International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 5 (2013) pp. 260-268

Comprehensive Analysis of Hierarchical Aggregation Functions Decision Trees, SVD, K-means Clustering, PCA and Rule Based AI Optimization in the Classification of Fuzzy based Epilepsy Risk Levels from EEG Signals

R.Harikumar [1], T.Vijayakumar[2]

[1]Bannari Amman Institute of Technology,
Sathyamangalam, India.
e-mail: harikumarrajaguru@gmail.com

[2]Bannari Amman Institute of Technology,
Sathyamangalam, India.
e-mail: vishal_16278@yahoo.co.in

***Abstract*— A comprehensive analysis for the performance of post classifiers such as Hierarchical Soft Decision Trees, Singular value decomposition(SVD), k-means clustering, Principal Component Analysis (PCA) and Rule based AI techniques in optimization of fuzzy outputs for the classification of epilepsy risk levels from EEG (Electroencephalogram) signals is presented in this paper. The fuzzy pre classifier 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. Hierarchical Soft decision tree (post classifiers with max-min criteria) four types, SVD, K-means clustering, PCA and AI optimization are applied on the classified data to identify the optimized risk level (singleton) which characterizes the patient's risk level. The efficacy of the above methods is compared and analyzed based on the bench mark parameters such as Performance Index (PI), and Quality Value (QV).**

Keywords: EEG Signals, Epilepsy Risk Levels, Fuzzy Logic, Hierarchical Decision Trees, SVD, K-means clustering, PCA,AI Techniques

I. Introduction

Epileptic seizures are a principal brain dysfunction with important public health implications, as they affect 0.8% of humans [1]. Electroencephalograms (EEGs) are recordings of electrical potentials produced by the brain. Analysis of EEG activity has been achieved principally in clinical settings to identify pathologies and epilepsies since Hans Berger's recording of rhythmic electrical activity from the human scalp [2]. In the past, interpretation of the EEG was limited to visual inspection by neurophysiologist, an individual trained to qualitatively make a distinction between normal EEG activity and abnormalities contained within EEG records. Epilepsy is a chronic disease characterized from recurrent seizures that cause sudden but revertible changes in the brain functions [3]. The Classification of epilepsy risk levels, according to international standard is difficult because individual laboratory findings and symptoms are often inconclusive [4]. Approximately 1% of the people in the world suffer from epilepsy. The electroencephalogram (EEG) signal is used for the purpose of epileptic detection as it is a condition related to the brain's activity. A common form of EEG recording used for this purpose is an ambulatory recording which contains EEG data for a long duration of even up to a week [5]. It involves an expert's effort in analyzing the entire data to detect traces of epilepsy. The traditional methods of analysis being tedious and time consuming, many automated epileptic EEG systems have emerged in recent years [3]. K.P.Adlassnig (1986) characterized the epilepsy disorder as sudden recurrent and transient disturbances of mental function and/or movements of body that results in excessive discharge group of brain cells [2].The presence of Epileptiform activity in the EEG confirms the diagnosis of epilepsy, which sometimes confused with other disorders producing similar seizure like activity. Between seizures, the EEG of a patient with epilepsy may be characterized by occasional epileptic form transients-spikes and sharp waves.

A. General Techniques

Today, in the mass storage era, knowledge acquisition represents a major knowledge engineering bottle neck.

Computer programs extracting knowledge from data successfully attempt to alleviate this problem. Among such systems, inducing symbolic decision trees for decision making, or classification, are very popular [6]. The resulting knowledge, in the form of decision trees and inference procedures, has been praised for comprehensibility. This appeal to a wide range of users who are interested in domain understanding, classification capabilities, or symbolic rules that may extract from the tree and subsequently used in rule based decision system. Decision trees were popularized by Quinlan with the ID3 program as identified by Cezary.Z (1998) [7]. Perhaps, the most important feature of decision tree is it's capability to break a complex decision making process into a collection of simpler decisions, thus providing a solution which is easier to interpret [6]. Alison (1993) noticed that the different types of epileptic seizures are characterized by different EEG waveform patterns [8]. 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. One of them is a classification of risk level of epilepsy by using Fuzzy Techniques examined in ref [1].

