

Article

A Voting-Based Ensemble Approach for Brain Disorder Detection Using Random Forest

Dina Aboelzahab ^{1,*}, Nawal Zaher ², Abdel Hamid Soliman ^{3,*} and Claude Chibelushi ⁴

¹ Department of Computer Engineering, Arab Academy for Science and Technology and Maritime Transport, Cairo P.O. Box 2033, Egypt

² Department of Electronics and Communications Engineering, Arab Academy for Science and Technology and Maritime Transport, Cairo P.O. Box 2033, Egypt

³ School of Digital, Technologies and Arts, Staffordshire University, Stoke-on-Trent ST4 2DF, UK

⁴ Semantics 21 Ltd., Staffordshire ST18 0WL, UK

* Correspondence: dinaabouldahab@aast.edu (D.A.); a.soliman@staffs.ac.uk (A.H.S.)

Abstract

Background: Automatic detection of abnormal electroencephalogram (EEG) signals is essential for supporting clinical screening and reducing human error in EEG interpretation. Although deep learning architectures such as CNN–LSTM have shown promising performance in EEG classification, challenges related to feature variability, non-stationarity, and sensitivity to pathological patterns remain. Our previous work with windowing-based CNN-LSTM architecture achieved strong performance but it did not achieve sufficient sensitivity for reliable clinical application. **Methods:** To overcome these limitations, we propose an enhanced voting-based ensemble framework that combines five CNN-LSTM base classifiers with a Random Forest (RF) meta-classifier, evaluated using 10-fold cross-validation. **Results:** The proposed ensemble model achieved a sensitivity of 92.86%, a specificity of 72.3%, and an overall accuracy of 83%, demonstrating competitive and clinically meaningful sensitivity for abnormal EEG detection under the adopted evaluation protocol. **Conclusions:** These findings demonstrate that integrating multi-model feature extraction with an RF-based voting ensemble improves diagnostic reliability, reduces false negatives, and supports early and accurate detection of brain disorders. This framework not only surpasses existing approaches but also provides a flexible foundation for future advancements in clinical decision support systems.

Keywords: EEG; abnormal EEG detection; CNN; LSTM; ensemble learning; Random Forest; voting classifier; cross-validation; SVM



Academic Editor: Rafiqul Chowdhury

Received: 15 November 2025

Revised: 23 December 2025

Accepted: 24 December 2025

Published: 4 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](#).

1. Introduction

1.1. Research Background

Electroencephalography (EEG) is a non-invasive neuroimaging technique that records electrical activity of the brain and provides valuable insights into neurological and cognitive processes. Despite its widespread clinical use, EEG signal interpretation remains challenging due to the non-stationary, noisy, and high-dimensional nature of the recorded signals [1]. Traditional signal processing and machine learning approaches often rely on handcrafted features and struggle to efficiently capture complex spatiotemporal patterns inherent in EEG data [2]. Recent advances in deep learning (DL) and artificial intelligence (AI) have significantly improved automated EEG analysis by enabling hierarchical feature learning and enhanced classification performance [3]. Convolutional Neural Networks (CNNs) [4],

Recurrent Neural Networks (RNNs) [5], transformers [6], and hybrid architectures [7] have demonstrated promising results across a wide range of EEG-based applications, including epilepsy and seizure-related disorders [8,9], Alzheimer's disease (AD) and mild cognitive impairment (MCI) [10,11], Parkinson's disease (PD) [12], schizophrenia [13], major depressive disorder (MDD) [14], and autism spectrum disorder (ASD) [15]. CNN-based models have been widely adopted for EEG analysis due to their ability to extract spatial features from multi-channel recordings [16]. Schirrmeister et al. [17] demonstrated that deep CNNs can outperform traditional methods in motor imagery classification, while Lawhern et al. [18] introduced EEGNet, a compact CNN architecture designed for generalizable EEG decoding across multiple brain–computer interface paradigms. To capture temporal dependencies, RNN-based models particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks have shown strong performance in modeling sequential EEG dynamics [19]. Tsiouris et al. [20] developed an LSTM-based framework for epileptic seizure prediction, achieving high sensitivity, while Ksibi et al. [21] employed LSTM networks for classifying depressive and healthy EEG recordings. More recently, hybrid CNN–RNN architectures have been proposed to jointly exploit spatial and temporal characteristics, such as CNN–GRU and bidirectional RNN-based models for emotion recognition and Parkinson's disease detection [19,22]. Transformers, originally developed for natural language processing, have also been adapted for EEG analysis due to their self-attention mechanisms, which enable efficient modeling of long-range temporal dependencies. Dai et al. [23] introduced a transformer-based architecture for sleep stage classification, demonstrating improved performance over conventional RNNs. In parallel, hybrid architectures combining CNNs, RNNs, and attention mechanisms have reported encouraging results in epilepsy detection, AD diagnosis, and PD classification tasks [24–26].

1.2. Research Motivation

Despite these advances, several challenges continue to limit the practical deployment of deep learning models for clinical EEG analysis. EEG recordings exhibit substantial inter-subject variability, heterogeneous pathological patterns, and sensitivity to recording conditions, particularly in large-scale real-world datasets. These factors often hinder model generalization and reduce robustness when transitioning from controlled experimental settings to clinical environments. In clinical EEG screening applications, sensitivity is a critical performance metric, as missed abnormal EEG recordings (false negatives) may delay diagnosis or intervention. Although many deep learning models achieve competitive overall accuracy, they may still exhibit suboptimal sensitivity when applied to heterogeneous clinical datasets. In our previous work, a windowing-based CNN–LSTM framework demonstrated strong classification performance; however, its sensitivity was insufficient for reliable abnormal EEG screening, highlighting the need for further improvement. Another limitation of many existing approaches is their reliance on a single feature representation or model architecture. Different temporal processing and downsampling strategies emphasize different characteristics of EEG signals, leading to complementary sensitivity–specificity trade-offs. Leveraging this diversity through ensemble learning provides a promising avenue for improving robustness and reducing missed abnormal cases without introducing excessive architectural complexity.

1.3. Research Objectives

Motivated by these challenges, the objective of this study is to develop a sensitivity-oriented ensemble framework for abnormal EEG detection that enhances robustness and reduces false negatives in clinical screening scenarios. Specifically, this work aims to:

- Employ multiple complementary temporal feature extraction strategies to generate diverse EEG representations while maintaining computational efficiency.
- Evaluate the performance of individual CNN-LSTM models trained on these complementary feature sets and analyze their sensitivity–specificity trade-offs.
- Integrate the outputs of the base CNN-LSTM models using a Random Forest-based voting ensemble to improve overall reliability and sensitivity; and assess the proposed framework on a large-scale, heterogeneous clinical EEG dataset using a consistent preprocessing and evaluation protocol.

Through these objectives, this study seeks to demonstrate that ensemble learning can effectively enhance abnormal EEG screening performance by combining complementary model behaviors within a unified and interpretable framework.

2. Methods

The presented framework shown in Figure 1 illustrates the methodological framework adopted for abnormal EEG detection using electroencephalogram (EEG) signals. Initially, raw EEG data are acquired and subjected to a feature selection process, wherein the most discriminative features are identified. Subsequently, a multi-model feature extraction strategy, comprising five complementary approaches, is employed to capture diverse signal characteristics. These extracted features are then processed by a deep learning model, designed to enhance classification performance. To further improve robustness and reliability, a voting system integrates the outputs from different models, thereby reducing bias and variance. The outcome of the framework is the classification of EEG signals into two categories: Normal or Abnormal, enabling accurate and efficient detection of brain disorders.

