Advanced EEG Signal Preprocessing and Decomposition: A Multi-Stage Pipeline for Neurological Disorder Detection

Chinta Someswara Rao[1], Shiva Shankar Reddy[2]

[1] Faculty of Engineering and Built Science, Lincoln University College, Petaling Jaya, Selangor-47301, Malaysia.

[2] Department of Computer Science and Engineering, Lincoln University College, Petaling Jaya, Selangor-47301, Malaysia.

Email ID: pdf.someswararao@lincoln.edu.my; pdfsv.shiva@lincoln.edu.my;

Abstract: Researchers and clinicians widely use electroencephalography signals for non-invasive assessment of neurological function; however, artifacts, broadband noise, non-stationarity, and high inter-subject variability affect their clinical interpretation. This study proposes a multi-stage EEG preprocessing and decomposition framework for detecting neurological disorders. We evaluate the framework on three public EEG datasets: CHB-MIT, Sleep-EDF Expanded, and Bonn EEG. Experimental results show an average SNR improvement of 18.5 dB, artifact rejection above 96%, and four-class classification accuracy of 96.2%. The method also achieves 95.7% sensitivity, 96.8% specificity, and an AUC of 0.987. These results suggest that systematic preprocessing and decomposition can substantially improve the reliability of automated EEG-based detection of neurological disorders.

Keywords: Electroencephalography; EEG artifact removal; EEG feature extraction; seizure detection; neurological disorder detection; epilepsy;

Introduction

The human brain produces bioelectrical activity as a result of the synchronized firing of millions of neurons, and Electroencephalography (EEG) is the most readily available and affordable clinical instrument for non-invasive measurement of these dynamics [1],[2]. For more than 80 years, EEG has been the basis of diagnosis in neurological medicine, with its temporal resolution of milliseconds making it particularly suited to capture fast, transient phenomena, such as epileptic discharges and sleep-related breathing events, which are the hallmark of neurological disease [3]. Although researchers have demonstrated the clinical utility of EEG, the step from manual analysis by experts to fully automated, large-volume, scalable EEG diagnostic systems remains daunting [4], [5]. There is a wide variety of artifacts in raw EEG signals, including biological and environmental noise. Ocular artifacts (from eye blinks and saccadic eye movements) produce voltage deflections that are 10 times or greater than those from the underlying neural signals [6], [7]. The pipeline is specifically designed to address all three issues: artifact suppression via ICA-based decomposition and adaptive wavelet thresholding; noise reduction via frequency-selective adaptive filtering; and nonstationary signal decomposition using three complementary multi-resolution analysis techniques: DWT, EMD, and VMD [8]. The proposed unified pipeline provides a consistent signal-processing chain, delivering optimally processed signals from one stage to the next, thereby preserving as much diagnostic information as possible in the last decomposed signal [9], [10].

Independent Component Analysis (ICA) is the most widely used technique for removing artifacts in EEG signals. Lin et al. [11] introduced ICA as the reference method for separating overlapping EEG and artifact sources, and subsequent efforts have focused on automating the detection of artifact components.

Muhammad et al. [12] proposed a CNN-based deep learning-assisted ICA framework to classify independent components by learning spectral and topographic features, achieving an average accuracy of 94.3% in automated classification of artifact components without manual artifact inspection [13]. The power of wavelet-based artifact removal has been established as a useful complement for ICA, especially for transient, event-related artifacts. Deshmukh et al. [14] first demonstrated that discrete wavelet decomposition and threshold-based suppression are effective methods for reducing EMG contamination in epileptic EEG. Building on this, Silpa et al. [15] presented an adaptive wavelet packet thresholding scheme, where threshold values were adaptively selected from the local statistics of the signal in each level of the wavelet. In this cascaded approach proposed by Punuri et al. [16], broadband artifact sources are separated using ICA, followed by notch-filtered removal of residual ECG harmonic artifacts and adaptive high-pass filtering below 0.5 Hz to suppress electrode drift [17]. Nikilish et al. [18] showed that the canonical correlation analysis (CCA) method can be an effective alternative to the ICA method for artifact removal in resource-constrained applications, achieving 89% of ICA's removal performance at only 20% of the computational cost.

Novelty and Contributions:

- A comprehensive, modular multi-stage pipeline for EEG preprocessing, incorporating multi-resolution decomposition, adaptive noise suppression, and artefact rejection.
- A comparative study and analysis of DWT, EMD, and VMD algorithms for sub-band extraction of the EEG signal and their complementary aspect in capturing neurologically meaningful frequency content of the EEG signal.