Many of these patients (20%) are resistant to treatment with drugs. The ability to anticipate the onset of seizures in such cases would permit clinical interventions [9]. Traditional signal analyses, such as the count of focal spike density, the frequency coherence or spectral analyses are not reliable predictors [10]. Many decisions are based on the determination of available alternatives enduring the relevant criteria. In these types of problems, the measurement of the satisfaction to the individual criteria is available [8]. The constructions of overall decision functions are complicated [11]. First, the construction of decision function requires a specification from the responsible decision maker of the relationship between the criteria for aggregation [12]. Once this specification of relationship and the criteria are obtained, the analyst is then facing with the problem of rendering this information into a form that can be evaluated in terms of the satisfaction to the individual criteria, which leads to the formulation of associated Multi Criteria Aggregation function [13]. This situation puts a premium of knowledge representation structures that allow for both a specification aggregation functions [14]. Based on the theory of fuzzy measures and the OWA operators, we introduce a hierarchical structure that allows for the construction of decision functions, which meets the above mentioned needs [15]. This paper addresses the application of hierarchical structured decision trees, SVD, K-means clustering, PCA and AI Techniques towards optimization of fuzzy outputs in the classification of epilepsy risk levels. We also present a comparison of these classifiers based on their Performance Indices and Quality values.

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. 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, we had selected artifact free EEG records with distinct features. These records were scanned by Umax 6696 scanner with a resolution of 600dpi.

A. Acquisition of EEG Data

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 [16]. 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.

B. Fuzzy System as a Pre Classifier

Fig 1 enumerates the overall epilepsy risk level (Fuzzy-Post optimization) classifier system. The motto of this research is to classify the epilepsy risk level of a patient from EEG signal parameters. This is accomplished as [17], Fuzzy classification for epilepsy risk level at each channel from EEG signals and its parameters. Fuzzy classifier results from each channel are optimized using four types of HDT, SVD, K-means clustering, PCA, and hand rule optimization methods. A comprehensive performance of fuzzy classifier and post classifiers optimization methods are analyzed.

The following parameters are extracted From EEG signals.

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

$$E = \sum_{i=1}^{n} x_i^2 \tag{1}$$

Where xi 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].

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

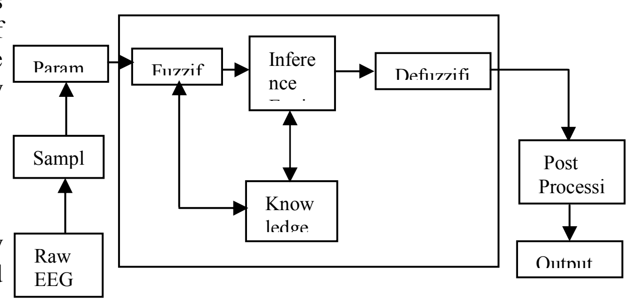


Figure 1. Fuzzy and Post Processing Classification System

5. The variance is computed as σ given by

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

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

6 .The average duration is given by

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

Where ti 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

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

C. Fuzzy Membership functions

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* [18]. 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* [19].

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.

D. 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 [15]. 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 [18]. 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 I.

Table I Representation of Risk Level Classifications

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

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

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%.

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

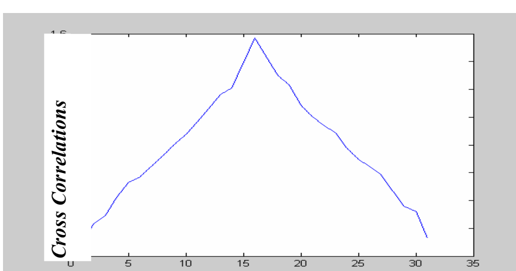
Let the fuzzy outputs as shown in Fig 2 is coded with appropriate numerical values. These numerical values are associated with the probability of each coded epilepsy risk level patterns. The five risk levels are encoded as Z>Y>X>W>U in binary strings of length five bits using weighted positional representation as shown in Table II. 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, [1]. Now the each input patterns are encoded in the numerical form of the range 0-1. The nonlinearities associated with fuzzy outputs in describing the epilepsy risk levels were identified by cross correlation. Thus the cross correlation function $r_{xy}(m)$ of the epochs x(n) and y(n) is defined 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,[17] The cross correlation r_{xy} (m) plot obtained through the equation (6) is shown in the "Fig.3",which emulates the occurrence of highly non periodic patterns in the fuzzy outputs. Therefore any closed solution will be failed for this purpose of optimization. Hence, it is advisable to prefer non linear techniques instead of linear one, such a one type is

HDT. Since, HDT is a common way to solve a wide variety of ill-posed problems which is not necessarily treated as hard constraint one.