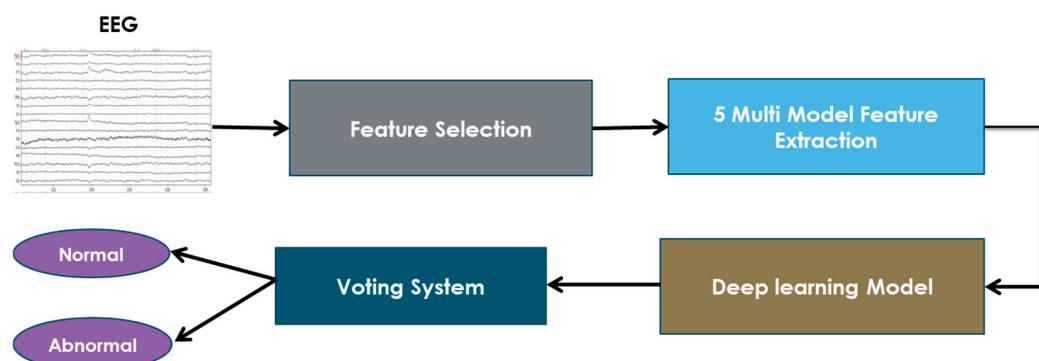


Figure 1. Framework for Abnormal EEG Detection Based on Multi-Model Feature Extraction and Voting Ensemble Learning.

2.1. Data Preprocessing

The Temple University Hospital (TUH) EEG Corpus is the largest publicly accessible clinical EEG database in the world, including more than 30,000 EEG recordings from more than 14,000 patients collected since 2002. An excellent resource for neurological research and machine learning applications, this large dataset spans a wide range of patient ages, diagnoses, electrode configurations, and sampling rates [27]. About 2500 new sessions are added to the corpus per year, and it is still growing. By offering highly annotated, real-world EEG data that represents the heterogeneity inherent in clinical situations, its size and clinical diversity make it possible to construct reliable automated EEG interpretation systems, especially for seizure detection and other brain illnesses. Because of this, the TUH EEG Corpus is a vital resource for developing data-driven neuroscience and enhancing diagnostic tools [27]. All EEG recordings were band-pass filtered to remove baseline drift and high-frequency noise, and notch filtering was applied to

suppress power-line interference. Signals were normalized on a per-channel basis to reduce inter-subject amplitude variability, and a consistent channel configuration and referencing scheme were maintained across all recordings. Table 1 summarizes the distribution of normal and abnormal EEG recordings across the training and testing sets, along with their corresponding gender proportions. The dataset is balanced in terms of class representation, with comparable male-female ratios across both splits, ensuring unbiased model evaluation.

Table 1. Distribution of Normal and Abnormal Cases in Training and Testing Datasets by Gender [28].

Training		Testing	
Normal	Abnormal	Normal	Abnormal
1150	1150	126	148
49.4% Male	43.9% Male	50% Male	43.2% Male
50.6% Female	56.1% Female	50% Female	56.8% Female

2.2. EEG Windowing Technique

In this study, the EEG data windowing process builds upon our previously proposed time–frequency analysis method based on Continuous Wavelet Transform (CWT) using Generalized Morse Wavelets (GMWs) [28]. GMWs are analytic wavelets well-suited for analyzing non-stationary signals with time-varying amplitude and frequency, such as EEG, due to their excellent time-frequency localization and minimal interference artifacts. Specifically, we set the wavelet parameters to $\gamma = 3$ and a time-bandwidth product of 60, balancing spectral resolution and computational efficiency to optimally capture EEG oscillatory behavior.

The CWT is applied independently to each electrode signal to generate magnitude scalograms, where signal energy is distributed across time–frequency representations. Importantly, this analysis is performed in a fully data-driven and label-agnostic manner, without using class information (normal or abnormal) at any stage of the window selection process. For each EEG recording, time instants corresponding to maximum wavelet energy are identified across electrodes, and a patient-specific average event time is computed after excluding outliers beyond three standard deviations. A temporal window centered around this average time is then selected for each recording. The same windowing criterion is applied uniformly to both normal and abnormal EEG recordings, ensuring a fair and unbiased comparison between classes. In this study, a fixed window length of 16,000 samples was used for all recordings, centered on the patient-specific average event time. This window length was chosen to capture a sufficiently long EEG segment containing informative signal dynamics while maintaining computational tractability. By focusing on this consistently defined temporal segment, the proposed windowing strategy reduces data dimensionality and suppresses irrelevant signal portions, enabling subsequent feature extraction and classification stages to operate on the most informative EEG segments without introducing label-dependent selection bias.

2.3. Multi-Feature Selection Block

EEG signals are characterized by high temporal resolution and complex non-stationary dynamics, making feature extraction a critical step in optimizing detection performance for brain disorders. To address this challenge, we propose an ensemble-based multi-resolution feature extraction framework that systematically reduces the original EEG signal length from 16,000 samples to 8000 samples while preserving diagnostically relevant information. This reduction is achieved through five distinct down sampling techniques, each designed to capture complementary aspects of neural activity. The techniques are as follows:

2.3.1. Averaging (Mean Pooling)

This method computes the arithmetic mean of adjacent samples within a sliding window, effectively smoothing high-frequency noise and baseline fluctuations while preserving the overall signal morphology. Averaging is widely used in EEG preprocessing to enhance the signal-to-noise ratio (SNR) by suppressing random artifacts without distorting underlying neural oscillations [29]. By reducing stochastic variability, this technique improves the robustness of subsequent machine learning models to intersession and inter-subject variability [30].

Given an EEG signal segment $X = [X_1, X_2, \dots, X_N]$, where $N = 16,000$, we apply non-overlapping sliding windows of size $k = 2$ and compute the mean of each pair:

$$X_{\text{avg}[i]} = (X_{2i-1} + X_{2i})/2 \quad i = 1, 2, \dots, N/2 \quad (1)$$

This approach effectively smooths high-frequency noise such as muscle artifacts and amplifier interference while preserving important low-frequency trends like delta and theta oscillations. It functions equivalently to a moving-average finite impulse response (FIR) filter with a boxcar kernel [31].

2.3.2. Max-Pooling

Max-pooling retains the highest amplitude value within each segment, thereby emphasizing transient events such as epileptic spikes, sharp waves, and other pathological discharges. This technique is particularly justified by prior studies in seizure detection, which have shown that max-pooling enhances the visibility of these high-amplitude transient events that are often clinically significant. By preserving abrupt changes in neural activity during down sampling, max-pooling ensures that critical features necessary for detecting disorders like epilepsy are effectively maintained [32,33].

$$X_{\text{max}[i]} = \max(X_{2i-1}, X_{2i}) \quad i = 1, 2, \dots, N/2 \quad (2)$$

Max-pooling emphasizes peak amplitudes, which is crucial for detecting epileptiform spikes, while also reducing sensitivity to baseline drifts.

2.3.3. Min-Pooling

Conversely, Min-pooling extracts the lowest amplitude value within each window, effectively capturing inhibitory phases, suppression bursts, and troughs in oscillatory activity. This approach is particularly useful for representing deep brain states characterized by cortical silencing, such as burst-suppression patterns observed during anesthesia or coma. Although direct references to min-pooling in EEG analysis are limited, related pooling and down sampling methods have been shown to preserve critical signal components associated with inhibitory neural dynamics and low-amplitude events [34]. By complementing max-pooling, min-pooling ensures that both excitatory and inhibitory neural dynamics are preserved within the feature set, providing a more comprehensive representation of brain activity.

$$X_{\text{min}[i]} = \min(X_{2i-1}, X_{2i}) \quad i = 1, 2, \dots, N/2 \quad (3)$$

2.3.4. Even Decimation

Even decimation subsamples the EEG signal by selecting every even-indexed sample, effectively reducing the temporal resolution by half while maintaining a sparse yet evenly distributed representation of the data. This approach is justified by its ability to significantly reduce computational overhead without substantial loss of diagnostic information, especially in scenarios where high-frequency details are less critical. Consequently, even

decimation offers a computationally efficient technique that is particularly advantageous for real-time processing applications requiring faster data handling and analysis [35].

$$X_{\text{even}[i]} = X_{2i} \quad i = 1, 2, \dots, N/2 \quad (4)$$

2.3.5. Odd Decimation

Similarly, odd decimation retains every odd-indexed sample, producing an alternative subsampled version of the original EEG signal. This approach, when combined with even decimation, helps mitigate the risk of losing temporally localized features that may occur exclusively at either even or odd sampling points. By preserving this redundancy, the method enhances feature retention and provides two slightly different yet structurally similar representations of the same signal, which can improve model generalization and robustness [36].

$$X_{\text{odd}[i]} = X_{2i-1} \quad i = 1, 2, \dots, N/2 \quad (5)$$

2.4. Feature Extraction and Classification

The proposed framework employs a hybrid CNN–LSTM architecture to jointly capture the spatial and temporal characteristics of EEG signals. The CNN–LSTM configuration follows the same design reported in our previous work published in MDPI AI [28], ensuring architectural consistency and enabling fair comparison across studies. Briefly, the model consists of two 1D convolutional layers using five filters of size 50, followed by two stacked LSTM layers with 100 and 120 hidden units, respectively. The CNN component processes the down sampled multi-resolution EEG segments to extract localized spatial features, such as spectral and morphological patterns, using ReLU activations and pooling operations to reduce dimensionality while preserving discriminative information. The resulting feature maps are reshaped into sequential representations and passed to the LSTM component, which models long-range temporal dependencies via gated memory mechanisms, supporting the analysis of non-stationary EEG dynamics.

