WEP Warwick Evans Publishing
International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 6, Number 1, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v6n1.01
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Epilepsy Detection and Prediction Using a High-Performance Ensemble Learning Algorithm Based on Feature Fusion

Hanbo Zhang, Zhisen Li, Min Liu, Guomou Zeng, Yahui Liu, Jiaming Shi, Jincan Zhang *

College of Information Engineering, Henan University of Science and Technology, Luoyang, China
*Corresponding Author: zjc850126@163.com

ABSTRACT

Epilepsy detection and prediction remain among the most challenging data analysis tasks for chronic brain disorders. Manual detection is time-consuming, and the inability to predict seizures artificially has driven the development of numerous automatic algorithms. However, most electroencephalogram (EEG) classification methods rely on a single feature, resulting in poor accuracy, and typically address either detection or prediction rather than both. In this study, we propose an integrated model based on feature fusion and selection. First, EEG signals were segmented into interictal, preictal, and ictal phases. Next, we applied the Discrete Wavelet Transform (DWT) for signal decomposition and extracted Standard Deviation (STD), Root Mean Square (RMS), Approximate Entropy (ApEn), and Fuzzy Entropy (FuzzyEn) features from each subband. A Random Forest (RF) algorithm then selected the ten most important features, and Ensemble (ENS) learning algorithms were employed for both detection and prediction. Validation on the CHB-MIT dataset showed that, for detection, the average accuracy, sensitivity, precision, F1-score, specificity, and Matthews correlation coefficient across 24 patients were 99.07%, 98.71%, 99.40%, 99.05%, 99.40%, and 0.98, respectively. For prediction, the corresponding metrics were 95.41%, 96.74%, 94.50%, 95.55%, 94.02%, and 0.94. The proposed model thus enables high-precision automatic detection and prediction of epileptic EEG signals, allowing timely preventive measures to mitigate seizure impact.

KEYWORDS

Epilepsy Detection; Feature Fusion; Ensemble Learning; Epilepsy Prediction

1. INTRODUCTION

Epilepsy, a neurological disorder affecting over 1% of the global population [1, 2], can cause irreversible brain damage and impair development in infants and young children when seizures occur uncontrolled. Although electroencephalography (EEG) is used for diagnosis, its resource intensity, susceptibility to false positives, and heavy reliance on clinical expertise for seizure prediction have driven the development of automatic detection and prediction models. Epilepsy detection helps physicians identify various epilepsy states by distinguishing ictal from interictal EEG signals, while prediction forecasts seizures in diagnosed patients by differentiating preictal from interictal signals [3]. Given the subtler differences between preictal and interictal features, prediction is inherently more challenging and requires higher robustness than detection.

To improve performance and robustness, accurately distinguishing interictal, preictal, and ictal states is critical, with feature extraction playing a key role. Previous studies have explored various methods—such as time-frequency analysis via the Fourier transform [4, 5] (limited to smooth signals) versus the Discrete Wavelet Transform (DWT) [6–9] for handling non-smooth, complex EEG signals—and highlighted that reliance on single features (time-domain statistics [10], Power Spectrum Density [11], nonlinear features like Approximate Entropy, Renyi Entropy, Shannon Entropy, Distribution Entropy, Lempel-Ziv Complexity [8, 9, 12, 13], or spatiotemporal features from the Common Spatial Pattern [14]) results in low accuracy. Feature fusion (e.g., combining spectral power with time-domain features [15] or integrating various nonlinear and spatiotemporal measures [16–19]) offers a comprehensive representation but may introduce irrelevant features and increase processing time. To address these challenges, we propose an automatic EEG detection and prediction model based on feature fusion and ensemble learning that: (1) denoises the EEG signal using a band-pass filter and segments each state; (2) applies DWT to extract standard deviation, root mean square, approximate entropy, and fuzzy entropy from each subband; (3) screens 24 features (four from each of six subbands) with a random forest to select the 12 most important; and (4) uses K Nearest Neighbor (KNN), Support Vector Machine (SVM), and Ensemble (ENS) classifiers for final detection and prediction. The paper is organized as follows: Section 2 reviews related literature, Section 3 presents the dataset and proposed model, Section 4 details the experimental results, and Section 5 concludes the article.

2. LITERATURE SURVEY

Most current studies rely on single-feature extraction for classifying epileptic EEG, with Deivasigamani S et al. [10] decomposing signals via DT-CWT to extract time-domain features and Gao Y et al. [11] using PSDED with DCNNs achieving over 90% accuracy. Nonlinear features, which better capture the EEG's inherent complexity, have also been extensively employed: Aung S T et al. [12] introduced mDistEn yielding a 92% accuracy, Raghu S et al. [28] applied five-level wavelet packet decomposition with log entropy, and AlSharabi K et al. [8] decomposed EEG into standard subbands via DWT to extract ShanEn for classification. Kumar Y et al. [9] combined DWT with ApEn to achieve 95% accuracy in distinguishing ictal from interictal states, while Tripathi D et al. [23] used EMD to extract FuzzyEn from subbands, enabling a nonlinear SVM to reach 97% accuracy. Moreover, several studies have fused nonlinear features with others to enhance performance; for example, Lu Y et al. [17] combined entropy measures with HHT and TQWT to achieve 98.2% accuracy, Aayesha et al. [18] integrated temporal, spectral, nonlinear, and pattern features with multiple classifiers (with FLR achieving 97.33% accuracy), and Kavya B S et al. [29] fused statistical and entropy features. Seizure prediction has also been advanced, as Zhong L et al. [30] constructed optimal spatiotemporal feature sets from time, nonlinear, and brain network dimensions to achieve a peak accuracy of 98.1%, while Wang X et al. [24] fused time-frequency and energy features—optimizing a random forest via grid search—to obtain 96.7% accuracy in distinguishing three epileptic EEG states.

3. MATERIAL AND METHOD

3.1. CHB-MIT Dataset

The study employs the CHB-MIT dataset, collected at Children's Hospital Boston by MIT. This dataset contains scalp EEG recordings from 23 epilepsy patients, comprising 24 recordings labeled chb01 to chb24—where chb21 and chb01 are from the same patient at different times, and chb24 is a supplementary recording without patient details. The data were recorded at 256 Hz with 16-bit resolution using the international 10-20 electrode system over 18 to 23 channels, primarily utilizing a dual-electrode stage. In total, the dataset includes 967.85 hours of continuous recordings and 178 seizure events.

3.2. Data Preprocessing

The CHB-MIT dataset is large, so we first segment the data. Epilepsy is classified into ictal, preictal, and interictal states (see Figure 1). Rectangular windows intercept these states: for ictal versus interictal, variable-length windows match the ictal period, while for preictal versus interictal, a fixed window of 76800 samples (300 s) is used. Afterwards, the segmented EEG data is filtered with a 2nd-order Butterworth bandpass filter (0.5 dB to 40 dB) and decomposed by DWT with a 'db4' wavelet at 5 levels to capture the complete epileptic EEG information.

