

Article

EEG-Based Eye State Recognition Using VMD-CWT and DCAM

Ming-Tao Ge ¹, Yue-Hong Gong ^{2,3,*}, Xin Liu ², Wang-Yang Zhang ⁴ , Jun-Jie Wang ⁴ and Ling-Han Zhang ⁴¹ School of Information Engineering, Pingdingshan University, Pingdingshan 467000, China; 6508@pdsu.edu.cn² School of Artificial Intelligence, Pingdingshan University, Pingdingshan 467000, China; 20230005@pdsu.edu.cn³ Henan International Joint Laboratory for Multidimensional Topology and Carcinogenic Characteristics Analysis of Atmospheric Particulate Matter PM2.5, Pingdingshan 467000, China⁴ School of Malaga Engineering, Pingdingshan University, Pingdingshan 467000, China; 16696316228@163.com (J.-J.W.); linghanzhang0822@163.com (L.-H.Z.)

* Correspondence: 4392@pdsu.edu.cn; Tel.: +86-186-3817-2709

Abstract

Electroencephalography (EEG)-based eye state recognition often suffers from low accuracy due to noise interference in the collected signals. To address this issue, this study proposes a novel method combining cascaded variational mode decomposition (VMD-CWT) and a dynamic context-aware mask (DCAM). First, variational mode decomposition is applied to decompose EEG signals, and the Pearson correlation coefficient is used to identify and extract sensitive feature components. Subsequently, a multi-band adaptive weighting strategy based on continuous wavelet transform (CWT), together with feature cascading, is employed to generate time–frequency feature maps, achieving both noise suppression and precise extraction of key features. Furthermore, a frequency-adaptive pooling granularity mechanism is integrated into the traditional context-aware mask to construct the DCAM, enabling an adaptive attention mechanism tailored to the target EEG features. Finally, a DCAM network is built with DCAM as its core. Experimental results indicate that the proposed method achieves an accuracy of 95.67% on the EEG Eye State dataset, outperforming the best-performing baseline method by 8.34 percentage points. In addition, the proposed framework reduces the number of parameters and computational complexity, demonstrating its effectiveness and efficiency for EEG-based eye state recognition.

Keywords: EEG-based eye state recognition; cascaded variational mode decomposition; dynamic context-aware mask; continuous wavelet transform

MSC: 68T10



Academic Editor: Jüri Majak

Received: 25 May 2026

Revised: 12 July 2026

Accepted: 20 July 2026

Published: 30 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The human brain serves as the central hub for regulating physical and cognitive activity, and its functional state can be effectively monitored using electroencephalography (EEG). EEG signals provide a direct reflection of neural activity, encompassing ionic exchanges and metabolic processes. Key signal characteristics, such as frequency and amplitude, offer critical insights into brain health and have been widely used for diagnosing neurological disorders, including epilepsy, brain tumors, head trauma, sleep disorders, Parkinson’s disease, and depression [1–3]. Beyond clinical applications, EEG signal analysis has increasingly been applied to eye state recognition in real-world scenarios, such as pain detection, driver mental state assessment, and drowsiness detection [4–8].

EEG signals are inherently dynamic, varying significantly across different conscious states. Brain–computer interfaces facilitate the collection of these signals, enabling in-depth analysis of cerebral activity and contributing to the prevention and treatment of mental illnesses. EEG analysis generally involves three key components: signal filtering, feature extraction, and classification. In feature extraction, techniques such as differential entropy, Fourier transform [9–11], power spectral density [12], autoregressive models [13], and common spatial patterns [14] are commonly employed. For EEG-based recognition and classification, prior studies have primarily relied on stochastic models [15], machine learning approaches [16–19], and deep learning techniques [20–23]. Despite extensive efforts in disease prediction using EEG, eye state recognition has received comparatively less attention. For example, Dharmendra Gurve et al. [24] integrated EEG spectrograms and nonnegative matrix factorization-derived features into a neural network. Wang et al. [25] proposed an incremental attribute learning method to progressively incorporate features during network training. Abolfazl Saghafi and colleagues [26,27] combined common spatial patterns with multiple empirical mode decomposition and machine learning classifiers, such as support vector machines, for eye state recognition. Similarly, Ashima Khosla et al. [28] employed genetic algorithms for feature extraction and systematically evaluated machine learning algorithms in the context of eye state recognition.

However, several challenges remain in EEG-based eye state recognition. First, EEG signals exhibit significant dynamic variations across frequency bands and are susceptible to interference from other bioelectrical signals, making the balance between feature extraction and noise suppression particularly challenging. Second, achieving high recognition accuracy often requires complex network architectures with large numbers of network parameters, resulting in high computational overhead that limits practical deployment in resource-constrained scenarios.

To address these challenges, various solutions have been proposed. Variational mode decomposition (VMD) has been applied in EEG feature extraction [18,22], where a filter bank is used to separate signal components across distinct frequency bands. While effective, traditional VMD uses uniform filtering parameters across the entire frequency range, limiting its adaptability to frequency-specific features. For EEG recognition, references [20,23,29] introduce the context-aware mask (CAM), which models feature dependencies to implement a lightweight attention mechanism, achieving favorable recognition performance with low computational complexity. Despite these advantages, CAM is constrained by its fixed receptive field structure in deep modeling, limiting its capacity to capture the dynamic variations of EEG signals across different frequency bands.