For training, the network is optimized using Adam with a learning rate of 1×10^{-4} , L2 regularization of 0.001, and a gradient threshold of 1, over a maximum of 30 epochs. The input to the network comprises 19 EEG channels, and the final LSTM output is fed to a fully connected layer with SoftMax activation to produce probabilistic predictions for normal vs. abnormal EEG classification. Previous studies have shown that CNN–LSTM models can provide improved generalization compared to standalone CNN or LSTM architectures in EEG classification tasks [37,38]. Full architectural and training details are provided in [28].

2.5. Random Forest Voting Ensemble

In this study, a Random Forest (RF) voting ensemble is employed to aggregate classification outputs from multiple CNN–LSTM models, enhancing the robustness and accuracy of abnormal EEG detection. Random Forest, an ensemble of decision trees, is well-suited for EEG signal classification due to its ability to handle high-dimensional, noisy, and non-stationary data while resisting overfitting and balancing classification errors across imbalanced datasets [39]. By using a voting mechanism over probabilistic outputs from diverse base classifiers, the RF ensemble effectively fuses complementary information, improving generalization and mitigating individual model biases. Prior research has demonstrated the efficacy of RF-based ensemble methods in various EEG classification tasks, including seizure detection and emotion recognition, achieving superior performance compared to single classifiers and other ensemble strategies [40,41]. Moreover, RF's inherent feature importance measures provide interpretability, which is valuable for clinical applications. The integration of RF voting in this framework thus leverages its strengths in ensemble learning to deliver reliable and accurate EEG classification outcomes.

3. Performance and Evaluation

All simulations are carried out using MATLAB (version 9.12.0.1884302 (R2022a), Math-Works, Natick, MA, USA), running on an Intel I core™ i7-8700 CPU @3.20 GHz with 128 GB of RAM (Intel Corporation, Santa Clara, CA, USA). The data is originally sampled at a frequency of 250 Hz [28]. Figure 2 presents the architecture of the proposed EEG classification framework. Initially, raw EEG signals are segmented using a windowing technique, where an optimized window size of 16,000 samples is employed to retain the most discriminative features. Subsequently, five distinct feature extraction strategies namely even decimation, max pooling, min pooling, average pooling, and odd decimation are applied to the segmented signals in order to capture complementary aspects of the underlying neural dynamics. Each feature set is then independently processed by the proposed CNN-LSTM deep learning model, which leverages convolutional layers to extract spatial patterns and recurrent layers to model temporal dependencies. The CNN layers employ 1D convolutions to extract localized spatial features (e.g., transient spikes, spectral shifts), followed by ReLU activation and max pooling to further condense the representation. The LSTM then models temporal dynamics, with gated memory cells capturing long-range dependencies critical for detecting non-stationary events like seizures. The five sub-models generate independent outputs, fed to a Random Forest (RF) ensemble of randomized decision trees with weighted voting. To prevent information leakage, the meta-classifiers are trained using out-of-fold predictions generated by the base CNN-LSTM models during a 10-fold cross-validation procedure. Specifically, the dataset is partitioned into ten stratified folds; in each iteration, nine folds are used to train the base CNN-LSTM models, while the remaining fold is used to generate validation predictions. These out-of-fold predictions are then aggregated and used as input features to train the Random Forest (RF) meta-classifier. This process is repeated until each fold has served as the validation set once, ensuring that predictions used for training the RF are obtained from data unseen by the base models. The RF ensemble is subsequently evaluated within the same cross-validation framework, providing a robust decision mechanism that balances sensitivity and specificity. The final output of the framework is a binary classification distinguishing between normal and abnormal EEG patterns.

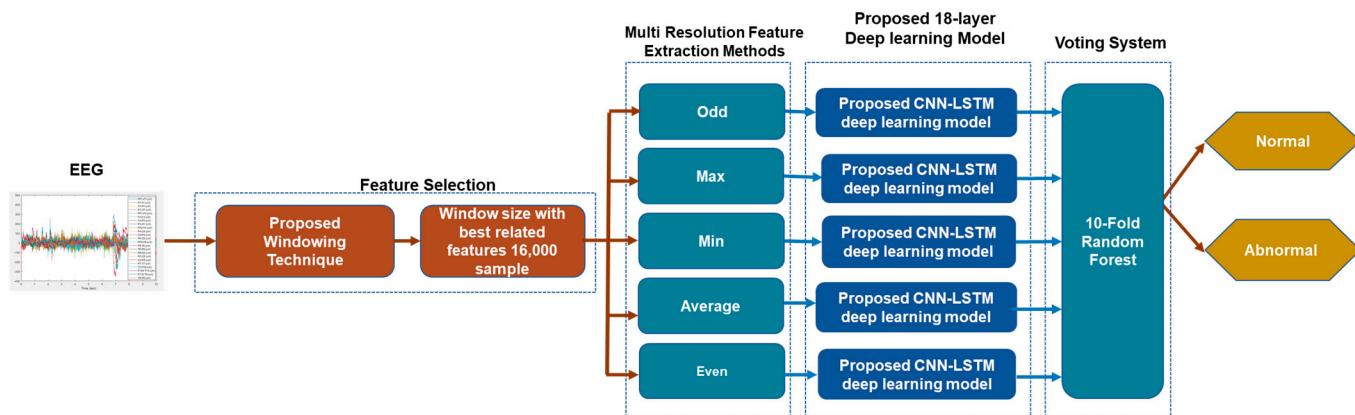


Figure 2. Proposed EEG classification framework. Five models (Max-CNN-LSTM, Min-CNN-LSTM, Avg-CNN-LSTM, Odd-CNN-LSTM, and Even-CNN-LSTM) are derived from different feature extraction strategies, and their outputs are combined through a Random Forest voting ensemble to classify EEG signals as normal or abnormal.

Importantly, the proposed integration of multiple complementary feature-extraction strategies with a hybrid CNN-LSTM architecture and ensemble learning represents a sensitivity-oriented framework for improving the accuracy and reliability of abnormal EEG detection. The TUH dataset [27] is used to assess the performance of the proposed

model. A total of 2574 patient records were utilized, with 2400 allocated for training and 274 reserved for testing.

3.1. Ablation Study and Ensemble Framework for EEG Classification

An ablation study was conducted to systematically evaluate five distinct feature extraction approaches applied to adjacent EEG samples. The study compared models employing: (1) average pooling (mean of adjacent samples), (2) max pooling (maximum value), (3) min pooling (minimum value), (4) odd decimation (odd-indexed samples), and (5) even decimation (even-indexed samples). Each method reduced the input dimensionality from 16,000 to 8000 points per epoch while preserving different signal characteristics. The comprehensive evaluation assessed performance across multiple clinically relevant metrics, including overall accuracy, F1-score, and individual precision/recall rates. When evaluated independently, these techniques demonstrated complementary performance profiles: max pooling achieved superior sensitivity for detecting transient pathological events like epileptic spikes but suffered from reduced specificity, while min pooling showed the inverse pattern. Average pooling provided more balanced performance but with less distinctive feature separation.

To overcome the limitations of the five individual feature selection approaches, we evaluated their performance by applying three distinct voting system techniques for classification. Support Vector Machine (SVM) was examined known for its optimal margin separation, Linear Regression for capturing linear trends, and Random Forest (RF) for modeling nonlinear relationships. Each technique was applied independently to the outputs of the five multi model features extraction, allowing us to assess their individual contributions and comparative effectiveness. This systematic evaluation provides a clear basis for identifying the most reliable technique for classification. This ensemble framework leveraged the complementary strengths of each constituent model while mitigating their individual weaknesses. The voting ensemble consistently demonstrated improved performance compared to individual models, achieving a more favorable trade-off between sensitivity and specificity.