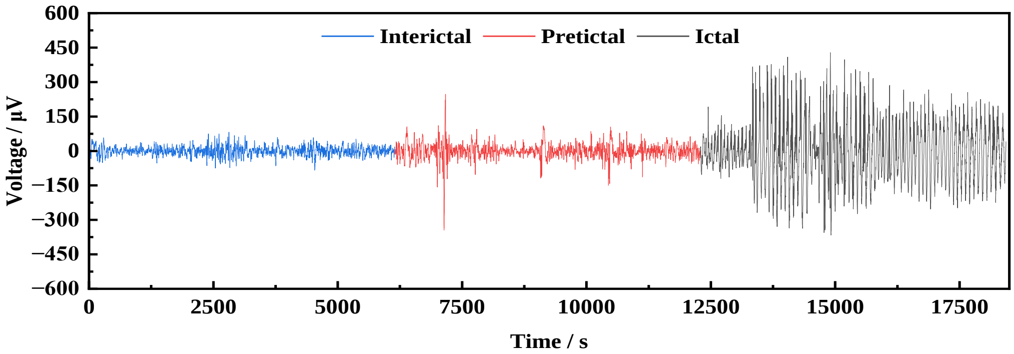


Figure 1. EEG state division

3.2.1. Discrete Wavelet Transform

The EEG signal is complex and benefits from time-frequency domain methods. In this study, the CHB-MIT EEG is preprocessed using DWT, which decomposes the signal into different frequency subbands. Leveraging the multi-resolution analysis of wavelets, DWT achieves time-frequency preservation by performing multiple decomposition levels. Essentially, DWT discretizes samples from the Continuous Wavelet Transform (CWT) as described below.

Firstly, for one-dimensional EEG signal f(t) ∈ L2(R), it is convolved with the wavelet basis function Ψu,s(t) to obtain CWT, which is express as equation (1).

$$CWT(u,s) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\Psi^*\left(\frac{t-u}{s}\right)dt \tag{1}$$

Where, Ψ* is the complex conjugate of the mother wavelet function Ψ with volatility and the mean value is zero. s and u denote the scale factor and translation factor, respectively.

Then, the discretization (s=2m, m∈Z) is completed in the form of exponential powers for the scale factor, and the discrete wavelet transform is obtained from the uniform discretization (u=k2m) of the translation factor. The formula of DWT can be expressed as:

$$DWT(m,k) = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{2^m}}\Psi^*\left(\frac{t-k2^m}{2^m}\right)dt \tag{2}$$

When using DWT, the mother wavelet function and decomposition levels must be specified. This study employs the 'db4' basis to decompose the signal into 5 levels. At each level, the input signal passes through a high-pass filter (producing the detail factor) and a low-pass filter (producing the approximation factor), forming the structure [Aj, Dj]. Figure 2 shows that after 5 levels, the frequency subbands are: [A1 (0–64 Hz), D1 (64–128 Hz)], [A2 (0–32 Hz), D2 (32–64 Hz)], [A3 (0–16 Hz), D3 (16–32 Hz)], [A4 (0–8 Hz), D4 (8–16 Hz)], and [A5 (0–4 Hz), D5 (4–8 Hz)].

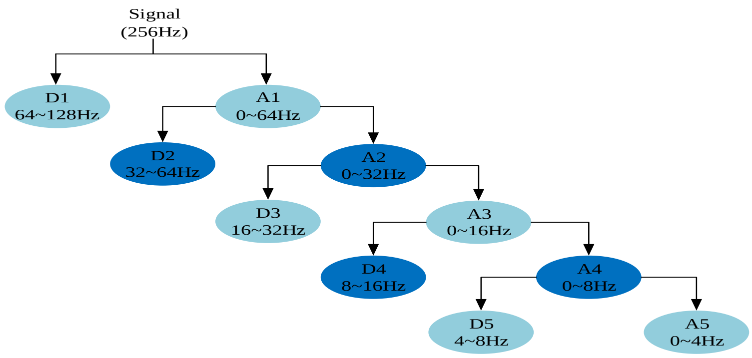


Figure 2. 5-level DWT decomposition of EEG signal

3.3. Feature Extraction

3.3.1. Nonlinear features

EEG signals in epilepsy patients are nonlinear and complex, measurable by features such as approximate entropy and fuzzy entropy. Approximate entropy is widely used to quantify time series regularity and assess data complexity even in noisy, short- or long-term recordings. Fuzzy entropy, which incorporates an exponential function for fuzzified similarity, offers smooth transitions, a wide parameter range, minimal dependence on sequence length, and strong noise robustness—making it well suited for biomedical signal analysis.

3.3.2. Standard deviation

Standard deviation (STD) is a common time-frequency feature, which can also be applied to the feature identification of EEG signal. It is simple and easy to calculate, and has a good identification effect. The calculation formula of STD is as follows:

$$STD = \sqrt{\frac{\sum(x-\mu)^2}{N}} \tag{3}$$

Where μ is the expectation value, N is the overall sample size, and x is the variable.

3.3.3. Root Mean Square

Root Mean Square (RMS) is a common method for calculating and measuring fluctuations or changes of a dataset, and it can be used to measure the deviation of an observation from its true value. Its calculation is as follows:

$$RMS = \sqrt{\frac{x_0^2+x_1^2+\ldots+x_{N-1}^2}{N}} \tag{4}$$

3.4. Feature Selection

The Random Forest (RF) algorithm was used to select EEG features because of its high parallelism, low variance, strong generalization, and insensitivity to missing data. As an ensemble method based on Bagging and CART decision trees, RF generates classification results from multiple trees and determines the final output by majority voting (see Figure 3).

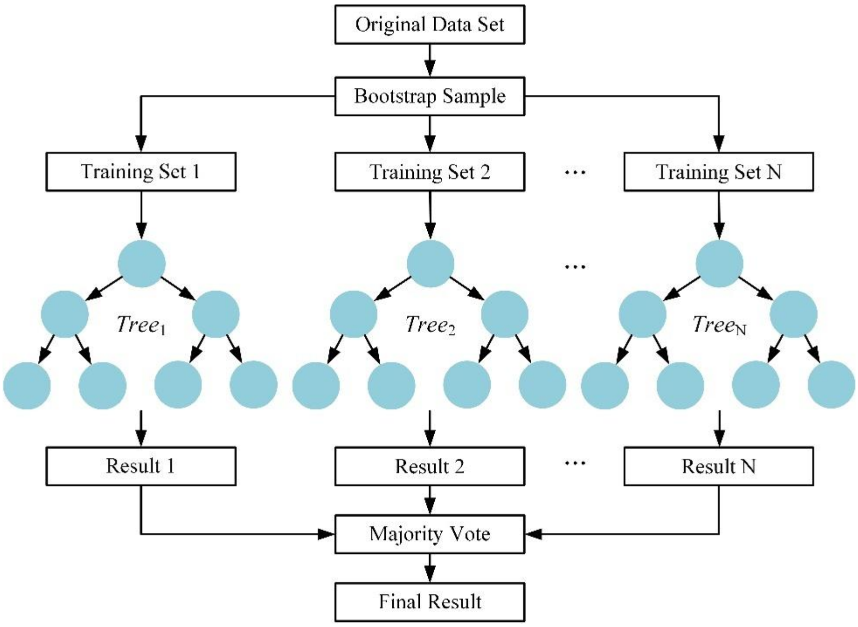


Figure 3. Random forest algorithm diagram

The EEG signal was band-pass filtered and decomposed by 5-level DWT into six subbands (D1–D5, A5). Four features were extracted from each subband to yield 24 features (see Table 1), and the RF algorithm then ranked these features, selecting the top 12 (see Table 2).

