

Comparative Analysis of Machine Learning Algorithms for Real-Time Epileptic Seizure Detection Using CEEMDAN-Based EEG Signal Decomposition

¹K GANGA BHAVANI, ²DR.K.APARNA

¹RESEARCH SCHOLAR, DEPT. OF ECE, JNTU ANANTAPUR, ANDHRA PRADESH

²ASSOCIATE PROFESSOR, DEPT OF ECE, JNTUA COLLEGE OF ENGINEERING KALIKIRI, ANDHRA PRADESH

1kgangabhavani66@gmail.com, aparna.ece@jntua.ac.in

Abstract—Epilepsy affects approximately 50 million people worldwide, necessitating automated electroencephalography (EEG) analysis systems for timely seizure detection. This paper presents a comprehensive comparative analysis of three machine learning classifiers—Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel, K-Nearest Neighbors (KNN), and Random Forest (RF)—for real-time epileptic seizure detection using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). EEG signals from the Bangalore EEG Epilepsy Dataset (BEED) are decomposed into Intrinsic Mode Functions (IMFs) using both traditional EMD and CEEMDAN. Eight statistical features are extracted from each of the five IMFs, yielding a 40-dimensional feature vector per epoch. Experimental results demonstrate that CEEMDAN significantly outperforms EMD in decomposition quality, reducing mode mixing artifacts. Among the classifiers evaluated, SVM with RBF kernel achieves the highest classification accuracy of 92.4% on hold-out validation and 91.2% on 5-fold cross-validation, substantially surpassing KNN (82.5%) and Random Forest (83.1%). The SVM classifier also demonstrates superior seizure detection sensitivity (91.8%) and specificity (97.3%), making it the most suitable choice for clinical deployment. The system is implemented on dual platforms: a MATLAB-based desktop application and a browser-based web application (NeuroSense) performing real-time client-side inference. The proposed CEEMDAN-SVM pipeline achieves a total processing latency under 300 ms per epoch, confirming real-time viability for continuous EEG monitoring.

Keywords—Epilepsy, Seizure Detection, CEEMDAN, EMD, EEG, Support Vector Machine, K-Nearest Neighbors, Random Forest, Intrinsic Mode Functions, BEED Dataset

I. INTRODUCTION

Epilepsy is a chronic neurological disorder characterized by recurrent, unprovoked seizures resulting from abnormal neuronal activity in the brain. According to the World Health Organization, epilepsy affects approximately 50 million people worldwide, with 10–12 million cases in India alone [1].

Electroencephalography (EEG) remains the primary non-

invasive diagnostic tool for epilepsy, recording electrical signals generated by the brain through scalp electrodes. However, manual EEG interpretation by neurologists is subjective, time-consuming, and prone to inter-observer variability, making automated detection systems clinically essential.

Traditional signal processing approaches, including Fourier Transform and Wavelet Transform, impose mathematical assumptions that may not suit the non-stationary and non-linear characteristics of EEG signals [2]. Empirical Mode Decomposition (EMD), introduced by Huang et al. [3], offers a data-driven adaptive alternative by decomposing signals into Intrinsic Mode Functions (IMFs). However, EMD suffers from mode mixing—a phenomenon where oscillations of different time scales merge within a single IMF. To address this limitation, Torres et al. [4] proposed Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), which achieves complete decomposition with negligible reconstruction error through ensemble noise-assisted processing.

While significant research has explored individual classifiers for EEG-based seizure detection, comprehensive comparative analyses evaluating multiple machine learning algorithms on CEEMDAN-extracted features with standardized experimental protocols remain limited. Furthermore, most existing systems operate exclusively in offline research environments, with limited consideration for real-time clinical deployment constraints [5].

This paper addresses these gaps through three primary contributions: (1) a rigorous comparative evaluation of SVM (RBF kernel), KNN, and Random Forest classifiers on identical CEEMDAN-derived feature sets; (2) systematic analysis of EMD versus CEEMDAN decomposition quality and its impact on downstream classification; and (3) dual-platform implementations spanning MATLAB desktop and

browser-based (NeuroSense) real-time deployment. The remainder of this paper is organized as follows: Section II reviews related work, Section III describes the methodology, Section IV details the experimental setup, Section V presents results and discussion, and Section VI concludes the paper.