This study achieved three primary objectives: First, it quantified the relative contribution of different extraction techniques to diagnostic performance. Second, it demonstrated the superiority of ensemble approaches over individual models for EEG classification tasks. Finally, the study established an optimized framework for clinical applications, with particular emphasis on achieving high sensitivity, which is critical for minimizing missed detections of neurological disorders. While specificity remains important, the results underscore that optimized feature extraction combined with intelligent model fusion is especially effective in enhancing sensitivity, thereby ensuring reliable early detection in clinical practice.

3.2. Results

The analysis for the five models based on true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), precision, recall (sensitivity), and F1 score are given in Table 1. The evaluation was performed on a testing dataset of 274 patients, including 126 abnormal and 148 normal cases. The results reveal notable trade-offs between sensitivity and precision across the models. Model 1 demonstrates the highest recall (0.8889), indicating strong ability to identify positive cases, which is further supported by the highest F1 score (0.7782). However, this comes at the cost of a relatively high false positive count ($FP = 50$), which lowers its precision (0.6914). Model 2 shows a modest balance with lower precision (0.6494) and recall (0.7937), resulting in a lower F1 score (0.7141). Model 3, on the other hand, achieves the highest precision (0.7946) with a significantly lower FP rate (23),

but suffers from the lowest recall (0.7063) among all models, reflecting a tendency to miss more actual positive cases ($FN = 37$). Model 4 presents the most balanced precision-recall combination (0.7049 and 0.6825), but with a slightly reduced overall F1 score (0.6945). Model 5 offers a middle ground, with moderate precision (0.6812) and recall (0.7460), producing an F1 score of 0.7118. Overall, some models show high sensitivity but lower specificity, while others exhibit the opposite pattern. This variation highlights the strengths and weaknesses of individual models and motivates the proposal of a voting system to leverage the complementary advantages of different models. Table 2 provides a comparison of the models in terms of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), precision, recall, and F1 score.

Table 2. Comparison of model performance of the five models (Model 1: Odd-CNN-LSTM, Model 2: Min-CNN-LSTM, Model 3: Max-CNN-LSTM, Model 4: Avg-CNN-LSTM, and Model 5: Even-CNN-LSTM), including true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), precision, recall, and F1-score. The evaluation was conducted on a testing set comprising 274 patients. The results highlight the trade-offs between sensitivity (recall), specificity, and overall predictive balance across models.

	TP	TN	FP	FN	Precision	Recall	F1 Score	Total Abnormal	Total Normal
Model 1	112	98	50	14	0.6914	0.8889	0.7782	126	148
Model 2	100	94	54	26	0.6494	0.7937	0.7141	126	148
Model 3	89	124	24	37	0.7946	0.7063	0.7470	126	148
Model 4	86	112	36	40	0.7049	0.6825	0.6945	126	148
Model 5	94	104	44	32	0.6812	0.7460	0.7118	126	148

The confusion matrices for the five models (Model 1–Model 5) is presented in Figure 3. Model 1 (Odd-CNN-LSTM) demonstrates strong sensitivity with 112 true positives, but also a relatively high number of false positives (50). Model 2 (Min-CNN-LSTM) shows slightly reduced sensitivity compared to Model 1, with 100 true positives, and a similar level of false positives (54). Model 3 (Max-CNN-LSTM) achieves the best balance, with 89 true positives and the lowest false positives (23), while also obtaining the highest true negatives (125), highlighting strong specificity. Model 4 (Avg-CNN-LSTM) reflects a higher false negative count (40), though it compensates with a larger number of true negatives (112). Finally, Model 5 (Even-CNN-LSTM) provides moderate performance, with 94 true positives and 104 true negatives, but also a noticeable level of misclassifications (32 false negatives and 44 false positives). Overall, the comparison across models highlights varying trade-offs between sensitivity and specificity, suggesting that no single model dominates across all metrics, and ensemble integration may provide a more robust classification outcome.

The performance across the five models is shown in Figure 4, which demonstrates a clear trade-off between sensitivity and specificity. Model 1 (Odd-CNN-LSTM) achieves the highest sensitivity (88.9%), making it the most effective at detecting positive cases, but this comes with a lower specificity (66.2%), indicating a higher rate of false positives. Model 2 (Min-CNN-LSTM) shows a relatively high sensitivity (79.4%) but with the lowest specificity (63.5%) among all models, suggesting a similar bias toward identifying positives. In contrast, Model 3 (Max-CNN-LSTM) stands out with the highest specificity (84.5%), reflecting its strength in correctly identifying negative cases, and it also achieves the highest accuracy (78.1%), despite having the lowest sensitivity (70.6%). Model 4 (Avg-CNN-LSTM) offers a balanced performance with moderate sensitivity (68.3%) and specificity (75.7%), leading to an accuracy of 76.4%. Model 5 (Even-CNN-LSTM), while showing the lowest accuracy (72.26%), maintains relatively balanced sensitivity (74.6%) and specificity (70.6%), suggesting a more neutral classification behavior.

