Comprehensive Performance Analysis of Classifiers in Diagnosis of Epilepsy

1. Introduction

The electroencephalogram (EEG) is the measurement of combined neuron firing in different portions of the brain. It includes information derived from a series of monitoring electrodes about fluctuations in the brain’s electrical potential. EEG variations are said to have been influenced by a broad variety of factors, including circulatory, biochemical, hormonal, metabolic, neuro-electrical, and behavioral factors [1]. The most significant behavior that can be identified from the EEG record is epilepsy. Epilepsy is characterized by either a segment or all of the central nervous system with uncontrolled excessive activity or potential discharge. Various EEG waveform patterns describe the several forms of epileptic seizures. Among the most substantial works, the description of EEG signals for epilepsy is as follows. The epileptic seizure classification, with the aid of neural networks, was implemented [2]. The multiclass epileptic seizure classification uses the time-frequency study of EEG signals which was performed [3]. An important function for predicting seizure onset using line length has been achieved [4]. The fuzzy mutual information (FMI), as a strategy for minimizing dimensions, is categorized with the aid of the firefly algorithm [5]. The classification of epilepsy through the multilabel dimensionality reduction techniques uses the genetic algorithm [6]. The feature extraction and recognition of ictal EEG were performed using EMD and SVM [7]. The electroencephalogram uses the empirical mode decomposition and Fourier–Bessel expansion (FB). In this, the EEG signal decomposes the EMD into the finite band limited...
signals as intrinsic mode function (IMF), and the FB calculates the mean frequency for each IMF. It improves the performance of the ictal and the seizure-free signals [8]. In this, a new fractional-order calculus is used for the EEG signal classification, and the technique of fractional linear prediction is used to model the ictal and seizure-free signals. It results that the ictal signals have more energy than the seizure-free signals [9].

The organization of this paper is as follows. The description of the dataset is defined in Section 2. The EM-based dimensionality reduction technique and the classifiers are analyzed are discussed next. The performance of the classifiers is accounted in Section 3. Finally, the paper is concluded for the future scope.

2. Materials and Methods

The database of the EEG signal is taken from the University of Bonn (UoB), Germany, is widely available for researchers to use online, and is used in this work. There are 5 separate datasets that are open to the public as well as to the academic community. The research was performed for 5 patients, and the progress of the patients was examined for ictal seizure disorder. Epileptic seizure (ictal) signals are obtained from the intracranial electrodes, which are placed within the epileptogenic zone and are represented in dataset E. In the ictal seizure activity, there were 100 epochs and time duration of each epoch is about 23.6 seconds with the 4096 samples. EEG readings were recorded with the assistance of a 128-channel amplifier module. The sampling rate considered at this point was 176.31 Hz, and there were 12 bits in analog-to-digital resolution for digitization purposes [10]. These ictal patterns could be used to localise recurring seizures in epilepsy, which could help with diagnosis, treatment, and prognosis. Further consideration is given to the preprocessing stage of the EEG signals since it is essential to use the technique to extract all the beneficial information hidden in the nonlinear biomedical signals. The architecture of the comprehensive epilepsy diagnostic system is shown in Figure 1.

The raw EEG signals are taken, and then, expectation maximization features are extracted, using the dimensionality reduction technique. The reduced features are given as input to the classifiers. Eventually, the diagnosis of epilepsy is carried out from EEG signals by involving the nonlinear models, and the hybrid model as the classifiers and the performance metrics classifier are analyzed.

