

An optimized framework for epileptic seizure detection using DWT-based feature extraction and hybrid dimensionality reduction

Rabel Guharoy^{1,2,3,*},
Nanda Dulal Jana¹, Suparna
Biswas⁴, Lalit Garg⁵ and Subhayu
Ghosh¹

¹Department of Computer Science
and Engineering, National Institute
of Technology, Durgapur, India

²Department of Information
Technology, School of Information
Technology, AI and Cyber Security,
Rashtriya Raksha University,
Gandhinagar, India

³Department of Computer Science
and Engineering, Amity University,
Kolkata, India

⁴Department of Electronics and
Communication Engineering, Guru
Nanak Institute of Technology,
Kolkata, India

⁵Department of Information and
Communication Technology,
University of Malta, Msida, Malta

*E-mail: rabelrock@gmail.com

Received for publication
July 15, 2025.

Abstract

Epileptic seizure detection using electroencephalography (EEG) offers a non-invasive means to monitor brain activity and detect neural anomalies. However, extracting detailed time and frequency features from EEG signals is challenging due to their inherent complexity. This study introduces a detection framework leveraging discrete wavelet transform (DWT) to extract robust features across multiple frequency bands, followed by dimensionality reduction to create compact and informative feature representations by using sophisticated feature extraction techniques such as principal component analysis (PCA), independent component analysis (ICA), and DWT. These methods retain essential signal characteristics, which are subsequently classified using K-nearest neighbor (KNN), Naive Bayes (NB), or support vector machine (SVM). Empirical evaluation on the Bonn EEG dataset reveals that the linear discriminant analysis (LDA)-NB configuration achieved a perfect accuracy of 100%, outperforming other combinations, including LDA-SVM at 89.17%, LDA-KNN at 80.42%, PCA-NB at 89.92%, and ICA-SVM at 90.42%. The LDA-NB model's ability to achieve optimal performance across all metrics underscores its effectiveness for seizure detection. Most of the existing work has used a large dimension of feature space, which indirectly increases the time complexity. However, in this study, we have integrated feature extraction with feature selection to reduce the dimensionality of the features. The proposed method can detect epilepsy with very high accuracy.

Keywords

discrete wavelet transform, electroencephalogram, epilepsy, independent component analysis, linear discriminant analysis, principal component analysis

1. Introduction

Epilepsy is a chronic neurological disorder characterized by abnormal electrical activity in the brain, known as seizures, which can be described metaphorically as an “electrical storm” in the brain [1, 2]. The main physiological basis of epilepsy is a disturbance in the brain's electrical activity, which can arise from a range of

causes, including low blood sugar levels and reduced oxygen supply during childbirth [3, 4]. Globally, around 50 million people are affected by epilepsy, and approximately 100 million individuals experience at least one seizure in their lifetime [3, 5]. The disorder accounts for about 0.5% of the global disease burden, with control rates remaining as low as 0.5%–1% [6]. Epileptic seizures can be detected by monitoring brain activity

using electroencephalogram (EEG) or electrocorticography (ECoG), which record electrical signals generated by neurons in the brain. These signals are inherently complex, non-linear, non-stationary, and often contain significant noise, presenting substantial challenges for seizure detection and classification. Effective seizure detection relies on identifying patterns within the EEG data without compromising performance, which has been approached through various machine learning (ML) classifiers. The primary challenge, however, lies in selecting optimal features and classifiers to maximize detection accuracy. In recent years, there has been a focus on utilizing ML algorithms with statistical feature representations, including both “black-box” and “non-black-box” methods, to improve seizure detection accuracy [7].

Epileptic seizures are commonly classified into two main types based on symptoms: partial seizures, affecting a specific area of the brain, and generalized seizures, involving multiple brain regions simultaneously [5, 8]. For feature extraction, the discrete wavelet transform (DWT) [9] has demonstrated substantial effectiveness in analyzing EEG signals across different frequency bands, enabling the identification of key features that facilitate epilepsy detection. DWT provides a robust framework for extracting and isolating relevant features, thereby reducing the data dimensionality and enhancing the efficiency of classifier performance. To further enhance the classification process, several dimensionality reduction techniques, including principal component analysis (PCA) [10], independent component analysis (ICA) [11], and linear discriminant analysis (LDA) [12], are used to remove irrelevant features while retaining key signal characteristics.

In the literature, numerous studies have explored the use of statistical functions—such as standard deviation, average power, mean absolute value, mean, variance, Shannon entropy, and skewness—as features for classifying EEG signals with various ML methods. A comparative analysis of classifiers, including support vector machine (SVM) [13], Naive Bayes (NB) [14], and K-nearest neighbors (KNN) [15], has shown that SVM generally excels in accuracy, whereas NB and KNN produce favorable results under certain conditions [1]. However, the high-dimensional feature space in these models often leads to increased time complexity, posing a significant limitation for real-time seizure detection applications.

Another common challenge in EEG signal processing is noise from sources such as electrooculogram (EOG) and electrocardiogram (ECG) artifacts, which can distort the analysis of brain activity. To address

this, recent methods have applied a combination of ICA, common spatial pattern (CSP), and wavelet transformation to decompose EEG signals into independent components, isolate wavelet coefficients, and remove noise. By applying CSP to denoise the EEG data, these approaches maintain the neural activity patterns essential for seizure detection [16].

This study aims to address these challenges by proposing a two-step feature reduction process that improves classifier’s accuracy while reducing time complexity. First, we apply DWT for feature extraction, followed by three distinct dimensionality reduction techniques (PCA, ICA, and LDA) to reduce feature dimensions. Second, a feature-level fusion technique is applied to achieve further dimensionality reduction. Finally, we employ three different classifiers (SVM, NB, and KNN) to assess the performance of this hybrid model in detecting epilepsy.

Despite extensive research, real-time seizure detection remains hindered by noisy EEG signals and high-dimensional feature spaces. We propose a novel hybrid framework that integrates DWT-based feature extraction with layered dimensionality reduction and classifier optimization. Our model not only achieves 100% accuracy with LDA-NB but also demonstrates robustness across multiple EEG subsets, outperforming traditional approaches.

The main contributions of this study are as follows:

1. A two-step feature reduction process for epileptic seizure detection using DWT for feature extraction is proposed, followed by dimensionality reduction techniques, namely PCA, ICA, and LDA, to effectively reduce the feature space.
2. A feature-level dimension reduction technique is applied to optimize feature dimensions, enhancing computational efficiency without compromising classification accuracy.
3. The performance is evaluated using three different classifiers—SVM, NB, and KNN—in detecting epilepsy based on the proposed feature extraction and reduction pipeline.
4. The proposed approach achieves high accuracy in seizure detection, with the LDA and NB combination yielding 100% accuracy on the Bonn dataset, demonstrating the effectiveness of this hybrid model over existing methods.
5. Extensive experiments on EEG data with noise removal techniques are conducted, demonstrating the robustness of the proposed model in handling noise while preserving essential neural activity patterns for accurate seizure detection.