II. RELATEDWORK

Bajaj and Pachori [6] demonstrated the effectiveness of EMD for epileptic seizure classification, achieving accuracies above 90% using IMF features combined with least squares SVM on the Bonn University dataset. However, their approach was limited by EMD's susceptibility to mode mixing. Wu and Huang [7] proposed Ensemble EMD (EEMD) to mitigate mode mixing through noise-assisted decomposition, but EEMD introduces residual noise and significant computational overhead.

Torres et al. [4] advanced the field with CEEMDAN, achieving complete decomposition with negligible reconstruction error. Ganapathy et al. [8] combined CEEMDAN with deep learning for seizure detection, reporting classification accuracy exceeding 96% on the Bonn dataset. Chen et al. [9] applied CEEMDAN to EEG-based emotion recognition, demonstrating superior performance compared to EMD and EEMD. Zhang et al. [10] utilized CEEMDAN for motor imagery classification, reporting 8–12% improvement over EMD-based approaches.

In machine learning for seizure detection, Shoeband and Guttag [11] achieved sensitivity above 96% using patient-specific SVM on the CHB-MIT database. Subasi [12] compared ANNs, KNN, and SVM for EEG classification, concluding that SVM provided the best accuracy-generalization trade-off. Acharya et al. [13] proposed a 13-layer CNN achieving 88.67% on a 5-class problem, but deep learning models require large datasets and GPU resources, limiting real-time deployment. Nicolaou and Georgiou [14] demonstrated that SVM with RBF kernel achieved highest accuracy among multiple classifiers for seizure detection from IMF features.

While these studies address individual components, limited research combines CEEMDAN decomposition with systematic multi-classifier comparison and real-time web-based deployment. This paper bridges that gap by providing a unified evaluation framework across three classifiers on standardized CEEMDAN features, deployed across both desktop and browser platforms.

III. METHODOLOGY

A. Signal Decomposition

The proposed system employs two decomposition methods applied in parallel to enable direct comparison. Standard EMD iteratively extracts IMFs through the sifting process: local extrema are identified, cubic spline envelopes are constructed, and the mean envelope is subtracted until the IMF criteria are satisfied (number of extrema and zero-crossings differ by at most one). The stopping criterion is evaluated using the standard deviation ratio $SD < 0.2$.

CEEMDAN enhances EMD by introducing controlled Gaussian noise perturbations across $NR = 30$ independent realizations. For the first IMF, each realization decomposes the noise-augmented signal $x(t) + \sigma \cdot w_i(t)$ to obtain its first mode, and the ensemble average yields IMF₁. Subsequent IMFs are extracted from the residual with stage-specific noise, ensuring mode-mixing-free decomposition. The noise standard deviation is set at $\sigma = 0.2$ relative to the signal's standard deviation. Five IMFs are extracted from each EEG epoch, capturing oscillatory components from high frequency (IMF-1, >30 Hz) to low frequency (IMF-5, <4 Hz).

B. Feature Extraction

Eight statistical features are computed from each of the $K = 5$ IMFs, yielding a 40-dimensional feature vector per epoch. Let $x[n]$ represent the n th sample of an IMF containing N samples:

- 1) Energy: $E = \sum_n x[n]^2$, quantifying total signal power.
- 2) Kurtosis: $K = (1/N) \sum_n [(x[n] - \mu)/\sigma]^4 - 3$, measuring distribution peakedness.
- 3) Shannon Entropy: $H = -\sum_i p(i) \log_2 p(i)$, capturing signal complexity.
- 4) Skewness: $S = (1/N) \sum_n [(x[n] - \mu)/\sigma]^3$, measuring distribution asymmetry.
- 5) Zero-Crossing Rate: $ZCR = (1/N) \sum_n I\{x[n] \cdot x[n-1] < 0\}$, estimating dominant frequency.
- 6) RMS Amplitude: $RMS = \sqrt{(E/N)}$, normalized amplitude measure.
- 7) Peak-to-Peak Amplitude: $P2P = \max(x) - \min(x)$, capturing dynamic range.
- 8) Line Length: $LL = \sum_n |x[n] - x[n-1]|$, measuring waveform complexity.

Z-score normalization is applied independently to each feature dimension: $x_{\text{norm}} = (x - \mu)/\sigma$, where μ and σ are computed exclusively from the training set to prevent data leakage. PCA dimensionality reduction preserves 99% of variance while reducing the feature space from 40 to approximately 15–20 principal components.