In response to these challenges, this study proposes an EEG-based eye state recognition method leveraging cascaded variational mode decomposition (VMD-CWT) and a dynamic context-aware mask (DCAM). The primary contributions are reflected in the following aspects. In the signal preprocessing and feature extraction stage, EEG signals are first decomposed using VMD, and sensitive feature components are identified through the Pearson correlation coefficient. Subsequently, a multi-band adaptive weighting strategy based on continuous wavelet transform (CWT), combined with a feature cascade mechanism, is employed to construct time–frequency feature maps, enabling effective noise suppression and precise extraction of discriminative EEG features. To further enhance EEG state recognition, this study introduces a frequency-adaptive pooling granularity mechanism into the traditional CAM to develop the DCAM, which achieves adaptive attention modeling of EEG target features. A DCAM-driven recognition network is then constructed to achieve adaptive attention learning for EEG signals. Experimental results demonstrate that the proposed framework achieves superior recognition performance while maintaining low computational complexity.

2. Related Work

2.1. VMD

VMD is an adaptive and non-recursive signal decomposition technique that has demonstrated strong capability in processing non-stationary signals. Compared with traditional decomposition methods, VMD can effectively alleviate the non-stationarity inherent in time series that exhibit substantial complexity and pronounced nonlinearity [22]. Therefore, this study adopts VMD to process EEG signals. In the VMD framework, the input EEG signal is decomposed into K VMD modes; unlike intrinsic mode functions (IMFs) in empirical mode decomposition, VMD modes are obtained through a constrained variational optimization process and are not required to satisfy the strict IMF definition. Subsequently, exponential correction $e^{-j\omega_k t}$ is introduced to shift the frequency spectrum of each mode function to the corresponding baseband. A Gaussian smoothing index is then applied for demodulation, enabling estimation of the bandwidth associated with each mode. The resulting constrained variational problem is formulated as follows:

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta_t + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{k=1}^K u_k(t) = f \end{cases} \quad (1)$$

where $u_k = \{u_1, u_2, \dots, u_k\}$ denotes each mode function, $\omega_k = \{\omega_1, \omega_2, \dots, \omega_k\}$ represents the corresponding center frequency, and $*$ denotes the convolution operation. To solve the constrained variational problem, an augmented Lagrangian function is introduced to transform it into an unconstrained optimization problem. Specifically, the formulation incorporates both a quadratic penalty term and a Lagrange multiplier, which can be expressed as follows:

$$\begin{aligned} L(\{u_k(t)\}, \{\omega_k\}, \lambda(t)) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| x(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), x(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (2)$$

where α denotes the quadratic penalty factor, and ∂_t represents the partial derivative of t with respect to $\mu_k(t)$.

2.2. CAM

To improve the recognition performance under low computational overhead, CAM introduces an attention mechanism that models dependencies among EEG feature representations, enabling adaptive emphasis on informative regions [23,29]. The detailed mechanism of CAM is described as follows.

(1) CAM first extracts spatial information from the input features via a local convolutional layer to generate feature y .

$$y = \text{Conv}(x) \quad (3)$$

where the $\text{Conv}()$ serves as the convolution operation.

(2) The resulting feature map is further processed by a second convolutional layer to generate the mask [29].

$$\text{mask} = \sigma(\text{ContextConv}(y)) \quad (4)$$

where σ denotes the Sigmoid activation function, and ContextConv is the contextual convolution operation.

(3) The mask is then multiplied element-wise with the feature, whereby the enhanced feature y' is obtained as follows:

$$y' = y \odot \text{mask} \quad (5)$$

where operator $\odot$ denotes element-wise multiplication.

Due to its fixed receptive field architecture, CAM exhibits limited flexibility when processing complex EEG signals and is therefore unable to adaptively emphasize discriminative features across varying frequency bands. To address this limitation, the proposed DCAM improves attention precision by introducing dynamic receptive field adaptation and integrating multi-scale contextual information.

3. The Proposed Method

3.1. VMD-CWT

VMD is an adaptive signal decomposition method grounded in rigorous mathematical reasoning and has been widely applied in non-stationary signal processing. However, when applied to EEG analysis, VMD mainly serves as a mode decomposition method and does not directly generate a two-dimensional time–frequency representation. Therefore, an additional time–frequency analysis step is required to characterize the temporal and spectral variations of the decomposed mode components. In contrast, CWT provides a time–frequency analysis framework capable of adaptively adjusting the window size according to signal frequency characteristics, enabling effective characterization of both high- and low-frequency features [29–31]. The reason for selecting CWT is that EEG signals are nonlinear and non-stationary, and their discriminative information for eye state recognition is distributed across both temporal and spectral domains. Compared with fixed-window time–frequency analysis methods, CWT provides a multi-resolution representation, in which short windows are used for high-frequency components and long windows are used for low-frequency components. This property enables CWT to better capture transient changes and rhythm-related variations in EEG signals. Although CWT has inherent resolution trade-offs, such as relatively poor time resolution at low frequencies and relatively poor frequency resolution at high frequencies, this limitation is acceptable in the present task because the proposed framework focuses on extracting robust time–frequency patterns from selected sensitive VMD mode components rather than estimating exact instantaneous frequency changes. Moreover, the subsequent correlation-weighted reconstruction and DCAM-based feature enhancement further help emphasize discriminative time–frequency regions for eye state classification. To address the limitation that VMD mainly provides one-dimensional mode decomposition but does not directly generate discriminative time–frequency representations for EEG eye state recognition, this study integrates VMD, Pearson-correlation-based sensitive mode selection, and CWT-based time–frequency reconstruction into the proposed VMD-CWT framework. The proposed method can process EEG signals by adaptively weighting the high- and low-frequency components obtained from CWT and combining them through feature concatenation to generate time–frequency feature maps. The detailed procedure of the VMD-CWT decomposition method is described as follows:

(1) EEG signal decomposition: The collected EEG signal is first decomposed using VMD, yielding a set of VMD modes.