The remainder of this paper is organized as follows: Section II reviews existing epilepsy detection methods, Section III presents relevant techniques, Section IV details our proposed methodology and algorithm, Section V provides the dataset description, data pre-processing, evaluation metrics, and experimental setups, and Section VI highlights and discusses our findings. We conclude the paper in Section VII.

II. Literature Review

This section provides an overview of previous research on epilepsy prediction techniques, highlighting the role of ML in health and biological datasets. According to Prochazka et al. [17] and Çınar and Acir [18], numerous ML applications have emerged within healthcare, facilitating improved outcomes. Researchers, particularly in data mining and ML, have proposed solutions to enhance seizure detection accuracy, employing various classifiers such as artificial neural networks (ANN), SVM, decision trees, and random forests [19, 20]. ML has proven instrumental in extracting meaningful patterns from health-related datasets and contributes to addressing challenges in healthcare [19–26]. Applications of these techniques to brain datasets include seizure detection, epilepsy lateralization, differentiation of seizure states, and localization [19, 20].

Amin et al. [27] proposed a feature extraction method using DWT on EEG signals, in which relative wavelet energy was calculated from both the approximation and detail coefficients at the final decomposition level. Their study demonstrated high classification accuracy, achieving 98% with SVM for EEG signals recorded under cognitive and resting conditions, highlighting DWT's efficacy in distinguishing complex cognitive states. Al-Qerem et al. [1] also employed the DWT in feature extraction, leveraging differential evolution for feature selection. Their model, tested with the Bonn dataset, achieved optimal classification by testing seven types of wavelets, including Discrete Meyer, Reverse Biorthogonal, and Daubechies, and proved efficient for EEG classification. Another novel method for classifying date fruits by analyzing their texture using scattering wavelet transforms, which extract robust features unaffected by shape or color similarities. These features are then processed through a stacking ensemble model combining random forest, SVM, and logistic regression to achieve highly accurate classification across diverse date varieties [28].

Epilepsy, a chronic neurological condition, is often diagnosed through EEG and ECoG data, which are

highly complex, noisy, and non-stationary, presenting challenges for accurate seizure detection. ML methods are increasingly applied to process these signals, effectively identifying patterns while addressing noise and non-linearity. Siddiqui et al. [7] reviewed various ML techniques in seizure detection, categorizing methods based on statistical features and classifier types. Their study highlighted the importance of selecting suitable classifiers and features in this domain, distinguishing between black-box and transparent models.

Classification framework for Arabic biomedical questions using transformer-based contextual embeddings, enhancing semantic understanding and domain-specific accuracy. By leveraging deep language models, the system achieves improved performance in categorizing complex medical queries in Arabic [29]. This study proposed a deep learning model using convolutional neural network (CNN) and Transfer Learning to classify breast cancer from ultrasound images with high accuracy. By pre-processing the breast ultrasound images (BUSI) dataset and deploying the model in a desktop application, it achieved >90% accuracy, highlighting its clinical potential [30]. Another way evaluates how different activation functions—rectified linear unit (ReLU), Leaky ReLU, Sigmoid, and Tanh—impact the performance of CNN architectures (ResNet50, VGG16, and GoogleNet) in classifying images affected by Poisson noise. It finds that ResNet50 combined with ReLU consistently delivers the highest accuracy across datasets with 3, 5, and 10 classes, demonstrating strong resilience to noise [31]. Several studies have utilized dimensionality reduction methods to enhance classifier performance for seizure detection. Martis et al. [32] used DWT for ECG signal analysis and applied PCA, LDA, and ICA for feature reduction, achieving an accuracy of 99.28% with SVM and neural networks. Shi et al. [33] introduced a binary harmony search (BHS) algorithm for selecting optimal EEG channels, significantly improving classification accuracy over conventional CSP methods in motor imagery classification.

Additional studies have developed robust feature extraction and denoising methods to address the noisy nature of EEG signals. For example, Geng et al. [16] integrated ICA, DWT, and CSP for EEG signal processing, achieving improved accuracy in identifying EOG and ECG artifacts. Kapoor et al. [34] developed a hybrid model using AdaBoost, random forest, and decision tree classifiers, achieving >96% accuracy on the children's hospital boston Children's Hospital Boston (CHB-MIT) dataset. Amin et al. [27] similarly

reported 98% accuracy using DWT and SVM, multi-layer perceptron and the kNN on the Bonn dataset.

Furthermore, dimensionality reduction techniques for brain-computer interface (BCI) applications were explored by Tan et al. [35], who utilized the Bonn dataset with classifiers such as SVM and KNN, attaining high classification accuracy. Priyanka et al. [36] also employed ANN on the Bonn dataset, achieving 96.9% accuracy.

These studies underscore the growing potential of ML approaches in improving epilepsy diagnosis and advancing seizure detection methodologies. This research builds upon prior work by integrating DWT for feature extraction, dimensionality reduction, and classifier optimization, further contributing to the field of automated epilepsy prediction.

III. Background

In this section, we discuss the various tools and techniques utilized in our proposed methodology, focusing on feature extraction using DWT and dimensionality reduction methods, namely PCA, ICA, and LDA.

a. DWT

DWT [9] is widely used in signal processing due to its ability to provide time–frequency localization, making it well suited for analyzing non-stationary signals, such as EEG and ECG [37]. DWT decomposes a one-dimensional signal into two sub-bands: high-frequency (detail) and low-frequency (approximation). Given a signal x , DWT applies a low-pass filter g and a high-pass filter h to generate these sub-bands. This decomposition is mathematically described by:

$$y[n] = (x \times g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k] \quad (1)$$

The signal passes through filters in successive levels, with downsampling by a factor of 2 at each stage. The resulting approximations and details capture information across different frequency bands (as illustrated in Figure 1), facilitating multi-level decomposition of the signal:

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k], \quad (2)$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k]. \quad (3)$$

The Daubechies wavelet family (db4) was selected due to its proven effectiveness in EEG signal analysis, particularly for seizure detection. This wavelet offers good time-frequency localization and has been widely used in previous studies.