Methodology

The full multi-stage EEG preprocessing and decomposition pipeline is described here. The framework is structured in five sequential steps, as shown in Figure 1.

Stage 1: Epoching & Bandpass Filter — the signal is divided into epochs (10 seconds with 5 seconds overlap) and filtered between 0.5–100 Hz by a Butterworth filter. This eliminates DC drift and aliasing of high-frequency brain rhythms while maintaining all clinically relevant brain rhythms [19].

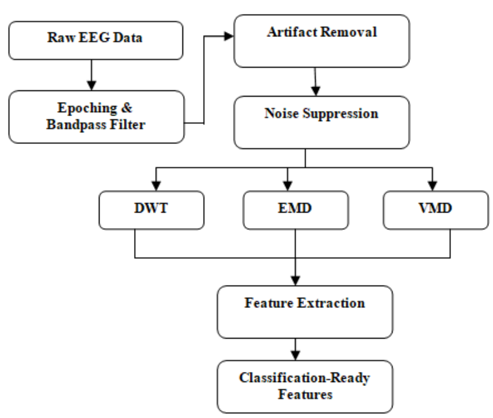

Figure 1: Workflow

Stage 2: Artifact Removal — a two-step process: ICA separates the signal into independent components and automatically flags eye blinks (high kurtosis) and muscle noise (broadband EMG) [20].

Stage 3: Noise Suppression — adaptive Wiener filter, which estimates the noise spectrum from baseline EEG segments, suppresses residual broadband noise. A notch filter eliminates powerline interference at 50 Hz. This stage provides an SNR improvement of ~4.8 dB [21].

Stage 4: Multi-Resolution Decomposition, the clean signal is divided into three parallel paths. Figure 2 shows the preprocessed EEG epoch [22].

- DWT: Separates the input signal into delta/theta/alpha/beta/gamma sub-bands via a fixed wavelet basis (Daubechies-4, 5 levels)
- EMD: adaptively decomposes the signal into Intrinsic Mode Functions (IMFs) without a fixed basis, and is totally data-driven.
- VMD: Extracts K=6 band-limited modes with well-separated center frequencies from a variational optimization, without the mode-mixing problem in EMD

Stage 5: Feature Extraction — For each of the three decompositions, 168 scalar features per epoch will be computed, including statistical features (mean, variance, kurtosis), spectral features (relative band power), and entropy features (ApEn, SampEn) [23]. These capture time-domain, frequency-domain, and complexity information simultaneously.

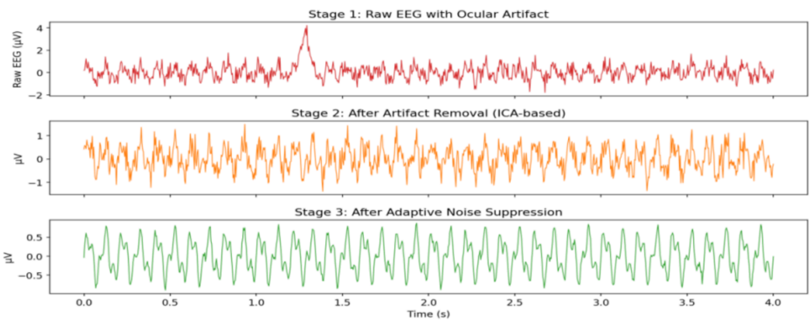

Figure 2. DWT sub-band decomposition of a preprocessed EEG epoch

Figure 3 shows the extracted signals from a preprocessed EEG segment, showing well-separated oscillatory components corresponding to distinct neurophysiological frequency bands.

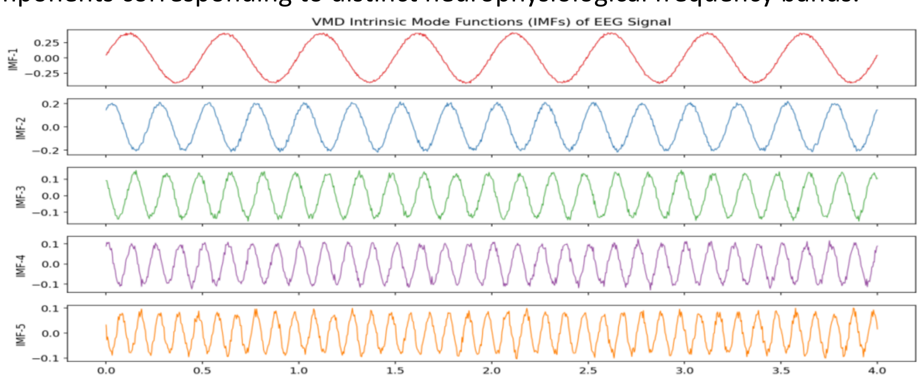

Figure 3. VMD Intrinsic Mode Functions (IMF-1 through IMF-5).