Table II. Binary Representation of Risk

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
Normal	U	00001	1/31= 0.03225	0.005376
		11111 = 31	Σ=1	

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



Epochs

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

III. Hierarchical Decision Trees for Optimization of Fuzzy Outputs

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. Hierarchical functions are non linear mapping from $(x_1, x_2, x_3, \dots x_n) \in R^n$ to $y \in R$ and this nonlinear mapping is general enough to approximate any non linear function with arbitrary accuracy. In contrast to conventional single stage classifiers where each data sample is tested against all classes, thereby reducing efficiency, in a decision tree a sample is tested against only certain subsets of classes, therefore unnecessary computations are eliminated[7],[13]. The main objectives of HDT 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.

A .Hierarchical Formulation

Let us review the hierarchical formulation in the R. Yager's perceptive. Again assume we have a set N={N1,…Nn} of directly measurable criteria, that is for each alternative x we can obtained $N_i(x)$, satisfaction of x to N_i. Now we describe the situation which inspires further generalization of our approach. Assume that in choosing an alternative we have two objectives or goals. Goal one, which has an incremental value of β_1= 0.6can be meet with the satisfaction of N1, N2, N3. Goal two which has an incremental value of β_2= 0.4 can be meet with the satisfaction of all N4 to N16 including max-min decisions. In order to model decision imperatives, we shall identify two types of aggregation, weighted average and OWA aggregation [15]. Let V be the q vector with components βi i= 1 to q, lying in the unit interval and summing to one. Then we denote $E_V(y_1,\dots,y_q)$

$$E_V\left(y_1,\dots y_q\right)=\sum_{i=1}^{q}\beta_i y_i = V^T Y \quad (7)$$

Let A be a p dimensional vector with components a_j, j=1to p, that also lies in the unit interval and sum to one. Here we shall denote $F_a(y_1,\dots,y_p)$ to be OWA average of the arguments

$$F_a\left(y_1,\dots y_p\right)=\sum_{j=1}^{p}a_j b_j = A^T B \quad (8)$$

Where b_j is the jth largest of the y_i. Using these two structures we can express the decision function needed to solve the preceding situation [20]. Let D(x) be the overall alternative x letting $N_i(x)=n_i$. We get $D(x) = G(n_1,\dots,n_n)$

$D(x) = E_{V1}(F_{a6}(n_1, n_2, n_3), \max(F_{a5}(n_4, n_5), \min(F_{a4}(n_6, n_7, n_8), \max(F_{a3}(n_9, n_{10}, n_{11}),\min(F_{a2}(n_{12}, n_3), F_{a1}(n_{14}, n_{15}, n_{16}))))))$. This formulation can be viewed as hierarchical structure [21]. In our approach we consider a decision frame work in which we have a collection, N={N1,N2,N3,…,N16}of primary attributes.. These first level concepts are decomposed into other concepts or primary attributes. We continue until we end up with all primary attributes.

B. Algorithm for HDT Optimization

The generic representation of HDT 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 and 48 (16x3) patterns are available. Now the optimization is a two stage process through HDT, which is explained as below,

1. Deduce the 16x3 matrix epilepsy risk level into 16x1 viz row wise optimization through two types of optimization viz, a) Hierarchical method of two level, and b) Maximum pattern in the particular row.
2. Deduce the 16x1 matrix into one optimum epilepsy risk level through HDT optimization with five levels.

Here also we have two decision methods at node level which are Max-min &Min-max combination. Therefore effectively we have four methods of HDT post classifier.

Stage I

1. The Hierarchical method converts the three column elements of i,j element into a single row element as N_{11}= Max (N_{I1}, N_{I2}) & N_I=Min (N_{11},N_{I3}) which is also depicted in the Fig..4.And the other method is self explained in nature. Now the row of three elements is converted into single element. This is repeated for all the 16 rows and the matrix is reduced into 16x1 matrixes.

