Journal of Information Systems Engineering and Management
2025, 10(28s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Improving the Accuracy of Epileptic Seizure Detection through EEG Analysis: A Comprehensive Classification Strategy

Manisha Sharma[1], Rajeev Pourush[2], Prateek Bhanti[3]
[1,2,3] *School of Engineering and Technology, Mody University of Science & Technology, Lakshmangarh, India.*
Email: [1] manasharma2000@gmail.com, [2] rajeevpourush.cet@modyuniversity.ac.in, [3] prateekbhanti.cet@modyuniversity.ac.in

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Epilepsy is a neurological disorder which impacts millions globally and continues to be a major public health challenge. The prompt identification of epileptic seizures is essential for effective treatment. In this study, we present an innovative methodology designed to enhance the accuracy of seizure detection through EEG data analysis. Our strategy involves creating a comprehensive EEG database that includes both healthy individuals and those experiencing seizures (ictal). We utilize a diverse range of classification models, including random forests, decision trees, XGBoost and k-nearest neighbors algorithm. For feature extraction, we have selected Linear Discriminant Analysis (LDA) as our preferred technique. The experimental results indicate that the random forest model is the most effective, achieving a perfect accuracy rate of 100% in detecting epileptic seizures. The decision tree model follows closely with an accuracy of 90.00%. Although the kNN algorithm has a slightly lower accuracy of 82.50%, it still plays a significant role in differentiating between normal and ictal EEG signals. Our results clearly demonstrate the effectiveness of our proposed method in reliably extracting spatial and temporal information from multi-channel EEG data, enabling accurate classification of epileptic seizures. This research highlights the robustness of our feature extraction approach and its potential to improve early diagnosis and treatment of epilepsy.

Keywords: Epilepsy, Linear Discriminant Analysis, Detection, EEG signals, Random Forest

INTRODUCTION

Epilepsy impacts around three million individuals globally, with estimates suggesting that approximately fifty million people have been diagnosed with this condition [1]. It ranks as the third most prevalent neurological disorder in every nation across the globe [2]. One of the most challenging features of epileptic seizures is their abrupt and unpredictable occurrence [3]. These seizures, which stem from irregular electrical activity present in the brain, pose a public health issue worldwide [4]. Timely identification of seizures is crucial for effective and improved patient outcomes. While traditional EEG methods have shown effectiveness, the emergence of deep learning and machine learning technologies presents promising opportunities for enhancing identification accuracy [5].

Electroencephalogram (EEG) recordings play a critical role in the analysis of brain disorders by mapping the electrical activity as shown in Figure.1 [8-10]. This is achieved through placing electrodes on the scalp and within the skull, as illustrated in figure 1. However, the process of obtaining prolonged EEG recordings can be both costly and time-intensive [11].

This research was conducted to improve the classifier's effectiveness in predicting epileptic seizures. Epilepsy is characterized by spontaneous seizures that can last from seconds to minutes. The following frequency bands are identified as shown in Figure.2. [12]:

- Delta waves (0.5–4 Hz): Connected to profound sleep and states of unconsciousness.
- Theta waves (4–8 Hz): Associated with light sleep, relaxation, and a sense of drowsiness.
- Alpha waves (8–13 Hz): Present during tranquil, wakeful relaxation with eyes closed.
- Beta waves (13–30 Hz): Corresponding to active thought processes, problem-solving, concentration.

- Gamma waves (30–100 Hz): Engaged in advanced cognitive functions and perception.

