A Novel Hybrid Method of Dimensionality Reduction with a Post Classifier for Assessing Epilepsy Risk Levels from EEG Signals

Sunil Kumar Prabhakar
Department of ECE
Bannari Amman Institute of Technology,
India
Email : sunilprabhakar22@gmail.com

Harikumar Rajaguru
Department of ECE
Bannari Amman Institute of Technology,
India
Email : harikumarrajaguru@gmail.com

Abstract—Epilepsy is a chronic condition where seizures are recurrent and spontaneous. Epileptic seizure is a transitory clinical condition caused by the over discharge of hyper synchronous neurons. Processing the information embedded within ElectroEncephaloGraphy (EEG) signals is our current challenge here. For the exact analysis, diagnosis and treatment of epilepsy it is vital to understand the classification of epileptic seizures. Classification schemes are significant and crucial to facilitate information sharing among health professionals. In this paper suitable dimensionality reduction techniques are used initially to reduce the dimension of the EEG data and then the classification of epilepsy risk levels from EEG Signals is done using Gaussian Mixture Model (GMM) Classifier. The benchmark parameters taken in this paper are Specificity, Sensitivity, Time Delay, Quality Value and Accuracy.

Keywords - Epilepsy, EEG, GMM

I. INTRODUCTION

Epilepsy is one of the most commonly occurring neurological disorders and about 1 to 2% of the world’s population is affected by this disorder [4]. Characterized by the sudden and recurrent onset of transient disturbances which is occurring due to the abnormal bursts of electrical discharges and due to the hyper synchronization of excessive neurons, epilepsy is a great menace to the health of the humans [5]. The main feature which characterizes the epilepsy is recurrent seizures. With the help of clinical manifestations, epileptic seizures can be classified into partial, focal, unilateral and generalized seizures. Therefore, epilepsy is a diverse set of neurological disorders characterized by the seizures and it is generally controlled with suitable medications. The brain’s electrical activity is found out for the detection of epilepsy using EEG signals. To analyse and diagnose the epileptogenic zone, EEG is used in clinical contexts [6]. The occurrence of seizures can never be easily predicted in a short interval of time and therefore incessant EEG recordings are necessary. The length of the EEG recordings is too huge and sometimes it takes a few days to weeks for the recordings of the EEG signals [7]. Since it is time consuming and there is a huge difficulty in processing it, automated epileptic EEG detection systems have been developed recently.

The organization of the paper is as follows: In Section 2, the materials and methods are discussed followed by the various dimensionality reduction techniques such as Independent Component Analysis (ICA), Linear Graph Embedding (LGE) and Fuzzy Mutual Information (FMI) in Section 3. The GMM is used here as a Post Classifier for the classification of epilepsy risk levels from EEG signals in Section 4 followed by the results and conclusion in Section 5.

II. MATERIALS AND METHODS

For the performance analysis of the epilepsy risk levels using ICA, LGE and FMI as Dimensionality Reduction technique followed by GMM as Post Classifiers, the raw EEG data of 20 epileptic patients who were under treatment in the Neurology Department of Sri Ramakrishna Hospital, Coimbatore in European Data Format (EDF) are taken for detailed study and analysis. The pre processing stage of the EEG signals is given more significance because it is important to use the best available technique in literature to extract all the useful information embedded in the non-stationary biomedical signals [4]. The EEG records which were obtained were continuous for about 30 seconds and each of them was divided into epochs of two second duration. Generally a two second epoch is long enough to avoid unnecessary redundancy in the signal and it is long enough to detect any significant changes in activity and to detect the presence of artifacts in the signal. For each and every patient, the total number of channels is 16 and it is over three epochs. The frequency is considered to be 50 Hz and the sampling frequency is considered to be about 200 Hz. Each and every sample corresponds to the instantaneous amplitude values of the signal which totals to 400 values for an epoch. The total number of artifacts present in the data is four. Chewing artifact, motion artifact, eye blink and electromyography (EMG) are the four numbers of artifacts present and approximately the percentage of data which are artifacts is 1%. No attempts
were made to select certain number of artifacts which are of more specific nature. The main objective to include artifacts is to differentiate the spike categories of waveforms from non spike categories. The figure 1 shows the block diagram of the procedure.