Stage II: Group (16x1) elements as the leaf nodes of the tree N1 to N16. These leafs are aggregated by the rectangular nodes named as A1 to A6. This structure is a mixed averaging hierarchical Decision tree which is depicted in Fig 5, we use rectangular box to indicate a weighted average aggregation and a circle used to indicate decision of MAX or MIN. The term inside the symbol indicates the associated vector. The outputs of A nodes are hierarchically combined by the circular B, Soft decision nodes of B1 to B4. The single node V1 (RECTANGULAR) is the root of the tree. In the case of Hierarchical method followed by hierarchical Max-min method, let N1, N2 … N16 leaf nodes are available.

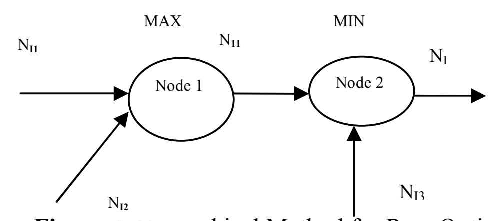


Figure 4. Hierarchical Method for Row Optimization

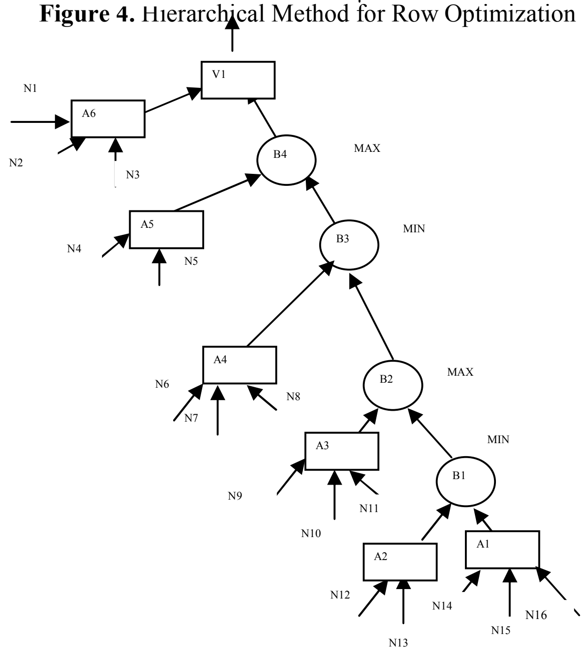


Figure 5. Optimization of Epilepsy Risk Levels through HTD (Max-min) Method

THE FINAL V1= $\begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$

The aggregate weights of A nodes are as,

For A1, A3, A4, &A6 = $\begin{bmatrix} 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}$ and A2 &A5= $\begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$;

And circular nodes B1=Min(A1,A2);B2=Max(A3,B1); B3=Min(A4,B2); B4=Max(A5,B3)In the case of Min –Max procedure the following decisions are taken at the nodes of Bi for i=1 to 4, when i=odd MAX & i=even Min and also at V1= 0.4(A6) +0.6(B4).The obtained singleton results are immensely helpful in devising the therapeutic procedure of the epileptic patients. Results from the four types of optimization methods are discussed in the next section.

IV. Singular Value Decomposition for Optimization of Fuzzy Outputs

The Singular Value Decomposition (SVD) is a well known approach that may be used for such tasks as dimensionality reduction, and determining the modes of a complex linear dynamical system [22]. SVD of a matrix has one or more columns that are identical, or that several groups of columns that are same which is useful in signal processing problems and applications. A SVD of an m × n matrix A=[a_1, a_2, a_3,…, a_n] is the composition of A into the product of three matrices as follows

$$A = U \sum V^T = \sum P \sigma_K \; u_K v_K^T \tag{9}$$

where p=min(m,n), U=[u_1, u_2, u_3,…, u_m] is an m × n ortho normal matrix, V=[v_1, v_2, v_3,…, v_n] is an n× m ortho normal matrix, and Σis an m × n matrix with elements σ_k along the diagonal and zeros everywhere else. Matrix U is called left singular matrix, V is called right singular matrix, and Σis the singular value matrix [23]. If the singular values are ordered so that $\sigma_1 \geq \sigma_{2,\ldots} \geq \sigma_p$, and if the matrix A has a rank r<p, then the last p-r singular values are equal to zero, and SVD becomes A=$\Sigma^{r} \sigma_k u_k v_k^{T}$