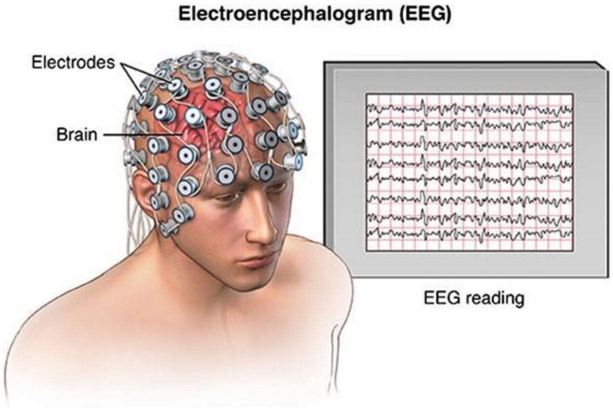

Figure 1 An overview of an Electroencephalogram (EEG) recording

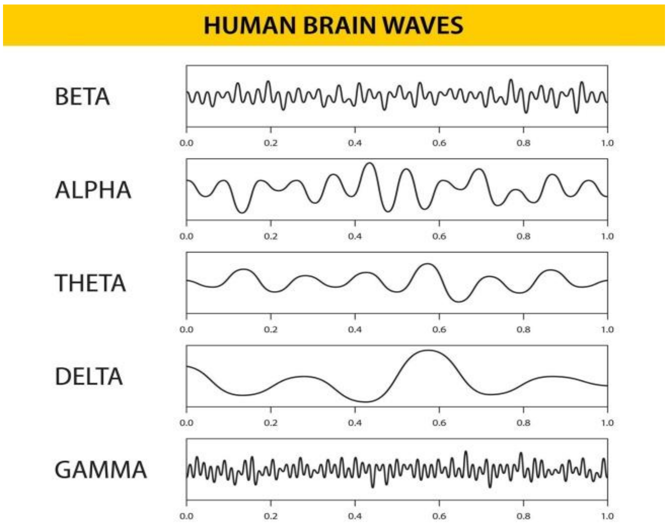

Figure 2. Human Brain Waves

In this research, we introduce a framework for preprocessing EEG data designed to improve the quality of the signals. The EEG data is subjected to filtering through a Butterworth band-pass filter, which effectively isolates and retains pertinent signal components within a designated frequency range while minimizing noise interference, thus facilitating a more accurate analysis of cerebral activity. To further refine detection precision, we apply Linear Discriminant Analysis (LDA) for feature extraction. LDA is recognized as a powerful method for capturing discriminative features from multi-channel EEG signals. To assess the effectiveness of our proposed methodology, we implement various classification models, including including random forests, decision trees, XGBoost and k-nearest neighbors algorithm. The experimental findings reveal that the Random Forest model surpasses the other classifiers, gaining an exceptional accuracy of 100% in the detection of epileptic seizures.

METHODS AND METHODOLOGY

(a) Database

The electroencephalogram dataset utilized in this research was found from the University Hospital in Bonn, Germany. This dataset includes total 200 single-channel EEG recordings. Each has duration of 23.6 sec, and is divided into two types: non-epileptic and epileptic. The dataset designated as S contains recordings of epileptic seizure events, while dataset Z comprises EEG recordings from 100 healthy subjects, collected using external surface electrodes under conditions where participants alternated between having their eyes open and closed.

Figure 3 illustrates a block diagram that outlines the proposed methodology. The process begins with the application of a Butterworth Bandpass Filter (BFF) to the input EEG brain signals, aimed at removing artifacts. The third block is dedicated to feature extraction, utilizing the Linear Discriminant Analysis (LDA) technique, which effectively reduces the dimensionality while retaining the most significant features of the raw EEG signals. In the fourth block, a variety of classifiers are utilized to categorize the extracted features. The final block presents the classification outcomes. Within this framework, the datasets Z and S are examined and classified into two distinct categories: normal (Z) and epileptic seizure (S).

Figure 3 Proposed Methodologies

The proposed methodology involves the analysis and categorization of two datasets, designated as Z and S, into two separate classifications. The class representing epileptic seizures (S) is incorporated within the normal category (Z) of epilepsy. The table 1 delineates the characteristics of EEG signals in normal individuals compared to those in patients with epilepsy, focusing on specific features:

1. *Alpha Waves*

- Normal Individual: In a healthy person, alpha waves, which signify a state of relaxation while awake, are typically observed.
- Epileptic Patient: In contrast, these waves may be either absent or markedly diminished in individuals suffering from epilepsy, especially during seizure events.

2. *Epileptiform Spikes*

- Normal Individual: A healthy EEG does not exhibit these abnormal, high-amplitude spikes.
- Epileptic Patient: The presence of epileptiform spikes, which reflect atypical neuronal activity.

