**IAES International Journal of Artificial Intelligence (IJ-AI)**
Vol. 15, No. 6, December 2026, pp. 5354~5366
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5354-5366 

# Seizure detection from biomedical signals using empirical decomposition and synthetic minority oversampling

**Lahcen Zougagh, Hamid Bouyghf, Mohammed Nahid**
Laboratory of Engineering Sciences and Biosciences, Faculty of Sciences and Techniques, Hassan II University, Mohammedia, Morocco

**Article Info**

***Article history:***

Received Jul 7, 2025
Revised Aug 6, 2026
Accepted Aug 31, 2026

***Keywords:***

Classification
Electrocardiogram
Electroencephalographic
Empirical mode decomposition
SMOTE

**ABSTRACT**

Early detection of epileptic seizures is crucial for clinical diagnosis and preventing disease propagation. This study presents a distinctive method for seizure detection, integrating electroencephalogram (EEG) and electrocardiogram (ECG) data in a healthcare setting. Using empirical mode decomposition (EMD) to analyze time-frequency data, the method effectively classifies seizures. The synthetic minority over-sampling technique (SMOTE) was employed to enhance the dataset by balancing the number of epileptic data points. Various machine learning classifiers, such as decision trees (DTs), support vector machines (SVMs), k-nearest neighbors (KNNs), and logistic regression (LR), were implemented to distinguish between normal and epileptic states. The study uses four recordings; each file contains approximately three hours of digitized EEG impulses, along with the ECG signal, recorded across 23 EEG channels and one ECG channel. both signals are preprocessed and decomposed using EMD. EMD together with SVM yielded the best results in terms of accuracy (97.22%), specificity (96.58%), and sensitivity (97.89%). These results confirm that the proposed method, which uses SMOTE and the ECG signal as additional information, improves identification performance and can effectively and efficiently detect epileptic seizures automatically.

*This is an open access article under the CC BY-SA license.*

***Corresponding Author:***

Lahcen Zougagh
Laboratory of Engineering Sciences and Biosciences, Faculty of Sciences and Techniques
Hassan II University
Mohammedia, Morocco
Email: lahcen.zougagh-etu@etu.univh2c.ma

## 1. INTRODUCTION

Epilepsy is the most common neurological disorder in the world and affects millions of people [1]. People with this condition have seizures that happen over and over again. These seizures happen when the brain's normal electrical activity suddenly changes, causing convulsions, sensory problems, and changes in consciousness [2]. Pyramidal neurons create localized electric fields at the cellular level by acting as intrinsic electrical dipoles [3]. When these dipolar fields add up in space and time, they travel to the skull, where they can be traced as electroencephalographic (EEG) signals without any surgery [4]. EEG is different from static neuroimaging methods because it gives us real-time, high-temporal-resolution information about how the brain works over long periods of time, from minutes to several days [5]. There are many different ways to record EEGs in a clinical setting, such as standard EEG, video-EEG monitoring, and ambulatory Holter EEG [6], [7]. Recent studies show how urgently smart diagnostic tools that can efficiently manage the complex character of physiological data while addressing problems in clinical datasets including unequal class distribution and variations in seizure types are needed.

Capable of adaptively decomposing EEG and electrocardiogram (ECG) into intrinsic mode functions (IMFs), empirical mode decomposition (EMD) has been extensively recognized as a potent tool for researching nonstationary biomedical signals [8], [9]. Without a priori basis functions, it is also able to. This adaptability provides better feature representations for classification models and lets EMD expose latent physiological dynamics across temporal ranges. Previous research has indicated that EMD is effective in identifying epileptic discharges; performance improvements have been noted when used in concert with feature selection and deep learning pipelines [10]–[12]. Relatively few studies, especially within a unified classification architecture, have investigated its synergistic use across both EEG and ECG domains, though.

Synthetic minority over-sampling technique (SMOTE) continues to be a pillar method for reducing the negative consequences of class imbalance in supervised learning applications. In the context of epilepsy, where the frequency of seizure events is small relative to the dataset, conventional classifiers sometimes cannot generalize across minority class boundaries. SMOTE redefines decision limits and improves classifier sensitivity to epileptic patterns by aggregating new examples of underrepresented classes [13], [14].

This work develops a complete framework integrating EMD for signal decomposition, SMOTE for minority class augmentation, and a quartet of classifiers: support vector machines (SVM), k-nearest neighbors (KNN), decision tree (DT), and logistic regression (LR) to address these challenges. Every classifier gain from the rebalanced data distributions following SMOTE augmentation and uses the enhanced feature space generated from EMD-transformed signals. Underlining the success of the framework in separating between epileptic and non-epileptic cases, the experimental results show a clear increase in classification accuracy, sensitivity, and specificity.

## 2. METHOD

This section presents a framework for identifying seizures in an EEG and ECG dataset. This framework is illustrated by Figure 1. The procedure is divided into four steps: data preparation, EMD, application of the SMOTE technique, and application of machine learning classifiers.

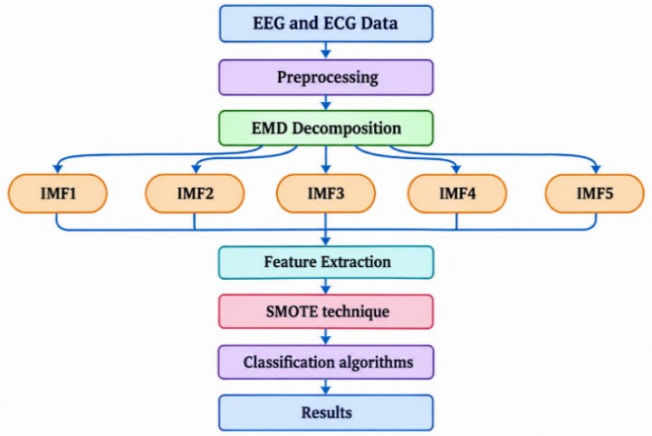


Figure 1. Proposed approach framework

### 2.1. Electroencephalographic and electrocardiogram data

The experimental data were collected from the Children's Hospital Boston–MIT (CHB–MIT) scalp EEG database, which contains recordings from 22 pediatric patients with medically intractable epilepsy. The EEG signals were acquired at a sampling frequency of 256 Hz with 16-bit resolution. Electrode placement and channel nomenclature followed the international 10–20 system [15], [16]. Although several channel configurations were used across the recordings, the most common configuration comprised 23 EEG channels. Most recordings lasted approximately one hour, while a smaller number had durations of two, three, or four hours. To obtain a reliable evaluation and minimize patient-specific bias, a subject-wise five-fold cross-validation strategy was adopted. Within each fold, all recordings belonging to a given patient were assigned exclusively to either the training or test set, ensuring that data from the same patient never appeared in both sets simultaneously. Moreover, the distribution of seizure and non-seizure segments was balanced as closely as possible in the training and test partitions. This patient-independent evaluation protocol gives a more reliable evaluation of the model's ability to perform on patients it has not seen before [17].