(2) Extraction of sensitive components: To identify the most informative components, the Pearson correlation coefficient between each VMD modes and the original EEG signal is computed. The specific formula is given as follows:

$$\rho(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (6)$$

$$\bar{\rho} = \frac{1}{K-1} \rho(u, f) \quad (7)$$

where u_k denotes each mode function, f represents EEG signal, and $\bar{\rho}$ is the Pearson correlation coefficient. After calculating the Pearson correlation coefficient between each VMD mode component and the original EEG signal, the components were ranked in descending order according to the absolute value of their correlation coefficients. In this study, the three mode components with the highest values were selected as sensitive components for subsequent CWT-based time–frequency reconstruction. This criterion was adopted to retain the components most strongly associated with the original EEG signal while reducing the influence of weakly related or redundant components.

(3) Wavelet basis function: The effectiveness of CWT depends on the proper selection of a wavelet basis function. In the present work, the Morlet wavelet is chosen to serve this role.

$$\psi(t) = \pi^{-\frac{1}{4}} e^{j\omega_0 t} e^{-t^2/2} \quad (8)$$

(4) Translation and scale transformation: The translation and scale transformation applied to the wavelet basis function is defined as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad (9)$$

where a is the scale factor controlling the lateral expansion of the wavelet basis function, and b is the time shift factor governing its temporal localization.

(5) CWT: The CWT of the EEG signal $f(t) \in L^2(R)$ is defined as

$$WT_\psi f(a, b) = \int_{-\infty}^{+\infty} f(t) \bar{\psi}_{a,b}(t) dt = a^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (10)$$

where $\psi(t)$ denotes the wavelet basis function, and $\bar{\psi}$ represents the complex conjugate of $\psi(t)$. The time–frequency representation of the EEG signal is obtained by aggregating the wavelet coefficients across different frequencies.

(6) Time–frequency feature reconstruction: After the sensitive VMD mode components are selected, CWT is applied to each selected component to obtain its corresponding time–frequency representation. Let $\mathcal{S}$ denote the set of selected sensitive mode components, and let ρ_k be the Pearson correlation coefficient between the k -th VMD mode component $u_k(t)$ and the original EEG signal $x(t)$. To ensure that the contribution of each selected component is comparable, the correlation coefficient is normalized as follows:

$$\eta_k = \frac{|\rho_k|}{\sum_{i \in \mathcal{S}} |\rho_i|}, \quad k \in \mathcal{S} \quad (11)$$

where η_k denotes the normalized reconstruction weight of the k -th sensitive mode component. The CWT coefficient matrix of the k -th selected component is calculated as

$$W_k(a, b) = \int_{-\infty}^{+\infty} u_k(t) \psi_{a,b}^*(t) dt \quad (12)$$

where a and b are the scale and translation parameters, respectively, and $\psi_{a,b}^*(t)$ denotes the complex conjugate of the scaled and shifted wavelet basis function. Then, the weighted

time–frequency feature map of the k -th sensitive component is obtained by applying the normalized Pearson weight to the magnitude of its CWT coefficients:

$$T_k(a, b) = \eta_k \cdot |W_k(a, b)| \quad (13)$$

Finally, the reconstructed time–frequency feature map is obtained by summing the weighted time–frequency feature maps of all selected sensitive components:

$$\begin{aligned} T_{\text{VMD-CWT}}(a, b) &= \sum_{k \in S} T_k(a, b) \\ &= \sum_{k \in S} (\eta_k \cdot |W_k(a, b)|) = \sum_{k \in S} \left(\frac{|\rho_k|}{\sum_{i \in S} |\rho_i|} \cdot \left| \int_{-\infty}^{+\infty} u_k(t) \psi_{a,b}^*(t) dt \right| \right) \end{aligned} \quad (14)$$

where $T_{\text{VMD-CWT}}(a, b)$ denotes the final reconstructed time–frequency feature map. Through this weighted summation process, the sensitive components with stronger correlations to the original EEG signal contribute more to the reconstructed time–frequency representation, while weakly related components are suppressed.

3.2. DCAM

Although the traditional CAM enhances feature representation capability, its fixed receptive field limits adaptability to the dynamic variations of EEG signals across different frequency bands [23,29]. To overcome this limitation, this study proposes a DCAM network based on CAM. It adaptively adjusts the receptive field according to the frequency characteristics of time–frequency features, enabling more precise attention allocation. The proposed DCAM consists of two parallel branches: a local branch and a dynamic context modeling branch (Figure 1). The dynamic context modeling branch further includes three core modules: frequency analysis, context pooling, and mask generation. Through this specialized architectural design, DCAM achieves effective EEG recognition while maintaining low computational complexity.

