



# Performance Exploration of Multiple Classifiers with Grid Search Hyperparameter Tuning for Detecting Epileptic Seizures from EEG Signals

C Ganesh Babu\*, M Gowri Shankar & Harikumar Rajaguru

Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu 638 401, India

*Received 19 August 2021; revised 03 June 2022; accepted 18 June 2022*

This study evaluates the performance of two-level classifications using dimensionality reduction methods to determine the risk level of epilepsy from EEG dataset. To diminish the complexity of EEG data, dimensionality reduction techniques such as Singular Value Decomposition (SVD), Independent Component Analysis (ICA), and Principal Component Analysis (PCA) are utilized. The risk level of epilepsy classification from EEG dataset would then be carried out using three classifiers: Hidden Markov Model (HMM), Naïve Bayesian Classifier (NBC) and Gaussian Mixture Model (GMM). The Grid Search (GS) process is employed to tune the hyperparameters of GMM and NBC classifiers. This study analyzed twenty patients' datasets. Performance evaluation of classifiers with and without GS hyperparameter tuning is examined, including performance index, sensitivity, specificity, and accuracy. The GMM classifier with the GS hyper-tuning approach for SVD dimensionality reduction technique achieved a higher accuracy of 98.18% than its counterpart classifiers.

**Keywords:** Epilepsy, GMM, Grid search, HMM, Hyperparameters

## Introduction

Seizures are symptom of certain functionality in the brain. Epileptic seizures are caused by neuronal activity disruptions in the brain.<sup>1,2</sup> The term "epilepsy" simply does not mention the cause of the occurrence and form of epilepsy. Epilepsy is a common term for an inflammation of the inclination. Epilepsy is often referred to as a seizure disease state. Although people with epilepsy follow a common pattern of symptoms called epilepsy, there is a variety of Epilepsy. Seizures may occur when a person is awake or asleep (nocturnal seizures).<sup>3</sup>

Epilepsy is a psychological condition of the nervous system that affects about 1% to 2% of the globe's inhabitants. This is due to excessive synchronization of neuronal cortical networks and has sudden repetitive and transient perceptive or behavioural interference.<sup>4</sup> It is a disorder in which a person suffers from excessive bursts of electric discharges. The word "epileptic seizures" is described for epilepsy. Seizures are categorized into partial or focal, general, unilateral and non-classified. Only part of the cerebral hemisphere and the related areas of the body are affected by epileptic seizures. The entire brain is involved in generalized epileptic seizures and

typically develops bilateral motors with a loss of consciousness.<sup>5</sup> Epileptic seizures of both types can occur at any age. The brain activity monitoring through the Electroencephalogram (EEG) is one of the most effective diagnostic methods of Epilepsy. The EEG signal is the temporary shape of the spike trains, spikes and wave complexes. It helps also to classify the epileptic syndrome.<sup>6</sup>

## Related Works

Signals from an EEG are analysed in order to track and record the electrical activities that occur in the brain. It is a method that does not entail any invasive procedures, and it involves placing several electrodes into the patient's scalp. The EEG records the changes in voltage that occur inside the brain as a result of the neurons electrical activity. The immediate brain activity is recorded by the various electrodes over the course of a certain amount of time.<sup>7</sup> Many researches have been undertaken in the literature to develop a system that reliably identifies irregular EEGs in humans, since proper identification of such EEGs may prime to the discovery of sleep disorders, epilepsy and other conditions.<sup>8</sup> Seizures are not only detected, but also predicted in several studies. Because EEG signals are non-stationary and non-linear, linear analysis approaches are unreliable and may not provide correct findings.<sup>9</sup> As a result, different investigations use multiple machine learning

\*Author for Correspondence  
E-mail: bits\_babu@yahoo.co.in

and non-linear analytic methodologies. Different characteristics are retrieved in machine learning-based investigations.<sup>10</sup> In the field of machine learning, several classifiers such as RF, KNN, SVM, Logistic Regression, Linear Regression, Expectation Maximization, Bagging Trees (BT), Perceptron, Neural Networks and others are used to identify signals according to their pattern or dimensionally reduced features.<sup>11</sup>

Lopez *et al.*<sup>12</sup> combined PCA with KNN and RF classifiers and attained 58% accuracy with the KNN classifier and 68% accuracy with the RF classifier. Veisi *et al.*<sup>13</sup> developed a reliable and speedy recognition system by making use of permutation entropy. This method was used to identify epilepsy based on EEG patterns. The study of Faust *et al.*<sup>14</sup> consisted of developing a computer-aided seizure identification approach for the purpose of identifying epilepsy with the assistance of EEG computation that was based on wavelets. Liu *et al.*<sup>15</sup> employed wavelet transformations and SVM to automatically identify seizures in the long-term intracranial EEG datasets. This allowed the seizures to be recognised in a manner that was completely automated. Wang *et al.*<sup>16</sup> developed a systematic technique of EEG segmentation for the identification of epileptic behaviour. This system is based on the principle of wavelet packet entropy to extract the features.

Prabhakar & Rajaguru<sup>17</sup> designed a patient remote observation technology for epilepsy identification using EEG recordings. Song *et al.*<sup>18</sup> developed an algorithm for the automated identification of epileptic seizures in EEG recordings. The algorithm is based on improved extreme learning machine and sample entropy. Zhou *et al.*<sup>19</sup> adopted approximation entropy to investigate the effects of epileptic disappearance seizures on the dynamic features of the patients' brains. An iterative filtering decomposition and HMM were employed for automated epileptic seizure identification from multi-channel EEG data, as suggested by Dash *et al.*<sup>20</sup> Riaz *et al.*<sup>21</sup> engaged SVM to classify seizure from EEG signals using empirical mode decomposition-based spectral and temporal Features.

EEG recordings were analysed by Pachori & Patidar<sup>22</sup> using a second-order difference plot of intrinsic mode function, which was used in conjunction with an ANN classifier to detect epileptic seizures. Srinivasan & Eswaran<sup>23</sup> suggested an automated seizure detection approach based on an Elman network, a combination of recurrent neural network. Analysis using Discrete Wavelet Transform (DWT) and Approximation Entropy (ApEn) with an

artificial neural network was the basis for Kumar *et al.*<sup>24</sup> suggested method. Since the system needs a lot of training and input data, it can be hard to use a real-time system to find an instant seizure. In order to circumvent this constrain, a research was carried out making use of fractal dimension values in order to get complex feature values. The benefit of fractal dimension values obtained using the box counting approach over other approaches is that repeated computation enables us to acquire more complex feature results. Iterative computations, on the other hand, require more time to complete.<sup>25</sup>

Most of the scientific literature on epileptic seizure identification appears at the level of classification and EEG dimensionality reduction. However, there are not many studies on the role of feature reduction on epileptic seizure extrapolation using Grid Search (GS) to determine the perfect machine learning model with hyperparameter tuning. To overcome these challenges with long-term EEG monitoring, we recommend a seizure identification technique that combines SVD, PCA, and ICA dimensionality reduction algorithms with HMM, NBC, and GMM classifiers based on EEG recordings. The GS approach is also used to tune the hyperparameters of the GMM and NBC classifiers. Since the primary goal of hyperparameter optimization is to attain the maximum possible accuracy while preventing overfitting, accuracy has been taken into consideration as the primary basis for evaluation in this process. Additionally, classification takes less time than previous approaches. The methodology for epilepsy risk level classification is shown in Fig. 1.