### 2.2. Preprocessing

In this study, the EEG and ECG data were divided into non-overlapping 10-second segments, providing a suitable compromise between temporal resolution and the stability of the extracted patterns. Very short segments may contain insufficient physiological information, whereas excessively long segments may mask transient changes associated with seizure activity. Before feature extraction, both EEG and ECG signals were preprocessed to reduce baseline noise and high-frequency interference. The EEG signals were filtered with a fourth-order Butterworth bandpass filter with cutoff frequencies of 0.5 and 40 Hz. This frequency range preserves the principal clinically relevant EEG rhythms while attenuating slow baseline fluctuations, high-frequency noise, and muscle-related artifacts. Similarly, the ECG signals were filtered within the 0.5–45 Hz frequency range to preserve the essential morphological characteristics of cardiac activity while reducing baseline wander and high-frequency interference, including contamination from the 50 Hz power-line frequency.

### 2.3. Feature extraction

#### 2.3.1. Empirical mode decomposition

Electroencephalograms (EEGs) provide a noninvasive method of monitoring cerebral activity and identifying seizures. However, EEG signals are notoriously non-stationary, noisy, and complex, which makes identifying seizures a significant challenge. Conventional signal processing methods, including Fourier and wavelet transforms, are limited when managing non-stationary data. EMD has received significant attention because it separates signals into intrinsic oscillatory modes already present in the data. This study explores applying EMD to EEG and ECG signal processing for seizure detection.

In EMD as illustrated by Figure 2, a signal is split into a set of IMFs, which are subsequently processed adaptively [18]. Every IMF satisfies two different criteria. First, either the number of extrema and zero crossings must be equal or vary by no more than one. The average of the upper and lower envelopes must also equal zero. By spotting local extremes, building envelopes, and separating the mean envelope from the signal, the EMD method gently extracts IMFs [19]. This process keeps on until the residual signal is monotonic or shows no more than two extremes.

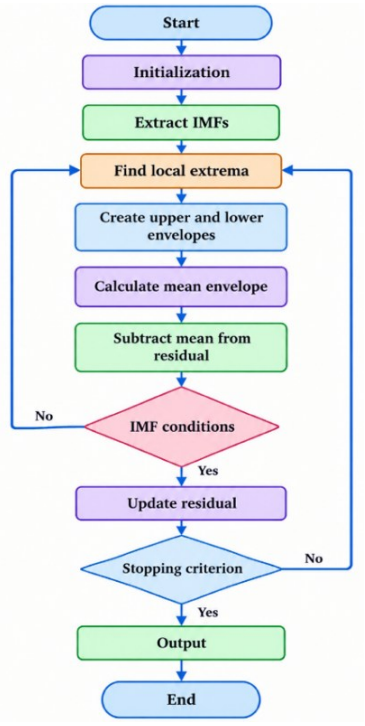


Figure 2. EMD diagram

This study decomposed each EEG and ECG segment through EMD. This process produces a series of IMFs by adaptively decomposing a non-stationary signal, $x(t)$, as shown in (1).

$$x(t) = \sum_{i=1}^{K} IMF_i(t) + r(t) \quad (1)$$

Where IMFs are oscillatory modes, and $r(t)$ is the residual. The EMD method is particularly suitable for epileptic signals because seizure dynamics manifest as nonlinear and nonstationary oscillations.

### 2.3.2. Application of empirical mode decomposition to EEG and ECG segments

To characterize seizure-related changes in brain and cardiac activity, representative seizure-free and ictal EEG–ECG segments were analyzed using EMD, with each signal transformed into five IMFs. The EEG results presented in Figures 3 and 4 show clear differences between the two physiological states. Compared with seizure-free activity, the ictal segment exhibits increased amplitudes and more pronounced high-frequency oscillations across the IMFs, consistent with the hypersynchronous neuronal activity associated with epileptic seizures. Similarly, the ECG decompositions shown in Figures 5 and 6 reveal changes in signal morphology and oscillatory behavior during the ictal period. These variations suggest that seizures also influence cardiac dynamics through autonomic pathways linking cerebral activity to heart-rate regulation. Overall, the observed differences indicate that ECG signals provide complementary information to EEG recordings, supporting the use of a multimodal EEG–ECG framework for seizure analysis.