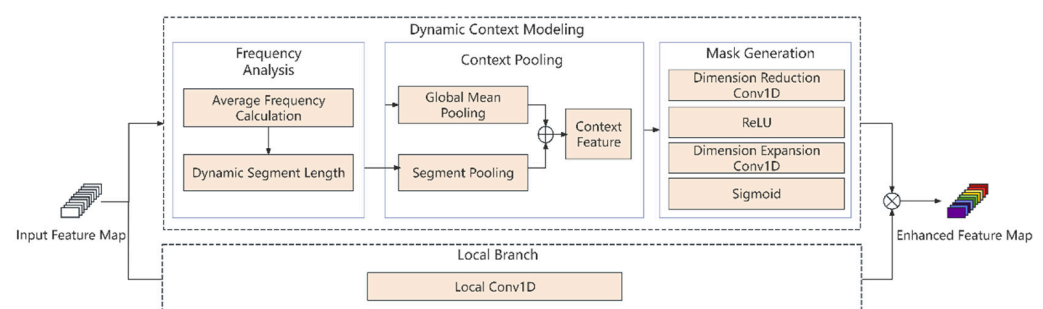


Figure 1. Structural framework of DCAM.

(1) In the local branch, DCAM adopts a local convolutional layer to extract the spatial information y_{local} from the input features:

$$y_{\text{local}} = \text{Conv}_{\text{local}}(x) \quad (15)$$

where x is the input feature map, and $\text{Conv}_{\text{local}}$ represents the local convolution operation.

(2) Within the dynamic context modeling branch, the frequency characteristics of the EEG signal are initially estimated by the frequency analysis module:

$$\omega = 2\pi f \quad (16)$$

$$\text{seg}_{\text{count}} = \beta \times \omega / (2\pi) \quad (17)$$

where ω , f , β , and seg_{count} denote the angular frequency, average frequency, basic segmentation coefficient, and number of segment pooling regions, respectively. This mechanism enables the network to adaptively ascertain the context window size.

(3) To capture both global average pooling and dynamic segment pooling, the context pooling module generates multi-scale contextual representations:

$$C = \text{Concat}(\text{GlobalMean}(X) + \text{SegPool}(X, seg_{count})) \quad (18)$$

where GlobalMean refers to global average pooling, and SegPool represents segment pooling based on the dynamic segment length. By combining these two operations, the network can simultaneously preserve global information and local context.

(4) In the mask generation stage, a reduce–activate–expand structure is employed to generate the adaptive mask:

$$M = \sigma(W_2 \times \text{ReLU}(W_1 \times C)) \quad (19)$$

where W_1 and W_2 denote the dimensionality reduction and expansion convolution, respectively. σ represents the Sigmoid function, and ReLU is the ReLU activation function. This design effectively reduces the number of parameters while preserving representational capacity. The final output feature y' is obtained as

$$y' = y_{local} \odot M \quad (20)$$

Compared with CAM, DCAM introduces frequency-aware adaptive context modeling, enabling improved characterization of the time–frequency properties of EEG signals. Furthermore, DCAM integrates multi-scale contextual information and adopts a reduce–expand mask generation strategy, which effectively decreases the parameter count while preserving feature representation capability. Consequently, DCAM consistently outperforms conventional methods in EEG state recognition tasks.

3.3. Eye State Recognition Model Based on VMD-CWT and DCAM

Severe noise interference in collected EEG signals degrades recognition performance, while traditional methods have difficulty extracting discriminative features across different frequency bands. To address this issue, this study proposes an eye state recognition framework that integrates VMD-CWT and DCAM. Specifically, the proposed framework combines VMD and CWT to adaptively extract time–frequency features from both high- and low-frequency bands. Furthermore, a frequency-adaptive pooling granularity mechanism is incorporated into the traditional context-aware masking architecture to construct the DCAM network, achieving EEG state recognition with improved recognition accuracy and reduced computational complexity. The overall structure diagram of the proposed method is illustrated in Figure 2. The recognition network used in the proposed framework is referred to as DCAMNet, which consists of a lightweight front-end convolutional network (LFCN) and a DCAM-based backbone network. The LFCN is designed to extract local shallow features from the reconstructed time–frequency maps with low computational cost, while the DCAM-based backbone further enhances the discriminative feature representation through adaptive context modeling. The proposed framework mainly includes EEG signal decomposition, sensitive mode selection, time–frequency feature reconstruction, DCAM-based feature enhancement, and final eye state classification.