Table 1. Extracted 24 features of EEG signals

D1 STD	D2 STD	D3 STD	D4 STD	D5 STD	A5 STD
D1 RMS	D2 RMS	D3 RMS	D4 RMS	D5 RMS	A5 RMS
D1 ApEn	D2 ApEn	D3 ApEn	D4 ApEn	D5 ApEn	A5 ApEn
D1 FuzzyEn	D2 FuzzyEn	D3 FuzzyEn	D4 FuzzyEn	D5 FuzzyEn	A5 FuzzyEn

Table 2. Selected 12 features by RF

D1 STD	D2 STD	D3 STD	D2 ApEn	D1 FuzzyEn	D2 FuzzyEn
D1 RMS	D2 RMS	D3 RMS	D4 RMS	D3 FuzzyEn	D4 FuzzyEn

3.5. Classification

3.5.1. K Nearest Neighbor

K Nearest Neighbor (KNN) is an instance-based algorithm that requires no training and yields stable classification. It predicts categories by weighting the labels of the K closest training samples based on distance, though comparing each sample to the entire dataset is computationally, time, and memory intensive.

3.5.2. Support Vector Machine

Support Vvector M machine (SVM) is widely used in epilepsy EEG classification due to its strong learning ability and performance. It minimizes generalization error by finding an optimal separating hyperplane. For nonlinear classification, features are transformed with appropriate kernel functions and projected into a high-dimensional space, while convex quadratic programming ensures a global minimum. However, SVM is highly sensitive to outliers and noise.

3.5.3. Ensemble

Ensemble (ENS) learning combines multiple classifiers to yield an integrated result based on certain rules. By constructing numerous weak learners and merging their biases or variances, it forms a strong, accurate, and robust model that improves classification accuracy, stability, and prevents overfitting. Ensemble learning models include Bagging, Boosting, and Stacking, and this study employs the Bagging model for classification.

Bagging (Bootstrap Aggregating) is a parallel integration algorithm that creates training sets by repeatedly sampling m samples from a population of size N, ensuring that the resulting base learners are independent. The following is the partial pseudo-code of the ensemble training process:

```
Input: Data set S={(xn, yn), n=1,…,N}
Output: Bagging ensemble learning model
Step1: Divide the original data into training set Strain={(xtrain, ytrain)}, test set Stest={(xtest, ytest)}
Step2: Training the base learner
    For i=1 to m
        | Training the base learner i
        | modeli = mi_train(Strain)
        | yi_train=model_predict(xtrain)
        | yi_test=model_predict (xtest)
    end
Step3: Building new datasets, Snew={( y1_train, y2_train, … , ym_train)}
Step4: Training Snew by meta-learner
Step5: Import the test set into the model to produce predictions, model_predict (xtest)
```

4. RESULTS AND DISCUSSIONS

4.1. Evaluation Indicators

The proposed model's performance is evaluated using Accuracy, Sensitivity, Specificity, Precision, F1-score, and the Matthews correlation coefficient. Accuracy measures how well the model distinguishes between different epilepsy states. Sensitivity indicates the likelihood of correctly diagnosing the disease, while Specificity measures the chance of avoiding misdiagnosis. Precision and F1-score assess the accuracy in detecting the seizure phase (for detection) and the pre-seizure phase (for prediction). The Matthews correlation coefficient, ranging from -1 to +1, is considered the best measure for describing the confusion matrix, with values closer to +1 indicating near-perfect detection. The specific formulas for the six indicators are as follows:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{5}$$

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{6}$$

$$Specificity = \frac{TN}{TN+FP} \tag{7}$$

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$F1 - score = \frac{2TP}{2TP+FP+FN} \tag{9}$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FN)*(TP+FP)*(TN+FN)*(TN+FP)}} \tag{10}$$

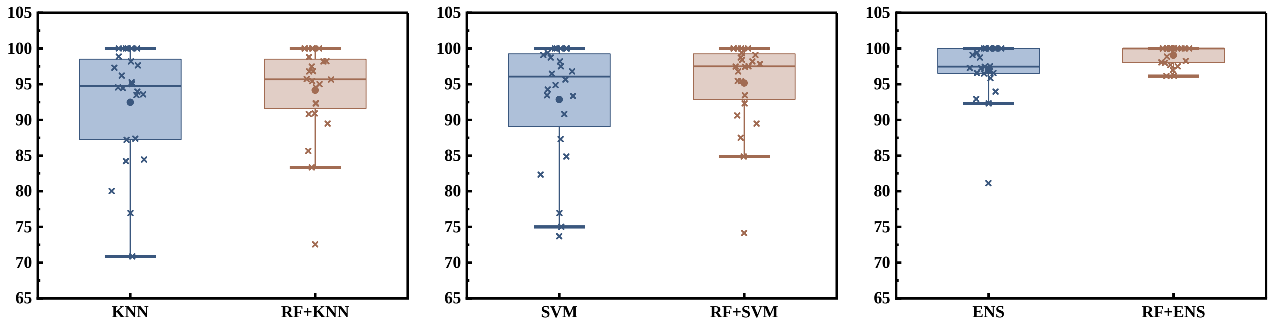
Here, TP represents the number of correctly classified positive cases, and TN represents the correctly classified negative cases. FP and FN denote the counts of data misclassified as positive and negative, respectively.

4.2. Experimental Results

The main purpose of our experiment is to compare the classification results of different classifiers with the feature selection algorithm for epileptic EEG, so as to obtain the best excellent model with the highest classification accuracy. After that, our work can be divided into two parts: detection and prediction of epileptic EEG. The detection is the two-classification task between the ictal and interictal, while prediction is the two-classification task between interictal and preictal.