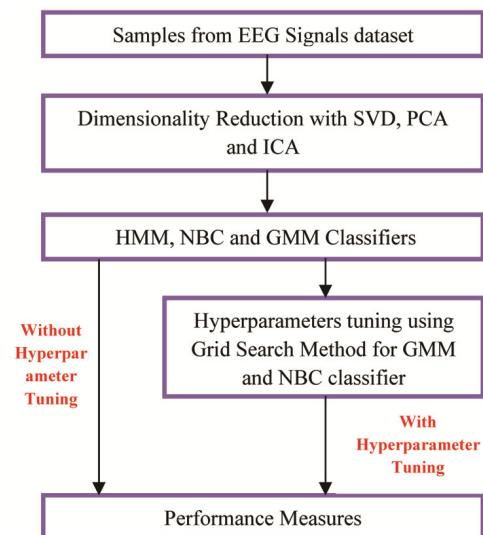


Fig. 1 — Methodology for Epilepsy Risk Level Classification

**Materials and Methods**

The EEG data Acquisition, dimensionality reduction approaches, classification, and hyperparameter tuning procedures that we have evaluated are briefly discussed in this section.

**EEG Acquisition**

The training phase involves the acquisition of EEG signals and pre-processing. For 20 epileptic patients, EEG signals were collected from the Department of Neurology at Sri Ramakrishna Hospital in Coimbatore, India.<sup>26</sup> The signal is obtained by means of a non-invasive electrode system, in which sixteen channel EEG signals are used. The continuous EEG signal is divided into a smaller duration signal with duration of two seconds.<sup>27</sup> The international electrode system incorporates elements such as muscle noise, eye motion, heart signals and linear noise due to the low conductivity of skull and synchronization of electrical activity.<sup>28</sup> With an EEG signal that is devoid of artefacts, it is feasible to identify epilepsy rather precisely. With the aid of a neurologist, an artefact-free EEG is acquired. The Epileptic EEG is diagnosed by acquiring the Epileptic EEG, extracting the features and processing the extracted features. The different steps involved in the processing of the features are elaborated. In this study, 20 patients are taken into consideration. The EEG signal is divided into four epochs, each with a sampling frequency of 50 Hz, with the duration of two seconds. Using an Independent Analysis method, EEG samples are reduced to 12 independent components per epoch per channel, while the remaining components are correlated with the others.

**Dimensionality Reduction (DR) Techniques**

**Singular Value Decomposition (SVD)**

To optimize average signal-to-noise ratio, SVD is widely employed in multi-signal processing, in which the initial data set is generally divided into extra subspaces known as signal and noise subspaces.<sup>29,30</sup> A popular method common technique for extracting signal and noise is SVD. Consider the findings' factual  $M \times N$ , matrix  $X$ , which could have been dissected as follows:<sup>31</sup>

$$X = USV^T \quad \dots (1)$$

where,  $S$  seems to be a  $M$  indicates zero with non-square matrix, and even on the diagonal, and  $s_i$  essentials prescribed in magnitude descendent order.

Every single  $s_i$  is equivalent toward  $\sqrt{\lambda_i}$ , it is square root of eigen values  $C = X^T X$ . The smallest eigenvalues are taken into account regardless of noise.<sup>32,33</sup> The eigenvectors of  $C$  are the columns of  $V$ . The projections of  $X$  onto eigenvectors of  $C$  are represented by the  $M \times M$  matrix  $U$ .

A strategy of SVD performance is as follows:

- (i) Calculate non-zero eigenvalues ( $N$ ), and  $\lambda_i$  of the matrix  $C = X^T X$ .
- (ii) Calculate the eigenvectors of such matrix  $X^T X$  that are orthogonal.
- (iii) Calculate the matrix of column vectors  $U: u_i = s_i^{-1}$ .
- (iv) And used the Gram Schmidt Orthogonalization method, add the remaining  $M \times N$  vectors to the matrix.

**Principal Component Analysis (PCA)**

A multivariate approach is PCA. It examines a given data in which occurrences are defined by a set of predictor variable that are all interconnected.<sup>34,35</sup> The goal is to pull out the most important information from the table and put it into a range of variables labelled principal components.<sup>36</sup> The major components are found by computing the eigenvectors and eigenvalues of the data covariance matrix. The PCA's goals include:

- (i) Get the most particular data from the table of data;
- (ii) Reduce the amount of the data collection by maintaining just the information that is the most relevant.
- (iii) To facilitate data set's description simpler; and
- (iv) Examine the structure of the data and variables

PCA generates new parameters called principle components as linear combinations of the original variables in order to accomplish these criteria.<sup>37</sup> The highest feasible variance is required for the first main component. The second component is chosen with the restriction that it must be orthogonal to the very first.<sup>38</sup> The mathematical characterization of the stochastic process is used in principal component analysis. Assume we have such a random vector  $e$ , where  $e = (e_1, \dots, e_n)^T$  and the population's average is expressed as,

$$\mu_e = F\{e\} \quad \dots (2)$$

The identical data set's covariance matrix is as follows:

$$C_e = F\{(e - \mu_e)(e - \mu_e)^T\} \quad \dots (3)$$

The covariances between both the random variables are represented by the components of  $C_e$ . The variance  $g$  represents the range of component values it around total mean. We can compute the sample mean and sample covariance matrix as estimations of the mean and covariance matrix from a sample of vectors  $e_1, \dots, e_M$ . A symmetric matrix, such as the covariance matrix,<sup>39</sup> can be used to calculate the eigenvalues and eigenvectors. The solutions of the eigenvectors  $f_i$  and the corresponding eigenvalues  $\lambda_i$  is,

$$C_e f_i = \lambda_i f_i \quad i = 1, \dots, n \quad \dots (4)$$

Assume they're separate for the purpose of simplicity. The solutions to the characteristic may be found to obtain these values

$$|C_e - \lambda I| = 0 \quad \dots (5)$$

where,  $I$  indicates the identity matrix and the  $|\cdot|$  represents the determinant matrix.

**Independent Component Analysis (ICA)**

The ICA model is a multivariate dataset model with a large sample database. The independent components of the data seen are non-Gaussian and mutually independent variables that are included inside the model. These can also be referred to as sources or factors.<sup>40</sup> If traditional approaches like PCA fail entirely, ICA is a strong tool for classifying the underlying variables or sources.<sup>41,42</sup>

To identify individual source signals from such a succession of linear component mixes, independent component analysis techniques are utilised.<sup>43,44</sup> Assume we examine  $n$  linear mixes of independent components  $x_1, \dots, x_n$  as shown in Eq. (6):

$$X_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n, j = 1, n \quad \dots (6)$$

The independent component  $s_i$  and the variable  $x_j$  are random variables in this equation, whereas  $x_j(t)$  and  $s_i(t)$  are random variables samples.

The independent component and the variable are both assumed to have a zero mean, simplifying the issue to the model zero-mean<sup>45</sup>, indicated by,

$$\hat{X} = X - E(X) \quad \dots (7)$$

Let  $x$  and  $s$  represent the random vectors  $x_1, \dots, x_n$  and  $s_1, \dots, s_n$ , respectively. Let  $A$  stand for the matrix containing the components  $a_{ij}$ , which is written as follows

$$x = As \text{ or } x = \sum_{i=1}^n a_i s_i \quad \dots (8)$$

Though just variable  $X$  measurements are provided then both the matrix  $A$  and the independent components are identified, the following equation is termed an independent component analysis, or ICA. Independent and non-Gaussian components are predicted in the model.<sup>40,41</sup> The histogram for patient 13 in SVD (DR) method is depicted in Fig. 2. The histogram for patient 13 in PCA (DR) method is depicted in Fig. 3. Also the Fig. 2 indicates that the histogram is Gaussian in nature for a normal, open-eyed healthy person, whereas for epileptic patients during seizure, it is non-Gaussian signifying an intermittent nonlinear effect. Therefore the histogram is considered useful to distinguish between an epileptic patient's brain activities during a seizure from another state.