3. *Frequency of Alpha Waves*

- Normal Individual: The standard frequency range for alpha waves is between 8-13 Hz, which is deemed normal.
- Epileptic Patient: When alpha waves are present in epileptic patients, their frequency generally remains within the normal range.

4. Frequency of Epileptiform Spikes

- Normal Individual: Such abnormal spikes are absent in a healthy EEG.
- Epileptic Patient: The frequency of epileptiform spikes is typically observed between 3-30 Hz, which is considered abnormal and indicative of seizure activity.

This comparison underscores the distinct EEG patterns that serve to differentiate normal individuals from those with epilepsy, thereby facilitating the identification and diagnosis of the condition.

Table 1. The distinctions in EEG waveforms between a healthy individual and a patient with epilepsy

Feature	Normal person	Epileptic patient
Alpha waves	Present	May be absent or reduced
Epileptiform spikes	Absent	Present
Frequency of alpha waves	8-13 Hz	Normal
Frequency of epileptiform spikes	3-30 Hz	Abnormal

(b) Artifact Removal

The preprocessing of electroencephalography (EEG) data is an essential step to guarantee the reliability and clarity of brain activity signals. EEG recordings frequently suffer from contamination due to various forms of noise and artifacts, including electrical interference, muscle contractions, and eye blinks. Such artifacts can mask significant neural information, resulting in misleading interpretations. A commonly employed technique for mitigating these artifacts is the utilization of a Butterworth Band-Pass Filter, which is described in detail below.

Algorithm 1: Artifact Removal

Step 1: Definition of the Transfer Function

The transfer function H(s) of a Butterworth bandpass filter quantitatively characterizes the correlation between the input and output signals within the frequency domain. This relationship is articulated in equation 1.

$$H(s) = \frac{1}{s^N} \cdot \frac{1}{S^2 + \frac{S}{Q} + 1} \; ..... \qquad (1)$$

The transfer function of the filter is denoted as H(s), where s represents the complex frequency variable. The parameter N indicates the order of the filter, which influences the steepness of the filter's roll-off. Additionally, Q refers to the quality factor of the filter, which governs the sharpness of the bandpass response.

Step 2: Coefficients for Implementation

The transfer function H(s) undergoes discretization and is subsequently implemented in digital signal processing (DSP) through the utilization of filter coefficients b and a, as delineated in equation 2.

$$H(s) = B(s) / A(s) .... \qquad (2)$$

The polynomial B(s) serves as the numerator, encapsulating the coefficients denoted by b. Conversely, A(s) functions as the denominator, representing the coefficients labeled as a. The precise values of 'b' and 'a' are contingent upon several factors, including the filter orderm the lower cutoff frequency, the upper cutoff frequency and the sampling frequency.

(c)Feature Extraction

Linear Discriminant Analysis serves as a strategy for feature extraction that aims to enhance classification accuracy by pinpointing the most distinguishing features within a dataset. This research utilizes LDA to decrease the dimensionality of the feature set while maintaining the critical information necessary for differentiating between various classes. The application of LDA is intended to improve the effectiveness of detecting epileptic seizures:

Algorithm 2: Feature Extraction

1. Calculate the Mean Vectors: For each class, determine the mean vector as shown in equation.3.

$$\mu_c = \frac{1}{N_c} \sum_{i=1}^{N_c} x_i .... \qquad (3)$$

where $x_i$represents feature vectors for class c and N_c is the number of samples in that class.

2. Compute the Scatter Matrices as shown in equation.4.

$$S_w = \sum_{c=1}^{C} \sum_{i=1}^{N_c} (x_i - \mu_c)\, x_i - \mu_c{}^T .... \qquad (4)$$

where μ is the overall mean of all samples.

The Generalized Eigenvalue Problem is addressed by LDA, which determines a projection matrix W, through the resolution of the eigenvalue equation.5.

$$\mathbf{S_w^{-1} S_b\, W = \lambda W} ..... \qquad (5)$$

The transformation matrix W is constructed from the eigenvectors associated with the largest eigenvalues, enabling the projection of the dataset into a lower-dimensional space while maintaining the distinction between two classes.