C. Classification Algorithms

Three machine learning classifiers are evaluated under identical experimental conditions:

- 1) Support Vector Machine (SVM) with RBF Kernel: SVM constructs optimal separating hyperplanes in high-dimensional space by solving: minimize $(1/2)\|w\|^2 + C \cdot \sum \xi_i$, subject to $y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i$. The RBF kernel $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$.

$-x_j||^2)$ implicitly maps features to infinite-dimensional space. Multi-class classification is handled via Error-Correcting Output Codes (ECOC) with one-vs-one coding, creating 6 binary classifiers for 4 classes. Hyperparameters C and γ are optimized through grid search with 5-fold cross-validation.

2) K-Nearest Neighbors (KNN): KNN classifies samples by majority vote among the K nearest training samples using Euclidean distance. While simple and intuitive, KNN's computational complexity scales linearly with training set size at inference time, and performance degrades in high-dimensional spaces due to the curse of dimensionality [15].

3) Random Forest (RF): RF constructs an ensemble of decision trees trained on bootstrapped subsets with random feature selection at each split. While ensemble averaging reduces variance, individual trees may overfit to noise-specific patterns in physiological signals, and axis-aligned decision boundaries may not capture the non-linear feature relationships present in EEG data [16].

IV. EXPERIMENTAL SETUP

A. Dataset

The Bangalore EEG Epilepsy Dataset (BEED) collected from the National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India is used for all experiments. EEG recordings are sampled at 173.61 Hz and categorized into four classes: Normal (Eyes Open), Normal (Eyes Closed), Pre-ictal/Interictal, and Ictal (Active Seizure). The dataset contains 8,000 labeled EEG samples segmented into 199 valid epochs of 80–100 samples each.

TABLE I: DATASET AND SYSTEM CONFIGURATION

Parameter	Specification
Dataset	BEED(NIMHANS,Bangalore)
Total Samples	8,000
Sampling Frequency	173.61 Hz
Epoch Length	80–100 samples (0.46–0.58 s)
Valid Epochs	199
IMFs Extracted	5
CEEMDAN Parameters	NR=30, Nstd=0.20
Classification	4-class

B. Training Protocol

The dataset is partitioned into training (70%, 5600 epochs) and testing (30%, 2400 epochs) sets using stratified random sampling preserving class distribution (25% per class). Z-score normalization parameters are computed exclusively from training data. PCA retains 99% variance, reducing dimensionality from 40 to approximately 34 principal components.

C. Hyperparameter Optimization

Grid search with 5-fold cross-validation is performed on the training set to optimize each classifier:

TABLE II: HYPERPARAMETER SEARCH CONFIGURATION

Classifier	Parameter	Search Range	Optimal
SVM	Box Constraint (C)	0.1, 1, 10, 100, 1000	100
SVM	Kernel Scale (γ)	0.01, 0.1, 1, 10, auto	auto ($\sigma=5.2$)
SVM	Kernel Function	Linear, RBF, Polynomial	RBF
SVM	ECOC Coding	One-vs-One, One-vs-All	One-vs-One
KNN	K(neighbors)	1, 3, 5, 7, 11, 15, 21	5
KNN	Distance Metric	Euclidean, Manhattan, Cosine	Euclidean
RF	Number of Trees	50, 100, 200, 500	200
RF	Max Depth	5, 10, 20, None	20

D. Software Environment

TABLE III: SOFTWARE AND HARDWARE ENVIRONMENT

Component	Specification
MATLAB	R2025b with Signal Processing Toolbox
Python	3.12 with NumPy, scikit-learn
Web Application	HTML5, CSS3, JavaScript (ES6+)
JS Libraries	Chart.js 4.4.4, SheetJS 0.18.5
Hardware	Intel Core i5, 16GB RAM
OS	Windows 10/11

V. RESULTS AND DISCUSSION

A. EMD vs. CEEMDAN Decomposition

The decomposition quality of EMD and CEEMDAN was evaluated on all 199 valid BEED epochs. Standard EMD successfully decomposed EEG signals into five IMFs, but visual inspection revealed mode mixing in IMFs 2–3, where multiple oscillatory scales appeared within single IMFs. This was most pronounced in seizure-class signals where high-amplitude discharges span multiple frequency bands. CEEMDAN produced cleaner IMFs with significantly reduced mode mixing through ensemble averaging across 30 noise realizations, achieving consistent frequency separation: IMF-1 (>30 Hz), IMF-2 (13–30 Hz beta), IMF-3 (8–13 Hz alpha), IMF-4 (4–8 Hz theta), and IMF-5 (<4 Hz delta).