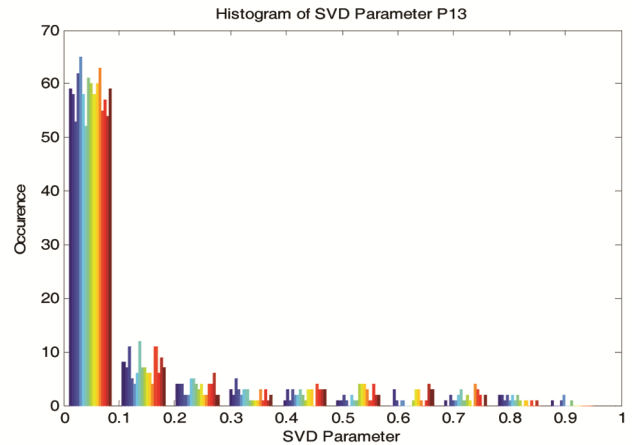


Fig. 2 — Histogram of SVD Parameter for P13

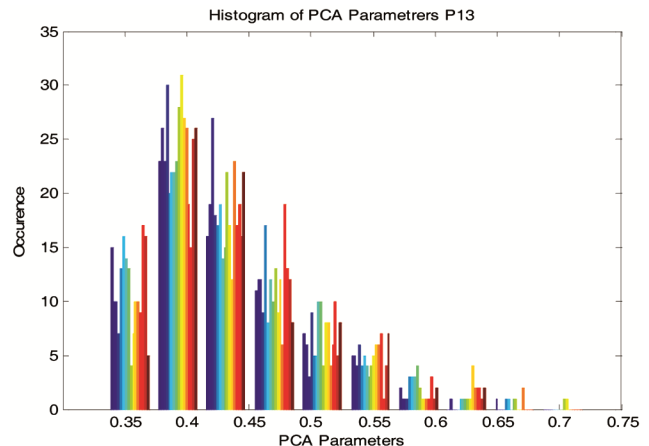


Fig. 3 — Histogram of PCA Parameter for P13

The histogram for patient 13 of the ICA (DR) method is enumerated in Fig. 4. It is conditioned in Fig. 4 that intermittent Gaussian nature of histogram established underlying inherited nonlinear nature of the distribution during epileptic seizure. Hurst is a time series measure of self-similarity, predictability, and long-term dependence. It is also a measure of the regularity of an asymptotic behaviour of the fractal time sequence's rescaled process. Hurst exponent  $H$  is well-defined as the following in accordance with Acharaya *et al.*<sup>46</sup> Hurst's general time series Eq. (9),

$$H = \frac{\log\left(\frac{R}{S}\right)}{\log(T)} \quad \dots (9)$$

where,  $T$  indicates the data sample duration and  $\left(\frac{R}{S}\right)$  denotes the rescaled range.  $R$  denotes the difference between the maximum and minimum deviation and  $S$  denotes the standard deviation. Plotting  $\left(\frac{R}{S}\right)$  versus  $T$  in the log-log axis yields the Hurst exponent. The regression slope line is the variable that is used to figure out the Hurst exponent.

The epileptic spikes of a distinct component are classified using the Hurst exponent. A time series is defined by the Hurst exponent. Hurst exponent estimate has been used in a variety of fields, from biology to computer networking. The Hurst Exponent value is shown below for several kinds of signals. A time series with a Hurst exponent of 0.5 corresponds to a random time series. In their seminal article, Arunkumar *et al.*<sup>47</sup> showed that the Hurst exponent of  $0 < H < 0.5$  shows so-called anti persistent characteristic. The time series must shift direction with every sample when  $H = 0$  is reached. A Hurst  $0.5 < H < 1$  exponent, on the other hand, denotes a temporally continuous time series. For many environmental, economic, and human events, a straight line with a nonzero slope will have a Hurst exponent  $H$  value of 0.70 to 0.76 at the limit. This study shows Hurst exponent values often range from 0.25 to 0.45 due to reduced dimensionality EEG results for epileptic seizures.

The antipersistent activity is apparent from the reduced dimensionality of SVD, PCA, and ICA technique applied to the EEG data set as evident from Table 1. However, the ICA method indicates the convergence of the EEG data set with the hyperchaotic situation due to the higher value of the Hurst exponent.

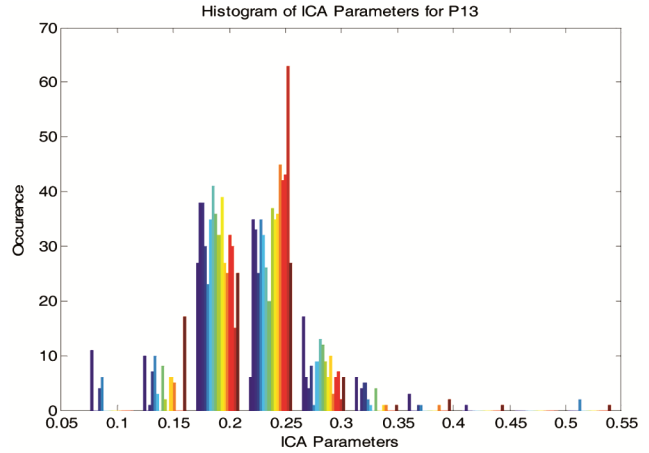


Fig. 4 — Histogram of ICA Parameter for P13

Table 1 — Average Hurst Exponent values for SVD, PCA and ICA based Dimensionality Reduced EEG data set

Average Hurst Exponent Values	DR Techniques		
	SVD	PCA	ICA
	0.416	0.432	0.454

**Classifiers for Epilepsy Risk Level Classification**

**Hidden Markov Model (HMM)**

A Hidden Markov Model (HMM) is a statistical Markov model designed especially for systems with probability changes. The state is clearly observable in Markov's model, which means the probabilities of state transition are the only parameters.<sup>45</sup> The State is not clearly observable in a hidden Markov model, but the output is dependent on the observable state. A probability distribution is established for output tokens.<sup>48</sup>The sequence of tokens therefore produces the information about the sequence of states. The HMM is a triple hidden Markov model  $(\pi, A, B)$  where  $\pi = \pi_i$  denotes the initial state probabilities of the vector,  $A = (a_{ij})$  denotes the transition matrix of state,  $p_r(x_i|x_j t-1)$  and  $B = (b_{ij})$  denotes the emission matrix,  $p_r(y_i|x_j)$

Every probability, such as that found in the state transition matrix and the emission matrix, is time-independent. This indicates that the matrix does not change as the system develops further over the course of time. In point of fact, it is among the most implausible inferences that can be made regarding the fundamental systems underlying Markov models.

**Vector Quantization**

The vector quantifier uses the Block Coding principle to translate  $k$ -dimensional vectors throughout the vector space  $R^k$  into some kind of

finite collection of vectors  $Y = \{y_i: i = 1, 2, \dots, N\}$ . A code vector or a code word is the name given to each vector in  $Y$ . A set of code words compensate the code book.<sup>41</sup> Each code word's Voronoi area is specified by Eq. (10).

$$V_i = \{x \in R^K: |x - y_i| \leq |x - y_j|, \text{ for all } j \neq i\} \quad \dots (10)$$

The following steps are involved in construction of a code book.