(d) Classification Models

In order to signify the core ideology of this methodology, it has been employed a battery of classification models, consists:

1. K-Nearest Neighbour (KNN)

KNN is a technique for the new data classification by evaluating its resemblance to previously encountered data. This algorithm retains the entire dataset and categorizes new data points into the most analogous class, thereby facilitating the classification process for incoming inputs. As a non-parametric and lazy learning method, KNN refrains from making any assumptions regarding the underlying data distribution and engages in classification only when required. During the training phase, the algorithm merely retains the dataset, and it classifies new instances by identifying and comparing them to the closest category. The operational mechanics of KNN can be delineated through the following algorithm.

The classification process utilizing K-Nearest Neighbors (KNN) commences with the selection of an appropriate value for K. The Euclidean distance is measured by equation.6.

$$D(y,z) = \sqrt{\sum_{i}^{n}(y_i - z_i)^2} \ldots.. \quad (6)$$

Following this calculation, the K nearest neighbors is identified based on the determined distances. The algorithm then assesses the distribution of categories among these neighbors, ultimately assigning the new data point to the class that has the greatest representation among them.

2. Random Forest

Random Forest operates by integrating the predictions of multiple decision trees rather than relying on a solitary tree. This ensemble method employs majority voting to ascertain the final output, whereby an increased number of trees contributes to a reduction in overfitting and an enhancement in accuracy. The process of Random Forest unfolds in two distinct phases: first, it constructs the forest by compiling N decision trees, and then it proceeds to generate predictions based on each tree developed during the initial phase.

The process of classification using Random Forest initiates with the random selection of K data points from the training dataset. These selected subsets are utilized to construct decision trees. A predetermined total number of decision trees (N) is established, and the initial two steps are iterated until the requisite number of trees is generated. For the classification of new data points, predictions are collected from each individual tree, and the ultimate classification is derived from the majority vote across all trees.

3. Decision Tree

A Decision Tree is a technique employed for classification purposes. It organizes data by systematically dividing it into smaller subsets according to the values of various features, resulting in a structure that resembles a tree. The construction of a Decision Tree follows several essential steps:

1. Identify the Optimal Feature: Determine which feature most effectively separates the data, utilizing metrics such as Gini Impurity or Information Gain.
2. Formulate Decision Nodes: Divide the dataset into branches according to the selected feature.
3. Continue the Splitting Process: Persist in dividing the subsets until a predetermined stopping criterion is satisfied, such as when all instances belong to a single class or a specified maximum depth of the tree is achieved.
4. Label the Terminal Nodes: The concluding nodes, or leaves, signify the class labels or the predicted outcomes in the case of regression.

Although Decision Trees are easy to interpret and visualize, they are prone to overfitting. To improve their generalization, techniques like pruning and ensemble methods, such as Random Forests, are commonly used.

4. *XGBoost (Extreme Gradient Boosting)*

XGBoost is an advanced and effective machine learning algorithm that utilizes gradient boosting techniques, making it highly suitable for both classification and regression problems. Its optimization for speed and performance has contributed to its popularity in data science competitions as well as in practical applications across various industries.

XGBoost Mechanism:

- Model Initialization: Begin with a basic weak learner, typically a simple decision tree.
- Residual Calculation: Determine the discrepancies between the actual outcomes and the predicted values.
- Tree Training: Construct a new tree aimed at reducing the residual errors.
- Weight Adjustment: Increase the weights assigned to incorrectly classified instances.
- Iteration: Persist in adding trees until a predetermined stopping condition is satisfied.
- Final Output: Combine the predictions from all trees to produce the ultimate result.