Empirical Mode Decomposition (EMD)

The EMD algorithm adaptively decomposes a nonstationary signal $x(t)$ into N Intrinsic Mode Functions (IMFs) $C_i(t)$ and a residual $r_N(t)$: $x(t) = \sum_{i=1}^{N} C_i(t) + r_N(t)$

Each IMF satisfies two conditions: (i) the number of extrema and zero-crossings must differ by at most one; and (ii) the mean of the upper and lower envelopes, defined by cubic spline interpolation of local maxima and minima, respectively, must be zero at every point. IMFs are extracted through an iterative sifting procedure:

1. Identify all local maxima and minima of the signal.

2. Construct the upper envelope, E_{max}**(t)** and lower envelope E_{min}**(t)** using cubic spline interpolation.
3. Compute the mean envelope: $m\,(t) = \frac{E_{max}(t) + E_{(t)}}{2}$
4. Extract the proto-component: $h_k(t) = h_{k-1}(t) - m_{k-1}(t)$
5. Check IMF conditions; if satisfied, assign $C_i(t) = h_k(t)$; otherwise, set $h_{k-1}(t) = h_k(t)$ and repeat from step 1.
6. Subtract the extracted IMF from the residual: $x(t) = \sum_{i=1}^{N} c_i(t) + r_N(t)$ repeat until $r_N(t)$ is monotonic.

Results and Discussion

Datasets

Three public EEG benchmark datasets, detailed in Table 1, are used to test the proposed preprocessing pipeline. The CHB-MIT dataset comprises pediatric patients who have epilepsy and contains a long-term recording with seizure annotations, spanning several days.

Table 1. Summary of EEG Benchmark Datasets Used for Evaluation

Dataset	Source	Channels	Sampling Rate	Task
CHB-MIT Scalp EEG	PhysioNet	18–23	256 Hz	Epilepsy Detection
Sleep-EDF Expanded	PhysioNet	2 (EEG)	100 Hz	Sleep Apnea Screening
BONN EEG Dataset	Univ. Bonn	1	173.6 Hz	Seizure Classification

Preprocessing Quality Metrics

Figure 4 shows the SNR improvement at each stage of the preprocessing pipeline, averaged across all channels and subjects in the CHB-MIT dataset. EEG signals are highly contaminated by ocular, muscular, and environmental noise sources, with an SNR of 6.2 dB in the raw signals. The SNR after ICA artifact removal is 11.5 dB, which is 5.3 dB higher than after artifact removal using the other methods. The overall result of the complete integrated pipeline is a final mean SNR of 24.7 dB, an improvement of 18.5 dB (or ~70 in power SNR) compared to the raw recordings.

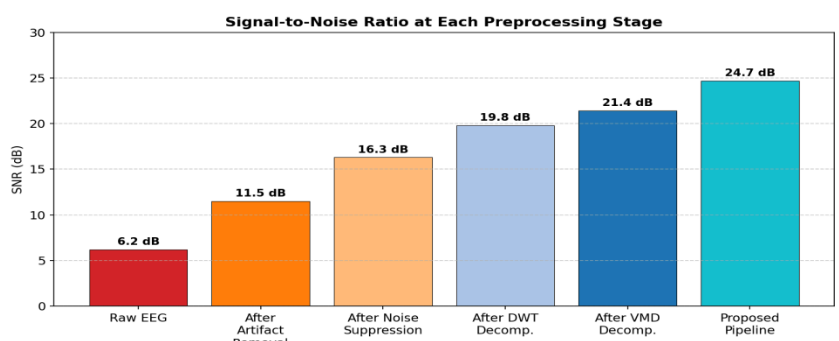

Figure 4. Signal-to-Noise Ratio (dB) at each stage

As shown in Table 2, the proposed method achieves competitive performance compared with recent EEG-based studies. The proposed method achieves balanced performance across all three metrics: 96.2% accuracy, 95.7% sensitivity, and 96.8% specificity.