- (i) The number of code words  $N$  is determined.
- (ii) The initial code book is the  $N$  code words chosen randomly.
- (iii) The vectors around the code words are clustered using the Euclidean distance.
- (iv) Either of every input vector and code word's Euclidean distance is calculated. Clustering is used to group input vectors that are closest to the code word.
- (v) As demonstrated in Eq. (11), the new set of code words is determined by averaging the cluster.

$$y_i = \frac{1}{m} \sum_{j=1}^m x_{ij} \quad \dots (11)$$

where,  $i$  denotes the vector components ( $x, y, z$  directions),  $m$  denotes the number of cluster vectors. Steps 2 and 3 should be repeated until the code words do not change otherwise small change. Encoder and Decoder are both part of a vector quantizer and encoder yields the code word index which gives the lowest distortion of the input vector.<sup>49,50</sup> If the nearest code is identified, the code word index is transmitted through a channel. The code word index is substituted with the code word associated with it when the encoder gets a code word index.<sup>51,52</sup> A block diagram of the encoder and decoder functionalities is presented in Fig. 5.

**Naïve Bayesian Classifier**

A well-known statistical analysis and supervised classification technique is Bayesian classification. The fundamental probabilistic distribution is required here, as well as the model's uncertainty may be simply conveyed by determining the probabilities of a occurrences. As a result, Naïve Bayes is utilised to tackle prediction as well as diagnostic issues. The Bayesian classification method is highly beneficial for assessing and comprehending learning systems. Because it is more resistant to noise, precise

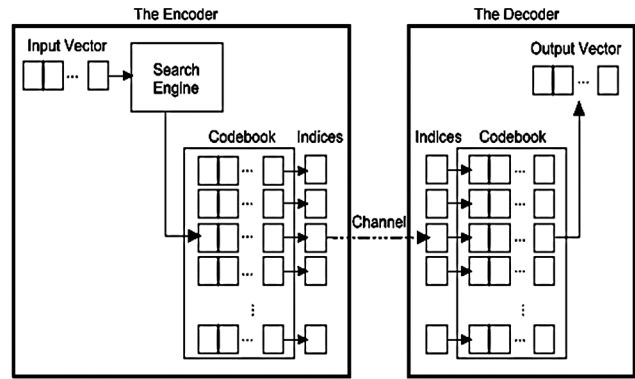


Fig. 5 — Block Diagram of Encoder and Decoder functionalities

probabilities for hypothesis are computed. Naïve Bayes, Decision Trees, SVM with Accuracy measures was compared by Huang *et al.*<sup>53</sup>

The NBC procedure has been created by Karim and Rahman for the purpose of categorization and the generation of realistic information for marketing strategy.<sup>54</sup> The Bayesian theorem is the foundation of the NBC, which is a straightforward probabilistic predictor that calls for very stringent independence conditions. One way to express this is as  $p(d|H_1 \dots \dots H_n)$  on the dependent variable of the classifier, which is reliant on numerous feature parameters  $H_1$  via  $H_n$ . When depending on the likelihood of this form of model, it becomes troublesome whenever the collection of features is quite vast. As an outcome, the model must be modified to make it more traceable, and described as follows in the context of the Bayesian theorem:

$$p(d|H_1 \dots \dots H_n) = \frac{p(d)p(H_1 \dots \dots H_n|d)}{p(H_1 \dots \dots H_n)} \quad \dots (12)$$

$$Posterior = \frac{priori \times likelihood}{evidence} \quad \dots (13)$$

Because the denominator doesn't always respond to  $d$ , and feature values  $H_i$  are taken into account, the denominator is continued product. The numerator represents  $p(d, H_1 \dots \dots H_n)$ , which is identical to the joint probability model. It may be restated as using the notion of conditional probability for repeated applications.

$$p(d, H_1 \dots \dots H_n) = p(d)p(H_1 \dots \dots H_n|d) \quad \dots (14)$$

$$p(d, H_1 \dots \dots H_n) = p(d)p(H_1|d)p(H_2 \dots \dots H_n|d, H_1) \quad \dots (15)$$

$$p(d, H_1 \dots H_n) = p(d)p(H_1|d)p(H_2|d, H_1)p(H_3 \dots H_n|d, H_1, H_2) \dots \quad (16)$$

$$p(d, H_1 \dots H_n) = p(d)p(H_1|d)p(H_2|d, H_1)p(H_3|d, H_1, H_2)p(H_4 \dots) \dots \quad (17)$$

$$p(d, H_1 \dots H_n) = p(d)p(H_1|d)p(H_2|d, H_1)p(H_3|d, H_1, H_2) \dots p(H_n) \dots \quad (18)$$

In this optimized model of the scenario, the assumptions of "naive" independence assertions have been included. Reassuring that entirely feature  $H_i$  is conditionally independent of every single further feature  $H_j$ , assuming  $j \neq i$ , indicates that

$$p(H_i|d, H_j) = p(H_i|d), \text{ for } i \neq j \quad \dots \quad (19)$$

As a result, the combined model can be written as

$$p(d, H_1 \dots H_n) = \frac{1}{X} p(d) \prod_{i=1}^n p(H_i|d) \quad \dots \quad (20)$$

where,  $X$  is known as a scaling factor that is reliant on  $H_1 \dots H_n$ , (i.e., a constant number). Because the component is treated as a class prior  $p(d)$  and an unrelated probability distribution  $p(H_i|d)$ , this model is considerably easier to maintain. It is considered that there have been  $K$  classes in total, and that if a model for each  $p(H_i|d = D)$  is described in terms of  $R$  factors, then the corresponding Bayes model contains  $(K - 1 + nRK)$  factors.  $K = 2$  is frequently used for binary classification, whereas  $R = 1$  is typically used for Bernoulli variables. The total number of variables in the Nave Bayesian model is  $(2n + 1)$ , where  $n$  indicates the total number of features of binary, it is utilized for classification.

*Parameter Estimation*

The majority of the attribute values, such as class priors and the probability distribution of attributes may be easily approximated by utilising frequency distributions from the training phase. These predictions of probabilities are sometimes referred to as optimum probability assessments of possibilities. The conception of equally likely classes is used in the computation of a class prior ( $priors = \frac{1}{\text{number of classes}}$ ). The class prior was therefore computed using the training set's estimated class probability. The prior for a particular class is determined by taking the proportion of the measured

values in the subject to the total sample amount. The parameters should be calculated for the feature distribution, and as such the distribution or development of non-parametric modeling techniques for the training features set is considered to be required. Whenever engaging with continuous data, Huang *et al.*<sup>53</sup>, assumes that the continuous values are connected to each class according to the Gaussian distribution. The training data, for illustrate, may be found in the continuous property  $y$ . The data will be first segregated by class, and then  $\mu_D$  is used to represent the mean and variance of  $y$ , which are connected to class  $D$ . The class  $D$  is related to value of variance  $y$ , therefore class is characterized as  $\sigma_D^2$ . Finally the probability is  $p(y = v|D)$ , where  $v$  denotes normal distribution.

$$p(y = v|D) = \frac{1}{\sqrt{2\pi\sigma_D^2}} e^{-\frac{(v-\mu_D)^2}{2\sigma_D^2}} \quad \dots \quad (21)$$

Some other significant approach for dealing with continuous data is to utilise binary for deconvolution. The distribution technique is a better choice there might be a little quantity of training data and a small amount of data distribution is known. The Discretization approach performs better when there is a lot of training data since it tries to figure out how to match the data distribution. In the event that a certain class and attribute score rarely appear together during training set, the frequency-dependent likelihood prediction score will be equal to zero. As a result, it is extremely troublesome, and when compounded, it will remove all or nearly all of the information in the other probabilities and no probability is essentially zero, therefore it is frequently beneficial to incorporate a tiny link in probability estimations. It is crucial to design a probability model classifier. The nave Bayesian probability model is a stand-alone feature model. This model is used in conjunction with the NBC classifier's decision rule. Maximum Aposteriori (MAP) decision rule are mostly utilized. It is selecting the hypothesis. Therefore classify function is,

$$classify(y(H_1 \dots H_n)) = arg \max_D p(d = D) \prod_{i=1}^n p(H_i = h_i|d = D) \quad \dots \quad (22)$$