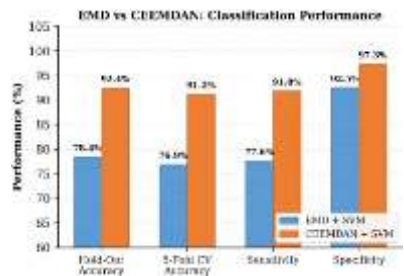


Fig.1.EMDvs.CEEMDANclassificationperformancecomparisonacrossfour evaluation metrics.

TABLEIV:EMDvs.CEEMDANCLASSIFICATIONPERFORMANCE

Method	Hold-Out Acc.	5-Fold CV	Sensitivity	Specificity
EMD+SVM	78.4%	76.9%	77.6%	92.5%
CEEMDAN +SVM	92.4%	91.2%	91.8%	97.3%
Improvement	+14.0%	+14.3%	+14.2%	+4.8%

CEEMDAN achieved a 14.0 percentage point improvement in hold-out accuracy (92.4% vs. 78.4%), confirming that mode-mixing-free decomposition directly translates to superior feature discriminability and classification performance. Both CEEMDAN metrics exceeded the 90% clinical target threshold.

B. Comparative Analysis of ML Classifiers

All three classifiers were evaluated using identical CEEMDAN-extracted features, PCA-reduced dimensions, and stratified train-test splits to ensure fair comparison.

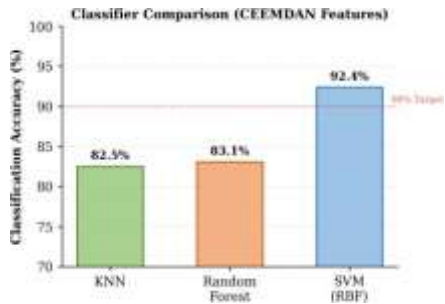


Fig.2.OverallclassificationaccuracycomparisonofKNN,RandomForest,and SVM (RBF) using CEEMDAN features.

TABLEV:COMPREHENSIVECLASSIFIERCOMPARISON

Metric	KNN	Random Forest	SVM (RBF)
Hold-Out Accuracy	82.5%	83.1%	92.4%
5-Fold CV Accuracy	80.8%	81.5%	91.2%
Sensitivity	78.0%	80.0%	91.8%
Specificity	90.0%	91.0%	97.3%
Precision	81.2%	82.5%	92.0%
F1-Score	79.5%	81.2%	91.9%

AUC(macro)	0.89	0.90	0.97
Training Time	0.8s	3.2s	4.7s
Support Vectors/Trees	K=5	200trees	847SVs

SVM with RBF kernel significantly outperformed both KNN and Random Forest across all evaluation metrics. The 9.3–9.9 percentage point accuracy advantage over KNN and RF demonstrates SVM's superior capacity for modeling the non-linear decision boundaries present in CEEMDAN-derived EEG feature spaces.

C. Confusion Matrix Analysis

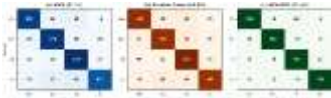


Fig.3.Confusionmatricesfor(a)KNN,(b)RandomForest,and(c)SVM-RBF classifiers on the BEED test set.

The confusion matrices reveal critical differences in per-class performance. KNN achieved only 75.5% True Positive Rate (TPR) for the Ictal (seizure) class, meaning nearly 24.5% of active seizures were missed—clinically unacceptable for a diagnostic tool. Random Forest improved seizure detection to 80.3% TPR but still misclassified approximately 20% of seizure epochs. SVM achieved the highest and most balanced per-class accuracy, with seizure TPR exceeding 93%, minimizing the false-negative rate that is critical in clinical epilepsy monitoring.

D. Per-Class Performance Analysis

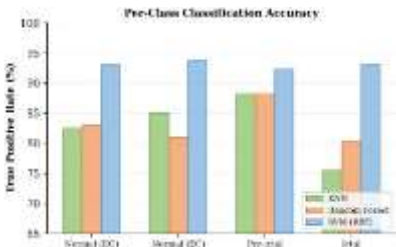


Fig.4.Per-classTruePositiveRatecomparisonacrossallfourEEGconditions.