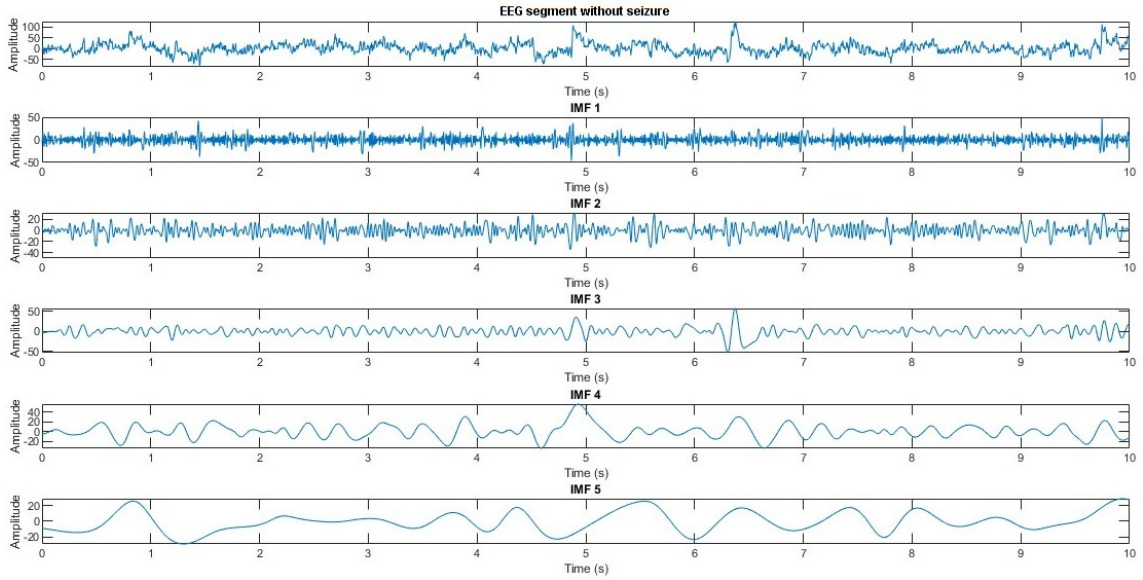


Figure 3. EEG segment without seizure

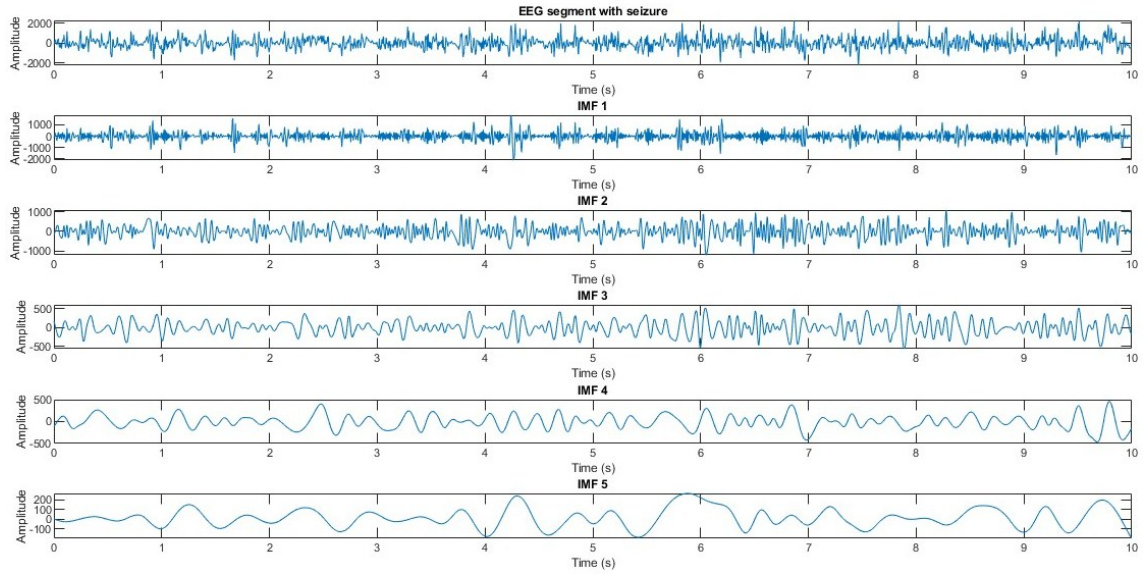


Figure 4. EEG segment with seizure

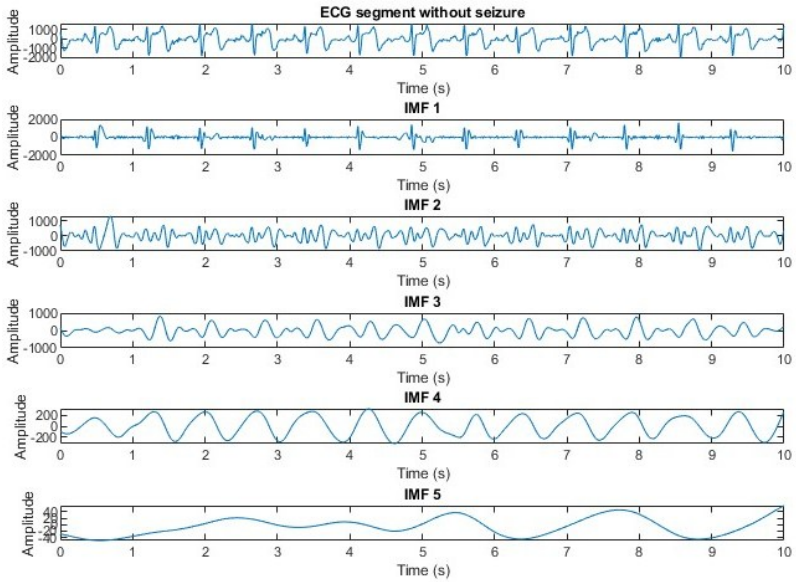


Figure 5. ECG segment without seizure

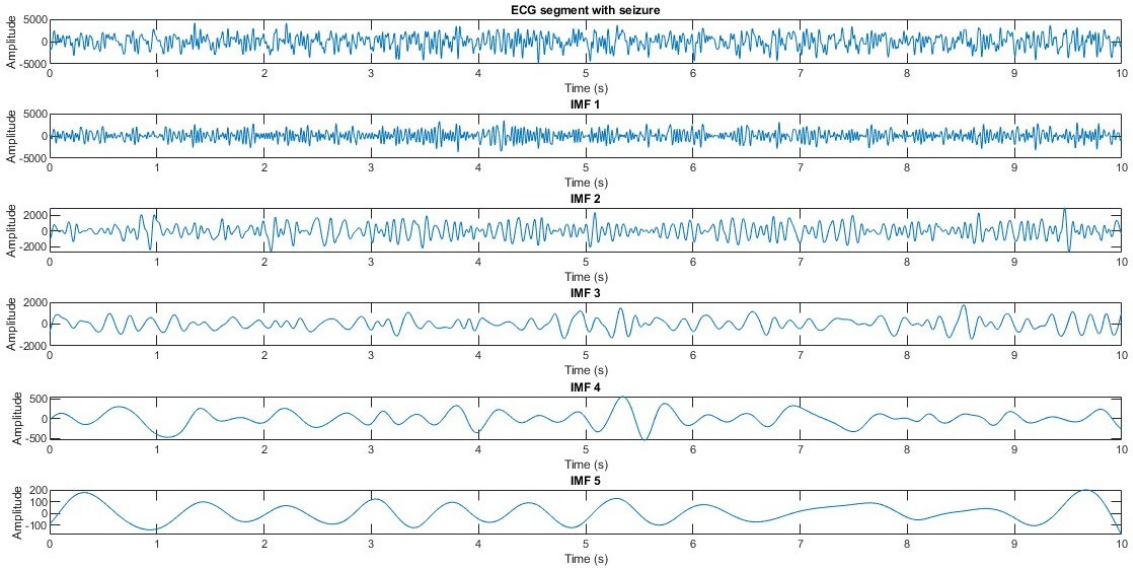


Figure 6. ECG segment with seizure

### 2.3.3. Features

Using EMD, each EEG and ECG segment was partitioned into IMFs to capture the nonlinear and nonstationary dynamics associated with seizure activity. Then, a compact set of statistical and energy-based characteristics was acquired from the IMFs, as summarized in (2) to (11). The maximum, minimum, and absolute mean describe extreme amplitudes and overall signal magnitude, whereas the variance and standard deviation quantify amplitude variability. Skewness and kurtosis characterize the asymmetry and peakedness of the amplitude distribution, making them sensitive to spikes, sharp waves, and other transient epileptiform patterns. EEG energy reflects changes in signal strength associated with neuronal recruitment and synchronization during seizures. Similarly, ECG energy and standard deviation capture seizure-related variations in cardiac activity resulting from autonomic modulation [20]. Together, these interpretable features provide a concise representation of complementary changes in EEG and ECG amplitude, variability, morphology, and energy.