**Gaussian Mixture Model (GMM)**

GMM is simply a distribution, consisting of linear finite number of Gaussian distributions. In practical cases, GMM is used for its high applicability. Firstly,

because of their good computational properties, a straightforward mathematical analysis is permitted for GMM. Secondly, GMM is the most effective approach for the approximation of numerous kinds of noise in physical systems. If there are a large number of unknown variables and instances, the GMM is perfect choice. GMM has further extended the use of the Gaussian distribution.<sup>55</sup>

GMM is most widely used for unsupervised learning because it tracks data clusters and patterns that share a similar behaviour. The Expectation Maximization (EM) algorithm's fundamental insights are it is derived with closed forms of the probability distribution, based on the assumption of the targeted data, although there are several unknown parameters. The expectation of the hidden values is calculated using the previously estimated values as a first and foremost step in the expectation algorithm. In case if it is the initialized phase, then the former estimated values should be any initialized values.<sup>56</sup> Once the current estimation of the hidden variables expectation is done, then it is used to finish the closed form of posteriori or likelihood. The parameters will be modified depending on the Maximum Likelihood or Maximum A posteriori (MAP) conditions. Hidden variables are widely preferred, and the reason to use them is that the closed form of posteriori is hardly computed and easy to use when dealing with hidden variables. If the hidden variables are not unknown, the expectation of the hidden variable is used. This method is iterative and the best solution for both parameters and likelihood is considered, as it converges to one particular point under certain settings.<sup>55</sup>

The following is often a statistical formulation of the Expectation Maximization (EM) algorithm: Here  $A$  and  $B$  are considered as the unobserved data and observed data which corresponds to  $A$  respectively, to calculate the likelihood  $f(B)$ ,  $\theta$  is the parameter needed and the main intention is, to compute the Maximum Likelihood  $\theta_{ML}$  therefore  $L(\theta) = \log f(B|\theta)$  is maximized. Generally the  $\log(f(A, B|\theta))$  has a well-defined form and so it is pretty easy to compute the maximum likelihood but it requires the unobserved data  $A$ . EM algorithm statistics obtain a sequence  $\theta'$  and  $\theta''$  in a manner represented as  $L(\theta') > L(\theta'')$ . The two important steps are<sup>56</sup>

Estimation Step: The expectation (EM) of the unnoticed data is expressed as

$$E_{f(A|B,\theta')}[\log f(A, B|\theta'')] \quad \dots(23)$$

Maximization Step:  $\theta''$  is found out such that

$$\theta'' = \arg \max(E_{f(A|B,\theta')}[\log f(A, B|\theta'')]) \quad \dots (24)$$

The following theorem holds good if

$$E_{f(A|B,\theta')}[\log f(A, B|\theta'')] > E_{f(A|B,\theta')}[\log f(A, B|\theta')] \quad \dots (25)$$

Then it's valid that  $L(\theta') > L(\theta'')$  in order to achieve goal of machine learning.

*Supervised Learning based on GMM*

It is assumed that there are  $N$  instances  $\{a_i\}_{i=1}^N$  is considered as a training data, here each  $a_i$  is a  $d$ -dimensional attribute vector, and for every instance the label is given by  $q_i \in (0,1)$ . A GMM is explained as follows,  $f(a; \mu, \Sigma) = \sum_{i=1}^k w_i N(a; \mu_i, \Sigma_i)$  where,  $a$  denotes an example from the data space, the total number of Gaussian components is given by  $k$ , and the weight of each Gaussian component is explained as  $w_i$  such that,

$$\sum_{i=1}^k w_i = 1 \text{ and } \forall_i: w_i \geq 0 \quad \dots (26)$$

$N(a; \mu_i, \Sigma_i)$  denotes the probability density function of normal distribution and is enlightened as

$$N(a; \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{|d|}{2}} \sqrt{|\Sigma_i|}} \exp\left(-\frac{1}{2}(a - \mu_i)\Sigma_i^{-1}(a - \mu_i)\right) \quad \dots (27)$$

where,  $\mu_i, \Sigma_i, i = 1, 2, \dots, k$  represent the Gaussian distribution parameters. It is highly useful and with the help of GMM and EM algorithm, classification according to likelihood is done performed.

*Classification According to Likelihood*

A strategy for supervised learning is explained here with a close observation of the basic principles and characteristics of the EM algorithm. The idea is to make modeling of GMM data within a specific class feasible, as GMM is capable of describing the original data that is complicated. Within each class, the probability density function of GMM can be used to compute the likelihood of any new instances, and thus



the maximum likelihood generated by the probability density function can be determined.<sup>56,57</sup> The first stage of classification is made clear; if the training data  $\{a_i\}_{i=1}^N$  is considered, then for each and every instance  $a_i$ , the respective correspondent label is  $q_i \in \{0,1, \dots, C\}$ . For each and every class  $c \in \{0,1, \dots, C\}$ , the  $K$  Gaussian component is utilized to model the distribution of data, thereby ending in the estimation of 3 groups of parameters namely;  $\theta = \{w_i, \mu_i, \Sigma_i, i = 1,2, \dots, K\}$ , where  $\sum w_i = 1$ . The indicator variable denotes the hidden variable only and is represented as  $f_i^j = \begin{cases} 1 \\ 0 \end{cases}$ , where 1 is satisfied for the  $j^{th}$  Gaussian component generated  $a_i$  and 0 for otherwise. Therefore the likelihood function oriented in terms and its hidden variables is represented as

$$H_0(\theta) = \Pr(a, f|\theta) = \prod_{i=1}^n \sum_{j=1}^k f_i^j w_j f(a_i; \mu_j, \Sigma_j) \quad \dots (28)$$

The hidden variables is depicted as an 2-valued indicator variable and hence the log likelihood function is written as

$$H(\theta) = \log H_0 = \prod_{i=1}^n \sum_{j=1}^k f_i^j \log [w_j g(a_i; \theta)] \quad \dots (29)$$

As a first step, the expectation is computed as

$$E(f_i^j | a_i, \theta') = 1 \times \Pr(f_i^j = 1 | a_i, \theta') + \theta \times \Pr(f_i^j = 0 | a_i, \theta') = \Pr(f_i^j | a_i, \theta') \quad \dots (30)$$

$$E(f_i^j | a_i, \theta') = \frac{\Pr(f_i^j | a_i, \theta') \Pr(a_i | \theta')}{\sum_l w_l g(a_i | \mu_l, \Sigma_l)} \quad \dots (31)$$

The maximization of the likelihood function is done by replacing  $f_i^j$  with  $E(f_i^j | a_i, \theta')$ , therefore

$$L(\theta, \theta') = \sum_{i=1}^n \sum_{j=1}^k E(f_i^j | a_i, \theta') \log [w_j g(a_i; \theta)] \quad \dots (32)$$

The conditional probability is maximize in the second stage by taking into consideration the derivatives of the parameters  $\theta$  and so the point  $\theta = \theta''$  is computed which helps in generating the zero derivate as

$$\theta'' = \arg \max_{\theta} (H(\theta, \theta'')) \quad \dots (33)$$

In a detailed manner it is expressed as follows

$$w_j'' = \frac{1}{n} \sum_{i=1}^n E(f_i^j | a_i, \theta') \quad \dots (34)$$

$$\mu_j'' = \frac{\sum_{i=1}^n (E(f_i^j | a_i, \theta') a_i)}{\sum_{i=1}^n E(f_i^j | a_i, \theta')} \quad \dots (35)$$

$$\Sigma_j'' = \frac{\sum_{i=1}^n (E(f_i^j | a_i, \theta') (a_i - \mu_j'')) (a_i - \mu_j''))^T}{\sum_{i=1}^n E(f_i^j | a_i, \theta')} \quad \dots (36)$$

Unless the likelihood becomes convergent, the iterations run continuously. For each and every class  $C$ , a set of parameters  $\theta_c$  is found and so for new condition  $\hat{a}$ , the likelihood is computed and then the class  $c$  is chosen as

$$\hat{C} = \arg \max_c (H(a, \theta_c)) \quad \dots (37)$$

Then the class  $\hat{C}$  is determined as the label of the newly formed instances. The pseudo code is explained as follows:

Input: Assumed examples  $\{a_i\}_{i=1}^N$ , and the label for each example  $q_i \in \{0,1, \dots, C\}$  and the example  $\hat{a}$  without label.