2.1. Expectation Maximization for Dimensionality Reduction Technique. The primary goal of the dimensionality reduction is for the transformation of the higher dimensional data to diminished dimensional data with deeper insight. For a single patient, the amount of data is about [4096 \times 100] samples, which is too intense to handle. Expectation-maximization (EM) algorithm is an iterative method for finding the maximum likelihood and maximum a posteriori estimates of parameters in models that typically depend on hidden variables. This method is based on the finite mixture model. Therefore, using EM, the dimensions are reduced to [4096 \times 10]. A particular set of features in the field of pattern recognition can easily represent each pattern, and it can be visualized as the n-dimensional feature space [11]. Consider there are ‘n’ observations, in which each observation is presumed with the random variable that is realized with m dimensions and is expressed in (1), and the mean A(x) is derived from the observations as shown in (2):

\[ x = (x_1, x_2, \ldots, x_m)^T, \]  

\[ A(x) = \mu = (\mu_1, \mu_2, \ldots, \mu_m)^T. \]

The corresponding covariance matrix and the observation matrix X is shown in (3) and (4)

\[ A\{(x - \mu)(x - \mu)^T\} = \Sigma_{(m \times m)}, \]  

\[ X = \{x_{ij}; 1 \leq i \leq m, 1 \leq j \leq n\}. \]

The arithmetic mean is symbolized as \( \mu_i \), and therefore, SD of \( i^{th} \) random variable is designated as \( \sigma_i \) and expressed as (5), and the observation matrix \( x_{ij} \) is standardized:

\[ \sigma_i = \sqrt{\Sigma_{i,j}}. \]

The linear dimension reduction technique is applied, and it is taken into account because the best is within the logic of mean square error. The principal component analysis (PCA) reduces the dimension of the data and trains the original variables with linear combination, which is orthogonal. The initial principal component \( c_1 \) has maximum variance and is expressed as
The empirical covariance matrix is $\Sigma = YY^T$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ correspond to the diagonal matrix; it describes the eigenvalue $\lambda_1 \leq \ldots \leq \lambda_m$ and the orthogonal matrix $Y = mXm$, which includes the eigenvectors too. The principal component is simplified by $m \times n$ rows of $m \times n$ matrix $C$ and arranged as $C = Y^T X$.

where $Y^T$ confer the weight matrix $W$.

2.2. Procedure for the Expectation Maximization Algorithm. In PCA, training the principal component from larger dataset is somewhat difficult, so an expectation maximization (EM) algorithm is preferred all along with the PCA. The EM-PCA comprises of E-step and M-step and is shown in (10) and (11):

$$E - \text{step} : R = (S^T S)^{-1} S^T Q,$$

$$M - \text{step} : S^{\text{new}} = QR^T (RR^T)^{-1},$$

where $Q$ denotes a $(m \times n)$ matrix with the entire observed data and $R$ represents $(h \times n)$ matrix with unknown states of large number. The EM learning algorithm is applied for PCA; it follows an iterative algorithm and discovers the subspace with ‘h’ leading eigenvectors. The reduced EM features are checked for nonlinearity and non-Gaussian through the process of probability plot, scatter plot, and the statistical parameters.

Figure 2 shows the normal probability plot for EM clusters in dimensionally reduced epilepsy set $E$, and it is observed that the EM cluster values are highly nonlinear and overlapping in nature.

Figure 3 depicts the scatter plot for adjacent clusters in EM features in dimensionally reduced epilepsy set $E$. Based on the projection and statistical test, this approach works well in simple cases in which the clusters are well separated; from Figure 3, it is noted that the clusters are overlapped and nonlinear, as well as non-Gaussian. Therefore, the classification will be performed by the nonlinear or heuristics-based classifiers.

Table 1 indicates the average statistical parameters for EM clusters in dimensionally reduced epilepsy set $E$. The mean and variance values of EM clusters are at the lower end. The skewness value indicates the right skewed condition. The Kurtosis demonstrates the flatness in the EM clusters.

The entropy exhibits the nonlinearity in the EM clusters. The PCC indicates no linear relation among the adjacent EM clusters in dimensionally reduced epilepsy set $E$.

2.3. Nonlinear and Hybrid Classifiers for Diagnosis of Epilepsy from EM Features. The Gaussian mixture model (GMM), logistic regression (LR), and firefly algorithm (FA) are the nonlinear model classifiers used for the diagnosis of the epilepsy from the dimensionally reduced EEG signals. Nonlinear classification approaches may be limited by the high-dimensional and noisy features of EEG. To improve the performance, the hybrid models preferred are the Cuckoo search with GMM and firefly algorithm with GMM for diagnosis of the epilepsy from EEG signals.