\[
\begin{align*}
\text{Raw EEG Signals} & \quad \rightarrow \quad \text{Samples} \\
\text{ICA, LGE and FMI as Dimensionality Reduction Technique} & \quad \rightarrow \quad \text{GMM as Post Classifier} \\
\text{Epilepsy Risk Levels:} & \\
1) & \quad \text{PI} \\
2) & \quad \text{QV} \\
3) & \quad \text{Time Delay} \\
4) & \quad \text{Accuracy} \\
5) & \quad \text{Specificity} \\
6) & \quad \text{Sensitivity}
\end{align*}
\]

Figure 1 Block Diagram of the Procedure

Initially the raw EEG signals are taken and then it is sampled. Since the dimension of the original data is too large, certain dimensionality reduction techniques such as ICA, LGE and FMI are employed to reduce the dimension of the data. The dimensionally reduced values are given as inputs to the GMM which acts the post classifier here and finally the epilepsy risk levels are measured.

### III. DIMENSIONALITY REDUCTION TECHNIQUES

Reducing the dimensionality of the original EEG data seems to be a very significant pre-processing step because it contains a huge data to be processed. The essential dimensionality reduction techniques discussed in this paper are ICA, FMI and LGE.

#### A. Independent Component Analysis

In the last few decades, much attention has been given to the concept of ICA especially in the fields of pattern recognition, signal processing, imaging applications and machine learning [3]. For the blind source separation, it is widely regarded as a standard statistical tool. The ICA model in its classic form is written as follows:

\[
X = As
\]

where \( X = [X_1, \ldots, X_m]^T \) is nothing but a random vector of observations. The random vector of hidden sources is defined as follows:

\[
S = [S_1, \ldots, S_m]^T
\]

The random vector of the hidden sources is always mutual with that of the independent components with \( 'A' \) being represented as non-singular mixing matrix [3]. Then \( W = A^{-1} \) is defined and it is termed as unmixing matrix. Only if any one of \( S \)'s components is Gaussian, then \( 'A' \) (thus \( W \) ) would be easily identifiable. Having \( 'n' \) identical and independently distributed samples of \( X \), for instance \( \{X(j) : 1 \leq j \leq n\} \), the aim of ICA is to estimate the unmixing matrix \( W \) and thus each hidden source can be easily recovered using [3]

\[
S_k = W_k X
\]

where \( W_k \) represents the \( k^{th} \) row of \( W \).

#### B. Fuzzy Mutual Information

To compensate the curse of dimensionality, irrelevant features and redundant dimensions are reduced using FMI technique [2]. Between each feature \( f \) and class label \( L \) and for all the \( 'n' \) training trials, the intrinsic relation's estimation between the EEG attributes and FMI technique is given as follows:

\[
I(f; L) = H(f) + H(L) - H(f, L)
\]

where

\[
H(f) = -\sum_{i=1}^{2} ((\Sigma_{k} \mu_{i,k} / n) \log(\Sigma_{k} \mu_{i,k} / n))
\]

is referred as the marginal entropy for each and every feature. The marginal class entropy is given as follows

\[
H(L) = -\sum_{i=1}^{2} (n_i / n) \log(n_i / n)
\]

where \( n_i \) represents the number of training trials corresponding to \( i^{th} \) class. Also the joint fuzzy entropy [2] is given as follows

\[
H(f, L) = -\sum_{i=1}^{2} ((\Sigma_{k \in A} \mu_{i,k}(n) \log(\Sigma_{k \in A} \mu_{i,k}(n))
\]
C. Linear Graph Embedding

Supposing there are 'm' data points stated as \( \{x_i\}_{i=1}^{m} \). If a graph \( G \) is given with 'm' vertices, with each vertex representing a specific data point. Assume \( W \) to be a symmetric \( m \times m \) matrix with weights \( W_{ij} \), whose vertex \( i \) is joined with vertex \( j \). The \( G \) and \( W \) always defines and characterizes the geometrical properties of a particular data set. Each vertex of a graph is represented as a low-dimensional vector \([1]\). Each vector of a graph must preserve the similarity in between the vertex pairs. If \( y = (y_1, \ldots, y_m)^T \) is assumed to be the map from a particular graph to a real line, the optimal \( y \) is given as follows \([1]\)

\[
\sum_{i,j} (y_i - y_j)^2 W_{ij} \quad \text{such that} \quad y^T D y = 1,
\]

where \( D \) represents the diagonal matrix. An arbitrary scaling factor is removed with the help of constraint \( y^T D y = 1 \). A heavy penalty is incurred if the neighbouring vectors \( i \) and \( j \) are mapped far apart. With the aid of algebraic formulation, we have

\[
\sum_{i,j} (y_i - y_j)^2 W_{ij} = 2y^T L y
\]

where \( L = D - W \) is the graph Laplacian.