Output: The label ' $q^*$ ' of the example  $a^*$

a) Based on the labels, the training data is split into ' $C$ ' subsets.

b) For every subset in a certain class  $C$ , perform the following

(i) Initialization is done:  $\theta_c = \{w_i, \mu_i, \Sigma_i, i = 1,2, \dots, K\}$

(ii) Unless the convergence is achieved, the process is initiated.

(iii) The expectation step is computed as  $E(f_i^j | a_i, \theta')$

(iv) The parameters are stored as  $\theta_c^\theta$

(v) The expectation of  $E(f_j)$  for  $a^*$  is computed and then the likelihood for every class is computed.

(vi) The  $\hat{C} = \arg \max_c (H(a, \theta_c))$  is found out and the output  $\hat{C}$  is the label of  $a^*$ .

End;

**Hyperparameters Tuning for GMM and NBC Classifiers**

Hyperparameter tuning is the next process in using the machine learning technique we recommend for

finding the ideal model and optimizing performance using that algorithm. There are many ways to change hyperparameters, but Grid Search (GS) is used in this study because it gives each solution the best possible value.<sup>57</sup> GS is a tuning method that seeks to determine the most optimal hyperparameter values. Model parameter values are searched exhaustively in this method. Estimators are another name for the model.<sup>58,59</sup> In order to determine global optimum; the following three steps must be carried out manually:

- (i) Begin by searching in a large area with a large phase scale.
- (ii) Reduce the search area and phase size based on past results of high-performing hyperparameter settings.
- (iii) Step 2 should be repeated multiple times until the desired result is achieved.

The quality of the solution improves in direct proportion to the increase in the fitness value of the hyperparameters. After that, GS chooses the hyperparameters that will result in the strongest fitness values. The process is going to be carried out again and again until the intended optimum or perfect outcome is reached. GS produces the optimal solution by optimizing hyperparameters. The procedure for carrying out GS in order to determine the GMM hyperparameters is shown in the following algorithm. *gs* and *gs-1* reflect the current and previous GS optimizer iterations. The optimal value for maximum iteration GS (*maxit\_GS*) has been discovered to be 240. This iterative technique yields the lowest error rate and also the optimal hyperparameter values. Similarly, the GS method can be used to update hyperparameters in NBC. The hyperparameters of NBC and GMM classifiers are shown in Table 2 together with their limitation values.

**Algorithm**

```

Initialization: maxit_GS, maxit_GMM, wi, μi, Σi
For gs = 1: maxit_GS
  Build GS search space
  Approximate each mixture model's fitness value
    
```

```

  Determine the optimal weight for the mixture model
    
```

```

  For gs_GMM=1: maxit_GMM
    
```

```

    Update the parameters using Eqs (23) to (37).
    
```

```

    Again approximate each mixture model's fitness value
  
```

```

  End for
    
```

```

  Analyze the rate of error
    
```

```

  Again analyze GS search space
    
```

```

  Analyze the new GS hyperparameter tuning range
    
```

```

  End for
    
```

GS hyperparameters tuning with the GMM classifier yielded the lowest optimum values of  $w_i = 0.421$ ,  $\mu_i = 0.398$  and  $\Sigma_i = 1E-08$  (Diagonal covariance matrices), as shown in Table 2.

**Results and Discussion**

In this study, mathematical formulas are given for calculating the Performance Index (PI), sensitivity, specificity, accuracy, and Mean Square Error Rate (MSE). These formulas were used to get the results. The Performance Index is calculated as follows<sup>60</sup>

$$PI = \left( \frac{PC - MC - FA}{PC} \right) \times 100 \quad \dots (38)$$

The sensitivity, sometimes termed as the true positive rate, is calculated as follows:

$$Sensitivity = \left( \frac{PC}{PC + FA} \right) \times 100 \quad \dots (39)$$

The expression for specificity is

$$Specificity = \left( \frac{PC}{PC + MC} \right) \times 100 \quad \dots (40)$$

The expression for accuracy is

$$Accuracy = \left( \frac{Sensitivity + Specificity}{2} \right) \times 100 \quad \dots (41)$$

As stated in the Eq. (42), the Mean Square Error (MSE) is determined by subtracting the Observed Value from the original Target Value.

$$MSE = \frac{1}{N} \sum_{i=1}^N (T_i - O_j)^2 \quad (i, j = 1, \dots, k) \quad \dots (42)$$

where, N indicates the twenty patients,  $O_j$  indicates target value at j and  $T_i$  indicates the observed value at i.

Table 2 — Evaluation of hyperparameters by grid search

Classifier	Hyperparameters	Lower Value	Upper Value	Best Value
GMM	$w_i$	0	1	0.421
	$\mu_i$	0	1	0.398
	$\Sigma_i$	0	1	1E-08
NBC	Var smoothing	0	1	1.163E-09

Table 3 — Average Performance Analysis for HMM, GMM and NBC Classifier with SVD, PCA, and ICA DR methods (Before Hyperparameters Tuning)

Classifiers	DR Methods	Parameters (%)								
		PC	MC	FA	PI	Sensitivity	Specificity	Accuracy	Error Rate	MSE
HMM	SVD	<b>89.95</b>	<b>10.05</b>	<b>0</b>	<b>88.82</b>	<b>100</b>	<b>89.95</b>	<b>94.97</b>	<b>10.05</b>	<b>1.21E-06</b>
	PCA	89.65	10.35	0	88.45	100	89.65	94.82	10.35	2.84E-06
	ICA	89.90	10.1	0	88.76	100	89.9	94.95	10.10	1.89E-06
GMM	SVD	<b>90.11</b>	<b>9.89</b>	<b>0</b>	<b>89.18</b>	<b>100</b>	<b>90.11</b>	<b>95.06</b>	<b>9.89</b>	<b>1.16E-06</b>
	PCA	79.69	0	20.31	78.43	79.71	100	89.86	20.31	1.52E-05
	ICA	82.29	17.71	0	78.47	100	82.29	91.15	17.71	1.09E-05
NBC	SVD	65.23	34.76	0	46.66	100	65.23	82.62	34.76	4.76E-05
	PCA	54.69	0	45.31	17.11	54.69	100	77.35	45.31	0.000193
	ICA	<b>71.68</b>	<b>28.32</b>	<b>0</b>	<b>60.48</b>	<b>100</b>	<b>71.68</b>	<b>85.84</b>	<b>28.32</b>	<b>0.000026</b>

Table 4 — Average Performance Analysis for GMM and NBC Classifier with SVD, PCA, and ICA DR methods (After GS Hyperparameters Tuning)