The EEG signals were decomposed up to the fifth level using db4 wavelet. This level was selected to capture the frequency bands relevant to epileptic seizure activity, particularly the delta, theta, alpha, beta, and gamma bands.

b. PCA

PCA [10] is a linear dimensionality reduction method that projects data along directions of maximum variance, preserving essential information while reducing dimensionality. It is computed through the covariance matrix C of the data, followed by eigenvalue decomposition:

$$C = (X - \bar{x})(X - \bar{x})^T. \quad (4)$$

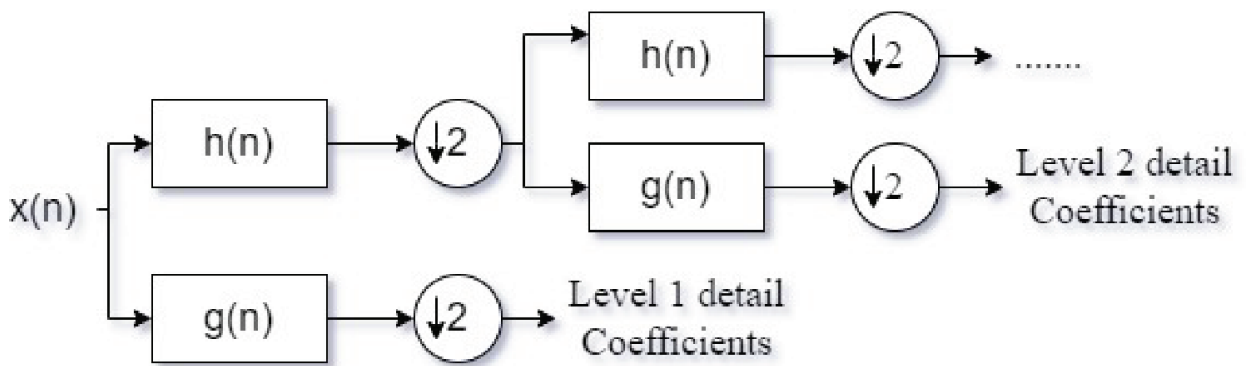


Figure 1: Second level of coefficients.

Eigenvectors and eigenvalues of C determine the principal components. The transformed data are then obtained by projecting the original data onto these components, emphasizing variance [38]:

$$\text{Projected data} = \left[V^T (X - \bar{X})^T \right]^T. \quad (5)$$

c. ICA

ICA [11] separates statistically independent components from mixed signals, often used for noise reduction in EEG signal processing. Given a vector x of mixed signals and a matrix A representing weights, the ICA model is formulated as:

$$x = As \text{ or } x = \sum_{i=1}^n a_i s_i. \quad (6)$$

The goal is to estimate both A and the source signals s under the assumption of non-Gaussian, statistically independent sources [7].

d. LDA

LDA [12] maximizes separation between classes by finding a linear combination of features that best discriminates among them. For a class mean vector μ_i and pooled covariance Σ , the linear score function is defined as:

$$s_i^L(X) = -\frac{1}{2} \mu_i \sum^{-1} \mu_i + \mu_i \sum^{-1} x + \log P(\pi_i) \quad (7)$$

LDA is commonly used for dimensionality reduction and classification in high-dimensional data.

IV. Proposed Approach

The proposed methodology involves a multi-step approach for epileptic seizure detection, as depicted in Figure 2. The process involves:

- Applying DWT for multi-level decomposition to capture relevant features.
- Reducing feature dimensionality through PCA, ICA, and LDA.
- Classifying the features using SVM, KNN, and NB classifiers.

The fifth level of DWT decomposition provides six subbands of EEG signals (Figure 3), enabling a robust classification. Each subband is then processed using PCA, ICA, or LDA, and classified for epileptic seizure detection with high accuracy.

Furthermore, Algorithm 1 presents the workflow of the proposed approach. The algorithm for epileptic seizure detection is designed to process EEG signal data by following a structured approach that includes pre-processing, feature extraction, dimensionality reduction, and classification. First, the EEG data, denoted as X , undergo pre-processing. This step involves normalizing the data to bring all sample values to a consistent scale, which helps improve the stability and performance of classifiers.

In the preprocessing step, 0.53 40 Hz filter has been used to remove the noise or artifacts, followed by normalization in the subsequent step.

Normalization is particularly important in EEG signal processing as it enhances the model's ability to detect subtle variations that could indicate a seizure, while also reducing the influence of noise or artifacts in the data.

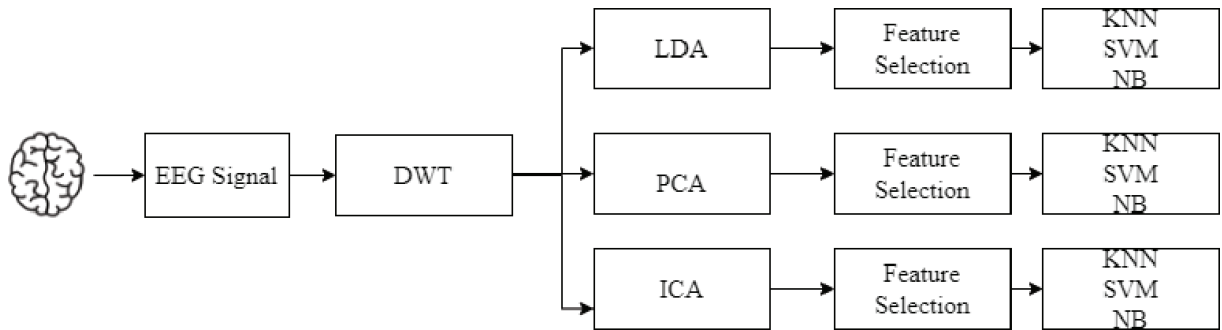


Figure 2: Block diagram of the proposed method. DWT, discrete wavelet transform; ICA, independent component analysis; KNN, K-nearest neighbor; LDA, linear discriminant analysis; NB, Naive Bayes; PCA, principal component analysis; SVM, support vector machine.

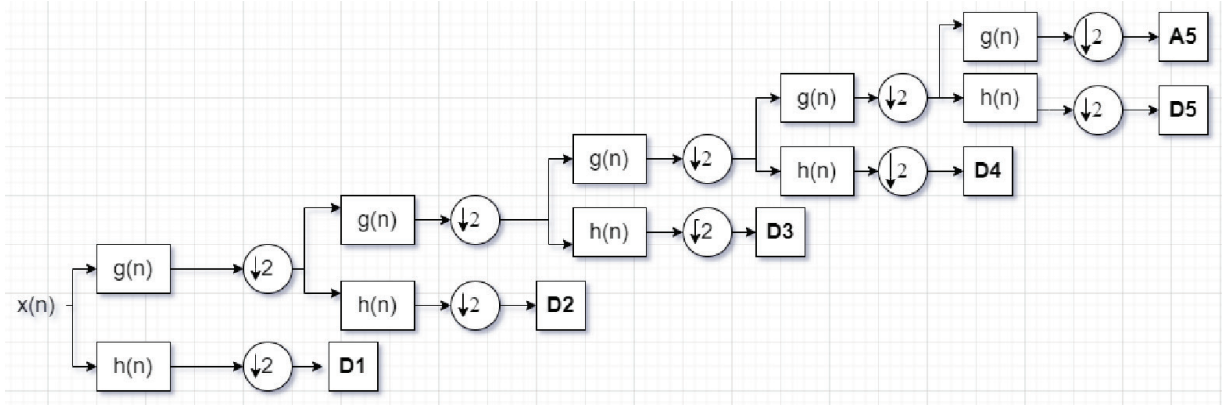


Figure 3: Block diagram of fifth level decomposition of the EEG signal. EEG, electroencephalogram.