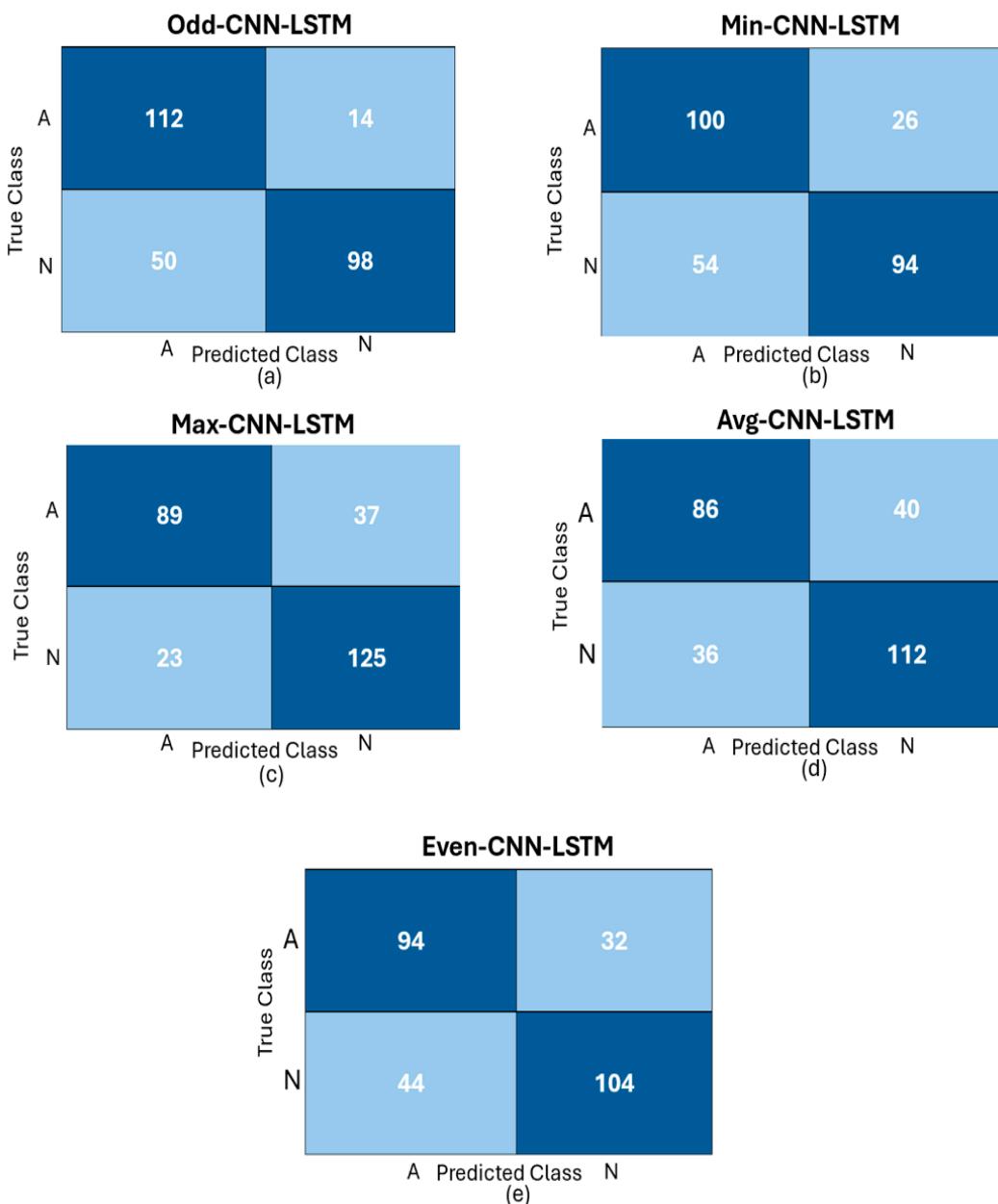


Figure 3. Confusion matrices of the five models: (a) Model 1: Odd-CNN-LSTM, (b) Model 2: Min-CNN-LSTM, (c) Model 3: Max-CNN-LSTM, (d) Model 4: Avg-CNN-LSTM, and (e) Model 5: Even-CNN-LSTM) illustrating classification performance on abnormal (A) and normal (N) EEG signals. The matrices highlight variations in true positive, true negative, false positive, and false negative counts across models, reflecting their trade-offs between sensitivity and specificity.

It is obvious from these results reveal some models are more sensitive but less specific, while others are more conservative in detecting positives. To benefit from the strengths of each, a voting system is proposed, where different ensemble strategies will be tested and evaluated to achieve a more robust and balanced classification performance. Figure 4 illustrates a bar chart that visually compares the accuracy, sensitivity, and specificity trends across all five models, helping to highlight the differences in their behavior.

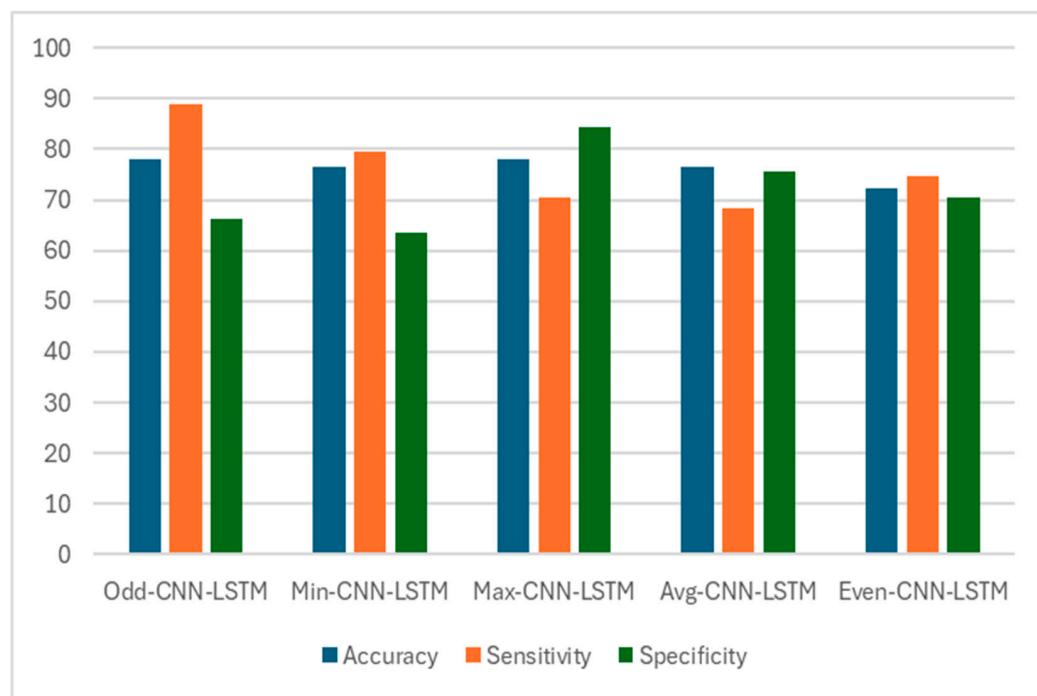


Figure 4. Bar Chart for comparison of accuracy, sensitivity, and specificity across the five models (Odd-CNN-LSTM, Min-CNN-LSTM, Max-CNN-LSTM, Avg-CNN-LSTM, and Even-CNN-LSTM).

To enhance EEG-based abnormal vs. normal EEG classification, we propose a two-stage meta-learning framework, where the predictions of five base models (Model 1–Model 5) serve as inputs to three distinct meta-classifiers: Random Forest (RF), Support Vector Machine (SVM), and Linear Regression (LR). In the first stage, each base model generates its classification output for the EEG data. In the second stage, these outputs are aggregated using three separate voting systems: (1) RF-based voting, which learns non-linear relationships between the base models' predictions; (2) SVM-based voting, which optimizes decision boundaries for maximal margin separation; and (3) LR-based voting, which applies a linear weighting scheme to combine predictions. Each meta-classifier is trained and evaluated on accuracy, sensitivity, and specificity to determine which approach best improves upon the individual models' performance. This systematic comparison will identify whether a non-linear (RF/SVM) or linear (LR) meta-learner is most effective for consolidating diverse EEG classification outputs in clinical settings. Table 3 presents a comparison of three voting system architectures SVM, logistic regression, and Random Forest applied to EEG-based abnormal vs. normal EEG classification, using TP, TN, FP, FN, precision, recall (sensitivity), and F1 score as evaluation metrics. Among the models, the Random Forest-based voting system achieved the highest recall (92.86%), indicating superior ability to correctly identify patients with abnormal EEG recordings. This is particularly critical in medical diagnosis, where missing a true positive case (i.e., a false negative) can lead to delayed or incorrect treatment. Although the Random Forest model exhibited the lowest precision (74.05%), it yielded the highest F1 score (82.37%), reflecting a strong balance between precision and recall. The logistic regression-based system achieved the highest precision (83.33%) but exhibited the lowest recall (75.39%), indicating a higher rate of false negatives. This characteristic may limit its suitability for clinical EEG screening applications, where failing to detect abnormal EEG activity has more serious clinical consequences than incorrectly classifying normal EEG recordings as abnormal. The SVM-based system offered a balanced performance, with a precision of 77.14%, recall of 85.71%, and an F1 score of 81.2%. Overall, although the Random Forest-based voting system exhibits slightly lower precision, its

higher sensitivity makes it particularly suitable for clinical EEG screening scenarios, where minimizing missed abnormal EEG recordings is of primary importance.

Table 3. Performance comparison of three voting-based ensemble models SVM, LR, and RF evaluated using true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), precision, recall, and F1-score. The results demonstrate varying trade-offs between precision and recall, with the RF-based voting model achieving the highest recall, while the LR-based voting model provides the highest precision.

	TP	TN	FP	FN	Precision	Recall	F1 Score
SVM-based voting Model	108	116	32	18	0.7714	0.8571	0.812
LR-based Voting Model	95	129	19	31	0.8333	0.7539	0.7917
RF-based Voting Model	117	107	41	9	0.7405	0.9286	0.8237

The confusion matrices reveal distinct performance characteristics for the three ensemble voting models, as illustrated in Figure 5. The SVM-based model achieves 108 true positives and 116 true negatives, corresponding to a sensitivity of 77.1% and a specificity of 86.6%. This reflects balanced performance, though the 32 false positives highlight some limitations in sensitivity. The Linear Regression-based model records 95 true positives and 129 true negatives, yielding a sensitivity of 75.4% and the highest specificity of 87.2% among the models. While it is strong at correctly identifying negative cases, its lower sensitivity indicates a higher risk of missed detections. In contrast, the Random Forest-based model achieves the highest sensitivity at 92.9% with 117 true positives and only 9 false negatives, making it the most effective at minimizing missed cases. However, its specificity is lower at 72.3% due to a higher false positive rate of 41. Overall, the Random Forest model is most suitable for sensitivity-critical clinical applications, the Linear Regression model is preferable when specificity is prioritized, and the SVM model provides a moderate balance between the two.