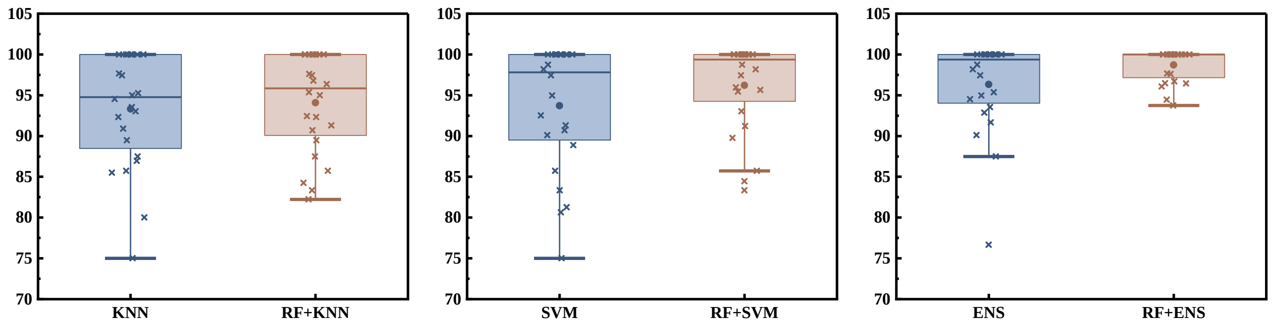
4.2.1. Feature selection preferences

Figures 4 to 9 display box plots of Accuracy, Sensitivity, Specificity, Precision, Matthews correlation coefficient, and F1-Score. The comparison before and after RF feature selection shows improved median and mean values for all classifiers. Additionally, the RF-selected boxes have fewer outliers and flatter shapes, indicating smaller fluctuations and more stable data. Hence, RF feature selection significantly enhances epileptic EEG classification performance.



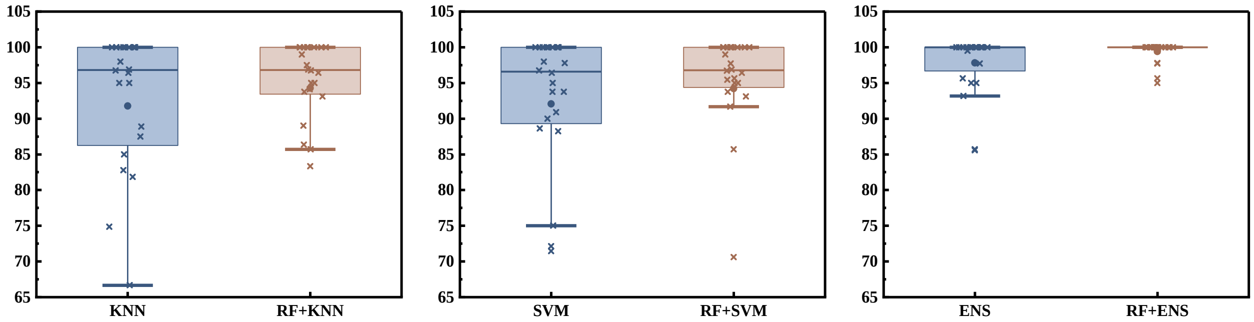
(a) (b) (c)

Figure 4. Comparison of Accuracy results with RF feature selection and without RF feature selection for different classifiers (a) KNN (b) SVM (c) ENS



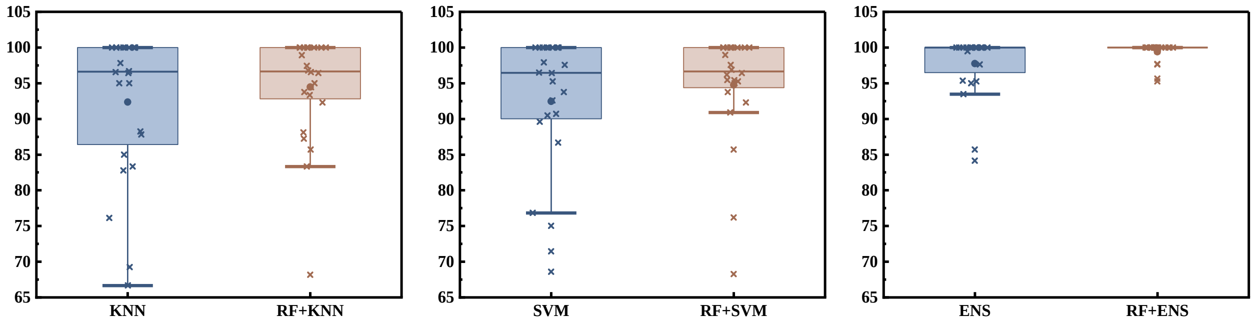
(a) (b) (c)

Figure 5. Comparison of Sensitivity results with RF feature selection and without RF feature selection for different classifiers (a) KNN (b) SVM (c) ENS



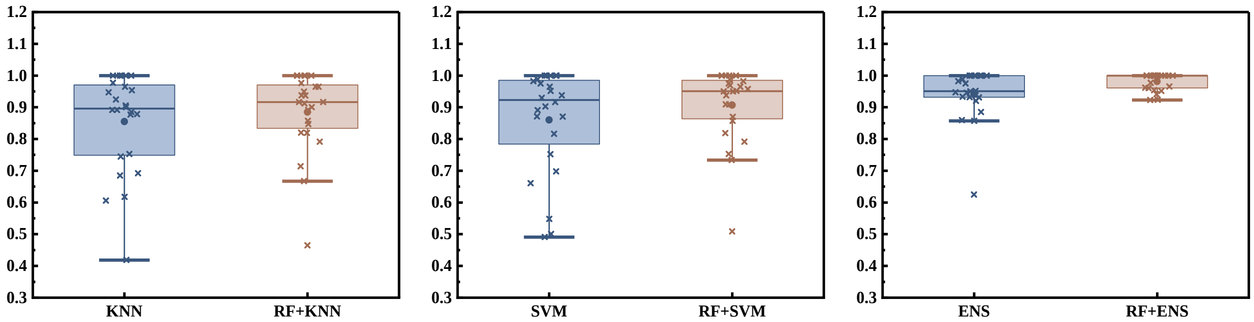
(a) (b) (c)

Figure 6. Comparison of Specificity results with RF feature selection and without RF feature selection for different classifiers (a) KNN (b) SVM (c) ENS



(a) (b) (c)

Figure 7. Comparison of Precision results with RF feature selection and without RF feature selection for different classifiers (a) KNN (b) SVM (c) ENS



(a) (b) (c)

Figure 8. Comparison of Matthews results with RF feature selection and without RF feature selection for different classifiers (a) KNN (b) SVM (c) ENS