2.4. Gaussian Mixture Model. The initialization of the probability density of GMM is expressed as

$$p(b) = \sum_{n=1}^{N} v_n P_n(b) \sum_{n=1}^{N} v_n Q(b|\mu_n, \sum_n),$$

where $N$ is the GMM model fixed number, the model weighting coefficient is represented as $v_n$ and $\sum v_n = 1$, and the $n^{\text{th}}$ single Gaussian probability density function is represented as $Q(b|\mu_n, \sum_n)$, and it is expressed as

$$Q = \frac{e^{-(1/2) (b-\mu)^T \sum^{-1} (b-\mu)}}{(2\pi)^{d/2} \sum^{1/2}}$$

The estimate of the parameters is accomplished since the probability density function is evaluated by the maximum probability [12]. The evaluation of the GMM is carried out according to the maximum expected value algorithm.

Step 1. initiation of the parameters:

(i) The mean values are set at random
(ii) Covariance matrix is set to be the unit matrix
(iii) The previous likelihood for each model is allocated by the weighting coefficient

Step 2. the calculation of the prior probability model for each item is as

$$P_r(i b_t, \theta^r) = \frac{v_n Q(b_t | \mu_n, \sum_n)}{\sum_{n=1}^{N} v_n Q(b_t | \mu_n, \sum_n)}$$

Step 3. the parameters are updated using the prior probability.

Step 4. repeat Steps 2 and 3, until the convergence is attained. It is represented as
\[ |\theta^{t+1} - \theta^t| < \varepsilon. \]

The estimation of the parameters at the previous stage is represented as \(\theta^{t+1} \); the current step is denoted as \(\theta^t\). \(\varepsilon\) is the threshold set at \(10^{-7}\).

2.5. Logistic Regression. In statistical modeling, LR is one of the most utilized techniques, where the probability of \(P_1\) of an outcome is closely related to variables as follows:

\[
\text{logit}(P_1) = \ln\left(\frac{P_1}{1 - P_1}\right) = \beta_0 + \beta_1 y_1 + \cdots + \beta_q y_q, \quad (16)
\]

\[
\text{logit}(P_1) = \beta_0 + \sum_{i=1}^{q} \beta_i y_i, \quad (16)
\]

where \(\beta_0\) signifies the interrupt and the coefficients associated with the explanatory variables and \(y_0, y_1, \ldots, y_q\) are represented by \(\beta_0, \beta_1, \ldots, \beta_q\). A discrete, dichotomous, or continuous combination measure may be used for independent variables. The probability of a nonoccurring event’s natural log ratio is expressed through the dependent
variable. There is no constraint between the response variable and the explanatory variable for the independent distribution formed by the Gaussian. The nonlinear relationship between these variables exists, so the likelihood of the occurrence is expressed as

$$P_i (y) = \frac{1}{1 + e^{-\beta_0 \log(P_i (y))}}$$  

(17)

$$= \frac{1}{1 + e^{-\sum_{i=1}^{q} \beta_i y_i}}.$$  

The probability values are established in the range of 0 and 1 by the logistic regression.

To determine the coefficients of $\beta_0, \beta_1, \ldots, \beta_q$, the maximum likelihood estimation (MLE) technique is used. MLE maximizes the probability of the log and clears the way for the dependent variable prediction. A new estimate is computed with the aid of an improved function, unless the convergence criterion is reached.

2.6. Firefly Algorithm. The metaheuristic algorithm is versatile and has the techniques as follows:

(a) Each firefly tends to attract each other with the help of brightness

(b) Fireflies that have greater brightness are attributed to greater attractiveness

(c) Lower-bright fireflies are moved to higher-bright fireflies

Every firefly has the optimal solution and will have its own brightness in relation to optimal solution fitness function [13]. Assume that each $i (P_i)$ is a firefly $i$ located at the current iteration. If solution $i$ is greater than function $j$, the distance between the corresponding fireflies $i$ and $j$ is obtained from

$$d_{ij} = \sqrt{(P_i - P_j)^2}.$$  

(18)