– Maximum value, as in (2).

$$max = X_{max} \tag{2}$$

– Minimum value, as in (3).

$$min = X_{min} \tag{3}$$

– Absolute mean, as in (4).

$$m_x = \frac{1}{A}\sum_{a=1}^{A}|X_a| \tag{4}$$

– Variance, as in (5).

$$s_x^2 = \frac{1}{A-1}\sum_{a=1}^{A}(X_a - m_x) \tag{5}$$

– Skewness, as in (6).

$$sk = \frac{1}{(A-1)\sigma_x^3}\sum_{a=1}^{A}(X_a - m_x)^3 \tag{6}$$

– Kurtosis, as in (7).

$$kr = \frac{1}{(A-1)\sigma_x^4}(\sum_{a=1}^{A}(X_a - m_x)^4) - 3 \tag{7}$$

– Energy for an EEG segment, as in (8).

$$E_{EEG10}(i,n) = \sum_{K=1}^{L}|EEG(i,K)|^2 \tag{8}$$

– Energy for an ECG segment, as in (9).

$$E_{ECG10}(n) = \sum_{K=1}^{L}|ECG(K)|^2 \tag{9}$$

– Standard deviations of the EEG, as in (10).

$$\sigma_{EEG10}(i,n) = \left(\frac{1}{L}\sum_{K=1}^{L}(EEG(i,K) - \overline{EEG(i)})^2\right)^{\frac{1}{2}} \tag{10}$$

– Standard deviations of the ECG, as in (11).

$$\sigma_{ECG10}(n) = \left(\frac{1}{L}\sum_{K=1}^{L}(ECG(K) - \overline{ECG(n)})^2\right)^{\frac{1}{2}} \tag{11}$$

### 2.4. Synthetic minority over sampling technique

In machine learning, the SMOTE is a data sampling approach that efficiently solves data imbalance. Class imbalance is the phenomena whereby one class dominates another in proportionate measure. This disparity can lead to biassed models favoring the dominant class, so distorting the whole outcomes and maybe leading to erroneous conclusions. SMOTE solves this by creating fresh minority class examples instead of replicating already existing ones [21]. Making a synthetic sample is obtaining a sample from the minority class sample $(x_i)$ then linear interpolation combines it with one of its $(K)$ nearest neighbors $(x_{zi})$. Equation $x = x_i + \lambda \cdot (x_{zi} - x_i)$, finds the synthetic sample x by means of a random variable ranging from 0 to 1. This procedure adds more data and preserves the links inside the data [22]. The following is a detailed outline of the SMOTE algorithm:

– Choose a minority group. The initial step in this process is the selection of a sample from a minority group.
– Determine the nearest neighbors. Identify the KNN in the characteristic domain.
– Random neighbor selection: randomly select one of these neighbors.
– Synthetic instance generating: generate a synthetic instance aligned with the segment linking the chosen minority class instance to its designated neighbor.

This technique balances the distribution of classes within the dataset and improves the model's ability to generalize. To address the prevalent class imbalance in seizure datasets, we applied SMOTE

exclusively to the training data. This study selected k = 5 nearest neighbors and ensured that no synthetic samples were included in the test set. Using IMF-based features with balanced training data improves classifier performance while maintaining the integrity of cross-validation and evaluation.

### 2.5. Seizure detection classifiers

Classification is a technique for categorizing a given set of data into different groups and applies to two types of data: structured and unstructured. The aim of classification predictive modelling is to estimate the relationship between input factors and discrete output values. In this research, we used four different supervised machine learning classifiers: SVM, LR, KNN, and DT.

#### 2.5.1. Support vector machine

The SVM is a frequent supervised learning method for categorizing EEG signals because it can accommodate high-dimensional and intricate data well [23]. SVMs work by identifying an ideal hyperplane that optimizes the separation between distinct classes. However, EEG data are inherently nonlinear due to their complex neurophysiological origins, making linear SVMs inadequate for precise categorization. The radial basis function (RBF) is often deployed in SVMs to transform the EEG feature vector into a hyperplane, which facilitates linear separation [24]. The RBF kernel is expressed in (12).

$$k(x_i, y_j) = e^{\left(-\gamma \|x_i - x_j\|^2\right)} \tag{12}$$

In this context, $x_i$ and $x_j$ are EEG feature vectors, $(x_i - x_j)$ signifies the squared Euclidean distance, and $y$ is a hyperparameter that regulates the impact of each training input. In biosignal applications, the RBF increases the accuracy of classification; hence, SVMs are a useful instrument for neurophysiological study and medical diagnosis.

#### 2.5.2. Decision tree

A DT is a machine learning method that organizes EEG and ECG data in a tree-like structure, breaking it down using specific features to make decisions. Based on the most significant EEG characteristics, DT divides data often into smaller groups that simplify the classification process. A DT comprises internal nodes and a root node storing all EEG and ECG data, offering both a starting point and features-based decision rules are shown as branches; the possible results of every choice are shown as leaf nodes. Final classification labels show as seizure against non-seizure [25]. Using an impurity metric such information gain or Gini impurity, each internal node is split into sections depending on the most valuable feature.

- Information gain, expressed in (13), quantifies the entropy of the dataset. Where $p_i$ is class *i*'s probability.

$$H(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \tag{13}$$

- Gini impurity, shown in (14), measures the degree of impurity in a split. A lower value of the Gini coefficient is indicative of a pure division.

$$Gini(S) = 1 - \sum_{i=1}^{c} p_i^2 \tag{14}$$

#### 2.5.3. Logistic regression

One type of machine learning extensively applied to solve problems including prediction is LR. Using biological signals—quantitative markers of how the body is functioning that are observable helps especially in predicting epileptic seizures. Depending on the features of the inherited EEG and ECG, LR is a statistical method that assesses the probability of classifying a given EEG segment as either a seizure or a non-seizure [26]. The method translates this data into a probability score ranging from 0 to 1 using a sigmoid function, so indicating the likelihood of a seizure as defined in (15). Where $P(Y = 1|X)$ is the possibility of a seizure occurring, $\beta_0$ is the intercept, $\beta_i$ are the learned weights, and $X_i$ represent the extracted features.