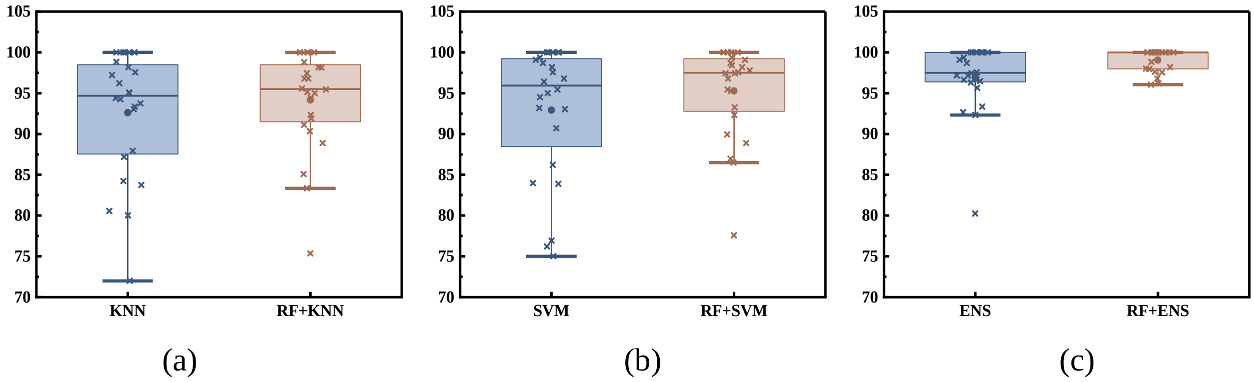


Figure 9. Comparison of F1-Score results with RF feature selection and without RF feature selection for different classifiers (a) KNN (b) SVM (c) ENS

4.2.2. Epilepsy detection (ictal vs interictal)

Epilepsy detection is a two-class task (ictal vs. interictal). With RF feature selection, three models—RF+KNN, RF+SVM, and RF+ENS—were applied to the CHB-MIT dataset (chb01 to chb24). As shown in Figure 10, which presents the averaged performance metrics for all three models, RF+ENS performed best with an average accuracy of 99.07%, sensitivity of 98.71%, precision of 99.40%, F1-score of 99.05%, and specificity of 99.40%. The RF+ENS model achieved 100% classification for chb02–chb07, chb09–chb11, chb16, chb18–chb19, and chb21–chb23, while chb08 and chb24 scored slightly above 96%, possibly due to data errors. Detailed performance metrics for the RF+ENS model are provided in Table 5.

Table 5. Detection results using RF+ENS algorithm

Number	Acc (%)	Sen (%)	Pre (%)	Matthews	F1-score (%)	Spe (%)
chb01	97.67	97.62	97.62	0.95	97.62	97.73
chb02	100.00	100.00	100.00	1.00	100.00	100.00
chb03	100.00	100.00	100.00	1.00	100.00	100.00
chb04	100.00	100.00	100.00	1.00	100.00	100.00
chb05	100.00	100.00	100.00	1.00	100.00	100.00
chb06	100.00	100.00	100.00	1.00	100.00	100.00
chb07	100.00	100.00	100.00	1.00	100.00	100.00
chb08	96.17	96.70	95.65	0.92	96.17	95.65
chb09	100.00	100.00	100.00	1.00	100.00	100.00
chb10	100.00	100.00	100.00	1.00	100.00	100.00
chb11	100.00	100.00	100.00	1.00	100.00	100.00
chb12	98.04	96.06	100.00	0.96	97.99	100.00
chb13	98.85	97.67	100.00	0.98	98.82	100.00
chb14	96.97	93.75	100.00	0.94	96.77	100.00
chb15	97.98	96.46	99.48	0.96	97.95	99.49
chb16	100.00	100.00	100.00	1.00	100.00	100.00
chb17	97.50	100.00	95.24	0.95	97.56	95.00
chb18	100.00	100.00	100.00	1.00	100.00	100.00
chb19	100.00	100.00	100.00	1.00	100.00	100.00
chb20	98.25	96.43	100.00	0.97	98.18	100.00
chb21	100.00	100.00	100.00	1.00	100.00	100.00
chb22	100.00	100.00	100.00	1.00	100.00	100.00
chb23	100.00	100.00	100.00	1.00	100.00	100.00
chb24	96.13	94.44	97.70	0.92	96.05	97.80
Average	99.07	98.71	99.40	0.98	99.05	99.40

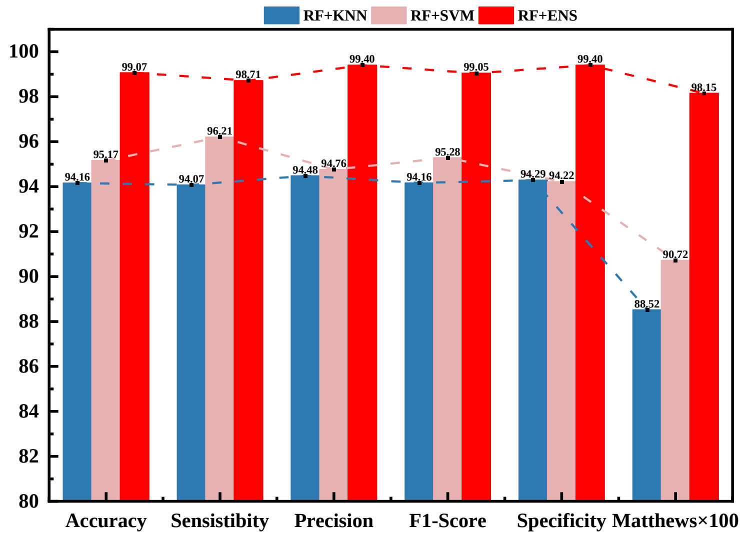


Figure 10. The average value of each evaluation index obtained using different classifiers

4.2.3. Epilepsy prediction (interictal vs preictal)

Table 8. Prediction results using RF+ENS algorithm

Number	Acc (%)	Sen (%)	Pre (%)	Matthews	F1-score (%)	Spe (%)
chb01	100.00	100.00	100.00	0.93	100.00	100.00
chb02	95.45	95.45	95.45	1.00	95.45	95.45
chb03	97.73	99.24	96.32	0.97	97.76	96.21
chb04	95.42	98.33	92.91	1.00	95.55	92.50
chb05	95.00	94.67	95.30	0.98	94.98	95.33
chb06	91.00	96.67	89.23	0.92	92.80	82.50
chb07	100.00	100.00	100.00	1.00	100.00	100.00
chb08	97.67	97.33	97.99	0.86	97.66	98.00
chb09	98.30	98.86	97.75	1.00	98.31	97.73
chb10	93.57	97.14	90.67	1.00	93.79	90.00
chb11	97.73	96.97	98.46	0.99	97.71	98.48
chb12	96.53	94.66	98.41	0.95	96.50	98.44
chb13	94.16	94.37	93.97	0.93	94.17	93.94
chb14	95.42	98.33	92.91	0.88	95.55	92.50
chb15	97.98	96.97	98.97	0.95	97.96	98.99
chb16	92.31	100.00	85.71	0.86	92.31	85.71
chb17	100.00	100.00	100.00	0.95	100.00	100.00
chb18	82.22	86.11	79.90	0.94	82.89	78.33
chb19	93.18	100.00	88.00	1.00	93.62	86.36
chb20	98.11	98.48	97.74	0.93	98.11	97.73
chb21	91.25	92.50	90.24	0.95	91.36	90.00
chb22	94.00	93.33	94.59	1.00	93.96	94.67
chb23	96.19	96.67	95.75	1.00	96.21	95.71
chb24	96.69	95.56	97.73	0.62	96.63	97.80
Average	95.41	96.74	94.50	0.94	95.55	94.02

Epilepsy prediction distinguishes interictal and preictal states. As shown in Figure 11—which now displays the averaged performance metrics for the three models—the RF+ENS model (detailed in Table 8) achieved the highest results, with an accuracy of 95.41%, sensitivity of 96.74%, precision of 94.50%, F1-score of 95.55%, specificity of 94.02%, and a Matthews correlation coefficient of 0.94.