SVD procedure takes vectors in one space and transforms them into another space. Advantages of using SVD are to combine two different uncertainty representations into a metric as total uncertainty. SVD decomposes uncertainty measures (possibility, belief, probability etc.,), combined as a collection of vectors of different units, into a principle space. We need this feature since our uncertainty measures cannot be added directly, they contain different units (epilepsy risk level codes). SVD has been applied successfully in many other technical disciplines as a tool to reduce coupled non linear behavior to uncoupled collections of linear behavior [24]. The fuzzy outputs are (16x3 matrix) considered as matrix A and SVD is taken for that matrix. The highest Eigen value is considered as the pattern of the known patient's epilepsy risk level. A group of twenty patients are analyzed in this study. The obtained singleton results are discussed in the following part of the paper.

V. K-Means Clustering Algorithm

During last decades, growing attention has been put on data clustering as robust technique in data analysis. Clustering or data grouping describes important technique of unsupervised classification that arranges pattern data (most often vectors in multidimensional space) in the clusters (or groups). Patterns or vectors in the same cluster are similar according to predefined criteria, in contrast to distinct patterns from different clusters [25], [26].K-means clustering algorithm proposed by Mac Queen in 1967 belongs to partitioning methods, which is widely used because of its simplicity and fast convergence. The primary process can be expressed as follows [27], [28].

1. Initialize K cluster centers chosen randomly.
2. Assign each x_i to its nearest cluster center c_k by Euclidean Distance (d).

$$KM(X,C) = \sum_{i=1}^{n} \min_{j\in\{1....k\}} \left\| x_i - c_j \right\|^2 \qquad (10)$$

$$d(p,q) = d(q,p) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \qquad (11)$$

3. Update each cluster center c_k as the mean of all x_i that belongs to it.
4. Repeat steps 2-4 until the cluster centers are stable.

K-means clustering is associated with computational expensiveness with respect to the cluster similarity distance measures. The results obtained through the K-Means clustering methods are discussed in the following section of the paper.

VI. Principal Components for Optimization of Fuzzy Outputs

All suboptimal transforms such as the DFT and DCT decomposes the signals into a set of coefficients, which do not necessarily represent in the constituent components of the signals. Moreover, as the transform kernel is independent of the data it is not efficient in terms of both de correlation of the samples and energy compaction. Therefore, separation of the signal and noise components is generally not available using these suboptimal transforms. Expansion of the data into a set of orthogonal components certainly achieves maximum de correlation of signals [29]. This can enable separation of the data into signal and noise subspace. PCA is widely used in data decomposition, classification, filtering, and whitening. Performing a Principal Component Analysis (PCA) is equivalent to performing an SVD on the covariance matrix. PCA uses the same concept as SVD and orthogonalization to decompose the data into constituent uncorrelated orthogonal components such that the auto correlation matrix is diagonalized [29]. Each Eigen vector represents a principle component and the individual eigen values are numerically related to the variance they capture in the direction of principle components. Principal Component Analysis (PCA) is a dimensionality –reduction technique that has been applied to many kinds of data. In fact, PCA is the optimal such linear transform-that is, for any choice for the number of dimensions, PCA returns the subspace that retains the highest variance [30]. In this section, we describe how to use PCA to optimize the code converters outputs.

PCA is a mathematical technique allows reducing the complex system of correlations in a smaller number of dimensions. χ being a table of P numeric variables (in columns) describing N individuals (in lines), w propose to seek a representation of N individuals (signals) e1, e2,…en in a subspace of initial space. In other words, we have to define K new variables, combination of P of initial space, which could make loss less possible information. These K variables will be called principal axes [31].