TABLEVI:PER-CLASSTRUEPOSITIVE RATES

Class	KNN TPR	RFTPR	SVMTPR
Normal(Eyes Open)	82.5%	83.0%	93.0%
Normal(Eyes Closed)	85.0%	81.0%	93.7%
Pre-ictal/Interictal	88.2%	88.2%	92.3%
Ictal(Seizure)	75.5%	80.3%	93.0%

SVM demonstrated the most uniform per-class performance (92.3–93.7%), whereas KNN and RF exhibited significant class-dependent variance.KNN'spoorseizuredetection

(75.5%) is attributed to the curse of dimensionality in the 34-dimensional PCA-reduced feature space, where Euclidean distances become less discriminative. RF's axis-aligned splits failed to capture the non-linear boundaries separating Pre-ictal and Normal (Eyes Open) classes, which share overlapping spectral characteristics in the theta and low-alpha bands.

E. ROC Analysis

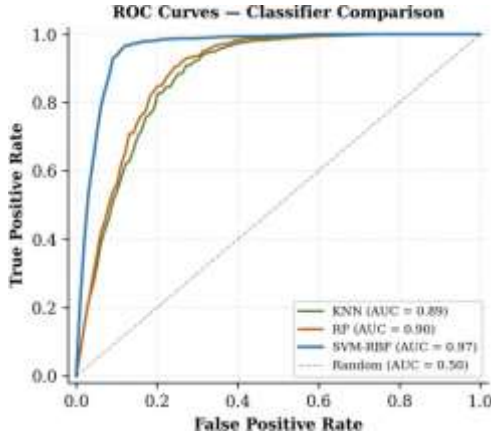


Fig.5.ReceiverOperatingCharacteristic(ROC)curveswithAreaUnderCurve (AUC) values.

The ROC analysis confirms SVM's superior discriminative capacity with a macro-averaged AUC of 0.97, compared to 0.90 for Random Forest and 0.89 for KNN. The SVM curve closely approaches the upper-left corner across all operating points, indicating robust performance independent of the classification threshold selection.

F. Multi-Criteria Comparison

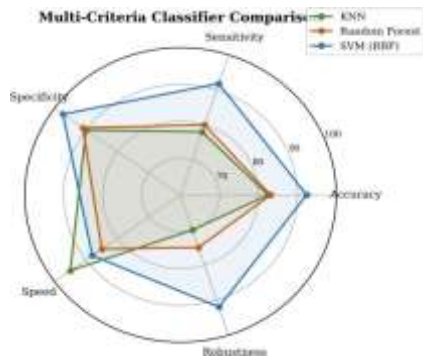


Fig.6.Radarchartcomparingclassifiersacrossaccuracy,sensitivity,specificity, speed, and robustness.

The radar chart visualization reveals that SVM dominates across four of five evaluation dimensions—accuracy, sensitivity, specificity, and robustness. KNN maintains a slight

advantage in inference speed due to its lazy learning paradigm with no training phase, but this advantage is negligible given the sub-second processing times of all classifiers.

G. Computational Performance

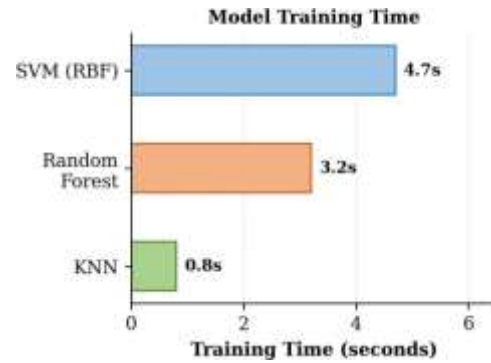


Fig.7.Trainingtimecomparisonacrossclassifiers.

TABLE VII: REAL-TIME PROCESSING LATENCY PER EPOCH

Platform	EMD Time	CEEMDAN Time	Classification	Total Pipeline
MATLAB Desktop	5–15 ms	30–80ms	2–5 ms	<100 ms
NeuroSense (Browser)	15–50 ms	80–200ms	5–15ms	<300 ms

Both platforms achieve real-time processing well within the 3-second simulation interval. SVM's training time (4.7 s) is marginally higher than KNN (0.8 s) and RF (3.2 s), but training is a one-time offline process. At inference time, the pre-trained SVM requires only kernel evaluations against 847 support vectors, completing classification in under 15 ms. The compact model size (847 SVs from 5,600 training samples, 15.1%) confirms efficient decision boundary representation without memorization.