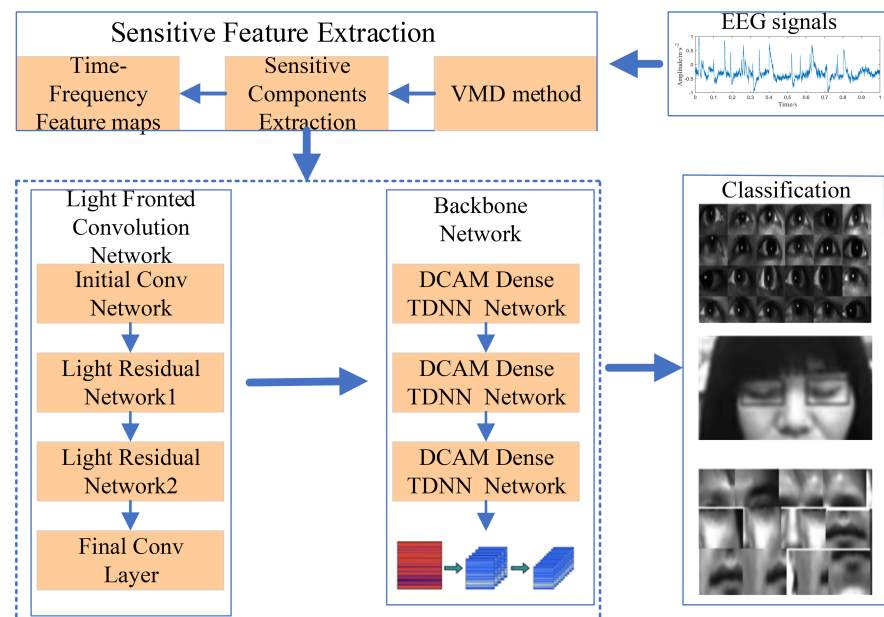


Figure 2. Overall framework of the proposed VMD-CWT-DCAM method.

4. Experimental Results and Discussion

4.1. Experimental Setup

The dataset used in this study came from the EEG Eye State dataset from the UCI machine learning warehouse (<http://archive.ics.uci.edu/ml/datasets/EEG+Eye+State> accessed on 9 June 2013). Data were collected by researchers at Baden-Württemberg State University (Germany) using an Emotiv EEG headset (San Francisco, CA, USA). The recording duration was 117 s, yielding 14,980 samples per channel. Eye states were tracked by a camera during EEG measurement. Each frame was manually labeled as either “1” for closed eyes or “0” for open eyes. Figure 3 presents an example signal fragment, sub1, which is labeled as 0 (open eyes) after outlier removal. The EEG signals were recorded from 14 channels of the Emotiv headset, including AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. All 14 channels were used in this study to preserve the spatial information of EEG signals for eye state recognition.

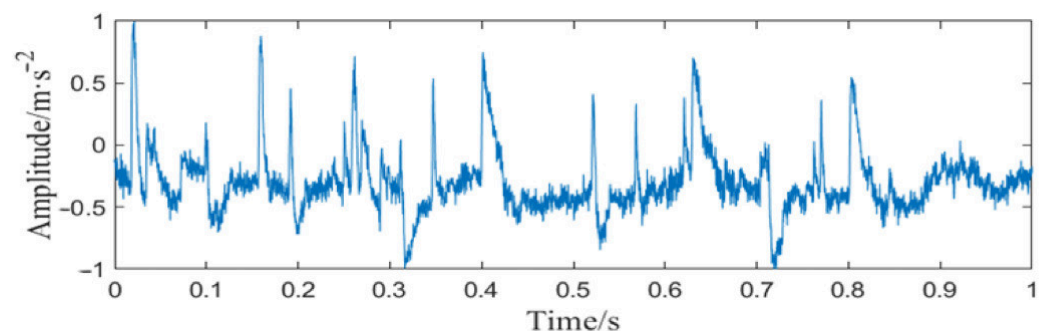


Figure 3. Original signal fragment.

4.2. Feature Extraction

In this study, the number of VMD modes was set to ($K = 5$). This setting was determined according to the prior frequency-band characteristics of EEG signals and the subsequent correlation-based mode selection strategy. EEG signals are commonly analyzed in several major rhythm bands, including delta, theta, alpha, beta, and gamma bands. Therefore, setting ($K = 5$) allows the VMD procedure to provide a moderate number of

candidate mode components corresponding to the main EEG rhythm characteristics, while avoiding excessive over-decomposition. After decomposition, the Pearson correlation coefficient was further used to select the sensitive modes that were more strongly related to the original EEG signal. In this way, the influence of weakly correlated or redundant modes could be reduced. The corresponding time-domain waveforms are shown in Figure 4. Furthermore, Pearson correlation coefficients between each VMD modes and the original signal were computed for correlation analysis (Figure 5).

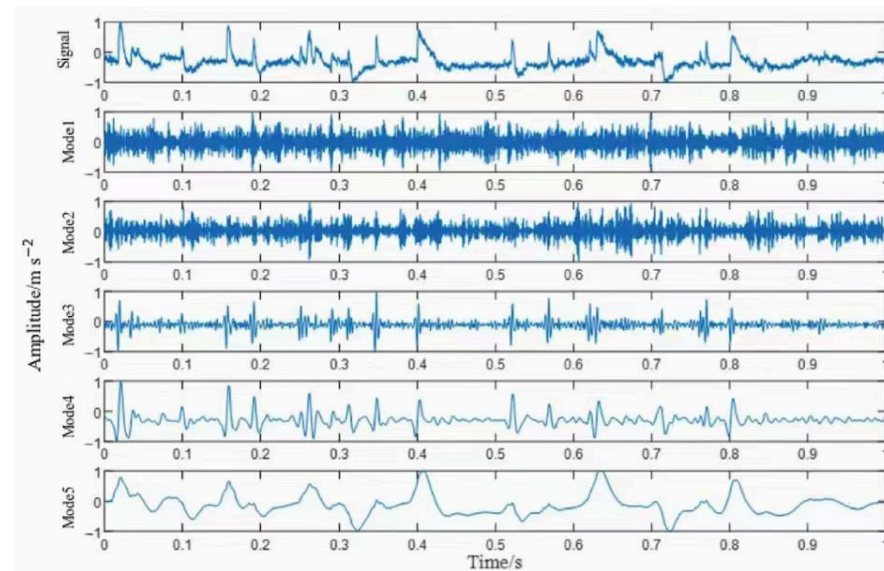


Figure 4. VMD decomposition results.