After pre-processing, the algorithm proceeds with feature extraction using the DWT. DWT is an effective tool for decomposing EEG signals into various frequency bands, which enables the analysis of distinct signal characteristics relevant to seizure detection. Here at first we have increased the wavelet decomposition up to seventh level and for each level change the mother wavelet Daubechies from db1 to db10. Computing performance for each case, it is noticed that level 5 and db1 provide the best result. Actually, during our experiment, one by one, we have increased the level up to 7. However, we observed that the proposed model provides the best performance at fifth level.

Algorithm 1: Epileptic seizure detection

Input: EEG data X

Output: Seizure or non-seizure classification

Step 1: Pre-processing

Normalize the EEG data X to a consistent scale.

Step 2: Feature Extraction

Apply DWT on X to decompose the EEG signal. Extract features from the fifth-level decomposition coefficients.

Step 3: Dimensionality Reduction

Choose one of the following methods for dimensionality reduction:

- PCA
- ICA
- LDA

Obtain reduced feature set $F_{reduced}$

Step 4: Classification

Choose one of the following classifiers:

- SVM
- KNN
- NB

Train the classifier on $F_{reduced}$

Step 5: Testing and Evaluation

Test the trained model on unseen data.

Compute evaluation metrics (e.g., accuracy, precision, and recall).

Return Classification result as seizure or non-seizure = 0

In Algorithm 1, we specifically use the fifth level of decomposition, which provides a balance between detail and approximation by capturing the primary subbands associated with seizure activity. The resulting DWT coefficients at this level serve as the extracted features, representing the signal's most informative frequencies in terms of seizure detection.

To reduce the high dimensionality of the extracted features, the algorithm applies a dimensionality reduction technique. Users can choose among three methods: PCA, ICA, or LDA. PCA reduces dimensions by finding orthogonal components that account for the most variance, which allows us to retain the most informative aspects of the data. ICA, on the contrary, identifies statistically independent components, making it effective in reducing noise and redundancy. LDA is particularly advantageous in supervised tasks like seizure detection as it maximizes the separability between seizure and non-seizure classes. The

outcome of this step is a reduced feature set $F_{reduced}$, which retains the essential signal characteristics while lowering computational complexity.

With the reduced feature set prepared, the algorithm then moves to the classification stage. The user can select from three classification methods: SVM, KNN, or NB. SVM is a powerful classifier that finds the optimal hyperplane to separate seizure and non-seizure classes, making it highly suitable for high-dimensional spaces. KNN, on the contrary, is a simpler approach that classifies samples based on their proximity to KNN in the feature space, providing an intuitive method for pattern recognition. NB employs a probabilistic framework based on Bayes' theorem, which is often effective for binary classification tasks such as this. The classifier is trained on labeled EEG data, allowing it to learn the distinctions between seizure and non-seizure patterns.

Finally, the trained model is tested on unseen data, where it predicts seizure or non-seizure labels for each sample. Evaluation metrics, such as accuracy, precision, and recall, are computed to measure the model's performance and assess its effectiveness in detecting seizures accurately. The output of the algorithm is a classification result indicating whether each sample is labeled as a seizure or non-seizure event. By combining DWT-based feature extraction with dimensionality reduction and reliable classification methods, this algorithm provides a comprehensive approach to accurately identify epileptic seizures from EEG data.

V. Experimental Setup

a. Dataset description

For the evaluation of the proposed epileptic seizure detection techniques, we have used the Bonn dataset [39], a widely recognized dataset for seizure

detection tasks. The Bonn dataset, which is publicly available, consists of electroencephalogram (EEG) signals recorded from healthy individuals and those diagnosed with epilepsy. It includes data captured in both pre-seizure and post-seizure intervals. The dataset is divided into multiple subsets, each containing EEG data from different subjects, recorded in various settings (e.g., resting, seizure, etc.).

The dataset features time-series EEG signals sampled at 173.61 Hz, with each sample being divided into 23.6-s segments. These segments contain a variety of seizure and non-seizure classes, making the dataset ideal for classification tasks. Each signal is represented by a series of 4,097 data points, making it a suitable choice for the analysis of temporal dynamics involved in epileptic seizures.

In this study, the Bonn dataset was pre-processed using DWT to extract meaningful features from the raw EEG signals. This feature extraction process helps in capturing both time and frequency-domain characteristics, which are essential for distinguishing between normal and seizure conditions. Further dimensionality reduction was performed using techniques such as PCA, ICA, and LDA to enhance the classification performance. For evaluation, we have used standard classification algorithms such as KNN, NB, and SVM.

Details of the explanation of the Bonn dataset are provided in Table 1. By applying the novel seizure detection techniques on this dataset, we aim to assess the effectiveness and robustness of the proposed framework, especially with respect to high classification accuracy and generalizability across different conditions.

b. Data preprocessing

The dataset used in this study is the Bonn dataset, which consists of EEG signals collected from healthy subjects and individuals with epilepsy. The raw EEG

Table 1: Samples of data in normal and seizure cases

Set name	Annotation of data	Size (KB)	Acquisition circumstances
Set A	Z000.txt—Z100.txt	564	Five healthy subjects with open eyes
Set B	O000.txt—O100.txt	611	Five healthy subjects with closed eyes
Set C	N000.txt—N100.txt	560	Five people with epilepsy with seizure-free status
Set D	F000.txt—F100.txt	569	Five people with epilepsy with seizure-free status inside five epileptogenic zones
Set E	S000.txt—S100.txt	747	Five subjects during seizure activity

data are pre-processed to remove any noise or artifacts, ensuring that only relevant signals are considered for feature extraction. The data are segmented into non-overlapping windows of 23.6 s, as this time window is sufficient to capture the underlying temporal dynamics of epileptic seizures. To standardize the signals across different subjects, each EEG signal is normalized to have a mean of zero and a standard deviation of one, ensuring consistency in the features extracted from different individuals.

c. Evaluation metrics

The performance of the proposed techniques is evaluated using standard classification metrics [40], including accuracy, sensitivity, specificity, precision, and F1-score. Accuracy is the proportion of correctly classified instances, while sensitivity and specificity measure the classifier's ability to detect seizures and non-seizures, respectively. Precision and F1-score provide further insights into the classifier's performance, especially in dealing with imbalanced datasets.