The bar chart in Figure 6 compares the performance of three ensemble voting techniques SVM, Linear Regression, and Random Forest used in EEG-based abnormal vs. normal EEG classification, evaluated by accuracy, sensitivity, and specificity. The Random Forest-based system outperformed the others in terms of sensitivity (92.86%) and accuracy (82.6%), making it the most effective at correctly identifying patients with abnormal EEG recordings. This high sensitivity is particularly valuable in medical applications where minimizing false negatives is critical to avoid missed diagnoses. SVM showed a balanced performance with an accuracy of 81.2%, sensitivity of 85.7%, and specificity of 78.38%, suggesting it maintains a reasonable trade-off between detecting true positives and avoiding false positives. In contrast, the Linear Regression model achieved the highest specificity (87.1%) indicating a strong ability to correctly identify healthy individuals but at the cost of the lowest sensitivity (75.4%), which could lead to more undiagnosed cases. Given the medical importance of capturing all true disorder cases, the Random Forest approach is the most clinically appropriate due to its superior sensitivity and overall robust performance. In EEG classification, sensitivity is a critical performance metric, particularly in clinical applications where missing pathological activity such as epileptic seizures or abnormal brain rhythms can have serious consequences. High sensitivity ensures that the model detects the majority of true neurological events, reducing the risk of false negatives that could delay diagnosis or intervention [42]. While specificity and accuracy remain important for minimizing false alarms and overall correctness, sensitivity is often prioritized in EEG analysis due to the high stakes of overlooking critical abnormalities. Striking a balance between these metrics is challenging, as overly sensitive systems may increase false positives, but in many medical contexts, the cost of missing a true event outweighs the cost

of additional verification. Thus, optimizing sensitivity without compromising practical utility is a key focus in EEG classification research.

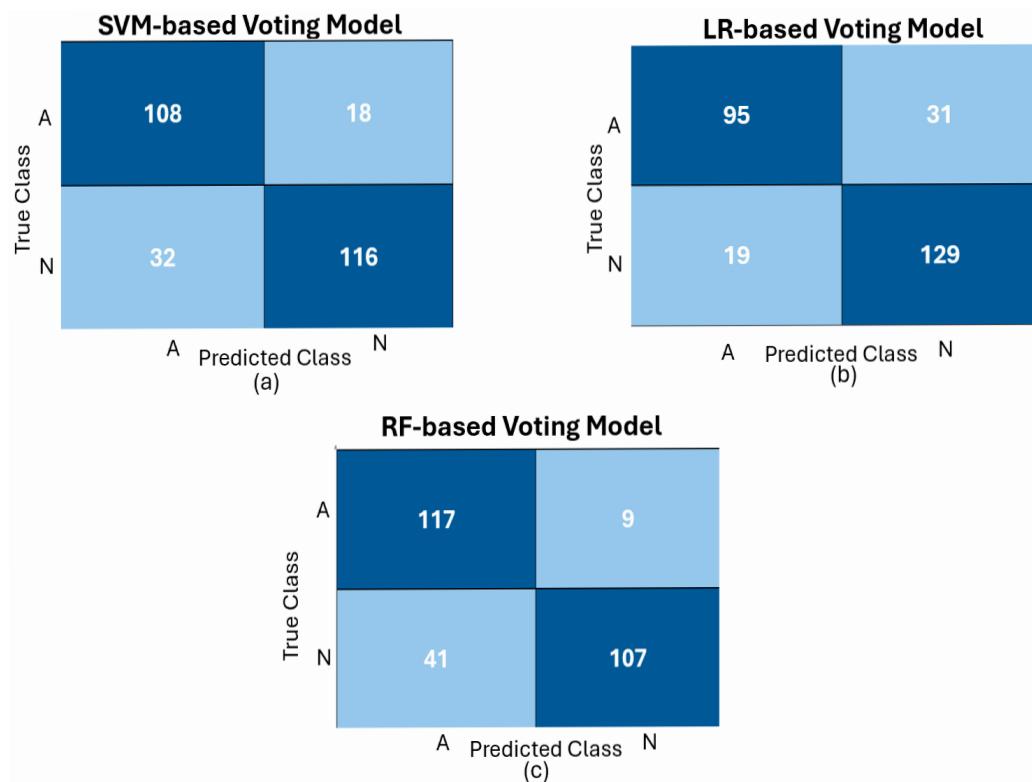


Figure 5. Confusion matrices of the three voting-based ensemble models: (a) SVM-based, (b) Linear Regression-based, and (c) Random Forest-based. The matrices illustrate classification outcomes in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), highlighting the trade-offs between sensitivity and specificity for each technique.

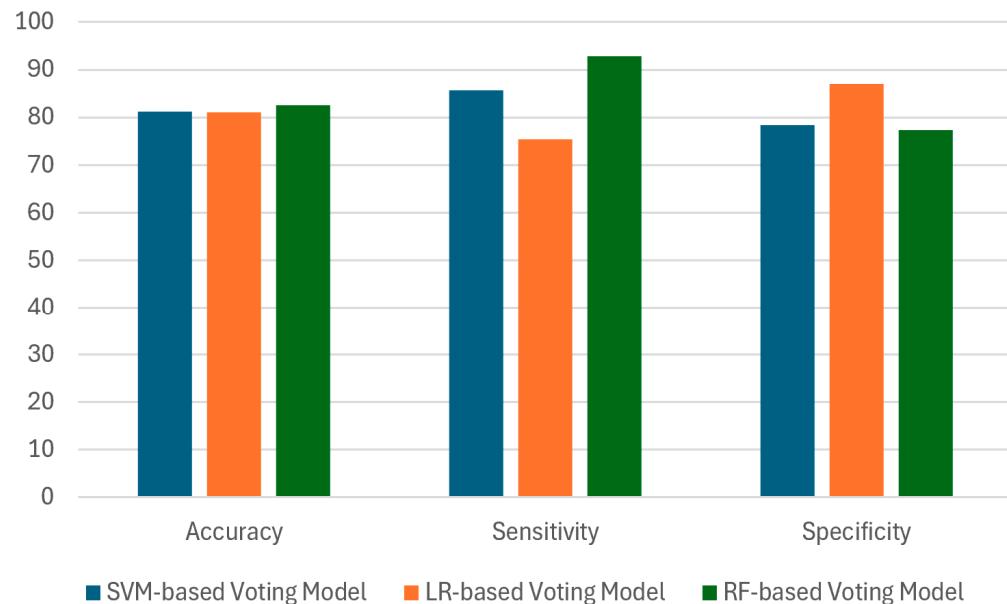


Figure 6. Bar Chart for the three voting-based ensemble models: SVM-based, Linear Regression-based, and Random Forest-based. The matrices illustrate classification outcomes in terms of sensitivity and specificity and accuracy for each technique.

Table 4 provides a comparison of the proposed model with representative EEG classification approaches reported in the literature, considering data selection techniques, feature

extraction methods, classification algorithms, accuracy, and sensitivity. All referenced studies were evaluated using the Temple University Hospital (TUH) EEG corpus; however, direct numerical comparison should still be interpreted cautiously due to differences in data partitions, preprocessing strategies, windowing methods, and evaluation protocols. Nevertheless, under the adopted evaluation setting, the proposed ensemble framework demonstrates competitive overall accuracy while achieving high sensitivity, highlighting its suitability for abnormal EEG screening tasks where minimizing false negatives is critical. Sharma [43] employed fuzzy entropy and fractal-based features combined with an SVM classifier, achieving an accuracy of 79.34% and a sensitivity of 77.54%, while Tomas et al. [44] reported a lower accuracy (68%) using an HMM with phase synchronization and energy features. Western et al. [45] achieved an accuracy of 81.88% using a CNN-based approach; however, sensitivity was not reported, limiting its clinical interpretability. T. Wu [46] attained the highest reported accuracy (89.13%) using a DWT–CatBoost framework, accompanied by a sensitivity of 84.92%. Similarly, Albaqami [47] leveraged wavelet packet decomposition (WPD) with LightGBM to achieve an accuracy of 86.59% and a sensitivity of 81.74%. In comparison, the proposed model achieves competitive accuracy (82.68%) while demonstrating notably higher sensitivity under the adopted evaluation protocol, underscoring its effectiveness in reducing false negatives. This characteristic is particularly important in clinical screening applications, where the cost of missed abnormal EEG detections often outweighs that of false positives. Accordingly, while maintaining competitive overall accuracy, the Random Forest-based voting framework exhibits improved sensitivity, reinforcing its practical value for abnormal EEG screening applications.