This updated feature now takes the place of the following function in order to assess the new attractiveness, and it is shown as

$$\beta = \beta_0 e^{-\nu d_{ij}}.$$  

(19)

By considering the new generation of the $i^{th}$ solution, the new location of the $i^{th}$ firefly can be easily calculated as follows:

$$P_{\text{new}} = P_i + \beta \cdot \text{rand} \cdot \Delta P_{ij} + \text{rand}.$$  

(20)

The attractiveness factor is denoted by $\beta_0$, the random number of the solution $i$ is named as rand, the value is 0.1, and the absorption coefficient $\nu$ value is 0.1. $P_i$ is a solution with a fitness function lower than $P_i$. The model calculates the updated size:

$$\Delta P_{ij} = (P_j - P_i).$$  

(21)

To conclude, for each solution, there is a probability of new solution depending on the comparative fitness levels $i$.

In this, the new set of solution $i$ is evaluated by fitness function parameters, and best one with the lowest fitness value is kept and the rest are to be discarded.

2.7. Cuckoo Search with GMM. Cuckoo search is a metaheuristic algorithm inspired by nature that focuses on some cuckoo species’ mandatory parasitic behavior [14]. This algorithm was recognized as Levy flights, rather than random walking techniques. The benefits of the random measures use Levy flights, enabling the optimum value in a shorter period. For cuckoo birds, there is a special breeding system in which the egg is laid in the nests of other host birds. This egg is believed by the host bird to be its own egg and takes possession of it. The host bird does not notice the egg or the host bird detects the egg. Two conditions are satisfied by this condition. If the host bird realizes that it does not belong to the egg, this egg is either killed or abandoned and three idealized laws are offered:

(1) A cuckoo lays an egg inside a randomly selected nest at a time

(2) The nests are passed on from one generation to the next generation with high egg content; the best nests are retained

(3) The number of host nests available is fixed in any generation

The host bird is likely to find the eggs a cuckoo has laid: $P_a \in [0, 1]$. The host bird can choose either to destroy the nest or to abandon it. The last hypothesis can be approximated by substituting the $P_a$ fraction of the nests for new nests. A Levy flight is performed when creating new solutions $x_{i+1}^t$ for a cuckoo $i$ as shown in (22) and (23):

$$x_{i+1}^t = x_i^t + \alpha \lambda \text{Levy}(\lambda),$$  

(22)

$$\text{Levy} \sim \mu = \tau \cdot \lambda,$$  

(23)

where $\lambda$ value is 4 and 0.2, the solution $i$ is given by $x_{i+1}^t$ at times $t+1$ and $x_i^t$ at times $t$, $a > 0$ is the step size that ought to be aligned with the rules of the problem, and the product $\sigma$ implies input-wise multiplications.

Basically, Levy flight provides a random walk, and its random steps are taken from a distribution of Levy that has an infinite variance with an infinite mean for large steps. The new solutions would be extracted by Levy Walk, round the best solution, and speed up the local search process. However, a fraction of the new solutions is generated spontaneously from the places, where the best solution was far from the best place at that time. It ensures that the system is not stuck to a minimal local standard.

The following steps can be summarized as a cuckoo search algorithm:

(1) Generate an initial nest population at random locations, $X = \{x_i^0, x_2^0, \ldots, x_n^0\}$, then evaluate their target values, and find the current best global $g_i^0$.

(2) Use Levy walk to update the most recent positions.
Table 2: Average performance analysis of classifier for EM clusters.