For the statistical analysis, we conducted cross-validation using a 10-fold cross-validation technique to ensure the reliability and generalizability of

the results. This process helps in mitigating overfitting and provides a more robust evaluation of the classifier's performance.

d. Implementation details

All experiments are implemented using the Python programming language and ML libraries. The experiments are run on a standard desktop machine with an Intel i7 processor and 16 GB of RAM.

VI. Results and Discussion

The performance metrics for the proposed approach are presented in Tables 2–4, where ICA is combined with KNN, SVM, and NB classifiers. The values obtained from each of the 100 runs are averaged and mentioned in Tables 2–4. These tables highlight the comparative metrics for each classification model. According to the data, the highest average accuracy (i.e., 100%) was achieved with the ICA and NB combination in the 2-class classification for the A–E dataset. Additional metrics, such as F-measure, recall, specificity, sensitivity, and precision, reported in Tables 2–4, further validate the model's performance.

Table 2: Performance metrics for ICA with KNN

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A–C	88.50	93.99	81.14	85.24	93.99	0.89
A–D	83.50	82.65	83.87	83.99	82.65	0.82
A–E	93.00	100.00	86.01	88.36	100.00	0.94
B–C	91.50	93.33	89.83	90.47	93.33	0.92
B–D	91.50	93.31	90.37	90.12	93.31	0.91
B–E	92.00	100.00	84.07	86.32	100.00	0.92

ICA, independent component analysis; KNN, K-nearest neighbor.

Table 3: Performance metrics for ICA with SVM

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A–C	88.00	92.68	83.67	86.33	92.68	0.89
A–D	85.50	96.03	75.96	79.14	96.03	0.86
A–E	97.50	100.00	94.96	95.55	100.00	0.98
B–C	86.50	83.63	89.65	91.07	83.63	0.87
B–D	90.50	89.63	92.59	91.72	89.63	0.90
B–E	94.50	100.00	88.78	90.69	100.00	0.95

ICA, independent component analysis; SVM, support vector machine.

The proposed method achieved a maximum average sensitivity of 100% across all three combinations (ICA + KNN, ICA + NB, and ICA + SVM) for the A–E dataset. As observed in Table 3, a maximum average sensitivity of 100% was also attained for the B–E dataset. A comparison across Tables 2–4 indicates that the maximum average specificity and F-measure values were obtained for the A–E dataset.

Tables 5–7 illustrate the performance metrics for the PCA dimensionality reduction technique with three classifiers: NB, SVM, and KNN. Based on the data in tables, the NB classifier achieved the highest

accuracy of 100% for the A–E dataset, with recall and F-measure values also reaching 100%. Specificity values of 100% were observed for datasets A–E and B–E using SVM and KNN, respectively.

Tables 8–10 provide performance metrics for the LDA technique using the NB, SVM, and KNN classifiers, respectively. Comparing Tables 2 and 10, the LDA + NB combination yields the best results across all metrics for all datasets. NB achieved 100% accuracy for datasets A–C, A–D, A–E, B–C, B–D, and B–E, with specificity and precision values also at 100% for all dataset combinations.

Table 4: Performance metrics for ICA with NB

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A–C	72.00	82.65	61.52	67.89	82.65	0.74
A–D	72.50	97.03	47.76	65.55	97.03	0.78
A–E	100.00	100.00	100.00	100.00	100.00	1.00
B–C	82.00	67.22	95.48	95.60	67.22	0.78
B–D	68.00	91.43	45.27	62.48	91.42	0.74
B–E	99.50	99.23	100.00	100.00	99.23	0.99

ICA, independent component analysis; NB, Naive Bayes.

Table 5: Results of PCA with KNN algorithm

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A–C	81.00	88.47	71.87	78.36	88.47	0.83
A–D	90.50	93.20	89.28	89.34	93.20	0.91
A–E	58.00	100.00	18.18	57.18	100.00	0.71
B–C	81.00	82.32	77.93	79.39	82.32	0.81
B–D	83.50	83.09	81.94	83.12	83.09	0.89
B–E	88.50	100.00	75.07	86.68	100.00	0.92

KNN, K-nearest neighbors; PCA, principal component analysis.

Table 6: Results of PCA with SVM algorithm

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A–C	77.50	96.87	55.57	71.09	96.87	0.81
A–D	84.50	97.07	71.43	77.77	97.07	0.86
A–E	93.50	100.00	86.87	89.34	100.00	0.94
B–C	83.00	93.65	71.89	77.89	93.65	0.85
B–D	85.00	93.85	74.36	80.27	93.85	0.86
B–E	90.00	100.00	80.09	83.98	100.00	0.91

PCA, principal component analysis; SVM, support vector machine.

Table 7: Results of PCA with NB algorithm

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A–C	80.50	94.45	67.12	74.17	94.45	0.83
A–D	80.00	96.26	63.67	72.64	96.26	0.82
A–E	100.00	100.00	100.00	100.00	100.00	1.00
B–C	90.50	84.28	95.71	97.50	84.28	0.90
B–D	89.00	90.31	88.38	87.95	90.31	0.89
B–E	99.50	99.00	100.00	100.00	99.00	0.99

NB, Naive Bayes; PCA, principal component analysis.

Table 8: Results of proposed model LDA with KNN algorithm

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A–C	77.50	82.95	70.30	76.89	82.95	0.78
A–D	66.50	64.45	67.84	69.40	64.45	0.66
A–E	92.00	100.00	84.65	86.49	100.00	0.92
B–C	76.50	64.08	87.32	82.89	64.08	0.72
B–D	80.00	73.44	82.98	85.54	73.44	0.77
B–E	90.00	100.00	79.61	84.92	100.00	0.91

KNN, K-nearest neighbors; LDA, linear discriminant analysis.

Table 9: Results of proposed model LDA with SVM algorithm

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A–C	100.00	100.00	100.00	100.00	100.00	1.00
A–D	72.00	72.48	73.36	72.75	72.48	0.71
A–E	96.00	99.09	90.63	95.56	99.09	0.97
B–C	91.00	86.70	94.02	93.57	86.70	0.90
B–D	100.00	100.00	100.00	100.00	100.00	1.00
B–E	76.00	88.38	63.88	74.31	88.38	0.80

LDA, linear discriminant analysis; SVM, support vector machine.

Figures 4–6 display the confusion matrices for PCA combined with KNN, SVM, and NB classifiers. Each matrix shows two primary classes: the actual (True) and the predicted. The layout provides insights into the classification accuracy and performance across these three models. Notably, the SVM-based matrix (Figure 5) differs slightly from those generated by KNN and NB, yet overall, all confusion matrices display robust class prediction

capabilities, highlighting the efficacy of PCA in feature reduction.