Classifiers	DR Methods	Parameters (%)								
		PC	MC	FA	PI	Sensitivity	Specificity	Accuracy	Error Rate	MSE
GMM	SVD	<b>96.35</b>	<b>3.64</b>	<b>0</b>	<b>96.21</b>	<b>100</b>	<b>96.35</b>	<b>98.18</b>	<b>3.64</b>	<b>1.6E-07</b>
	PCA	83.34	16.66	0	80.01	100	83.34	91.67	16.66	9E-06
	ICA	89.07	10.93	0	87.71	100	89.07	94.53	10.93	2.89E-06
NBC	SVD	72.66	27.34	0	62.32	100	72.66	86.33	27.34	2.401E-05
	PCA	65.88	34.12	0	48.16	100	65.88	82.94	34.12	4.624E-05
	ICA	<b>78.39</b>	<b>21.61</b>	<b>0</b>	<b>76.63</b>	<b>100</b>	<b>78.41</b>	<b>89.21</b>	<b>21.61</b>	<b>1.681E-05</b>

$$\text{Error Rate} = \left\{ \frac{(MC + FA)}{(PC + MC + FA)} \right\} \times 100 \quad \dots (43)$$

Here, PC, MC, and FA stands for Perfect Classification, Missed Classification, and False Alarm, respectively. The average performance analysis for before hyperparameter tuning of HMM, GMM and NBC classifiers with SVD, PCA and ICA DR methods is shown in Table 3. It is identified from Table 3 that the higher accuracy of 94.97% is attained in the SVD DR technique for HMM classifiers. The HMM classifier is plagued by the problem of more missed classifications. A higher accuracy of 95.06% is attained in the SVD DR technique for the GMM classifier. The GMM classifier performs better in terms of missed classification and false alarms among the dimensionality reduction techniques with lower MSE values. The higher accuracy of 85.84% is attained in the ICA DR technique for the NBC classifier. The NBC classifier is poorly performing in terms of missed classification and false alarms among the dimensionality reduction techniques with higher MSE values.

The average performance analysis for after GS hyperparameter tuning of GMM and NBC classifiers with SVD, PCA and ICA DR methods is given in Table 4. It is identified from Table 4 that the higher

accuracy of 98.18% is attained in the SVD DR technique for GMM classifiers. A higher accuracy of 89.21% is attained in the ICA DR technique for the NBC classifier. The GMM and NBC classifiers with GS hyperparameters tuning performs better in terms of missed classification and false alarms among the dimensionality reduction techniques with lower MSE values. As a consequence of the research results, the GS with GMM classifier performed much better than the other classifiers evaluated on the EEG features. It would appear that this work is successful in both the dimensionality reduction procedures and the classification of epilepsy.

**Conclusions**

Identifying a better classifier for a quick and efficient classification is very important in the selection of classifiers. The above study is also dealt with in the same way. While achieving better accuracy, we need to compromise on a certain error in classification that is a trade-off condition by fixing the MSE value to an appreciable one. Results demonstrate that the GMM Classifier with GS hyperparameter tuning achieved 98.18% accuracy for classifying epilepsy from EEG signals in SVD dimensionality reduction. Based on the results, the GMM classifier with GS hyperparameter tuning has been superior at classifying EEG signals than the

other classifiers. Further research will be conducted in the direction of bio-inspired classifiers with different hyperparameters tuning for the above-mentioned problem.

## References

- 1 Fisher R S, Blum D E, DiVentura B, Vannest J, Hixson J D, Moss R, Herman S T, Fureman B E & French J A , Seizure diaries for clinical research and practice: limitations and future prospects, *Epilepsy Behav*, **24(3)** (2012) 304–310.
- 2 Glauser T, Shinnar S, Gloss D, Alldredge B, Arya R, Bainbridge J, Bare M, Bleck T, Dodson W E, Garrity L & Jagoda A, Evidence-based guideline: treatment of convulsive status epilepticus in children and adults: report of the Guideline Committee of the American Epilepsy Society, *Epilepsy Curr*, **16(1)** (2016) 48–61.
- 3 Witte H, Iasemidis L D & Litt B, Special issue on epileptic seizure prediction, *IEEE Trans Biomed Eng*, **50(5)** (2003) 537–539.
- 4 Sarfo F S, Awuah D O, Nkyi C, Akassi J, Opare Sem O K & Ovbiagele B, Recent patterns and predictors of neurological mortality among hospitalized patients in Central Ghana, *J Neurol Sci*, **363** (2016) 217–224.
- 5 GoizMárquez G, Chacón S C, Ortiz H S & López H S, Canine epilepsy, *Vet Mex*, **39(3)** (2008) 279–321.
- 6 Anjum S M M, Käufer C , Hopfengärtner R, Waltl I, Bröer S & Löscher W, Automated quantification of EEG spikes and spike clusters as a new read out in Theiler's virus mouse model of encephalitis-induced epilepsy, *Epilepsy Behav*, **88** (2018) 189–204.
- 7 Bhati D, Pachori R B, Sharma M & Gadre V M, Automated detection of seizure and nonseizure EEG signals using two band biorthogonal wavelet filter banks, *Proc Biomed Signal Process* (Springer, Singapore) 2020, 137–155.
- 8 Acharya U R, Oh S L, Hagiwara Y, Tan J H, Adeli H & Subha D P, Automated EEG-based screening of depression using deep convolutional neural network, *Comput Methods Programs Biomed*, **161** (2018) 103–113.
- 9 Aarabi A & He B, Seizure prediction in patients with focal hippocampal epilepsy, *Clin Neurophysiol*, **128(7)** (2017) 1299–1307.
- 10 Acharya U R, Hagiwara Y & Adeli H, Automated seizure prediction, *Epilepsy Behav*, **88** (2018) 251–261.
- 11 Acharya U R, Hagiwara Y, Deshpande S N, Suren S, Koh J E W, Oh S L, Arunkumar N, Ciaccio E J & Lim C M, Characterization of focal EEG signals: a review, *Future Gener Comput Syst*, **91** (2019) 290–299.
- 12 López S, Suarez G, Jungreis D, Obeid I & Picone J, Automated identification of abnormal adult EEGs, *Proc IEEE Signal Process Med Biol Symp* (SPMB) 2015, 1–5.
- 13 Veisi I, Pariz N & Karimpour A, Fast and robust detection of epilepsy in noisy EEG signals using permutation entropy, *Proc IEEE 7<sup>th</sup> Int Conf Bioinform* 2007, 200–203.
- 14 Faust O, Acharya U R, Adeli H & Adeli A, Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis, *Seizure*, **26** (2015) 56–64.
- 15 Liu Y, Zhou W, Yuan Q & Chen S, Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG, *IEEE Trans Neural Syst Rehabilitation Eng*, **20(6)** (2012) 749–755.
- 16 Wang D, Miao D & Xie C, Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection, *Expert Syst Appl*, **38(11)** (2011) 14314–14320.
- 17 Prabhakar S K & Rajaguru H, Development of patient remote monitoring system for epilepsy classification, *Proc 16<sup>th</sup> Int Conf Adv Biomed Eng* (Springer, Singapore) 2017, 80–87.
- 18 Song Y, Crowcroft J & Zhang J, Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine, *J Neurosci Methods*, **210(2)** (2012) 132–146.
- 19 Zhou Y, Huang R, Chen Z, Chang X, Chen J & Xie L, Application of approximate entropy on dynamic characteristics of epileptic absence seizure, *Neural Regen Res*, **7(8)** (2012) 572–577.
- 20 Dash D P, Kolekar M H & Jha K, Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and Hidden Markov Model, *Comput Biol Med*, **116** (2020) 103571.
- 21 Riaz F, Hassan A, Rehman S, Niazi I K & Dremstrup K, EMD-based temporal and spectral features for the classification of EEG signals using supervised learning, *IEEE Trans Neural Syst Rehabilitation Eng*, **24(1)** (2015) 28–35.
- 22 Pachori R B & Patidar S, Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions, *Comput Methods Programs Biomed*, **113(2)** (2014) 494–502.
- 23 Srinivasan V & Eswaran C, Artificial neural network based epileptic detection using time-domain and frequency-domain features, *J Med Syst*, **29(6)** (2005) 647–660.
- 24 Kumar Y, Dewal M L & Anand R S, Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network, *Signal Image Video Process*, **8(7)** (2014) 1323–1334.
- 25 Yuan Q, Zhou W, Liu Y & Wang J, Epileptic seizure detection with linear and nonlinear features, *Epilepsy Behav*, **24(4)** (2012) 415–421.
- 26 Harikumar R, Ganesh Babu C & Vijayakumar T, Performance analysis of elman neural networks as post classifiers for wavelet transforms based feature extraction using hard and soft thresholding methods in the classification of epilepsy risk levels from EEG signals, *Eur J Sci Res*, **71** (2012) 221–232.
- 27 American electroencephalographic society guidelines for standard electrode position nomenclature, *J clin Neurophysiol*, **8(2)** (1991) 200–202
- 28 J Lu M, Pellouchoud E & Gevins A, A rapid method for determining standard 10/10 electrode positions for high resolution EEG studies, *Electroencephalogr Clin Neurophysiol*, **106** (1998) 554–558.
- 29 Stewart G W, On the early history of the singular value decomposition, *SIAM Rev*, **35** (1993) 551–566.
- 30 Freire S L M & Ulrych T J, Application of singular value decomposition to vertical seismic profiling, *Geophysics*, **53** (1988) 778–785.
- 31 Wall, Michael E, Andreas Rechtsteiner & Luis M Rocha, Singular value decomposition and principal component analysis, *Proc Pract Appr Micro Data Anal* (Springer, Boston, MA) 2003, 91–109.