Figure 2 illustrates that the differences between interictal and preictal EEG signals are subtler than those between ictal and interictal states, underscoring the inherent challenge of seizure prediction. Notably, a 100% prediction rate was achieved for channels chb01, chb07, and chb17.

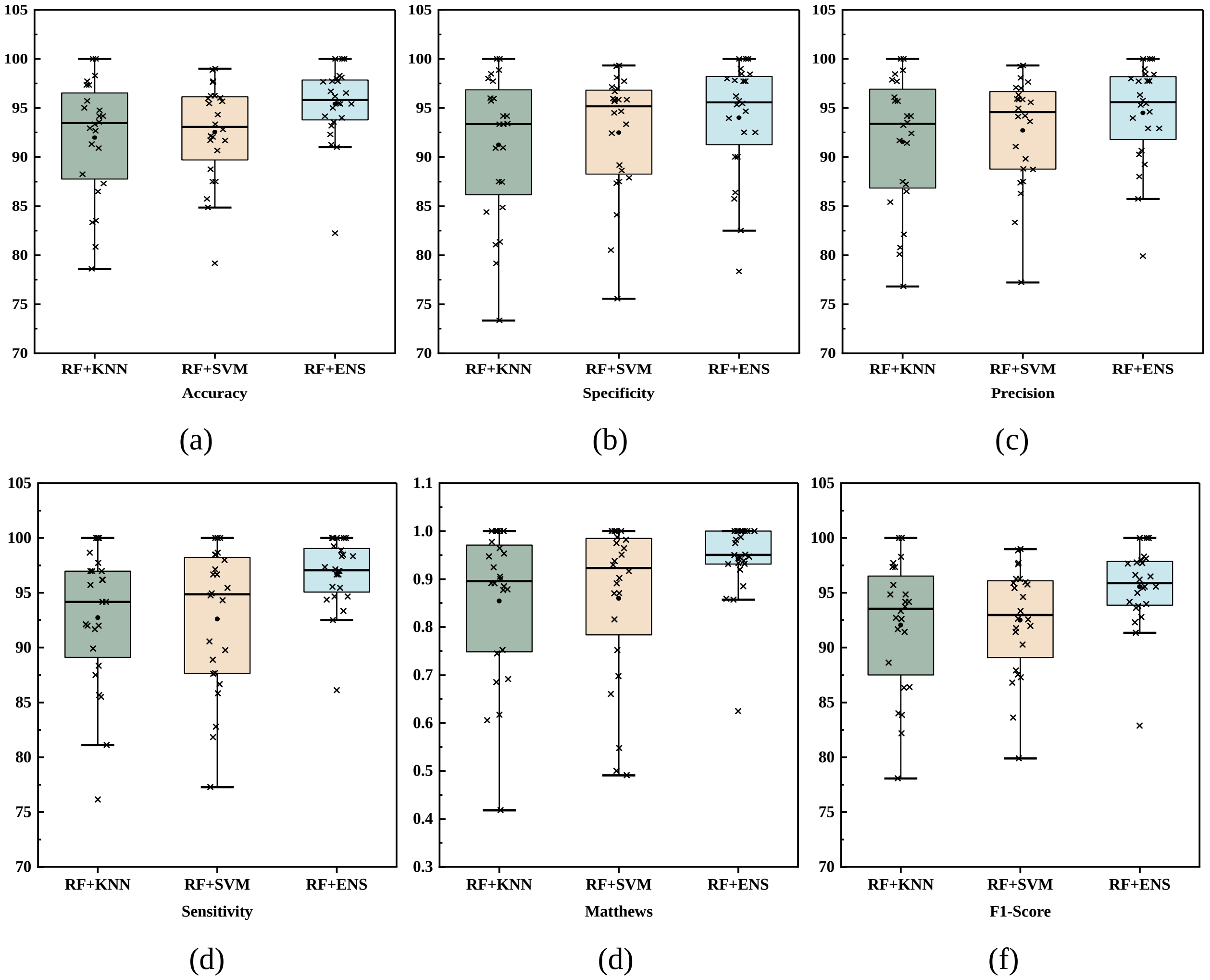


Figure 11. Box line diagram of each prediction result evaluation index after combining different classifiers with RF

4.3. Discussion

Epilepsy detection performance comparisons (Table 9) reveal that most studies relied on a single feature to differentiate interictal from ictal EEG signals. For example, Raghu S et al. [31] employed Sigmoid entropy with an SVM, Song K et al. [32] used Synchrosqueezed Wavelet Transform (SWT) decomposition with CNN, and both Zhang Y et al. [33] and Alharthi M K et al. [27] extracted single features with Bi-GRU and CNN, respectively, resulting in relatively poorer performance. Some approaches applied feature fusion: Wu J et al. [34] fused multiple time-domain features using CEEMD and XGBoost (95.79% accuracy), Zhong L et al. [30] combined FuzzyEn, PSD, and brain networks with an SVM (98.01% accuracy), and Xu X et al. [16] integrated ApEn, SampEn, permutation, spectral, and wavelet entropy with GBDT. In contrast, the proposed model overcomes these shortcomings with 99.07% accuracy, 98.71% sensitivity, and 99.40% precision, while uniquely assessing both detection and prediction.

Table 10 compares epileptic seizure prediction methods: Rasheed et al. [35], Hellar et al. [19], Singh et al. [36], and Gao et al. [11] achieved accuracy rates ranging from 90% to 92%. In contrast, our proposed model attained an average accuracy of 95.41%, sensitivity of 96.74%, precision of 94.50%, F1-score of 95.55%, specificity of 94.02%, and Matthews correlation coefficient of 0.94 across 24 datasets, with 100% prediction rates for chb01, chb07, and chb17. Notably, our method maintains high performance across all CHB-MIT patients, whereas existing methods typically demonstrate high performance only on selected subjects [16, 30] and generally focus exclusively on either detection or prediction tasks [32, 37].