Figures 7–9 further present the fold-wise accuracy for ICA, PCA, and LDA with SVM, NB, and KNN classifiers. In Figure 7, we observe that ICA paired with SVM yields higher accuracy than ICA combined with NB or KNN, emphasizing SVM's strong predictive performance with ICA. Similarly, Figure 8 illustrates the fold-wise accuracy for PCA, where SVM consistently outperforms both

Table 10: Results of proposed model LDA with NB algorithm

Case	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure
A-C	100.00	100.00	100.00	100.00	100.00	1.00
A-D	100.00	100.00	100.00	100.00	100.00	1.00
A-E	100.00	100.00	100.00	100.00	100.00	1.00
B-C	100.00	100.00	100.00	100.00	100.00	1.00
B-D	100.00	100.00	100.00	100.00	100.00	1.00
B-E	100.00	100.00	100.00	100.00	100.00	1.00

LDA, linear discriminant analysis; NB, Naive Bayes.

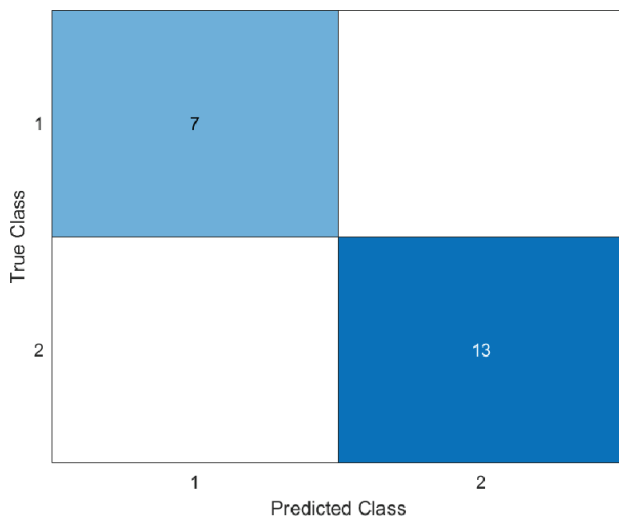


Figure 4: Confusion matrix for PCA with KNN algorithm. KNN, K-nearest neighbors; PCA, principal component analysis.

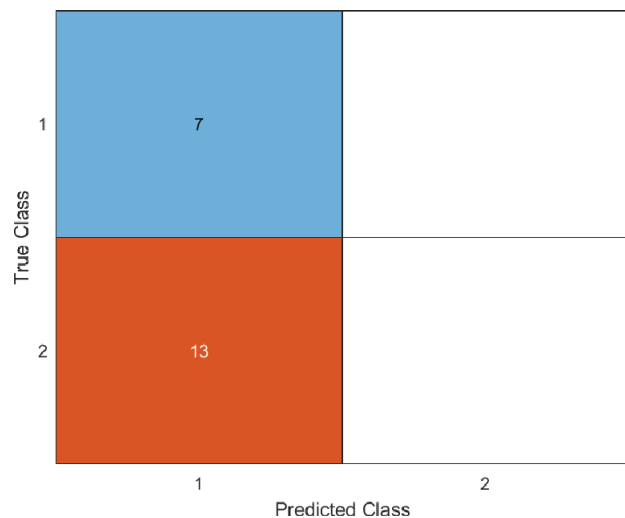


Figure 5: Confusion matrix for PCA with SVM algorithm. PCA, principal component analysis; SVM, support vector machine.

KNN and NB. Interestingly, in the case of LDA (Figure 9), the NB classifier surpasses both SVM and KNN in accuracy, suggesting that LDA benefits more from NB's probabilistic approach to classification.

Finally, Figure 10 presents the receiver operating characteristic curve (ROC) plots for the classifiers (SVM, NB, and KNN) with dimensionality reduction methods (PCA, ICA, and LDA). Each classifier-ROC plot remains nearly identical, with all models approaching 100% performance, indicating high true positive rates and minimal misclassification. This consistency suggests that the classifiers are well suited for the dataset and exhibit near-optimal performance across all evaluated dimensionality

reduction techniques, confirming their reliability and robustness in classifying EEG signal sets.

The results presented in this study demonstrate the effectiveness of various dimensionality reduction techniques (PCA, ICA, and LDA) combined with different classifiers (KNN, SVM, and NB) for EEG signal classification. The confusion matrices for each classifier revealed high classification accuracy across multiple folds. The ROC plots further confirmed the robustness of the classifiers, with almost identical curves approaching 100% accuracy for all methods. These findings suggest that dimensionality reduction techniques enhance the classification capabilities of ML models for EEG analysis.

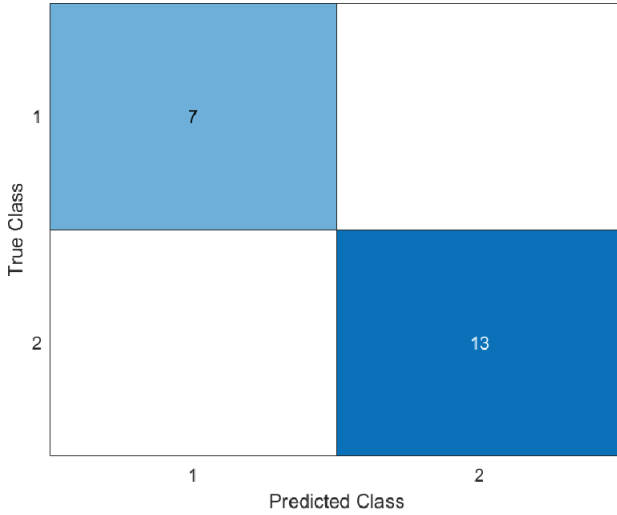


Figure 6: Confusion matrix for PCA with NB algorithm. NB, Naive Bayes; PCA, principal component analysis.

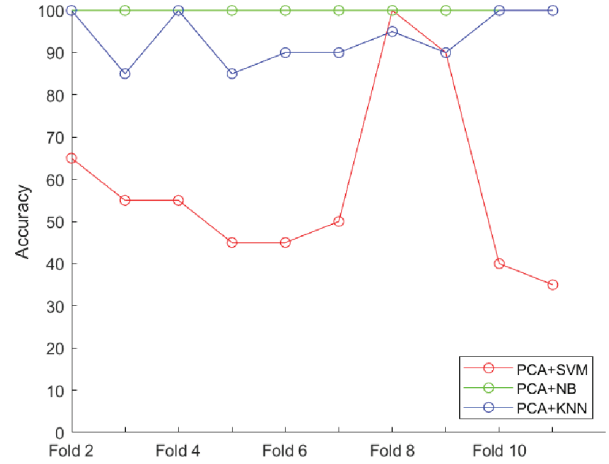


Figure 8: Fold-wise accuracy using PCA and SVM, NB, KNN. KNN, K-nearest neighbors; NB, Naive Bayes; PCA, principal component analysis; SVM, support vector machine.