| Parameters     | Gaussian mixture model | Logistic regression | Firefly algorithm | Cuckoo search with GMM | Firefly with GMM |
|----------------|------------------------|---------------------|-------------------|------------------------|-----------------|
| PC (%)         | 82.55625               | 73.4375             | 68.16425          | 84.38                  | 76.0425         |
| MC (%)         | 17.44375               | 0                   | 0                 | 0                      | 0               |
| FA (%)         | 0                      | 26.5625             | 31.83875          | 15.62                  | 23.9575         |
| PI (%)         | 78.85125               | 63.765              | 53.27363          | 81.4325                | 69.7275         |
| Sensitivity (%)| 100                    | 73.4375             | 68.16425          | 84.38                  | 76.05           |
| Specificity (%)| 82.55625               | 100                 | 100               | 100                    | 100             |
| Accuracy (%)   | 91.27813               | 86.71875            | 84.08213          | 92.19                  | 88.025          |
| MSE (%)        | 1.02E-05               | 2.21E-05            | 3.84E-05          | 7.84E-06               | 1.94E-05        |
| GDR (%)        | 78.87047               | 73.4375             | 84.16221          | 84.38                  | 76.0425         |
| Error rate (%) | 17.44375               | 26.5625             | 31.83779          | 15.62                  | 23.9575         |

(3) A uniform distribution [0, 1] produces a random number \( r \). If \( r > Pa \), the abandon parameter is less than \( r \), the update location of \( x_i^{\tau+1} \), and then evaluates and finds the new best \( g_i^\tau \) globally.

(4) If stopping condition is achieved, then \( g_i^\tau \) is the best global solution, else return to Step 2.

The cuckoo search will have a limitation of the local minima problem. This will reduce the performance of the classifiers. Therefore, it is suggested that, in a hybrid classifiers, which includes cuckoo search with the GMM model to improve the performance.

The EM features processed are given as the input to the cuckoo search; it finds the best global solution and is then processed with the Gaussian mixture model. Here, the GMM uses the maximum likelihood and evaluates the probability density function. The dimension of the features are further reduced from \([4096 \times 10]\) to \([4096 \times 1]\) using the GMM.

2.8. Firefly Algorithm with GMM. This algorithm is a metaheuristic optimization algorithm inspired by fireflies’ flashing actions, and it is hybrid model combined with the Gaussian mixture model. The hybridization approach is also used to enhance the optimization process. The GMM is used to adapt the FA parameters and improve algorithm in terms of global search and local search processes; the dimensions are reduced to \([4096 \times 1]\).

3. Results and Discussion

Considering the expectation maximization as dimensionality reduction techniques and GMM, LR, firefly algorithm, Cuckoo search with GMM, and firefly algorithm with GMM are the classifiers used in this work. Based on the performance index, mean square error (MSE) rate, accuracy, and error rate, the result obtained are shown in Table 2. The mathematical formulae for computing the performance index (PI), sensitivity, specificity, accuracy, good detection rate (GDR), error rate, and mean square error (MSE) rate are shown from (24) to (30).

Performance index (PI):

\[
PI = \left( \frac{PC - MC - FA}{PC} \right) \times 100,
\]

where PC is the perfect classification, MC is the missed classification, and FA is the false alarm.

The mathematical formulae define the measure of sensitivity, specificity, and accuracy are

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

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

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

The good detection rate (GDR) is represented as

\[
\text{GDR} = \frac{PC - MC}{PC + FA} \times 100.
\]

The mean square error (MSE) is expressed as follows:

\[
\text{MSE} = \frac{1}{N} \sum_{j=1}^{N} (O_j - T_j)^2.
\]

Figure 4 shows analysis of the performance index with accuracy and GDR.

Thus, the paper gives the performance analysis by taking into account the EM features for the dimensionality reduction technique; the nonlinear models and hybrid models serve as classifiers for the diagnosis of epilepsy from EEG signals. From Table 2, it is inferred that cuckoo search with GMM achieves 92.19% accuracy, performance index of 81.43%, and low error rate of 15.62% rather than other classifiers. The low error rate of cuckoo search with GMM depends on the MC, FA, and PC. In that, the MC value is zero and the FA value is 15.62%. The firefly algorithm results with low accuracy because the perfect classification agrees on the same epilepsy risk level by the physician and the messenger and was very low. Figure 5 shows the comparison of
all the five classifiers with respect to sensitivity, specificity, and accuracy.

