college level. He is pursuing his doctoral programme as a research scholar in Bio-Medical Engineering under guidance of **Dr. R.Harikumar** at Department of ECE, Bannari Amman Institute of Technology Sathyamangalam. Currently he is working as Assistant Professor in Department of IT at Bannari Amman Institute of Technology, Sathyamangalam. His area of interest is Bio signal Processing, Soft computing, Embedded systems and Communication Engineering. He is Member of ISTE.