RESULTS AND DISCUSSION

The comparative evaluation of various techniques for detecting epileptic seizures through EEG data underscores the efficacy of distinct feature extraction and classification methods, as illustrated in Table 2. The 1D-LBP and PCA-ICA-LDA techniques achieved an impressive accuracy of 99.5%, indicating their strong feature extraction capabilities. In a similar vein, the HHT-SVM method reached an accuracy of 99.13%, while the Time Frequency RD-STFT approach demonstrated a marginally higher accuracy of 99.8%. The Multiwavelet Transform (MWT) technique recorded an accuracy of 99.85%, and the AESD method achieved 99.6%, both reflecting significant potential for seizure detection. The Genetic Algorithm method yielded an accuracy of 99.2%, affirming its effectiveness in feature selection. Notably, the proposed system utilizing the Random Forest classifier surpassed all previously mentioned methods, attaining a flawless accuracy of 100%, which highlights its exceptional capability to differentiate between non-epileptic and epileptic signals. This finding emphasizes the proposed methodology's effectiveness in enhancing the precision of seizure detection and improving diagnostic and therapeutic outcomes for patients.

Table 2. Results pertaining to the normal and seizure classifications

Method	Accuracy (%)
1D-LBP [14]	99.5
PCA-ICA-LDA[15]	99.5
HHT-SVM[16]	99.13
Time Freq,RD-STFT[17]	99.8
MWT[18]	99.85
AESD[19]	99.6
Genetic Algorithm[20]	99.2
Proposed System (Random Forests)	100

The evaluation of various classifiers was conducted using metrics such as accuracy, precision, recall, and F-score to assess their efficacy in detecting epileptic seizures, as detailed in Table 3 and Figure.4. The Random Forest classifier was identified as the most proficient model, achieving a perfect score of 100% across accuracy, precision, recall, and F-score, thereby illustrating its strong capability in accurately recognizing seizure patterns. Both the Decision Tree and XGBoost classifiers reached an accuracy of 90%, exhibiting flawless precision (100%) but slightly reduced recall at 82.60%, which resulted in an F-score of 90.47%. In contrast, kNN classifier demonstrated the lowest accuracy at 82.5%, despite also achieving perfect precision (100%); however, its recall was notably lower at 73.07%, culminating in an F-score of 84.44%. In summary, the Random Forest model was established as the most dependable option for epileptic seizure detection, while the Decision Tree and XGBoost classifiers yielded competitive outcomes, and kNN exhibited moderate performance.

Table 3. Performance evaluation for proposed model

“Classifier”	“Accuracy in %”	“Precision in %”	“Recall in %”	“F-Score in %”
“Decision tree”	90.0%	100%	82.60%	90.47%
Random Forests	100%	100%	100%	100%
k-Nearest Neighbors (kNN)	82.5%	100%	73.07%	84.44%
XGBoost	90.0%	100%	82.60%	90.47%

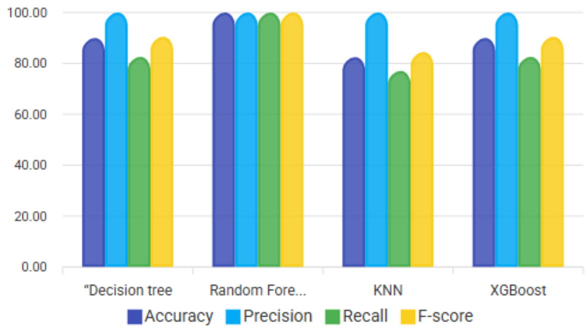

Figure 4 Performance evaluation for proposed model

The efficacy of Linear Discriminant Analysis (LDA) as a method for feature extraction was assessed by analyzing classifier accuracy across various classes both with and without the application of LDA as shown in Figure.5.