Table 9. CHB-MIT Detection Comparison

Author	Year	Feature	Classifier	ACC	SEN	SPE
Chen Z et al. [38]	2018	EEG	SVM	94	100	97.9
Raghu S et al. [31]	2019	Sigmoid entropy	SVM	94.38	94.21	NA
Wu J et al. [34]	2020	Multi-domain features	CEEMD+ XGBoost	95.79	NA	NA
Zanetti R et al. [21]	2020	Bandpower+entropy	RF	NA	96.6	92.5
Jiang Y et al. [39]	2021	Synchro Extracting Chirplet Operator (SECO)	SVM	99.29	98.71	99.22
Liu H et al. [40]	2021	Supervised Locality Preserving Canonical Correlation Analysis (SLPCCA)	LS-SVM	97.18	NA	NA
Zhang Y et al. [33]	2022	EEG	Bi-GRU	98.49	93.89	98.49
Alharthi M K et al. [27]	2022	EEG	1D-CNN	96.87	96.85	96.98
Xu X et al. [16]	2022	ApEn+ SampEn +permutation entropy+spectral entropy +wavelet entropy	GBDT	92.50	91.90	NA
Song K et al. [32]	2022	SWT	CNN	96.99	96.48	97.46
Zhong L et al. [30]	2023	FuzzyEn + PSD + Brain networks	SVM	98.01	NA	NA
This work	2023	STD+RMS+ApEn+FuzzyEn	RF+ENS	99.07	98.71	99.40

Table 10. CHB-MIT Prediction Comparison

Author	Year	Feature	Classifier	ACC	SEN	SPE
Gao Y et al. [11]	2020	PSDEDs	DCNNs	92.6	92.3	97
Tang L et al. [41]	2020	fractal spectrum+relative band energy+ synchronization modularity	Multi-view Convolutional Gated Recurrent Network (Mv-CGRN)	NA	94.5	NA
Rasheed K et al. [35]	2021	STFT	CESP	91.17	88.02	86
Hellar J et al. [19]	2022	EmDMD	SVM	NA	90.1	88.5
Yang X et al. [42]	2021	STFT	Resiual network (RDANet)	92.07	89.25	92.67
Rebello B C et al. [15]	2022	Spectral Power+ Mean + STD	SVM	76.92	50	66.66
Singh Y P et al. [36]	2023	Fourier transform Spectral Power	LSTM	91.99	91.84	94.95
Xu X et al. [43]	2023	EEG	Deep Residual Shrinkage Network (DRSN) and GRU	NA	90.54	NA
This work	2023	STD+RMS+ApEn+FuzzyEn	RF+ENS	95.41	96.74	94.02

5. CONCLUSION

The detection and prediction of epilepsy using EEG have attracted increasing attention in clinical applications. This paper proposes an automatic model for EEG detection and prediction based on feature fusion and an ensemble learning algorithm, validated using the CHB-MIT dataset. Comparative experiments reveal that the RF+ENS model excels in both tasks. For detection, it achieves an average accuracy of 99.07%, sensitivity of 98.71%, precision of 99.40%, F1-score of 99.05%, specificity of 99.40%, and a Matthews correlation coefficient of 0.98. For prediction, the model records mean values of 95.41% accuracy, 96.74% sensitivity, 94.50% precision, 95.55% F1-score, 94.02% specificity, and 0.94 for the Matthews correlation coefficient across 24 groups. These results lay a solid foundation for further epilepsy research and offer promising clinical applications. For patients, timely and accurate classification of interictal and preictal EEG signals can help prevent seizure damage, while for doctors, the model aids in diagnosing epilepsy and forecasting seizures. Future work will involve validating the proposed model on additional datasets.

ACKNOWLEDGEMENTS

This research is funded by the 2024 Annual Undergraduate Research Training Program (No. 2024146) and the 2024 Annual Undergraduate Research Training Program (No. 2024135).

REFERENCES

[1] Iasemidis L D. Seizure Prediction and its Applications [J]. Neurosurgery Clinics of North America, 2011, 22(4): 489–506.

[2] Iasemidis L D. Epileptic seizure prediction and control [J]. IEEE Transactions on Biomedical Engineering, 2003, 50(5): 549–558.

[3] Levin K, Klaus L, Richardson M P, et al. Seizure prediction — ready for a new era [J]. Nature Reviews Neurology, 2018.

[4] Morales S, Bowers M E. Time-frequency analysis methods and their application in developmental EEG data [J]. Developmental Cognitive Neuroscience, Oxford: Elsevier Sci Ltd, 2022, 54: 101067.

[5] Maguire M J, Abel A D. What changes in neural oscillations can reveal about developmental cognitive neuroscience: Language development as a case in point [J]. Developmental Cognitive Neuroscience, Oxford: Elsevier Sci Ltd, 2013, 6: 125–136.

[6] Alickovic E, Kevric J, Subasi A. Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction [J]. Biomedical Signal Processing and Control, Oxford: Elsevier Sci Ltd, 2018, 39: 94–102.

[7] Banupriya C, Devi A. Robust Optimization of electroencephalograph (EEG) Signals for Epilepsy Seizure Prediction by utilizing VSPO Genetic Algorithms with SVM and Machine Learning Methods [J]. Indian Journal of Science and Technology, 2021, 14: 1250–1260.

[8] AlSharabi K, Ibrahim S, Djemal R, et al. A DWT-Entropy-ANN Based Architecture for Epilepsy Diagnosis Using EEG Signals [A]. 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (atsip) [C]. New York: Ieee, 2016: 283–286.

[9] Kumar Y, Dewal M L, Anand R S. Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network [J]. Signal Image and Video Processing, London: Springer London Ltd, 2014, 8(7): 1323–1334.

[10] Deivasigamani S, Senthilpari C, Yong W H. Classification of Focal and Nonfocal EEG Signals Using ANFIS Classifier for Epilepsy Detection [J]. International Journal of Imaging Systems and Technology, Hoboken: Wiley, 2016, 26(4): 277–283.

[11] Gao Y, Gao B, Chen Q, et al. Deep Convolutional Neural Network-Based Epileptic Electroencephalogram (EEG) Signal Classification [J]. Frontiers in Neurology, 2020, 11: 375.

[12] Aung S T, Wongsawat Y. Modified-Distribution Entropy as the Features for the Detection of Epileptic Seizures [J]. Frontiers in Physiology, Lausanne: Frontiers Media Sa, 2020, 11: 607.

[13] Fathillah M S, Jaafar R, Chellappan K, et al. Multiresolution analysis on nonlinear complexity measurement of EEG signal for epileptic discharge monitoring [J]. Malaysian Journal of Fundamental and Applied Sciences, 2018, 14(2): 219–225.

[14] Alotaiby T N, Abd El-Samie F E, Alshebeili S A, et al. Seizure Detection with Common Spatial Pattern and Support Vector Machines [A]. 2015 International Conference on Information and Communication Technology Research (ictrc) [C]. New York: Ieee, 2015: 152–155.

[15] Rebello B C, Ramirez A R G, Heredia-Negron F, et al. A Machine Learning-Based Approach to Epileptic Seizure Prediction Using Electro-Encephalographic Signals [J]. Journal of Engineering Research, 2022, 2(8): 2–9.