- 32 Jiang F, Kannan R, Littman M & Vempala S, Efficient singular value decomposition via improved document sampling, *Tech Rep CS-99-5, Department of Computer Science* (Duke University), 1999.
- 33 Abdi H, Singular value decomposition (SVD) and generalized singular value decomposition (GSVD), in *Encyclopedia of Measurement and Statistics* edited by N J Salkind, (Thousand Oaks: Sage Publications) 2007, 907–912.
- 34 Abdi H & Williams L J, Principal component analysis, *Wiley Interdiscip Rev Comput Stat*, **2(4)** (2010) 433–459.
- 35 Wold S, Esbensen K & Geladi P, Principal component analysis, *Chemometr Intell Lab Syst*, **2(1–3)** (1987) 37–52.
- 36 Klema V C & Laub A J, The Principal Component Analysis : its Computation and Some Applications, *IEEE Trans Auto Control*, **25** (1980) 164–176.
- 37 Dray S, On the number of principal components: A test of dimensionality based on measurements of similarity between matrices, *Comput Stat Data Anal*, **52(4)** (2008) 2228–2237.
- 38 Saporta G & Niang N, Principal component analysis: application to statistical process control, in *Data Analysis* edited by G Govaert, (London: John Wiley & Sons) 2009, 1–23.
- 39 Abdi H, Eigen-decomposition: eigenvalues and eigenvectors, in *Encyclopedia of Measurement and Statistics* edited by N J Salkind (Thousand Oaks: Sage Publications) 2007, 304–308.
- 40 McSharry P E & Clifford G D, A comparison of nonlinear noise reduction and independent component analysis using a realistic dynamical model of the electrocardiogram, *Proc Fluct Noise Biol, Bioph, Biom Syst II* (SPIE) **5467**, (2004) 78–88.
- 41 Penny W D, Everson R M & Roberts S J, Hidden Markov independent component analysis, *Proc Adv Indepe Comp Anal* (Springer, London) 2000, 3–22.
- 42 Lee T W, Girolami M, Bell A J & Sejnowski T J, A unifying information-theoretic framework for independent component analysis, *Comput Math Appl*, **39** (2000) 1–21.
- 43 Xu L, Cheung C C, Yang H H & Amari S I, Independent component analysis by the information-theoretic approach with mixture of densities, *Proc Int Jt Conf Neural Netw* (ICNN'97-IEEE) **3**, 1997, 1821–1826.
- 44 Oja E, Nonlinear PCA criterion and maximum likelihood in independent component analysis, *Proc ICA'99 Int Work Indep Comp Anal Signal Separat*, (Aussois, France) 1999, 143–148.
- 45 Choudrey R A, Bayesian ICA with hidden Markov sources, *Proc Int Conf Indep Com Ana*, 2003.
- 46 Acharya U R, Sree S V, Swapna G, Martis R J & Suri J S, Automated EEG analysis of epilepsy: a review, *Knowl Based Syst*, **45** (2013) 147–165.
- 47 Arunkumar N, Balaji V S, Ramesh S, Natarajan S, likhita V R & sundari S, Automatic detection of epileptic seizures using independent component analysis algorithm, *Proc IEEE-Int Conf Adv Eng Sci Manage* (ICAESM -2012) 2012, 542–544.
- 48 Schuller B, Rigoll G & Lang M, Hidden Markov model-based speech emotion recognition, *Proc - IEEE Int Conf Acoust Speech Signal Process*, (ICASSP'03) 2003.
- 49 Makhoul J, Roucos S & Gish H, Vector Quantization in Speech Coding, *proc IEEE*, **73** (1985) 1551–1588.
- 50 Pan J S, Lu Z M & Sun S H, An efficient encoding algorithm for vector quantization based on subvector technique, *IEEE Trans Image Process*, **12(3)** (2003) 265–270.
- 51 Kekre H B & Kulkarni V, Speaker identification by using vector quantization, *Int J Eng Sci Technol*, **2(5)** (2010) 1325–1331.
- 52 Figueiredo M A, Scalar and vector quantization, *Instituto Superior Tecnico* (Lisboa, Portugal) 2008.
- 53 Huang J, Lu J & Ling C X, Comparing naive Bayes, decision trees, and SVM with AUC and accuracy, *Proc IEEE Int Conf Data Mining* (IEEE) 2003.
- 54 Karim M & Rahman R M, Decision tree and naive bayes algorithm for classification and generation of actionable knowledge for direct marketing, *J Softw Eng Appl*, **06** (2013) 196–206.
- 55 Chen Z & Ellis T, A self-adaptive Gaussian mixture model, *Comput Vis Image Underst*, **122** (2014) 35–46.
- 56 Xuan G, Zhang W & Chai P, EM algorithms of Gaussian mixture model and hidden Markov model, *Proc Int Conf Image Process* (Cat. No. 01CH37205) 2001, 145–148.
- 57 Shekar B H & Dagnev G, Grid search-based hyperparameter tuning and classification of microarray cancer data, *2<sup>nd</sup> Int Conf Adv Comput Commun Para* (ICACCP) 2019, 1–8.
- 58 CG S G & Sumathi B, Grid search tuning of hyperparameters in random forest classifier for customer feedback sentiment prediction, *Int J Adv Comput Sci Appl*, **11(9)** (2020) 173–178.
- 59 Wang X, Gong G & Li N, Automated recognition of epileptic EEG states using a combination of symlet wavelet processing, gradient boosting machine, and grid search optimizer, *Sensors*, **19(2)** (2019) 219–237.
- 60 Harikumar R, Ganesh Babu C & Gowri Shankar M, Extreme learning machine (ELM) based performance analysis and epilepsy identification from EEG signals, *IETE J Res*, (2021) 1–11.