For N observations, we will have a matrix of NxP size which is given by

$$e = [e_1 \quad e_2 \quad e_3 \quad .. \quad e_n] \qquad (12)$$

The average signal is defined by:

$$\psi = \frac{1}{N}\sum_{m=1}^{N} e_m \qquad (13)$$

For each element the difference:

$$\delta_i = e_i - \psi \qquad (14)$$

The computation of covariance matrix is:

$$C = \frac{1}{N}\sum_{m=1}^{n} \delta_m \delta_m^T = \frac{1}{N} A \times A^T \qquad (15)$$

With: $A = [\{\delta_1 \quad \delta_2 \quad ... \quad \delta_N]$ (16)

However, the determination of the Eigen vectors of covariance matrix will require an excessive calculation [32]; the size of this matrix is (PxP). If υ_i is the Eigen vector of AXA^T its Eigen values are:

$A^T A \upsilon_i = \mu\, \upsilon_i$

Then the Eigen vectors of C are calculated by:

$U_i = A\, \upsilon_i$ (17)

Finally the principal component of each signal e_i is given by:

$$w_k = u_k^T \times (e_i - \delta) \qquad (18)$$

The vector w_k represents the new parameters completely de correlated and optimized for classification. The results obtained through the PCA methods are discussed in the following section of the paper.

VII. Artificial Intelligence (A.I) Optimization

The AI optimization was a post classifier to classify risk levels of epilepsy based on a few simple rules. The proposed technique for optimizing the classified risk level of epilepsy is based on a few simple rules. The hand-rule optimization is a search process in which a maximized risk level output is obtained from a cluster of non-optimized fuzzy system outputs. In maximizing the risk levels by this optimization technique,

1) A simple weight age is given to the representations as follows, Z > Y > X > W > U
2) The variance is taken as a second parameter for optimization.
3) The repeated patterns with high risk levels are identified.
4) The occurrences of low risk level patterns are more than the high risked one then the weight age rule (1) is ignored. This process is carried out in three stages: Target stage,

Intermediate stage, and Output stage.

The target stage is the output epilepsy risk level obtained from the fuzzy data. In the target stage, the numerical values of the fuzzy logic outputs are converted into the corresponding encoded representations. Once each epoch is encoded as a string of 6 characters, each channel is divided into sets of three strings each. The intermediate stage is where each of the sets is optimized independently. A set of sixteen strings is obtained from the target stage by following row-wise optimization for every channel in each epoch. As an example, for the state WYYWYY, WYYWYY, WZYYWW the row wise optimized output is WYYXYX, which is obtained by using simple rules as mentioned above. This set of sixteen outputs is again divided into four sets of four strings each. A column-wise optimization procedure is followed wherein each of the set is optimized to a single string. The four strings obtained in this column-wise optimization are again grouped in sets of two and the output stage of operation is carried out. In the output stage, each group of two strings give one optimized string and another step in combining the two outputs yields the final optimized risk level of the epileptic patient.

VIII. Results and Discussion

To study the relative performance of the Fuzzy techniques, HTD systems (4 Types), SVD, K-means clustering, PCA, and AI optimization, 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 III. It is evident that the HTD Optimization with [MAX & h Max-min] method as well as [MAX & h Min-max] method gives a better performance than the AI optimization, fuzzy techniques and other two hierarchical techniques because of its lower missed classifications. SVD and PCA optimization methods are settled at PI of 95.88% and 95.12% respectively with low missed classification. K-means clustering method is pegged at low PI of 92.122% due to high missed classification of epilepsy risk levels.

B. Quality Value

In Order to compare different classifier we need a measure that reflects the overall quality of the classifier [13].Their quality is determined by three factors namely Classification rate, Classification delay, and False Alarm rate.

The Quality Value QV is defined by,

$$Q_V = \frac{C}{(R_{fa} + 0.2)*(T_{dly} * P_{dct} + 6 * P_{msd})} \qquad (19)$$

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 III Performance Index

Methods	**Perfect Classi fication**	**Missed Classi fication**	**False Alarm**	**Perfor mance Index**
Fuzzy logic	50	20	10	40
hier &h max-min	95.42	3.33	1.25	95.2
hier & h min-max	95.63	4.16	0.208	95.43
Max &hmax-min	96.84	0.416	2.17	96.77
Max& hmin-max	97.5	0.416	2.08	97.44
SVD Method	96.04	1.04	2.92	95.88
K-means clusters	92.79	4.33	2.875	92.122
PCA	**95.8**	**1.87**	**2.25**	**95.12**
AI optimization	80	5	10	81.25