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i X_i)}} \tag{15}$$

#### 2.5.4. K-nearest neighbors

KNN is one of the simplest and easiest machine learning algorithms based on supervised learning. It is a nonparametric method, indicating that the underlying data need no assumptions for categorization [27].

It suggests the resemblance between the new class and existing instances, categorizing the new class among the most closely related accessible categories. The estimate derived from the KNN system is susceptible to local noise and is often unsatisfactory. A higher value of k results in a smoother categorization border, while a lower k yields a more intricate boundary.

## 3. RESULTS AND ANALYSIS

This section analyses the results of the proposed seizure prediction module. It evaluates the effectiveness of the combined EMD, SMOTE, and machine learning classifiers in the system. The performance of the module is discussed, and comparisons are made to demonstrate its efficacy.

### 3.1. Performance measurements

Four measures were used to evaluate the effectiveness of the proposed detector. The area under the curve (AUC) is a common metric that evaluates the performance of a classification model over all its possible value ranges.

- The AUC is a common metric that evaluates the performance of a classification model over all its possible value ranges.
- Accuracy: classification accuracy is determined as the proportion of epochs accurately identified by the model, as presented in (16).

$$Accuracy = \frac{TN+TP}{TN+TP+FP+FN} \times 100(\%) \quad (16)$$

- Specificity: is determined as the proportion of marked non-seizure events accurately recognized as non-seizure by the model, as represented in (17).

$$Specificity = \frac{TN}{TN+FP} \times 100(\%) \quad (17)$$

- Sensitivity: is determined as the proportion of seizure events accurately recognized as seizure events by the model, as described in (18).

$$Sensitivity = \frac{TP}{TP+FN} \times 100(\%) \quad (18)$$

### 3.2. Result

The experimental dataset consisted of 160 EEG/ECG segments, including 100 epileptic and 60 non-epileptic segments. This uneven distribution of classes could cause the classifiers to favor the epileptic class. To mitigate this effect, SMOTE generated 40 synthetic samples for the minority non-epileptic class. The resulting balanced dataset contained 200 segments, with 100 in each class. This equal representation reduced class-related bias and enabled a fairer assessment of the classifiers' ability to distinguish between epileptic and non-epileptic EEG/ECG patterns. Confusion matrices were generated for both experimental conditions. Figure 7 shows the classification results from the original imbalanced dataset of 160 segments, and Figure 8 shows the results after balancing the dataset with SMOTE. Minor discrepancies between performance metrics calculated directly from the confusion matrices and those reported in the tables may result from numerical rounding and averaging results across cross-validation folds.

Table 1 shows the average results achieved for each performance metric by the four classification algorithms chosen in this research, with ECG data added to the EEG data as additional information. The first parameter examined was accuracy. The SVM model reached a value of 0.92, the DT model reached a value of 0.91, the KNN model achieved a value of 0.91, and the LR model yielded a value of 0.89. Regarding the specificity parameter, the SVM algorithm had the highest value (0.84), while the LR algorithm had the lowest (0.79). For the sensitivity metric, all algorithms yielded scores higher than 0.9. Regarding the AUC parameter, the SVM and DT algorithms yielded scores higher than 0.91, while the KNN and LR models achieved a score lower than 0.90.

The classification results presented in Table 2 were significantly improved by using the SMOTE technique on the combined ECG and EEG data for all classifiers. The first parameter examined was accuracy. The SVM model reached a value of 0.97, the DT model reached a value of 0.92, the KNN model reached a value of 0.88, and the LR model yielded a value of 0.86. Regarding the sensitivity parameter, the SVM algorithm had the highest score (0.978), while the LR algorithm had the lowest (0.866). For the specificity metric, the model with the best results was the SVM, with a score of 0.96. Regarding the AUC parameter, the

SVM, DT, and KNN algorithms yielded scores higher than 0.91, while the LR model achieved a score lower than 0.90.

As illustrated in Figures 9 to 12, the receiver operating characteristic (ROC) curves for the top four classifiers: DT, LR, SVM, and KNN; trained to discriminate between EEG and ECG signals showing seizures and non-seizure states are shown. Among the various algorithms examined, SVM demonstrated the most optimal performance, exhibiting an average AUC value of 0.958. Conversely, the LR algorithm exhibited the least favorable performance. The ROC curves demonstrate a trade-off between sensitivity and specificity. The SVM algorithm demonstrates a robust correlation with the upper left quadrant of the ROC space, thereby signifying its superior classification performance. Classifiers with ROC curves that are proximate to the upper-left corner are regarded as superior, while those with ROC curves that are proximate to the 45-degree diagonal are deemed to be less accurate in future prediction.