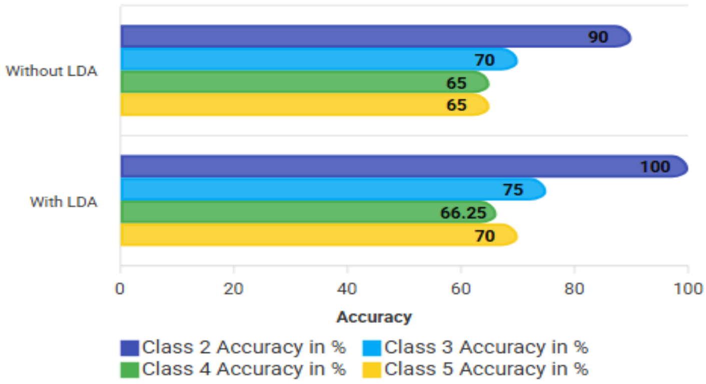

Figure 5 Various Class Comparisons Utilizing and Excluding LDA

The findings reveal a significant enhancement in classification performance when LDA is utilized. In the absence of LDA, Class 2 recorded the highest accuracy at 90%, whereas Classes 3, 4, and 5 exhibited lower accuracy rates of 70%, 65%, and 65%, respectively. Conversely, when LDA was employed, Class 2 achieved a perfect accuracy of 100%, underscoring the technique's capability in effectively differentiating features. Classes 3, 4, and 5 also experienced improvements, attaining accuracy levels of 75%, 66.25%, and 70%, respectively. These findings underscore the beneficial impact of incorporating LDA on classification accuracy, particularly for Class 2, which reached flawless classification. Although the enhancements in the other classes are modest, the overall influence of LDA on feature extraction and classification efficacy is clearly demonstrated.

CONCLUSION

This research illustrates the efficacy of utilizing EEG for the detection of epileptic seizures through Linear Discriminant Analysis (LDA) for feature extraction, alongside various machine learning classifiers. The Random Forest model proposed in this study achieved an impressive accuracy rate of 100%, surpassing conventional detection methods. A comparative analysis of classification outcomes with and without the application of LDA underscores its critical role in enhancing feature discrimination and overall detection precision. The findings suggest that the integration of machine learning with robust feature extraction methodologies can substantially improve the accuracy of seizure prediction and classification.

Moreover, the study emphasizes LDA's capability to reduce dimensionality while preserving essential discriminative features. This enhancement not only boosts the performance of classifiers but also increases computational efficiency, thereby facilitating real-time EEG analysis. The results indicate that the incorporation of optimization strategies, such as genetic algorithms or deep learning-based feature selection, may further elevate classification performance.

Future investigations could aim to strengthen the resilience of seizure detection models by integrating deep learning approaches, enabling real-time processing, and incorporating multi-modal EEG data. Additionally, enhancing computational efficiency for practical clinical applications and examining the effects of adaptive models for personalized seizure prediction could lead to improved patient outcomes. The exploration of explainable AI (XAI) techniques to enhance the interpretability of classification outcomes for clinical practitioners represents another promising avenue. Furthermore, validating the proposed system on larger and more diverse EEG datasets will be essential to ensure its generalizability and effectiveness in clinical environments.

Acknowledgement:

We sincerely thank Rajeev Pourush, and Prateek Bhanti for their valuable contributions to this work. Their support and insights have been instrumental in the completion of this study.

Funding Statement:

"No financing / There is no fund received for this article" or "The authors did not receive financing for the development of this research".

Data Availability:

“No new data were created or analyzed in this study. Data sharing is not applicable to this article”.

Conflict of interest:

"The authors declare that there is **no conflict of interest".**

REFERENCES

[1] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri , 2013, Automated EEG analysis of epilepsy: a review Knowledge-Based Systems, vol. 45, pp. 147–165.

[2] Zhou X., Ling B.W.-K., Zhou Y., Law N.F., “Phase space reconstruction, geometric filtering based Fisher discriminant analysis and minimum distance to the Riemannian means algorithm for epileptic seizure classification”, Expert Systems with Applications, Vol.219, 2023.

[3] Chavan P.A., Desai S., “Effective Epileptic Seizure Detection by Classifying Focal and Non-focal EEG Signals using Human Learning Optimization-based Hidden Markov Model”, Biomedical Signal Processing and Control, Vol. 83,2023.

[4] Qiu X., Yan F., Liu H., “A difference attention ResNet-LSTM network for epileptic seizure detection using EEG signal”, Biomedical Signal Processing and Control, Vol.83, 2023.

[5] Ashokkumar S.R., Anupallavi S., Premkumar M., Jeevanantham V., “Implementation of deep neural networks for classifying electroencephalogram signal using fractional S-transform for epileptic seizure detection”, International Journal of Imaging Systems and Technology, Vol. 31, 2021.