Table IV. Results of Classifiers Taken As Average of All Ten Patients

Methods	**Weighte d delay (s)**	**False-al arm rate/set**	**Perfor mance Index %**	**Qualit y value**
Fuzzy logic	4	0.2	40	6.25
hier&hmax-min	2.108	1.25	95.2	22.3
hier&hmin-max	2.1662	0.208	95.43	20.9
Max&hmax-min	1.962	2.71	96.77	22.4
Max&h min-max	1.975	2.08	97.44	22.9
SVD Method	1.9832	2.92	95.88	21.99
K-means clusters	2.1155	2.875	92.17	21.927
PCA	2.09	2.25	95.12	22.93
AI optimization	2.8	0.1	81.25	11.9

Table IV shows the Comparison of the fuzzy, HTD SVD, K-means clustering, PCA, and AI optimization techniques. It is observed from table IV, that HTD (Max& h max-min) (Max& h min-max), and PCA methods are performing well with the higher performance index and quality values. As such maximum pattern followed by decision trees are empowered with high false alarm rate and also low weighted delay. This indicates the lower threshold value of the classifiers. SVD and K-means clustering techniques are sustained with Quality value of 21.99 and 21.927 respectively.

IX. Conclusion

In this paper, we consider generic classification of the epilepsy risk level of epileptic patients from EEG signals and investigated the performance of six post classifiers in optimizing 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. HTD SVD, K-means clustering, PCA, and AI optimization techniques were chosen to optimize the risk level by incorporating the low false alarm and near nil missed classifications. HTD (max & hmin-max) has better performance index whereas HTD and PCA performs better than AI optimization techniques and Fuzzy Techniques with high Quality value and with moderate time delay. The performance of post classifiers such as SVD and K-means clustering methods are at the midway between the AI and PCA techniques. From this method we can infer the occurrence of High-risk level frequency and the possible medication to the patients.

The major limitation of this method is that if one channel has a high-risk level, then the entire group will be maximized to that risk level. This will affect the non-epilepsy spike region in the groups. A comparison of EM and SVM will be taken for further studies.

ACKNOWLEDGEMENTS

The authors express their sincere thanks to the Management and the Principal of Bannari Amman Institute of Technology, Sathyamangalam for providing the necessary facilities for the completion of this paper. **This research is also funded by AICTE RPS.:F No 8023/BOR/RID/Rps-41/2009-10 , dated 10th Dec 2010.**

References

[1] 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)*, 86 (1), pp.1-10, May 2005

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

[3] D.Zumsteg and H.G.Wieser, "*Presurgical evaluation: current role of invasive EEG*,"Epilepsia,.41(3),pp.S55-60,2000.

[4] Celement.C etal, "A Comparison of Algorithms for Detection of Spikes in the Electroencephalogram", *IEEE Transaction on Bio Medical Engineering,* , 50 (4),pp. 521-26, April 2003.

[5] W.De Clereq etal.,"Characterization of interictal and ictal scalp EEG signals with the Hilbert Transform", *Proc of the 25th Annual International conference of the IEEE EMBS Cancun, Mexico, September 17-21*, pp-2459-2462,2003.

[6] S.Rasoul Safavian, David Landgrebe, "A Survey of Decision Tree Classifier Methodology," *I.E.E.E Trans SMC*,.21(3), pp.660-674,May 1991.

[7] Cezary.Z Janikow, "Fuzzy Decision Tree: Issues and Methods," *I.E.E.E Trans SMC*,.28(1), pp. 1-14,February1998.

[8] Alison A Dingle et al, "A Multistage system to Detect epileptic form activity in the EEG",*IEEE Transactions on Biomedical Engineering*, , 40(12),pp.1260-1268, 1993.

[9] Mark van Gils, "Signal processing in prolonged EEG recordings during intensive care", *IEEE EMB Magazine* ,16(6) pp.56-63, 1997,

[10] L.D.Iasemidis, "Epileptic Seizure Prediction and Control," *IEEE Transactions on Biomedical Engineering,*50 (5), pp. 549-558 , May 2003.

[11] R.R.Yager,"Including Importance's in OWA Aggregations Using Fuzzy Systems Modeling", *IEEE Trans. on Fuzzy Systems*,6(2),pp. 286-294,1996.