Finally, the problem of minimization is reduced \([1]\) to find the following equation

\[
y^* = \arg \min_{y: y^T D y = 1} y^T L y
\]

\[
y^* = \arg \min_{y: y^T D y = 1} \frac{y^T L y}{y^T D y}
\]

By solving the minimum eigen value problems, the \( y^* \)'s optimum value can be found out easily as follows

\[
L y = \lambda D y
\]

The graph embedding is very closely associated with the various differential geometries.

IV. GAUSSIAN MIXTURE MODEL AS A POST CLASSIFIER

The probability density function (PDF) of a particular random variable \( (x) \) is represented in terms of Gaussian mixture model as follows

\[
x \in R^d
\]

It is always represented as a weighted total sum of \( 'k' \) Gaussian distributions \([8]\) and is mathematically expressed as follows

\[
P(x / \Theta) = \sum_{m=1}^{k} \alpha_m P(x / \Theta_m)
\]

where \( \Theta \) is known as the mixture model. The weight of the component is given as \( \alpha_m \). Each component’s density is usually represented as the normal probability distribution and is expressed as follows

\[
p(x / \theta_m) = \frac{1}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} (x - \mu_m)^T \Sigma_m^{-1} (x - \mu_m) \right\}
\]

In the training process, the parameters such as \( \sigma, \mu \) and \( \Sigma \) are optimized in an iterative manner throughout the EM algorithm \([8]\). EM algorithm is employed in order to obtain the maximum value of log-likelihood model. If there is a collection of \( n' \) independent samples which are identically distributed as \( X = \{x_1, x_2, \ldots, x_n\} \), then the log-likelihood function \([8]\) which corresponds to a particular mixture model \( \Theta \) is given as follows

\[
L(X; \Theta) = \log \prod_{i=1}^{n} P(x_i; \Theta)
\]

\[
L(X; \Theta) = \sum_{i=1}^{n} \log \sum_{m=1}^{k} \alpha_m p(x_i; \theta_m)
\]

Finally the likelihood estimates are joined together to yield the posterior probability value of the seizure and it is done with the help of Bayesian formula as follows

\[
P(S / x) = \frac{P(x / \Theta_S) P(S)}{P(x / \Theta_S) P(S) + P(x / \Theta_N) P(N)}
\]

where the prior probabilities of the seizures and non-seizures classes are expressed as \( P(S) \) and \( P(N) \) respectively. Therefore it is the combination of both GMM and the Bayesian formula and is referred to as GMM Classifier.

The probability density function (PDF) of a particular random variable \( (x) \) is represented in terms of Gaussian mixture model as follows

\[
x \in R^d
\]

It is always represented as a weighted total sum of \( 'k' \) Gaussian distributions \([8]\) and is mathematically expressed as follows

\[
P(x / \Theta) = \sum_{m=1}^{k} \alpha_m P(x / \Theta_m)
\]

where \( \Theta \) is known as the mixture model. The weight of the component is given as \( \alpha_m \). Each component’s density is usually represented as the normal probability distribution and is expressed as follows
\[ p(x / \theta_m) = \frac{|\Sigma_m|^{-1/2}}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} (x - \mu_m)^T \Sigma_m^{-1} (x - \mu_m) \right\} \]

In the training process, the parameters such as \( \sigma \), \( \mu \) and \( \sum \) are optimized in an iterative manner throughout the EM algorithm [8]. EM algorithm is employed in order to obtain the maximum value of log-likelihood model. If there is a collection of \( n \) independent samples which are identically distributed as \( X = \{x^1, x^2, ..., x^n\} \), then the log-likelihood function [8] which corresponds to a particular mixture model \( \Theta \) is given as follows

\[
L(X; \Theta) = \log \prod_{i=1}^{n} P(x_i; \Theta) = \log \sum_{m=1}^{k} \alpha_m p(x_i; \theta_m)
\]