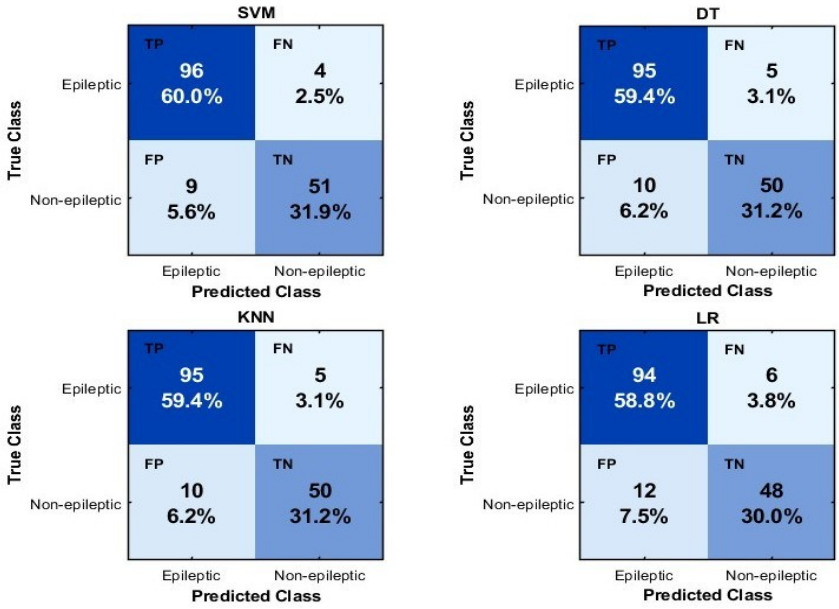


Figure 7. Confusion matrices of all classifiers before SMOTE

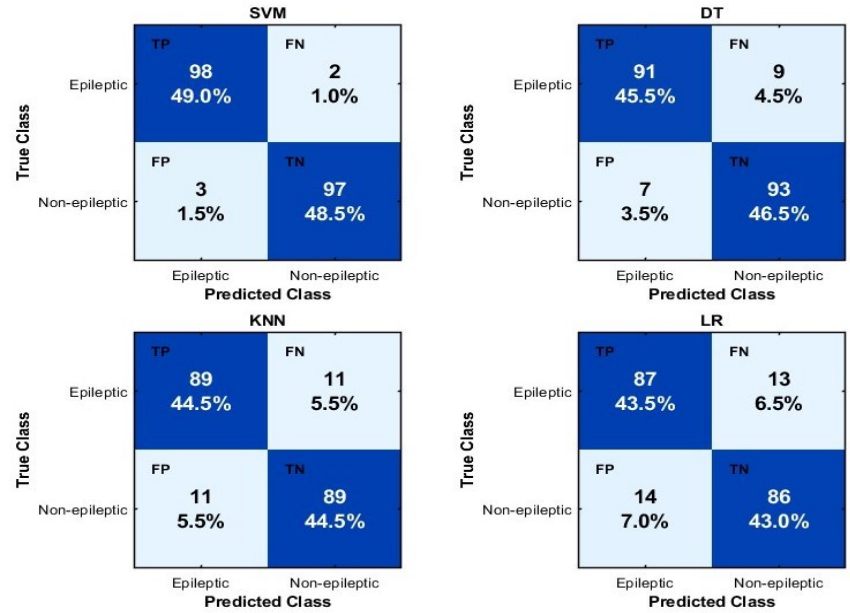


Figure 8. Confusion matrices of all classifiers after SMOTE

Table 1. Performance of algorithm without SMOTE on EEG and ECG data

| Results | Sensitivity (%) | Specificity (%) | Accuracy (%) | F1-score (%) | AUC |
|---|---|---|---|---|---|
| SVM | 95.12 | 84.54 | 92.45 | 95.21 | 0.910 |
| DT | 94.42 | 82.41 | 91.32 | 94.20 | 0.933 |
| KNN | 94.25 | 82.32 | 91.19 | 94.09 | 0.853 |
| LR | 93.10 | 79.07 | 89.89 | 92.78 | 0.733 |

Table 2. Performance of algorithm with SMOTE on EEG and ECG data

| Results | Sensitivity (%) | Specificity (%) | Accuracy (%) | F1-score (%) | AUC |
|---|---|---|---|---|---|
| SVM | 97.89 | 96.58 | 97.22 | 97.17 | 0.958 |
| DT | 91.16 | 92.91 | 92.01 | 92 | 0.938 |
| KNN | 89.38 | 88.40 | 88.87 | 88.89 | 0.918 |
| LR | 86.67 | 85.91 | 86.28 | 86.34 | 0.852 |

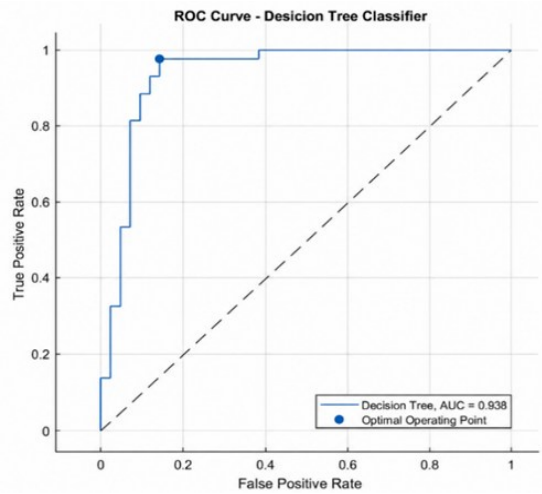


Figure 9. ROC curve derived from DT

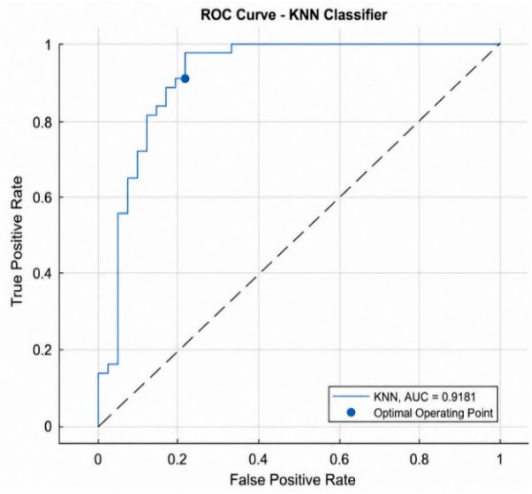


Figure 10. ROC curve derived from KNN

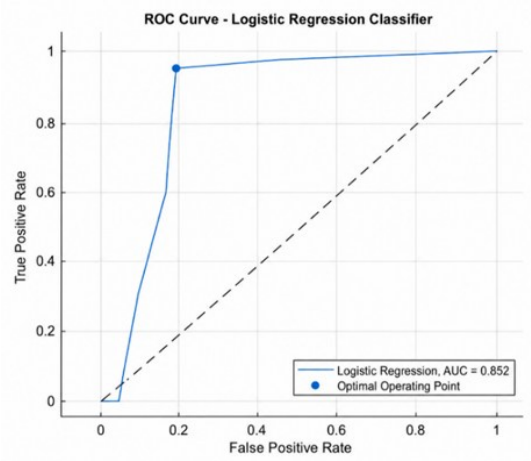


Figure 11. ROC curve derived from LR

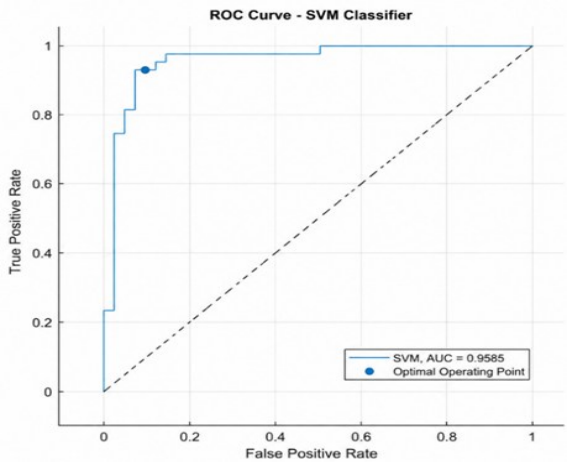


Figure 12. ROC curve derived from SVM

### 3.3. Discussion

Table 3 demonstrates the complementary contribution of every component of the designed framework. Compared with the EEG-only baseline, EMD improves performance by decomposing nonstationary EEG signals into oscillatory modes that facilitate the extraction of seizure-related multiscale patterns. The addition of ECG further enhances sensitivity and AUC, indicating that seizure-induced autonomic changes provide complementary information to cortical EEG activity. SMOTE produces the

greatest improvement by reducing class imbalance and limiting the classifier's bias toward the majority class. To avoid data leakage, SMOTE was applied exclusively within the training folds.