[12] Gleb Beliakov and Jim Warren, "Appropriate choice of Aggregation Operators in Fuzzy Decision Systems," *IEEE Transactions on Fuzzy Systems*,9 (6), pp. 773-784, December 2001.

[13] Harikumar.R Vijayakumar.T, "Comparison of Hierarchical Aggregation Functions Decision Trees and Rule Based AI Optimization in the Classification of Fuzzy based Epilepsy Risk Levels from EEG Signals,"*Proc.11th International Conference on Hybrid Intelligent Systems (HIS 2011) Malacca, Malaysia during December 05-08, 2011,*pp. 61-66,2011.

[14] P.Filev and R.R.Yager, "Context Dependent Information Aggregation", *Proceedings of IEEE International Conference on Fuzzy Systems*, pp. 672-677, 2003

[15] Ronald.R.Yager, "Hierarchical Aggregation Functions Generated From Belief structures," *I.E.EE Transaction of Fuzzy Systems*,8(5), pp. 481-490, October 2000.

[16] Hao Qu 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,* 44(2),pp.115-122 ,February 1997.

[17] R.Harikumar and B.Sabarish Narayanan, "Fuzzy Techniques for Classification of Epilepsy risk level from EEG Signals", *Proceedings of IEEE Tencon – 2003, Bangalore, India*, pp. 209-213, 14-17 October 2003.

[18] Donna L Hudson, "Fuzzy logic in Medical Expert Systems", *IEEE EMB Magazine*,13(.6),pp. 693-698, November/December 1994.

[19] 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 Transaction on Fuzzy Systems,,4(2), pp.124-31,1996.

[20] R.R.Yager, "On role of Anxiety in Decisions Under Possibilistic Uncertainty," *IEEE Trans on SMC*, v 34 (2,) pp.1224-1234, April 2004.

[21] Tomasa Calvo, Radko Mesiar, and R.R.Yager, "Quantitative Weights and Aggregation," *IEEE Transactions on Fuzzy Systems*, 12(1), pp. 62-69, February 2004.

[22] Sankeun lee, Monson, "Properties of the Singular Value Decomposition for Efficient Data Clustering," *IEEE Signal Processing Letters,*11(11), pp .862-866, November 2004.

[23] P.K.Sadasivam, D.Narayana Dutt, "SVD Based Technique for noise Reduction in Electroencephalographic signals", *Elsevier Signal Processing*, 55 ,pp. 179-189, 1996.

[24] Mu-chun Su, Chien –Hsing Chou, "A modified version of the k-means clustering algorithm with a distance based on cluster symmetry", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (6), pp. 674-680, June 2001.

[25] Mc. Culloch, C.E, " Maximum likelihood algorithm for generalized linear mixed models,". *Journal of the American Statistical Association*, 92, pp.162-170, 1997.

[26] Stephen J Redmond and Conor Heneghan, "A method for initializing the k-means clustering algorithm using kD-trees", *Pattern Recognition*, 28, pp.965-973,2007.

[27] R. Xu, D. Wunsch II, "Survey of clustering algorithms", *IEEE Transcations on Neural Networks*, 16, pp. 645-678, 2005.

[28] Jain, M. Murty, P. Flynn, "Data Clustering: A Review" *ACM Computing Surveys*, 31 , pp.264-323. 1999.

[29] E.Bair,T. Hastie, D.Paul,and R.Tibshirani," Prediction by supervised Principal Components," *Journal of the American Statistical Association* 101,(473), pp. 119-137,2006.

[30] M. Kheli, m. Boundraa, A. Kechida, and R.Drai, " Classification of Defects by the SVM Method and the Principal Component Analysis(PCA)", *World academy of science, Engineering and Technology,* 9, pp 226-231,2005.

[31] A.Sophian,G.Tian, D.Taylor and J.Rudin, "A Feature Extraction Technique Based on Principal Component Analysis for Pulsed Eddy Current NDT", *NDT&E International Journal,*36,pp. 37-41,2003.

[32] Alexandros. T. tzallas, Markos.G. Tsipouras and Dimitrios. I.Fotiadis, "A Time- Frequency Based Method for the Detection of Epileptic seizure in EEG Recording,*" Proc.12th IEEE International Symposium on Computer based Medical Systems (CBMS '07),* pp 23-27,2007.

Author Biographies

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, and 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.

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.