Finally the likelihood estimates are joined together to yield the posterior probability value of the seizure and it is done with the help of Bayesian formula as follows

\[ P(S / x) = \frac{P(x / \Theta_S)P(S)}{P(x / \Theta_S)P(S) + P(x / \Theta_N)P(N)} \]

where the prior probabilities of the seizures and non-seizures classes are expressed as \( P(S) \) and \( P(N) \) respectively. Therefore it is the combination of both GMM and the Bayesian formula and is referred to as GMM Classifier.

V. RESULTS AND CONCLUSION

For ICA, LGE and FMI as dimensionality reduction techniques and GMM as a Post Classifier, based on the Quality values, Time Delay and Accuracy the results are computed in Table I respectively. The formulae for the Performance Index (PI), Sensitivity, Specificity and Accuracy are given as follows

\[ PI = \frac{PC - MC - FA}{PC} \times 100 \]

where PC – Perfect Classification, MC – Missed Classification, FA – False Alarm,

The Sensitivity, Specificity and Accuracy measures are stated by the following

\[ Sensitivity = \frac{PC}{PC + FA} \times 100 \]

\[ Specificity = \frac{PC}{PC + MC} \times 100 \]

\[ Accuracy = \frac{Sensitivity + Specificity}{2} \times 100 \]

The Specificity and Sensitivity Analysis for the application of ICA, LGE and FMI as dimensionality reduction technique followed by the application of GMM as Post Classifiers is shown in Figure 2. The Time Delay and Quality Value Analysis for the application of ICA, LGE and FMI as dimensionality reduction technique followed by the application of GMM as Post Classifiers is shown in Figure 3. Similarly the Performance Index and Accuracy Analysis for the application of ICA, LGE and FMI as dimensionality reduction techniques followed by the application of GMM as Post Classifiers is shown in Figure 4.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>FMI + GMM</th>
<th>LGE + GMM</th>
<th>ICA + GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC (%)</td>
<td>95.69</td>
<td>89.51</td>
<td>87.5</td>
</tr>
<tr>
<td>MC (%)</td>
<td>4.3</td>
<td>9.65</td>
<td>9.236</td>
</tr>
<tr>
<td>FA (%)</td>
<td>0</td>
<td>0.83</td>
<td>3.26</td>
</tr>
<tr>
<td>PI (%)</td>
<td>95.49</td>
<td>87.03</td>
<td>84.32</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>100</td>
<td>99.16</td>
<td>96.73</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>95.69</td>
<td>90.34</td>
<td>90.76</td>
</tr>
<tr>
<td>Time Delay (sec)</td>
<td>2.17</td>
<td>2.36</td>
<td>2.30</td>
</tr>
<tr>
<td>Quality Value</td>
<td>23.02</td>
<td>20.86</td>
<td>20.09</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>97.84</td>
<td>94.75</td>
<td>93.75</td>
</tr>
</tbody>
</table>

Figure 2 Specificity and Sensitivity Measures

It is inferred from figure 2 that when FMI-ApEn is engaged, a specificity of about 95.69 is obtained when compared to the other two techniques.
Figure 3 Time Delay and Quality Value Measures

It is observed from figure 3 that the time delay measures are less for FMI-GMM combination as of 2.17 seconds rather than the ICA-GMM combination as of 2.30 seconds. A comparatively high time delay of 2.36 seconds is obtained in the LGE-GMM combination.

Figure 4 Performance Index and Accuracy Measures

It is inferred from figure 4 that an average PI of about 95.49% is obtained for FMI-GMM combination which is comparatively high than the LGE-GMM and ICA-GMM combination.

It is thus concluded that the FMI-GMM Combination has a highest Perfect Classification as of 95.69%, an average accuracy of about 97.84%, and a high quality value of about 23.02. It is performing better when compared to the LGE-GMM and ICA-GMM combinations respectively. Future works may incorporate the possible usage of different dimensionality reduction techniques followed by various other types of classifiers.

REFERENCES