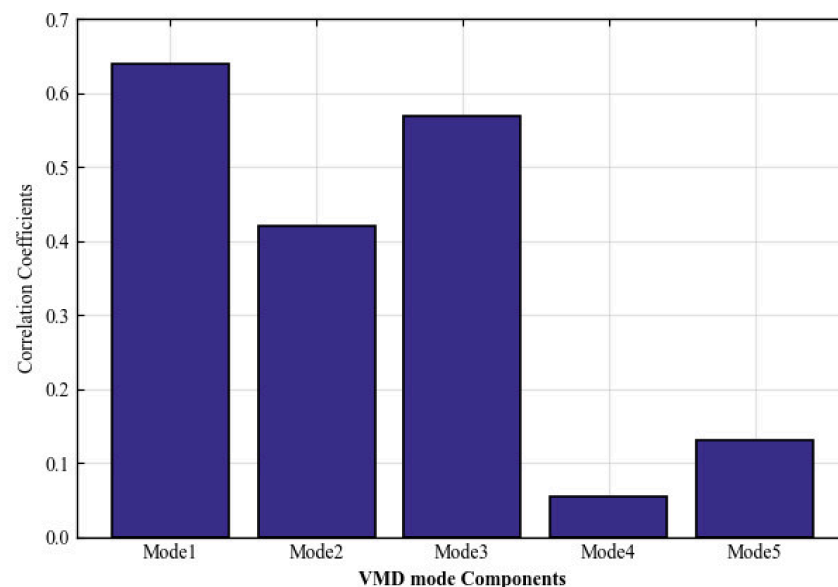


Figure 5. Correlation coefficients of the VMD-decomposed components.

As shown in Figure 5, VMD Mode 1, VMD Mode 2, and VMD Mode 3 exhibited the three highest absolute Pearson correlation coefficients with the original EEG signal. According to the sensitive component selection criterion defined in Section 3.1, the VMD mode components were ranked in descending order, and the top three components were retained. Therefore, VMD Mode 1, VMD Mode 2, and VMD Mode 3 were selected as the sensitive components for subsequent CWT-based time–frequency reconstruction.

The selected sensitive components VMD Mode 1, VMD Mode 2, and VMD Mode 3 were further processed using CWT to obtain their respective time–frequency representations. These representations were then reconstructed using correlation coefficients to generate the final reconstructed time–frequency feature map. For comparison, CWT was applied to the VMD-derived mode components and the VMD-CWT-selected sensitive mode components, respectively, to obtain the corresponding time–frequency representations, as shown in Figure 6.

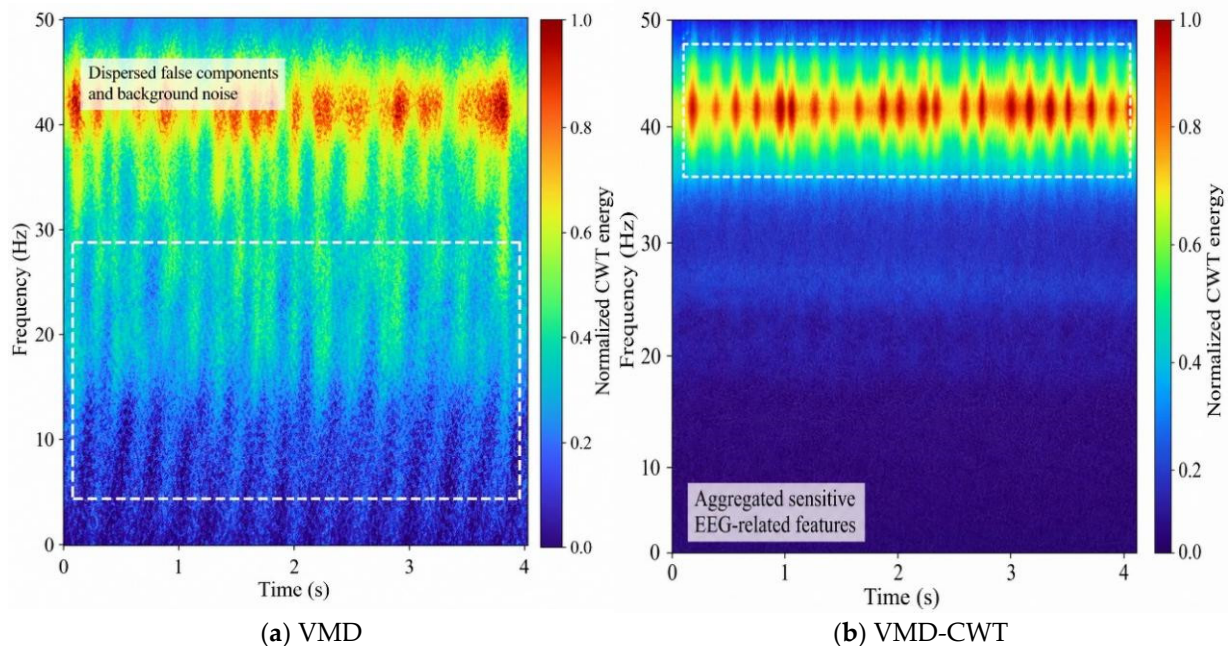


Figure 6. CWT-based time–frequency representations obtained from VMD-derived modes and VMD-CWT-selected sensitive modes.