Table 4. Comparison of Model Performance with different architectures in literature in terms of data selection technique, feature extraction, classification technique, accuracy and sensitivity.

	Data Selection Technique	Feature Extraction	Classification Technique	Accuracy	Sensitivity
Sharma [43]	1st minute	Fuzzy Entropy + Logarithmic Squared Norm + Fractal Dimension	SVM	79.34	77.54%
Tomas et al. [44]	-	PS + PLV + Energy	HMM	68%	68%
Western et al. [45]	2nd minute	-	CNN	81.88%	-
T Wu [46]	-	DWT	CatBoost	89.13%	84.92%
Albaqami [47]	-	WPD	LightGBM	86.59%	81.74%
Abooelzahab [28]	Windowing Technique	CNN	LSTM	82.68%	78.5%
Proposed Model	Windowing Technique	5 Multi model + CNN LSTM model	Random Forest Voting system	82.68%	92.86%

4. Discussion

The Comparative analysis with existing approaches demonstrates the effectiveness of the proposed Random Forest–based voting system for abnormal EEG detection. While earlier studies achieved varying levels of performance, many faced trade-offs between accuracy and sensitivity. For instance, Sharma [43] reported an accuracy of 79.34% with moderate sensitivity (77.54%), while Tomas et al. [44] achieved substantially lower performance (68% accuracy and sensitivity) using HMM-based methods. More advanced techniques such as DWT with CatBoost [46] and WPD with LightGBM [47] reached higher accuracies of 89.13% and 86.59%, with sensitivities of 84.92% and 81.74%, respectively. CNN-based methods, such as that of Western et al. [45], also showed promising accuracy (81.88%) but without explicitly reporting sensitivity. In contrast, the proposed model, though achieving a slightly lower overall accuracy (82.68%) compared to CatBoost and LightGBM, demonstrated a substantially higher sensitivity (92.86%). This finding underscores the model’s strength in correctly identifying positive cases, which is particularly critical in clinical applications where missed detections can have severe consequences. By prioritizing sensitivity while

maintaining competitive accuracy, the proposed framework addresses a key limitation observed in prior methods and establishes itself as a more reliable solution for practical neurological disorder detection.

The superior sensitivity achieved by the proposed framework can be attributed to the combination of a windowing-based data selection strategy and the integration of five complementary CNN-LSTM models through a Random Forest voting scheme. Windowing enhances the temporal resolution of EEG signals, allowing the extraction of more discriminative features, while the ensemble fusion mitigates the bias of individual models and captures diverse decision patterns. This synergy improves the system's ability to detect subtle EEG abnormalities, thereby reducing false negatives and ensuring higher reliability in clinical diagnosis.

Despite these encouraging results, several limitations should be acknowledged. First, although the proposed framework demonstrates strong performance on a large-scale clinical EEG dataset, it addresses a binary abnormal-versus-normal classification task and does not distinguish between specific neurological conditions. Second, the ensemble relies on fixed, non-learnable temporal transformations; incorporating learnable multi-scale or attention-based mechanisms may further enhance representational capacity. Third, while cross-validation provides a reliable estimate of generalization performance, external validation on independent datasets would further strengthen clinical applicability.

Future research will focus on extending the proposed framework to multi-class EEG classification, integrating attention-based or transformer architectures within the ensemble, and exploring patient-specific adaptation strategies. Additionally, improving model interpretability and validating the framework across multiple clinical datasets remain important directions for enhancing trust and deployment in real-world clinical environments.

5. Conclusions

This study proposed and evaluated an ensemble-based EEG classification framework that integrates multiple feature extraction strategies with a hybrid CNN-LSTM architecture and ensemble learning. By systematically analyzing five distinct feature extraction approaches and three ensemble voting techniques, the results demonstrated the inherent trade-offs between sensitivity and specificity across models. Notably, the Random Forest-based voting ensemble achieved the highest sensitivity of 92.86%, outperforming previously reported methods in the literature and establishing it as the most effective for sensitivity-critical clinical applications such as seizure detection, where minimizing false negatives is essential. In contrast, the Linear Regression-based ensemble provided the highest specificity, while the SVM-based ensemble offered a balanced trade-off between sensitivity and specificity. These findings highlight the importance of combining optimized feature extraction with intelligent model fusion to enhance both robustness and clinical reliability. Overall, the proposed framework not only outperforms comparable approaches in literature but also contributes a flexible and effective solution for EEG-based abnormal detection, paving the way for future research to refine ensemble integration strategies and extend their application to broader neurological diagnostic settings.

Author Contributions: D.A. writing the manuscript, conducting experiments & data analysis. N.Z.: Conducting experiments, data analysis, reviewing and approving the final manuscript. A.H.S.: Data analysis and approving the final manuscript. C.C.: Data analysis and approving the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The datasets used and analyzed during the current study are available in https://isip.piconepress.com/projects/nedc/html/tuh_eeg/ (accessed on 1 September 2020).

Conflicts of Interest: Claude Chibelushi was employed by the company Semantics 21 Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Parsa, M.; Rad, H.Y.; Vaezi, H.; Hossein-Zadeh, G.A.; Setarehdan, S.K.; Rostami, R.; Rostami, H.; Vahabie, A.H. EEG-based classification of individuals with neuropsychiatric disorders using deep neural networks: A systematic review of status and future directions. *Comput. Methods Programs Biomed.* **2023**, *240*, 107683. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Bhardwaj, S.; Kumar, A. A systematic review of EEG based automated schizophrenia diagnosis using AI techniques. *Front. Hum. Neurosci.* **2024**, *18*, 1347082. [\[CrossRef\]](#)
3. Abdelfattah, S.M.; Abdelrahman, G.M.; Wang, M. Deep learning in EEG: A survey of recent advances and challenges. *IEEE Access* **2022**, *10*, 36219–36244.
4. Mohan, R.; Perumal, S. Classification and Detection of Cognitive Disorders like Depression and Anxiety Utilizing Deep Convolutional Neural Network (CNN) Centered on EEG Signal. *Trait. Du Signal.* **2023**, *40*, 971–979. [\[CrossRef\]](#)
5. Najafi, T.; Jaafar, R.; Remli, R.; Wan Zaidi, W.A. A classification model of EEG signals based on RNN-LSTM for diagnosing focal and generalized epilepsy. *Sensors* **2022**, *22*, 7269. [\[CrossRef\]](#)
6. Vafaei, E.; Hosseini, M. Transformers in EEG analysis: A review of architectures and applications in motor imagery, seizure, and emotion classification. *Sensors* **2025**, *25*, 1293. [\[CrossRef\]](#)
7. Nour, M.; Senturk, U.; Polat, K. A novel hybrid model in the diagnosis and classification of Alzheimer's disease using EEG signals: Deep ensemble learning (DEL) approach. *Biomed. Signal Process. Control* **2024**, *89*, 105751. [\[CrossRef\]](#)
8. Qiu, X.; Yan, F.; Liu, H. A difference attention ResNet-LSTM network for epileptic seizure detection using EEG signal. *Biomed. Signal Process. Control* **2023**, *83*, 104652. [\[CrossRef\]](#)
9. Daoud, H.; Bayoumi, M.A. Efficient epileptic seizure prediction based on deep learning. *IEEE Trans. Biomed. Circuits Syst.* **2019**, *13*, 804–813. [\[CrossRef\]](#)
10. Aviles, M.; Sánchez-Reyes, L.M.; Álvarez-Alvarado, J.M.; Rodríguez-Reséndiz, J. Machine and Deep Learning Trends in EEG-Based Detection and Diagnosis of Alzheimer's Disease: A Systematic Review. *Eng* **2024**, *5*, 1464–1484. [\[CrossRef\]](#)
11. Kim, M.J.; Youn, Y.C.; Paik, J. Deep learning-based EEG analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset. *NeuroImage* **2023**, *272*, 120054. [\[CrossRef\]](#)
12. Göker, H. Automatic detection of Parkinson's disease from power spectral density of electroencephalography (EEG) signals using deep learning model. *Phys. Eng. Sci. Med.* **2023**, *46*, 1163–1174. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Sairamya, N.J.; Subathra, M.S.; George, S.T. Automatic identification of schizophrenia using EEG signals based on discrete wavelet transform and RLNDiP technique with ANN. *Expert Syst. Appl.* **2022**, *192*, 116230. [\[CrossRef\]](#)
14. Xia, M.; Zhang, Y.; Wu, Y.; Wang, X. An end-to-end deep learning model for EEG-based major depressive disorder classification. *IEEE Access* **2023**, *11*, 41337–41347. [\[CrossRef\]](#)
15. Mohi-ud-Din, Q.; Jayanth, A.K. Autism Spectrum Disorder classification using EEG and 1D-CNN. In Proceedings of the 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IMECON), Jaipur, India, 1–2 December 2021; pp. 1–5.
16. Amrani, G.; Adadi, A.; Berrada, M.; Souirti, Z.; Boujraf, S. EEG signal analysis using deep learning: A systematic literature review. In Proceedings of the 2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 20–22 October 2021; pp. 1–8. [\[CrossRef\]](#)
17. Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* **2017**, *38*, 5391–5420. [\[CrossRef\]](#)
18. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* **2018**, *15*, 056013. [\[CrossRef\]](#)
19. Tong, W.; Yang, L.; Qin, Y.; Che, Y.; Han, C. EEG-Based Emotion Recognition by Using Machine Learning and Deep Learning. In Proceedings of the 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 5–7 November 2022; pp. 1–5. [\[CrossRef\]](#)
20. Tsioris, K.M.; Pezoulas, V.C.; Zervakis, M.; Konitsiotis, S.; Koutsouris, D.D.; Fotiadis, D.I. A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput. Biol. Med.* **2018**, *99*, 24–37. [\[CrossRef\]](#)
21. Ksibi, A.; Zakariah, M.; Menzli, L.J.; Saidani, O.; Almuqren, L.; Hanafieh, R.A. Electroencephalography-based depression detection using multiple machine learning techniques. *Diagnostics* **2023**, *13*, 1779. [\[CrossRef\]](#) [\[PubMed\]](#)