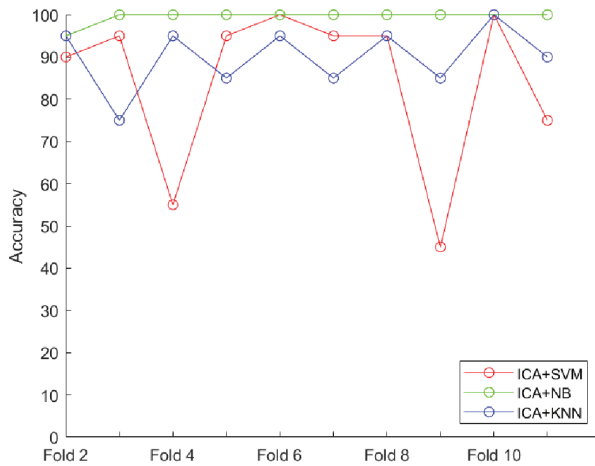


Figure 7: Fold-wise accuracy using ICA and SVM, NB, KNN. ICA, independent component analysis; KNN, K-nearest neighbors; NB, Naive Bayes; SVM, support vector machine.

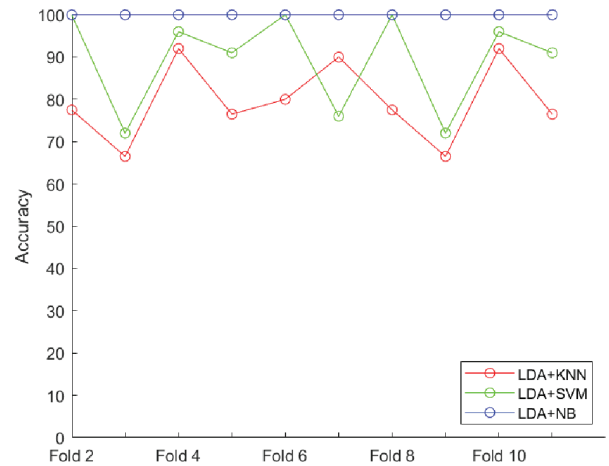


Figure 9: Fold-wise accuracy using LDA and SVM, NB, KNN. KNN, K-nearest neighbors; LDA, linear discriminant analysis; NB, Naive Bayes; SVM, support vector machine.

VII. Conclusion

This study presents a novel and efficient approach to epileptic seizure detection by combining DWT for feature extraction with three dimensionality reduction techniques: PCA, LDA, and ICA. The proposed method leverages feature-level fusion to reduce the feature space, ensuring both computational efficiency and high classification accuracy. Experimental results

demonstrate that the combination of LDA with NB achieves 100% accuracy, surpassing existing methods in performance. This low-dimensional feature space makes the method suitable for real-time, cost-effective clinical applications, addressing the challenges of time-consuming and expensive epilepsy diagnosis. The simplicity and accuracy of the approach highlight its potential as a practical solution for clinical use, enabling quick and reliable seizure detection.

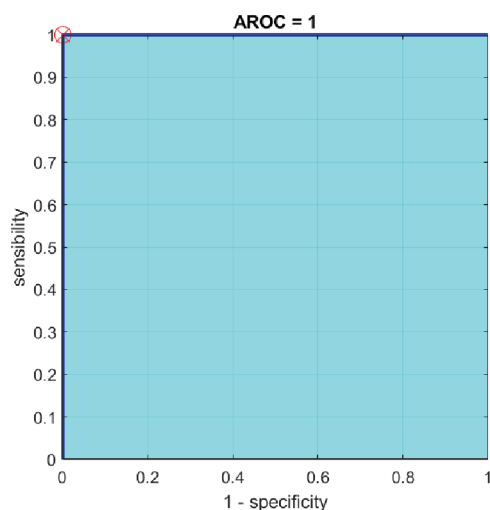


Figure 10: ROC plot for PCA, ICA, LDA and SVM, NB, KNN. ICA, independent component analysis; KNN, K-nearest neighbors; LDA, linear discriminant analysis; NB, Naive Bayes; PCA, principal component analysis; SVM, support vector machine.

To validate the robustness of the performance differences observed across models, we propose incorporating statistical significance testing in future experiments. Future work will validate the method using larger datasets to assess its robustness and generalizability across diverse populations and seizure types.

Declaration

Competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Authors' contributions

Rabel Guharoy, Nanda Dulal Jana, and Suparna Biswas contributed to the study conception and design. Coding implementation, coding analysis, and writing of the first draft of the manuscript were done by Rabel Guharoy. Subhayu Ghosh participated in the revision of the paper and completed the final draft. Lalit Garg contributed with various useful recommendations during the development of the paper. Nanda Dulal Jana and Suparna Biswas

guided throughout the research and supervised every aspect of this work.

Data availability and access

Data will be made available upon request.

Ethical and informed consent for data used

Not applicable.

Funding information

There is no funding from anywhere for this work.

References

- [1] Al-Qerem, A., Kharbat, F., Nashwan, S., Ashraf, S., Blaou, K.: General model for best feature extraction of EEG using discrete wavelet transform wavelet family and differential evolution. *International Journal of Distributed Sensor Networks* 16(3), 1550147720911009 (2020)
- [2] Qu, H., Gotman, J.: Improvement in seizure detection performance by automatic adaptation to the EEG of each patient. *Electroencephalography and clinical Neurophysiology* 86(2), 79–87 (1993)
- [3] Sirven, J.I.: Epilepsy: a spectrum disorder. *Cold Spring Harbor perspectives in medicine* 5(9) (2015)
- [4] Delanty, N., Vaughan, C.J., French, J.A.: Medical causes of seizures. *The Lancet* 352(9125), 383–390 (1998)
- [5] Caplan, R.: Adhd in pediatric epilepsy: fact or fiction? *Epilepsy Currents* 17(2), 93–95 (2017)
- [6] Sirven, J.I.: Sticks and stones: What's in the name of epilepsy. *Epilepsy Currents* 14(5), 257–258 (2014)
- [7] Siddiqui, M.K., Morales-Menendez, R., Huang, X., Hussain, N.: A review of epileptic seizure detection using machine learning classifiers. *Brain informatics* 7(1), 1–18 (2020)
- [8] Thurman, D.J., Beghi, E., Begley, C.E., Berg, A.T., Buchhalter, J.R., Ding, D., Hesdorffer, D.C., Hauser, W.A., Kazis, L., Kobau, R., et al.: Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia* 52, 2–26 (2011)