The complete EEG–ECG–EMD–SMOTE–SVM framework achieved a sensitivity of 97.89%, specificity of 96.58%, accuracy of 97.22%, and AUC of 0.958. Its accuracy exceeded those reported by Pattnaik *et al.* [28], Alotaiby *et al.* [29], Ullah *et al.* [30], and Acharya *et al.* [31] by 4.22%, 3.02%, 1.42%, and 2.02%, respectively. These findings highlight the notable discriminative capacity of the presented multimodal approach. Nevertheless, direct comparisons should be interpreted cautiously because the studies used several different types of data, class distributions, and validation strategies, while several did not report all evaluation metrics.

Despite these promising results, SMOTE-generated samples may not fully represent real physiological variability and may increase the risk of overfitting. Performance also depends on SVM hyperparameter tuning and may not generalize to independent patients. Furthermore, IMF-based statistical features may not capture all available time–frequency information, while clinical EEG and ECG recordings are affected by artifacts, missing channels, interpatient variability, and acquisition differences. Therefore, the proposed framework should be regarded as a clinical decision support tool rather than an independently operating diagnostic system. Future work should include patient-independent, prospective, and multicenter validation, reduced reliance on synthetic oversampling, and evaluation of false alarm rates, detection latency, and real-time performance.

Table 3. Performance comparison of the proposed framework with existing seizure detection techniques

| Method | Sensitivity (%) | Specificity (%) | Accuracy (%) | AUC |
|---|---|---|---|---|
| Pattnaik *et al.* [28] | 91.5 | — | 93.0 | — |
| Alotaiby *et al.* [29] | — | — | 94.2 | — |
| Ullah *et al.* [30] | — | — | 95.8 | — |
| Acharya *et al.* [31] | 95.0 | 90.0 | 95.2 | — |
| Proposed method | 97.89 | 96.58 | 97.22 | 0.958 |

## 4. CONCLUSION

This work combines advanced signal processing and machine learning methods with EEG and ECG signals to create a detailed and organized system for automatically detecting epileptic seizures. EMD is used in the proposed approach for the adaptive decomposition of nonlinear and nonstationary biomedical signals, which allows the extraction of physiologically relevant aspects. The SMOTE efficiently augments the minority (seizure) class, improving the sensitivity of classifiers and addressing the natural class imbalance unique to seizure datasets. With 97.22% accuracy, 96.58% specificity, and 97.89% sensitivity, the EMD-SVM combination of the four machine learning models assessed—SVM, KNN, DT, and LR yielded the most promising results. These results show how important it is to combine different types of signals and emphasize that ECG is a helpful extra signal for detecting seizures.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lahcen Zougagh | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | |
| Hamid Bouyghf | ✓ | | | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | | |
| Mohammed Nahid | | ✓ | | ✓ | | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

## DATA AVAILABILITY

The data supporting the findings of this study are openly available on PhysioNet at https://physionet.org/content/chbmit/1.0.0/.

## REFERENCES


[1] World Health Organization, *Global status report on alcohol and health 2018. Geneva, Switzerland*. Geneva, Switzerland: World Health Organization, 2018.

[2] U. R. Acharya, S. Vinitha Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: a review," *Knowledge-Based Systems*, vol. 45, pp. 147–165, Jun. 2013, doi: 10.1016/j.knosys.2013.02.014.

[3] A. I. Gulyás, T. F. Freund, and S. Káli, "The effects of realistic synaptic distribution and 3D geometry on signal integration and extracellular field generation of hippocampal pyramidal cells and inhibitory neurons," *Frontiers in Neural Circuits*, vol. 10, Nov. 2016, doi: 10.3389/fncir.2016.00088.

[4] M. O. Rahman and M. N. Karim, "Predicting epileptic seizure from electroencephalography (EEG) using Hilbert Huang transformation and neural network," Thesis, Bachelor of Science in Computer Science and Engineering, BRAC University, Dhaka, Bangladesh, 2015.

[5] M. Marino and D. Mantini, "Human brain imaging with high-density electroencephalography: techniques and applications," *Journal of Physiology*, vol. 604, no. 2, pp. 783–812, 2026, doi: 10.1113/JP286639.

[6] K. Zeng, J. Yan, Y. Wang, A. Sik, G. Ouyang, and X. Li, "Automatic detection of absence seizures with compressive sensing EEG," *Neurocomputing*, vol. 171, pp. 497–502, Jan. 2016, doi: 10.1016/j.neucom.2015.06.076.

[7] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition," *Frontiers in Neuroscience*, vol. 12, Mar. 2018, doi: 10.3389/fnins.2018.00162.

[8] N. E. Huang *et al.*, "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, Mar. 1998, doi: 10.1098/rspa.1998.0193.

[9] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, May 2007, doi: 10.1016/j.eswa.2006.02.005.

[10] R. G. Thangarajoo *et al.*, "Machine learning-based epileptic seizure detection methods using wavelet and EMD-based decomposition techniques: a review," *Sensors*, vol. 21, no. 24, 2021, doi: 10.3390/s21248485.

[11] Z. Liu, B. Zhu, M. Hu, Z. Deng, and J. Zhang, "Revised tunable Q-factor wavelet transform for EEG-based epileptic seizure detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1707–1720, 2023, doi: 10.1109/TNSRE.2023.3257306.

[12] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1106–1117, Feb. 2015, doi: 10.1016/j.eswa.2014.08.030.

[13] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.

[14] N. Vakitbilir *et al.*, "Multivariate modelling and prediction of high-frequency sensor-based cerebral physiologic signals: narrative review of machine learning methodologies," *Sensors*, vol. 24, no. 24, Dec. 2024, doi: 10.3390/s24248148.