22. Chinnathambi, D.; Ravi, S.; Dhanasekaran, H.; Dhandapani, V.; Rao, R.; Pandiaraj, S. Early detection of Parkinson's disease using deep learning: A convolutional bi-directional GRU approach. In *Intelligent Technologies and Parkinson's Disease: Prediction and Diagnosis*; IGI Global Scientific Publishing: Hershey, PA, USA, 2024; pp. 228–240.

23. Dai, Y.; Li, X.; Liang, S.; Wang, L.; Duan, Q.; Yang, H.; Zhang, C.; Chen, X.; Li, L.; Li, X.; et al. Multichannelsleepnet: A transformer-based model for automatic sleep stage classification with psg. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 4204–4215. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Pandey, S.K.; Janghel, R.R.; Mishra, P.K.; Ahirwal, M.K. Automated epilepsy seizure detection from EEG signal based on hybrid CNN and LSTM model. *Signal Image Video Process.* **2023**, *17*, 1113–1122. [\[CrossRef\]](#)

25. Kowshiga, A.; Pavithra, T.; Priyanka, V. Deep Learning Into the Future: Hybrid CNN-RNN for Early Detection of Alzheimer's Disease. In Proceedings of the 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 7–9 August 2024; pp. 940–946.

26. El-Sayed, R.S. A Hybrid CNN-LSTM Deep Learning Model for Classification of the Parkinson Disease. *IAENG Int. J. Appl. Math.* **2023**, *53*, 1427.

27. Harati, A.; Choi, S.; Tabrizi, M.; Obeid, I.; Picone, J.; Jacobson, M.P. The temple university hospital EEG corpus. In Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, USA, 3–5 December 2013; pp. 29–32.

28. Abooezhabab, D.; Zaher, N.; Soliman, A.H.; Chibelushi, C. A Combined Windowing and Deep Learning Model for the Classification of Brain Disorders Based on Electroencephalogram Signals. *AI* **2025**, *6*, 42. [\[CrossRef\]](#)

29. Sanei, S.; Chambers, J.A. *EEG Signal Processing*; John Wiley & Sons: Hoboken, NJ, USA, 2013.

30. Repovs, G. Dealing with noise in EEG recording and data analysis. *Inform. Medica Slov.* **2010**, *15*, 18–25.

31. Subasi, A. *Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques: A MATLAB Based Approach*; Academic Press: London, UK, 2019.

32. Roy, S.; Kiral-Kornek, I.; Harrer, S. ChronoNet: A deep recurrent neural network for abnormal EEG identification. In Proceedings of the Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, 26–29 June 2019; Proceedings 17; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 47–56.

33. Acharya, U.R.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adeli, H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput. Biol. Med.* **2018**, *100*, 270–278. [\[CrossRef\]](#) [\[PubMed\]](#)

34. Tveitstøl, T.; Tveten, M.; Pérez, T.A.S.; Hatlestad-Hall, C.; Yazidi, A.; Hammer, H.L.; Hebold Haraldsen, I.R. Introducing Region Based Pooling for handling a varied number of EEG channels for deep learning models. *Front. Neuroinform.* **2024**, *17*, 1272791. [\[CrossRef\]](#)

35. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001. [\[CrossRef\]](#)

36. Wang, Z.; Yan, W.; Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 1578–1585.

37. Liu, J.; Wu, G.; Luo, Y.; Qiu, S.; Yang, S.; Li, W.; Bi, Y. EEG-based emotion classification using a deep neural network and sparse autoencoder. *Front. Syst. Neurosci.* **2020**, *14*, 43. [\[CrossRef\]](#) [\[PubMed\]](#)

38. Khademi, Z.; Ebrahimi, F.; Kordy, H.M. A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals. *Comput. Biol. Med.* **2022**, *143*, 105288. [\[CrossRef\]](#)

39. Mary, G.; Chitti, S.; Vallabhaneni, R.B.; Renuka, N. EEG signal classification automation using novel modified random forest approach. *J. Sci. Ind. Res.* **2023**, *82*, 101–108. [\[CrossRef\]](#)

40. Hosseini, M.P.; Pompili, D.; Elisevich, K.; Soltanian-Zadeh, H. Random ensemble learning for EEG classification. *Artif. Intell. Med.* **2018**, *84*, 146–158. [\[CrossRef\]](#)

41. Molina, W.C.; Cavanagh, J.; Lin, C.Y. Application of Random Forest to classify EEG data of mTBI patients and control adults obtained during a Visuospatial Working Memory Task. *J. Vis.* **2022**, *22*, 3842.

42. Aslam, M.H.; Usman, S.M.; Khalid, S.; Anwar, A.; Alroobaea, R.; Hussain, S.; Almotiri, J.; Ullah, S.S.; Yasin, A. Classification of EEG signals for prediction of epileptic seizures. *Appl. Sci.* **2022**, *12*, 7251. [\[CrossRef\]](#)

43. Sharma, M.; Patel, S.; Acharya, U.R. Automated detection of abnormal EEG signals using localized wavelet filter banks. *Pattern Recognit. Lett.* **2020**, *133*, 188–194. [\[CrossRef\]](#)

44. Iešmantas, T.; Alzbutas, R. Convolutional neural network for detection and classification of seizures in clinical data. *Med. Biol. Eng. Comput.* **2020**, *58*, 1919–1932. [\[CrossRef\]](#) [\[PubMed\]](#)

45. Western, D.; Weber, T.; Kandasamy, R.; May, F.; Taylor, S.; Zhu, Y.; Canham, L. Automatic report-based labelling of clinical EEGs for classifier training. In Proceedings of the 2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 4 December 2021; pp. 1–6.

46. Wu, T.; Kong, X.; Zhong, Y.; Chen, L. Automatic detection of abnormal EEG signals using multiscale features with ensemble learning. *Front. Hum. Neurosci.* **2022**, *16*, 943258. [[CrossRef](#)]
47. Albaqami, H.; Hassan, G.M.; Subasi, A.; Datta, A. Automatic detection of abnormal EEG signals using wavelet feature extraction and gradient boosting decision tree. *Biomed. Signal Process. Control* **2021**, *70*, 102957. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.