4. Conclusions

This paper aims to identify the diagnosis of epileptic patients from the EEG signals. The EEG signal-derived parameters are stored as dataset. The EM feature is used as the dimensionality reduction technique. The reduced features are classified for epilepsy using five classifiers. The hybrid classifier, namely, cuckoo search with GMM slated at 92.19% of accuracy with the low error rate than the other four classifiers. For the diagnosis of epilepsy, the same technique will be applied to other EEG sets such as A, B, C, and D. Further research will be in the direction of deep learning (use the A to F datasets), convolutional neural network, and autoencoder techniques and acts as the postclassifier for the diagnosis and prognosis of epilepsy from the EEG signals.

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. Sukanesh and R. Harikumar, “Diagnosis and classification of epilepsy risk levels from EEG signals using fuzzy aggregation techniques,” Engineering Letters, vol. 14, no. 1, 2007.
[2] R. P. Costa, P. Oliveira, G. Rodrigues, B. Direito, and A. Dourado, "Epileptic seizure classification using neural networks with 14 features," in Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, pp. 281–288, Springer, Zagreb, Croatia, September 2008.

[3] P. P. Acharyee and C. Shahnaz, "Multiclass epileptic seizure classification using time-frequency analysis of EEG signals," in Proceedings of the 7th International Conference on Electrical and Computer Engineering - ICECE 12, pp. 260–263, Dhaka, Bangladesh, December 2012.

[4] R. Esteller, J. Echauz, T. Cheng, B. Litt, and B. Pless, "Line length: an efficient feature for seizure onset detection," in Proceedings of the 23rd International Conference IEEE Engineering Medicine Biology Soc., vol. 2, pp. 1707–1710, Berlin, Germany, July 2001.

[5] R. Harikumar and P. S. Kumar, "Principal component analysis as a dimensionality reduction technique and sparse representation classifier as a post classifier for the classification of epilepsy risk levels from EEG," Journal of Pharmaceutical Sciences and Research, vol. 7, no. 6, pp. 282–284, 2015.

[6] H. Rajaguru and S. K. Prabhakar, "Epilepsy classification through multi-label dimensionality reduction through dependence maximization and elite genetic algorithm," in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, pp. 594–597, Coimbatore, India, April 2017.

[7] S. Li, W. Zhou, Q. Yuan, D. Cai, and S. Geng, "Feature extraction and recognition of ictal EEG using EMD and SVM," Computers in Biology and Medicine, vol. 43, no. 7, pp. 807–816, 2013.

[8] R. B. Pachori, "Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition," Journal of Electrical and Computer Engineering, vol. 2008, Article ID 293056, 5 pages, 2008.

[9] D. Pandey, B. K. Pandey, and S. Wairya, "Hybrid deep neural network with adaptive galactic swarm optimization for text extraction from scene images," Soft Computing, vol. 25, pp. 1563–1580, 2021.

[10] S. Piyush, T. K. Gandhi, K. P. Bijaya, M. Tripathi, and S. Anand, "A novel robust diagnostic model to detect seizures in electroencephalography," Expert Systems with Applications, vol. 56, pp. 116–130, 2016.

[11] S. K. Prabhakar and H. Rajaguru, "Expectation maximization based PCA and Hessian LLE with suitable post classifiers for epilepsy classification from EEG signals," in Proceedings of the International Conference on Soft Computing and Pattern Recognition, pp. 364–374, Vellore, India, December 2016.

[12] R. Anand, T. Shanthi, R. S. Sabeenian, M. E. Paramasivam, and K. Manju, "GLCM feature-based texture image classification using support vector machine," in Proceedings of the 3rd EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing, pp. 143–154, Springer, Cham, Switzerland, 2022.

[13] A. M. Boopathi and A. Abudhahir, "Firefly algorithm tuned fuzzy set-point weighted PID controller for antilock braking systems," Journal of Engineering Research, vol. 3, no. 2, pp. 79–94, 2015.

[14] R. Anand, S. Veni, and J. Aravinth, "Robust classification technique for hyperspectral images based on 3D-discrete wavelet transform," Remote Sensing, vol. 13, no. 12, p. 1255, 2021.