[15] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.

[16] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-context learning approach for EEG epileptic seizure detection," *BMC Systems Biology*, vol. 12, no. S6, Nov. 2018, doi: 10.1186/s12918-018-0626-2.

[17] M. Agop, A. Gavriluţ, V. Ghizdovăţ, and L. Eva, "Fractalization through stochasticization processes in epileptic dynamics," in *Mathematical Modeling in Bioscience: theory and applications*, H. Dutta, Ed. Academic Press, 2025, pp. 123–139, doi: 10.1016/B978-0-44-315445-4.00012-3.

[18] L. Li, X. Chai, S. Zheng, and W. Zhu, "A denoising method for track state detection signal based on the statistical characteristic of noise," *Journal of Transportation Technologies*, vol. 4, no. 4, pp. 327–336, Oct. 2014, doi: 10.4236/jtts.2014.44029.

[19] L. El Bouny, M. Khalil, and A. Adib, "ECG signal filtering based on CEEMDAN with hybrid interval thresholding and higher order statistics to select relevant modes," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13067–13089, May 2019, doi: 10.1007/s11042-018-6143-x.

[20] Z. Lahcen, B. Hamid, and N. Mohammed, "A PSO-SVM-based approach for classifying ECG and EEG bio signals in seizure detection," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 4, pp. 1144–1157, Aug. 2025, doi: 10.35882/jeeemi.v7i4.1159.

[21] I. M. Alkhawaldeh, I. Albalkhi, and A. J. Naswhan, "Challenges and limitations of synthetic minority oversampling techniques in machine learning," *World Journal of Methodology*, vol. 13, no. 5, pp. 373–378, Dec. 2023, doi: 10.5662/wjm.v13.i5.373.

[22] G. H. L., F. Flammini, S. Srividhya, C. M. L., and S. Selvam, "Computer Science Engineering," in *Computer Science Engineering*, 2024, vol. 1, doi: 10.1201/9781032711157.

[23] D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine Learning: Methods and Applications to Brain Disorders*, A. Mechelli and S. Vieira, Eds. Cambridge, MA, USA: Academic Press, 2019, pp. 101–121, doi: 10.1016/B978-0-12-815739-8.00006-7.

[24] Y. Li, Q. Liu, S. R. Tan, and R. H. M. Chan, "High-resolution time-frequency analysis of EEG signals using multiscale radial basis functions," *Neurocomputing*, vol. 195, pp. 96–103, Jun. 2016, doi: 10.1016/j.neucom.2015.04.128.

[25] S. Das, S. P. Nayak, B. Sahoo, and S. C. Nayak, "Machine learning in healthcare analytics: a state-of-the-art review," *Archives of Computational Methods in Engineering*, vol. 31, no. 7, pp. 3923–3962, Sep. 2024, doi: 10.1007/s11831-024-10098-3.

[26] S. Nusinovici *et al.*, "Logistic regression was as good as machine learning for predicting major chronic diseases," *Journal of Clinical Epidemiology*, vol. 122, pp. 56–69, Jun. 2020, doi: 10.1016/j.jclinepi.2020.03.002.
[27] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Annals of Translational Medicine*, vol. 4, no. 11, Jun. 2016, doi: 10.21037/atm.2016.03.37.
[28] S. Pattnaik, N. Rout, and S. Sabut, "Machine learning approach for epileptic seizure detection using the tunable-Q wavelet transform based time–frequency features," *International Journal of Information Technology*, vol. 14, no. 7, pp. 3495–3505, Dec. 2022, doi: 10.1007/s41870-022-00877-1.
[29] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi, and S. R. Alrshoud, "Epileptic seizure prediction using CSP and LDA for scalp EEG signals," *Computational Intelligence and Neuroscience*, vol. 2017, Oct. 2017, doi: 10.1155/2017/1240323.
[30] I. Ullah, M. Hussain, E. ul H. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Expert Systems with Applications*, vol. 107, pp. 61–71, Oct. 2018, doi: 10.1016/j.eswa.2018.04.021.
[31] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, Sep. 2018, doi: 10.1016/j.compbiomed.2017.09.017.

## BIOGRAPHIES OF AUTHORS



**Lahcen Zougagh** received his license degree in Biomedical Instrumentation and Maintenance from the Institute of Health Sciences (ISSS) at the University Hassan I, Settat, Morocco in 2018 and then a master's degree in Biomedical Engineering from FMPM Marrakech in 2020. He is currently a Ph.D. student at Hassan II University in Mohammedia, Casablanca, Morocco, in the Laboratory of Engineering Sciences and Biosciences (LSIB). Under the direction of Pr. M. Nahid, Professor in the Department of Electrical Engineering at the same university, his research and interests are centered on the development, design, and optimization of electronic systems for biomedical engineering and health sciences. He can be contacted at email: lahcen.zougagh-etu@etu.univh2c.ma.



**Hamid Bouyghf** was born in Errachidia, Morocco, in 1982. He got his B.S. and M.S. degrees in Electrical Engineering and Telecom from the University of Science and Technology in Fez, Morocco, in 2007, and his Ph.D. in Electrical Engineering and Telecom from Hassan II University of Casablanca, Morocco, in 2019. From 2015 until 2019, he worked as a research assistant at the Princeton Plasma Physics Laboratory. Since 2019, he has worked as an assistant professor at the Department of Electrical Engineering of Hassan II University's FST Mohammedia in Casablanca, Morocco. In the topic of IC optimization, he has written numerous articles. His research interests include biomedical electronics, analog IC design, electromagnetic fields, low power design, and BLE applications. He can be contacted at email: hamid.bouyghf@gmail.com.



**Mohammed Nahid** studied electronics and received the B.S diploma in Electronics at ENSET Institute, Mohammedia, Morocco in 1994, received the aggregation diploma of Electronics and Electrical Engineering, from ENSET institute of Rabat, Morocco in 2000. He received the DESA diploma in Computer Engineering, Telecoms and Multimedia in 2004. He conducted doctoral research at images, videos coding and quality assessment under psychovisual quality criteria since 2004 at Mohamed V university of Rabat, Morocco, where he received his Ph.D. in 2010. He is research professor in the Faculty of Science and Technology of Mohammedia since 2011, conducts research on visual perception and its application to coding, understanding, and display of visual information and to human vision, image quality, and digital imaging. Research and projects are also focused on artificial intelligence, embedded systems, biomedical systems, smart grids, digital and analog IC design. He can be contacted at email: mohammed.nahid@gmail.com.