As shown in Figure 6, the decomposition process of the VMD method introduced spurious components, leading to dispersed time–frequency feature distributions in the CWT domain, which made it difficult to clearly distinguish different eye states. In contrast, by utilizing correlation coefficients to filter out spurious components and retain EEG-related sensitive components, the VMD-CWT framework followed by CWT enabled adaptive extraction of time–frequency features across different frequency bands. Consequently, the reconstructed time–frequency feature map exhibited more concentrated feature representations.

4.3. Model Evaluation

This section evaluated the performance of the proposed method on the eye state recognition task and compared it with existing methods. The recognition metrics of the proposed method on the EEG Eye State dataset were analyzed, as shown in Table 1.

Table 1. Performance evaluation results.

Dataset	Class	Precision (%)	Recall (%)	F1-Score	Support
EEG	open	95.97	95.33	0.9565	750
	closed	95.36	96.00	0.9568	750
	weighted avg	95.67	95.67	0.9567	1500

As shown in Table 1, for the open class of the EEG dataset, the proposed method achieved a precision of 95.97%, a recall of 95.33%, and an F1-score of 0.9565. For the closed

class, the precision, recall, and F1-score were 95.36%, 96.00%, and 0.9568, respectively. Overall, the weighted average precision and recall both reached 95.67%, with an F1-score of 0.9567. These results indicate that the proposed method achieves consistently balanced and high recognition performance across both classes. To further evaluate the effectiveness of the proposed method, several representative baseline models and recent EEG classification models were selected for comparison, including AGNet [29], ResNet-18 [30], ConvNeXt [31], Separable-CNN [32], GAT-CNN [33], and Dual-Gate Chebyshev Linkage Network [34] (Figure 7). AGNet was selected as a representative attention-guided neural network, which is suitable for evaluating the effectiveness of attention-based feature enhancement in EEG eye state recognition tasks. ResNet-18 was selected as a representative residual convolutional neural network, which has been widely used in image-based classification tasks. ConvNeXt was selected as a modern convolutional architecture to evaluate the performance of the proposed lightweight DCAM-based model against recent CNN-based designs. Separable-CNN was included as a lightweight convolutional baseline, since depthwise separable convolution can reduce computational cost while maintaining effective feature extraction capability. GAT-CNN was selected as a representative graph-attention-based convolutional model, which is effective for modeling spatial dependencies among EEG channels. Dual-Gate Chebyshev Linkage Network was included as a recent graph-based EEG classification model, which enhances EEG feature representation by capturing complex relationships among brain regions through Chebyshev graph convolution and dual-gate information interaction. For the same-dataset baselines, the models were evaluated on the EEG Eye State dataset using the same data partition and evaluation metrics.

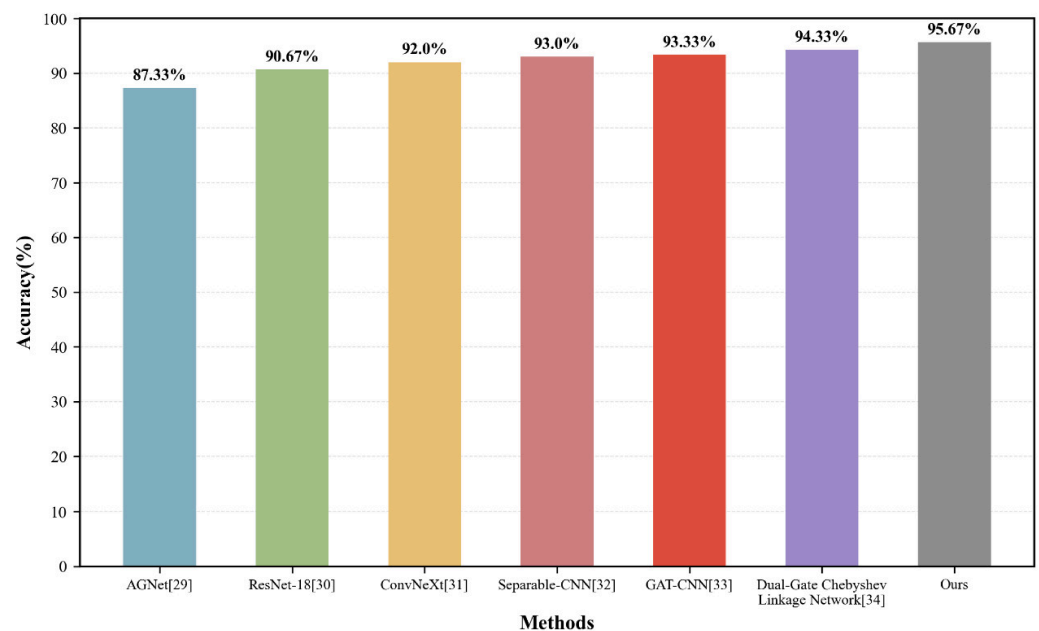


Figure 7. Performance comparison of different models.