[16] Xu X, Lin M, Xu T. Epilepsy Seizures Prediction Based on Nonlinear Features of EEG Signal and Gradient Boosting Decision Tree [J]. International Journal of Environmental Research and Public Health, Basel: Mdpi, 2022, 19(18): 11326.

[17] Lu Y, Ma Y, Chen C, et al. Classification of single-channel EEG signals for epileptic seizures detection based on hybrid features [J]. Technology and Health Care, Amsterdam: Ios Press, 2018, 26: S337–S346.

[18] Aayesha, Bilal Qureshi M, Afzaal M, et al. Fuzzy-Based Automatic Epileptic Seizure Detection Framework [J]. Computers, Materials & Continua, 2022, 70(3): 5601–5630.

[19] Hellar J, Erfanian N, Aazhang B. Epileptic electroencephalography classification using embedded dynamic mode decomposition [J]. Journal of Neural Engineering, Bristol: IOP Publishing Ltd, 2022, 19(3): 036029.

[20] Djemili R, Bourouba H, Amara Korba M C. Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals [J]. Biocybernetics and Biomedical Engineering, 2016, 36(1): 285–291.

[21] Zanetti R, Aminifar A, Atienza D. Robust Epileptic Seizure Detection on Wearable Systems with Reduced False-Alarm Rate [A]. 42nd Annual International Conferences of the Ieee Engineering in Medicine and Biology Society: Enabling Innovative Technologies for Global Healthcare Embc'20 [C]. New York: Ieee, 2020: 4248–4251.

[22] Jaiswal A K, Banka H. Epileptic seizure detection in EEG signal using machine learning techniques [J]. Australasian Physical & Engineering Sciences in Medicine, 2018, 41(1): 81–94.

[23] Tripathi D, Agrawal N. Epileptic Seizure Detection Using Empirical Mode Decomposition Based Fuzzy Entropy and Support Vector Machine [A]. International conference on green and human information technology [C]. 2018.

[24] Wang X, Gong G, Li N, et al. Detection Analysis of Epileptic EEG Using a Novel Random Forest Model Combined With Grid Search Optimization [J]. Frontiers in Human Neuroscience, 2019, 13: 52.

[25] Zhang S, Chen D, Ranjan R, et al. A lightweight solution to epileptic seizure prediction based on EEG synchronization measurement [J]. The Journal of Supercomputing, 2021, 77(4): 3914–3932.

[26] Song J-L, Li Q, Pan M, et al. Seizure tracking of epileptic EEGs using a model-driven approach [J]. Journal of Neural Engineering, 2020, 17(1): 016024.

[27] Alharthi M K, Moria K M, Alghazzawi D M, et al. Epileptic Disorder Detection of Seizures Using EEG Signals [J]. Sensors, Basel: Mdpi, 2022, 22(17): 6592.

[28] Raghu S, Sriraam N, Kumar G P. Classification of epileptic seizures using wavelet packet log energy and norm entropies with recurrent Elman neural network classifier [J]. Cognitive Neurodynamics, Dordrecht: Springer, 2017, 11(1): 51–66.

[29] Kavya B S , Dr S N Prasad. Identifying Epileptic Seizure by Optimized Feature Extraction Process Using the Method of Feature Fusion Technique [J]. Journal of Pharmaceutical Negative Results, 2022: 1092–1109.

[30] Zhong L, Wan J, Yi F, et al. Epileptic prediction using spatiotemporal information combined with optimal features strategy on EEG [J]. Frontiers in Neuroscience, 2023, 17: 1174005.

[31] Raghu S, Sriraam N, Temel Y, et al. Performance evaluation of DWT based sigmoid entropy in time and frequency domains for automated detection of epileptic seizures using SVM classifier [J]. Computers in Biology and Medicine, Oxford: Pergamon-Elsevier Science Ltd, 2019, 110: 127–143.

[32] Song K, Fang J, Zhang L, et al. An Intelligent Epileptic Prediction System Based on Synchrosqueezed Wavelet Transform and Multi-Level Feature CNN for Smart Healthcare IoT [J]. Sensors, 2022, 22(17): 6458.

[33] Zhang Y, Yao S, Yang R, et al. Epileptic Seizure Detection Based on Bidirectional Gated Recurrent Unit Network [J]. Ieee Transactions on Neural Systems and Rehabilitation Engineering, Piscataway: Ieee-Inst Electrical Electronics Engineers Inc, 2022, 30: 135–145.

[34] Wu J, Zhou T, Li T. Detecting Epileptic Seizures in EEG Signals with Complementary Ensemble Empirical Mode Decomposition and Extreme Gradient Boosting [J]. Entropy, 2020, 22(2): 140.

[35] Rasheed K, Qadir J, O'Brien T J, et al. A Generative Model to Synthesize EEG Data for Epileptic Seizure Prediction [J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 2322–2332.

[36] Singh Y P, Lobiyal D K. Automatic prediction of epileptic seizure using hybrid deep ResNet-LSTM model [J]. AI Communications, 2023, 36(1): 57–72.

[37] Ma M, Cheng Y, Wei X, et al. Research on epileptic EEG recognition based on improved residual networks of 1-D CNN and indRNN [J]. BMC Medical Informatics and Decision Making, 2021, 21(S2): 100.

[38] Chen Z, Lu G, Xie Z, et al. A Unified Framework and Method for EEG-Based Early Epileptic Seizure Detection and Epilepsy Diagnosis [J]. Ieee Access, Piscataway: Ieee-Inst Electrical Electronics Engineers Inc, 2020, 8: 20080–20092.

[39] Jiang Y, Chen W, Li M, et al. Synchroextracting chirplet transform-based epileptic seizures detection using EEG [J]. Biomedical Signal Processing and Control, Oxford: Elsevier Sci Ltd, 2021, 68: 102699.

[40] Liu H, Gao Y, Zhang J, et al. Epilepsy EEG classification method based on supervised locality preserving canonical correlation analysis [J]. Mathematical Biosciences and Engineering, Springfield: Amer Inst Mathematical Sciences-Aims, 2022, 19(1): 624–642.

[41] Tang L, Xie N, Zhao M, et al. Seizure Prediction Using Multi-View Features and Improved Convolutional Gated Recurrent Network [J]. Ieee Access, Piscataway: Ieee-Inst Electrical Electronics Engineers Inc, 2020, 8: 172352–172361.

[42] Yang X, Zhao J, Sun Q, et al. An Effective Dual Self-Attention Residual Network for Seizure Prediction [J]. Ieee Transactions on Neural Systems and Rehabilitation Engineering, Piscataway: Ieee-Inst Electrical Electronics Engineers Inc, 2021, 29: 1604–1613.

[43] Xu X, Zhang Y, Zhang R, et al. Patient-specific method for predicting epileptic seizures based on DRSN-GRU [J]. Biomedical Signal Processing and Control, Oxford: Elsevier Sci Ltd, 2023, 81: 104449.