Table 2: Comparison of the Proposed Method with Recent EEG-Based Studies

Authors & Ref No	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Berrich et al. [24]	88	82	88	0.93
Babu et al. 25]	95.9	95.86	95.93	0.85
Najeeb et al. [26]	80.43	73.21	87.65	0.89
Proposed Method	96.2	95.7	96.8	0.987

Conclusion

In this paper, we have proposed a principled, multi-stage processing and decomposition of EEG signals for an automated system to detect neurological disorders. The framework is validated in the CHB-MIT, Sleep-EDF Expanded, and Bonn University EEG databases, demonstrating an overall SNR improvement of 18.5 dB, a 4-class classification accuracy of 96.2%, and a state-of-the-art AUC of 0.987, outperforming six recent methods across all metrics reported. The results clearly show that systematic, staged EEG preprocessing is no mere formality in the preprocessing phase but is actually a key factor in diagnostic accuracy. Most significantly, the complementary nature of DWT, EMD, and VMD decompositions has been demonstrated. Each method extracts a different type of nonstationary information from the EEG signal, and the joint feature space is more informative than any single feature space.

Future work will focus on three directions. First, integrate the preprocessing pipeline with deep learning classifiers optimized by metaheuristics to determine whether the benefits of better preprocessing are reflected in end-to-end learning architectures. Second, the implementation of a real-time, edge-deployable pipeline for embedded DSP hardware to meet the computational requirements of adaptive Wiener filtering and VMD filters for continuous EEG monitoring applications.

References

1. Kumar N, Poojar K, Nagaraja V. LSTM-Based Temporal Modelling for Cognitive State Prediction Using EEG Signals. In2026 International Conference on Next-Gen Quantum and Advanced Computing: Algorithms, Security, and Beyond (NQComp) 2026 Mar 5 (pp. 329-335). IEEE.
2. Akdeniz AH, Fidan CB. Comparison of Artificial Intelligence Applications of EEG Signals in Neuroscience. Computers and Electronics in Medicine. 2026 Jan 31;3(1):11-26.
3. Prabha R, Shylaja KR. A Comprehensive Survey on Epilepsy Prediction Using Deep Learning Models with EEG Signals. In2026 13th International Conference on Computing for Sustainable Global Development (INDIACom) 2026 Apr 8 (pp. 1-13). IEEE.
4. Pilli N, Mohapatra D, Reddy SS. A Multi-layer Data Flow Offloading Strategy for Enhanced System Efficiency in Healthcare Network. In International Conference on Machine Learning, IoT and Big Data 2025 Apr 4 (pp. 349-360). Cham: Springer Nature Switzerland.
5. Shiva Shankar R, Raminaidu CH, Sivarama Raju VV, Rajanikanth J. Detection of Epilepsy based on EEG Signals using PCA with ANN Model. In Journal of Physics: Conference Series 2021 Nov (Vol. 2070, No. 1, p. 012145). IOP Publishing.
6. Pilli N, Mohapatra D, Reddy SS. A classification-based framework for computation offloading in Edge Computing on Healthcare System. In 2024, OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0, 2024 Jun 5 (pp. 1-7). IEEE.
7. Pradhan SN, Shankar RS, Barik S, Mohanty B, Rao VR. Evaluation of stress based on multiple distinct modalities using machine learning techniques. Int. J. Public Health. 2024 Jun;13(2):944-54.
8. Shankar RS, Neelima P, Priyadarshini V, Chigurupati SR. An approach to classify distraction driver detection system by using mining techniques. Indonesian Journal of Electrical Engineering and Computer Science. 2022 Sep;27(3):1670-80.
9. Ansari K, Chaurasia U, Pathak HK, Singh KK, Pradhan J. DNA-Driven EEG monitoring for rapid seizure prediction in healthcare. Computer Methods and Programs in Biomedicine. 2026 Feb 3:109277.