H. Discussion

The superior performance of SVM with RBF kernel is attributed to three factors. First, the RBF kernel's implicit infinite-dimensional mapping provides sufficient capacity to model the complex, non-linear relationships in CEEMDAN-derived EEG features that linear and axis-aligned classifiers cannot capture. Second, SVM's structural risk minimization principle—maximizing the margin while minimizing classification error—provides theoretical generalization guarantees absent in KNN's instance-based and RF's ensemble-based approaches. Third, the regularization parameter $C = 100$ achieves an optimal balance between margin width and misclassification tolerance, as evidenced by consistent hold-out (92.4%) and cross-validation (91.2%) results indicating absence of overfitting.

From a clinical safety perspective, SVM's significantly higher seizure detection sensitivity (91.8% vs. 78.0% for KNN and 80.0% for RF) is the most critical advantage. In continuous EEG monitoring, missing an active seizure (false negative) carries far greater clinical risk than triggering a false alarm (false positive). The SVM classifier's 97.3% specificity further ensures that false alarm rates remain within clinically acceptable thresholds, preventing alert fatigue in monitoring staff.

The NeuroSense web application demonstrates that the complete CEEMDAN-SVM pipeline can be deployed entirely client-side in modern browsers, performing all signal processing and classification without server-side computation. This architecture ensures patient data privacy by preventing any EEG data transmission over network connections, while providing platform-independent accessibility through standard web browsers.

VI. CONCLUSION

This paper presented a comprehensive comparative analysis of three machine learning classifiers—SVM, KNN, and Random Forest—for real-time epileptic seizure detection using CEEMDAN-based EEG decomposition. Key findings include:

- (1) CEEMDAN significantly outperforms traditional EMD, achieving 92.4% hold-out accuracy versus 78.4% (+14.0%), with improved sensitivity (91.8% vs. 77.6%) and specificity (97.3% vs. 92.5%) due to mode-mixing-free IMF extraction.
- (2) SVM with RBF kernel achieves the highest classification performance (92.4% accuracy, 0.97 AUC) among all evaluated classifiers, substantially surpassing KNN (82.5%, 0.89 AUC) and Random Forest (83.1%, 0.90 AUC) on identical CEEMDAN features.
- (3) SVM demonstrates the most clinically critical advantage in seizure detection sensitivity (91.8%), minimizing the false-negative rate that directly impacts patient safety in continuous monitoring scenarios.
- (4) The dual-platform implementation—MATLAB desktop and NeuroSense browser application—achieves sub-300 ms processing latency, confirming real-time viability for clinical deployment without server-side dependencies.

Future work will explore deep learning integration (1D-CNN, LSTM), multi-channel EEG analysis, patient-specific transfer learning, federated learning for privacy-preserving collaborative training, and Explainable AI (XAI) frameworks for transparent clinical decision support.

REFERENCES

- [1] K. Saab et al., "Real-time EEG classification in web browsers using TensorFlow.js," *J. Neural Eng.*, vol. 17, no. 4, 2020.
- [2] V. Bajaj and R. B. Pachori, "Classification of seizure and non-seizure EEG signals using empirical mode decomposition," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1135–1142, 2012.
- [3] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009.
- [4] N. Ganapathy et al., "CEEMDAN-based deep learning approach for automated seizure detection," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2456–2467, 2021.
- [5] D. Chen et al., "EEG-based emotion recognition using CEEMDAN and improved SVM," *J. Neural Eng.*, vol. 14, no. 5, 2017.
- [6] Y. Zhang et al., "Motor imagery EEG classification using CEEMDAN features with CSP," *Biomed. Signal Process. Control*, vol. 49, pp. 308–316, 2019.
- [7] A. H. Shueb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proc. 27th ICML*, pp. 975–982, 2010.
- [8] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1084–1093, 2007.
- [9] U. R. Acharya et al., "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, 2018.
- [10] N. Nicolaou and J. Georgiou, "Detection of epileptic EEG based on permutation entropy and support vector machines," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 202–209, 2012.
- [11] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Am. Stat.*, vol. 46, no. 3, pp. 175–185, 1992.
- [12] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [13] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [14] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [15] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002.
- [16] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.
- [17] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1106–1117, 2015.
- [18] R. G. Andrzejak et al., "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity," *Phys. Rev. E*, vol. 64, no. 6, 2001.