[6] Srinath R., Gayathri R., “Detection and classification of electroencephalogram signals for epilepsy disease using machine learning methods”, International Journal of Imaging Systems and Technology, Vol. 31, 2021.

[7] Collins C., Nuno A., Broderick A., Curnick D.J., de Vos A., Franklin T., Jacoby D.M.P., Mees C., Moir-Clark J., Pearce J., Letessier T.B., “Understanding Persistent Non-compliance in a Remote, Large-Scale Marine Protected Area”, Frontiers in Marine Science, Vol. 8, 2021 Wang Y., Zhou Q., Luo J., Lu Y., Wang H., Pang Y., Huang Z., “VLSI design of multiclass classification using sparse extreme learning machine for epilepsy and seizure detection”, IEICE Electronics Express, Vol. 19, 2022

[8] Varshney Y.V., Chandel G., Upadhyaya P., Farooq O., Khan Y.U., “Early onset/offset detection of epileptic seizure using M-band wavelet decomposition”, International Journal of Biomedical Engineering and Technology, Vol. 40, 2022.

[9] Yuan S., Mu J., Zhou W., Dai L.-Y., Liu J.-X., Wang J., Liu X., “Automatic Epileptic Seizure Detection Using Graph-Regularized Non-Negative Matrix Factorization and Kernel-Based Robust Probabilistic Collaborative Representation", IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 30, 2022.

[10] Abdelhameed A.M., Bayoumi M.,”Semi-supervised EEG signals classification system for epileptic seizure detection”, IEEE Signal Processing Letters, Vol. 26, 2019.

[11] Rajaguru H., Prabhakar S.K., “Hierarchical functions decision trees and minimum relative entropy in the detection of epilepsy risk levels from EEG signals”, International Journal of Mechanical Engineering and Technology, Vol. 9, 2018.

[12] Sharmila A., Aman Raj S., Shashank P., Mahalakshmi P., “Epileptic seizure detection using DWT-based approximate entropy, Shannon entropy and support vector machine: a case study”, Journal of Medical Engineering and Technology, Vol. 42, 2018.

[13] Satyajit Anand Sandeep Jaiswal, Pradeep Kumar, Ghosh, “Epileptic seizure detection in eeg signal using discrete stationary wavelet based stock Transform” Majlesi Journal of electrical engineering. Vol. 13, 2019.

[14] H. Hassanpour and M. Mesbah, ‘ New born EEG Seizuredetection based on Interspike space distribution in the Time- Frequency domain’, IJE Transactions A: Basics Vol. 20, No. 2 (June 2007) 137-146. Murugavel A.S.M., Ramakrishnan S., “Hierarchical multi-class SVM with ELM kernel for epileptic EEG signal classification” Medical and Biological Engineering and Computing, Vol. 54, 2016.

[15] Harikumar R., Vijayakumar T., “Analysis of wavelet transforms and RBF neural networks for epilepsy risk level classification from EEG signals”, International Journal of Applied Engineering Research Vol.10, 2015.

[16] Vijay Anand S., Shantha Selvakumari R., "Epileptic seizure detection of eeg signals based on support vector machine and DB4", International Journal of Applied Engineering Research, Vol. 10, 2015.
[17] Praveena J., Thiruvenkadam S., "Efficient seizure detection and classification using multi-class support vector machine", International Journal of Applied Engineering Research, Vol. 10, 2015.
[18] Bajaj V., Rai K., Kumar A., Sharma D., "Time-frequency image based features for classification of epileptic seizures from EEG signals", Biomedical Physics and Engineering Express, Vol.3, 2017.
[19] Kiran P.S., Komala G., Aneesh C.R.S.D., Rao S.K., "Electroencephalogram signal analysis using wavelet transform and statistical signal processing", Journal of Advanced Research in Dynamical and Control Systems, Vol. 9, 2017.
[20] Zhang Z., Wen T., Huang W., Wang M., Li C., "Automatic epileptic seizure detection in EEGs using MF-DFA, SVM based on cloud computing", Journal of X-ray science and technology, Vol. 25, 2017.