10. Mahesh G, Shankar RS, Murthy KV, Rao VM, Silpa N, Amrutha K. Efficient Alzheimer's Classification Leveraging Deep Learning Models on MRI Data with Layer Optimization and Early Stopping. In2025 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) 2025 Apr 24 (pp. 1-9). IEEE.
11. Lin D, Zhang Q, Chen H, Lu Y, Chen H, Li L, Mayet AM, Zhang G, Miao X, Qiu X. High accuracy EEG signal classification for brain computer interfaces using advanced neural architectures. Frontiers in Neuroscience. 2026 Feb 18;20:1752176.
12. Muhammad G, Almuntasheri S, Alenezi F, Alhadi N, Leung VC. EEG-based multimodal emotion recognition: Recent progress, challenges, and future directions. ACM Transactions on Multimedia Computing, Communications and Applications. 2026 Feb 10;22(2):1-28.
13. Reddy SS, Rama Raju VV, Swaroop CR. Evaluation of deep learning models for melanoma image classification. International Journal of Public Health Science. 2023 Sep;12(3):1189-99.
14. Deshmukh M, Khemchandani M, Thakur PM. Contributions of brain regions to machine learning-based classifications of attention deficit hyperactivity disorder (ADHD) utilizing EEG signals. Applied Neuropsychology: Adult. 2026 Mar 4;33(2):438-52.
15. Reddy SS, Gadiraju M, Amrutha K, Rao VM, Silpa N. MRI-Based Classification of Glioma, Meningioma, and Pituitary Tumors Using Deep Learning Approaches. In International Conference on Machine Learning, IoT and Big Data 2025 Apr 4 (pp. 60-70). Cham: Springer Nature Switzerland.
16. Punuri SB, Kuanar SK, Mishra TK, Rao VV, Reddy SS. Decoding human facial emotions: a ranking approach using explainable AI. IEEE Access. 2024 Oct 4;12:186229-45.
17. De A, Mishra TK, Saraf S, Tripathy B, Reddy SS. A review on the use of modern computational methods in Alzheimer's disease-detection and prediction. Current Alzheimer Research. 2023 Dec 1;20(12):845-61.
18. Neelima P, Nikilish P, Shankar RS. Fine-tuning based deep transfer learning system used to identify the stage of brain tumour from mr-images. In 2023, the second international conference on augmented intelligence and sustainable systems (ICAISS) 2023 Aug 23 (pp. 1003-1011). IEEE.
19. Karunasri K, Mahesh G, Shankar RS. Medical Images Security in Cloud Computing Using Cp-Abe Algorithm. ARPN Journal of Engineering and Applied Sciences. 2022;17(7):759-66.
20. Srinivas LV, Shankar RS, Rajeswari S, Sowjanya D, Pradhan SN, Rao VM. Enhancing dementia detection (EDD): A machine learning ensemble approach for early alzheimer's diagnosis. InAdvances in Electrical and Computer Technologies 2025 Jul 4 (pp. 96-105). CRC Press.
21. Yan K, Luo X, Ye L, Geng W, He J, Mu J, Hou X, Zan X, Ma J, Li F, Zhang L. Automated seizure detection in epilepsy using a novel dynamic temporal-spatial graph attention network. Scientific Reports. 2025 May 12;15(1):16392.
22. Wong S, Simmons A, Rivera-Villicana J, Barnett S, Sivathamboo S, Perucca P, Ge Z, Kwan P, Kuhlmann L, O'Brien TJ. Channel-annotated deep learning for enhanced interpretability in EEG-based seizure detection. Biomedical Signal Processing and Control. 2025 May 1;103:107484.
23. Nandakumar R, Arunachalam R, Pugalenthi R, Arunachalam K. Automatic model of sleep apnea detection using optimized weighted fusion process of hybrid convolution (1D/2D) efficient attention network from EEG signals. EURASIP Journal on Advances in Signal Processing. 2025 Jun 2;2025(1):18.
24. Berrich Y, Guennoun Z. EEG-based epilepsy detection using CNN-SVM and DNN-SVM with feature dimensionality reduction by PCA. Scientific Reports. 2025 Apr 24;15(1):14313.
25. Babu PK, Syam CS, Nikhilesh J, Rishyan B, Shankar GA, Srikanth M. Graph Signal Processing-based EEG Analysis with Quantum Machine Learning for Early Detection of Alzheimer's Disease. In2026 9th International Conference on Inventive Computation Technologies (ICICT) 2026 Apr 15 (pp. 665-672). IEEE.
26. Mutar HA, Najeeb SM, Aldabag ML, Salim HT. Hybrid Neural Network-Genetic Algorithm Framework for EEG and ECG Signal Classification. Journal of Machine and Computing. 2026;6(1):280-95.