- [9] Alessio, S.M., Alessio, S.M.: Discrete wavelet transform (dwt). Digital signal processing and spectral analysis for scientists: concepts and applications, 645–714 (2016)
- [10] Castells, F., Laguna, P., Sörnmo, L., Bollmann, A., Roig, J.M.: Principal component analysis in ECG signal processing. EURASIP Journal on Advances in Signal Processing 2007, 1–21 (2007)
- [11] James, C.J., Hesse, C.W.: Independent component analysis for biomedical signals. Physiological measurement 26(1), 15 (2004)
- [12] Sharma, A., Paliwal, K.K.: Linear discriminant analysis for the small sample size problem: an overview. International Journal of Machine Learning and Cybernetics 6, 443–454 (2015)
- [13] Suthaharan, S., Suthaharan, S.: Support vector machine. Machine learning models and algorithms for big data classification: thinking with examples for effective learning, 207–235 (2016)
- [14] Ting, S., Ip, W., Tsang, A.H., et al.: Is Naive Bayes a good classifier for document classification. International Journal of Software Engineering and Its Applications 5(3), 37–46 (2011)
- [15] Steinbach, M., Tan, P.-N.: KNN: K-nearest neighbors. In: The Top Ten Algorithms in Data Mining, pp. 165–176. Chapman and Hall/CRC, ??? (2009)
- [16] Geng, X., Li, D., Chen, H., Yu, P., Yan, H., Yue, M.: An improved feature extraction algorithms of EEG signals based on motor imagery brain-computer interface. Alexandria Engineering Journal 61(6), 4807–4820 (2022)
- [17] Prochazka, A., Kukal, J., Vysata, O.: Wavelet transform use for feature extraction and EEG signal segments classification. In: 2008 3rd International Symposium on Communications, Control and Signal Processing, pp. 719–722 (2008)
- [18] Çınar, S., Acir, N.: A novel system for automatic removal of ocular artefacts in EEG by using outlier detection methods and independent component analysis. Expert Systems with Applications 68, 36–44 (2017)
- [19] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R.: Advances in knowledge discovery and data mining. (1996)
- [20] Islam, M.Z., D'Alessandro, S., Furner, M., Johnson, L., Gray, D., Carter, L.: Brand switching pattern discovery by data mining techniques for the telecommunication industry in australia. Australasian Journal of Information Systems 20, 1–17 (2016)
- [21] Aljumah, A.A., Ahamad, M.G., Siddiqui, M.K.: Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University-Computer and Information Sciences 25(2), 127–136 (2013)
- [22] Aljumah, A., Siddiqui, M.: Data mining perspective: Prognosis of life style on hypertension and diabetes. International Arab Journal of Information Technology (IAJIT) 13(1) (2016)
- [23] Siddiqui, M.K., Menendez, R.M., Gupta, P.K., Hussain, F., Khatoon, K., Ahmad, S., et al.: Correlation between temperature and covid-19 (suspected, confirmed and death) cases based on machine learning analysis (2020)
- [24] Aljumah, A.A., Siddiqui, M.K.: Hypertension interventions using classification based data mining. Research Journal of Applied Sciences, Engineering and Technology 7(17), 3593–602 (2014)
- [25] Almazyad, A.S., Ahamad, M.G., Siddiqui, M.K., Almazyad, A.S.: Effective hypertensive treatment using data mining in saudi arabia. Journal of clinical monitoring and computing 24, 391–401 (2010)
- [26] Singh, G.A.P., Gupta, P.: Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Computing and Applications 31, 6863–6877 (2019)
- [27] Amin, H.U., Malik, A.S., Ahmad, R.F., Badruddin, N., Kamel, N., Hussain, M., Chooi, W.-T.: Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. Australasian physical & engineering sciences in medicine 38, 139–149 (2015)
- [28] Al-Kharaz, A.A., Alwahhab, A.B.A., Sabeeh, V.: Innovative date fruit classifier based on scatter wavelet and stacking ensemble. HighTech and Innovation Journal 5(2), 361–381 (2024) <https://doi.org/10.28991/HIJ-2024-05-02-010>
- [29] Ait Talghalit, I., Alami, H., El Alaoui, S.O.: Contextual semantic embeddings based on transformer models for arabic biomedical questions classification. High-Tech and Innovation Journal 5(4), 1024–1037 (2024) <https://doi.org/10.28991/HIJ-2024-05-04-011>
- [30] Armoogum, S., Motean, K., Dewi, D.A., Kurniawan, T.B., Kijsonporn, J.: Breast cancer prediction using transfer learning-based classification model. Emerging Science Journal 8(6), 2373–2384 (2024) <https://doi.org/10.28991/ESJ-2024-08-06-014>
- [31] Goh, K.W., Surono, S., Afiatin, M.Y.F., Mahmudah, K.R., Irsalinda, N., Chaimanee, M., Onn, C.W.: Comparison of activation functions in convolutional neural network for poisson noisy image classification. Emerging Science Journal 8(2), 592–602 (2024) <https://doi.org/10.28991/ESJ-2024-08-02-014>

- [32] Martis, R.J., Acharya, U.R., Min, L.C.: ECG beat classification using pca, lda, ica and discrete wavelet transform. *Biomedical Signal Processing and Control* 8(5), 437–448 (2013)
- [33] Shi, B., Wang, Q., Yin, S., Yue, Z., Huai, Y., Wang, J.: A binary harmony search algorithm as channel selection method for motor imagery-based BCI. *Neurocomputing* 443, 12–25 (2021)
- [34] Kapoor, B., Nagpal, B., Jain, P.K., Abraham, A., Gabralla, L.A.: Epileptic seizure prediction based on hybrid seek optimization tuned ensemble classifier using EEG signals. *Sensors* 23(1), 423 (2022)
- [35] Tan, P., Wang, X., Wang, Y.: Dimensionality reduction in evolutionary algorithms-based feature selection for motor imagery brain-computer interface. *Swarm and Evolutionary Computation* 52, 100597 (2020)
- [36] Priyanka, S., Dema, D., Jayanthi, T.: Feature selection and classification of epilepsy from EEG signal. In: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 2404–2406 (2017)
- [37] Agrawal, A., Garg, L., Dauwels, J.: Application of empirical mode decomposition algorithm for epileptic seizure detection from scalp EEG. *Transactions of Japanese Society for Medical and Biological Engineering* 51(Supplement), 207 (2013)
- [38] Salem, N., Hussein, S.: Data dimensional reduction and principal components analysis. *Procedia Computer Science* 163, 292–299 (2019)
- [39] Li, S., Zhou, W., Yuan, Q., Geng, S., Cai, D.: Feature extraction and recognition of ictal EEG using EMD and SVM. *Computers in biology and medicine* 43(7), 807–816 (2013)
- [40] Doma, V., Pirouz, M.: A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. *Journal of Big Data* 7(1), 18 (2020)