As shown in Figure 7, the proposed method demonstrated superior recognition performance, achieving an accuracy of 95.67% on the EEG dataset and outperforming existing methods by more than 8.34%. These results confirm the effectiveness of the proposed method for EEG-based eye state recognition.

t-SNE is a nonlinear dimensionality reduction technique that projects high-dimensional features into low-dimensional space, enabling the visualization of feature distributions. To reduce the subjectivity of visual inspection, the t-SNE distributions were further analyzed in terms of intra-class compactness and inter-class separability. Specifically, the cluster

separation ratio was used to describe the relationship between the inter-class distance and the intra-class distance:

$$R = \frac{D_{\text{between}}}{D_{\text{within}}} \quad (21)$$

where D_{between} denotes the Euclidean distance between the centers of the open-eye and closed-eye clusters in the t-SNE space, and D_{within} denotes the average distance between samples and their corresponding class centers. A larger value indicates better feature separability. To further validate the feature learning capability of the proposed method, EEG Eye State dataset were employed to perform t-SNE visualization of the features from the last layer, with the results presented in Figure 8.

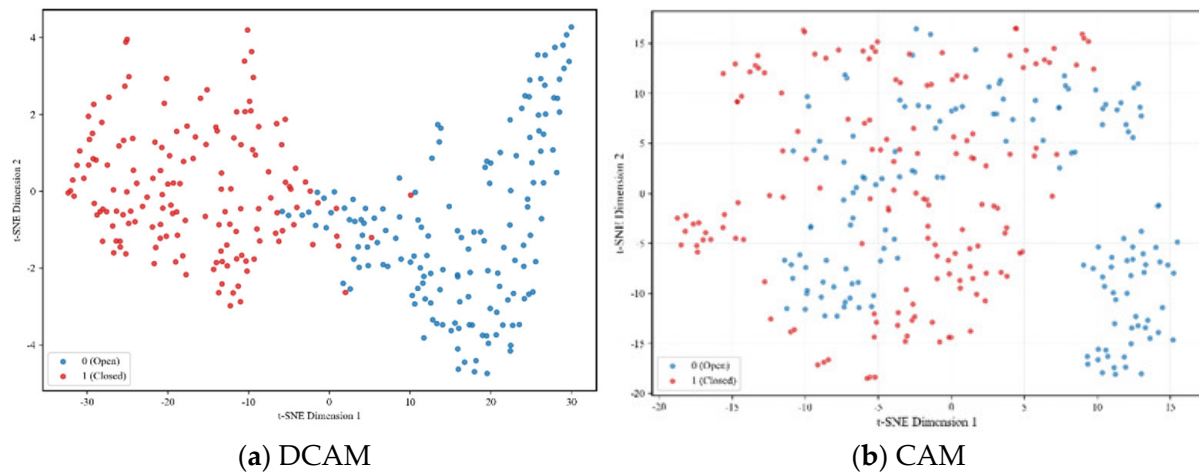


Figure 8. t-SNE visualization.

As shown in Figure 8, samples from both datasets formed relatively distinct feature clusters, indicating that the proposed method possessed strong feature discrimination capability. DCAM showed a larger cluster separation ratio than CAM, indicating that the proposed DCAM module produced more compact intra-class distributions and clearer inter-class separation. The proposed method generated highly discriminative feature representations, enabling distinct separation of samples from different categories in the feature space.

4.4. Recognition and Classification

To evaluate the effectiveness of each component in the proposed framework, a series of ablation experiments were conducted. During the experiments, different components were replaced or removed to quantitatively analyze their contribution to the overall performance. LFCN denotes the lightweight front-end convolutional network. The experimental results for different component combinations are summarized in Table 2.

Table 2. Comparison of ablation experiment results.

Methods	Feature	Modules	Accuracy (%)	F1-Score	Params (M)	FLOPs (G)
Ours	VMD-CWT	LFCN + DCAM	95.67	0.9567	0.24	0.08
1	VMD-CWT	LFCN + CAM	95.00	0.9500	0.29	0.15
2	VMD	LFCN + DCAM	94.33	0.9434	0.21	0.09
3	VMD	LFCN + CAM	92.67	0.9267	0.37	0.14
4	EMD	LFCN + DCAM	91.33	0.9134	0.55	0.13
5	EMD	LFCN + CAM	90.00	0.9000	0.68	0.16

As shown in Table 2, under the same feature extraction method, the proposed method (VMD-CWT + LFCN + DCAM) achieved accuracy improvements of 1.34% and 5.67% compared to the methods using VMD and EMD, respectively. Correspondingly, the F1-scores increased by 0.0133 and 0.0567, demonstrating the effectiveness of the VMD-CWT framework in extracting discriminative EEG features. From the perspective of the attention mechanism, the proposed method outperformed the method using CAM by 0.67% in recognition accuracy, while reducing the parameter count by 0.05 M and computational cost by 0.07 G. In addition, under the EMD-based feature extraction setting, EMD + LFCN + DCAM achieved an accuracy of 91.33% and an F1-score of 0.9134, outperforming EMD + LFCN + CAM by 1.33 percentage points in accuracy and 0.0134 in F1-score. Meanwhile, the parameter count and FLOPs were reduced from 0.68 M to 0.55 M and from 0.16 G to 0.13 G, respectively. This improvement can be attributed to the dynamic feature enhancement capability of DCAM, which can adaptively adjust the receptive field according to the frequency characteristics of the input signal. To further evaluate the classification capability of different methods, receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC) values were calculated (Figure 9).

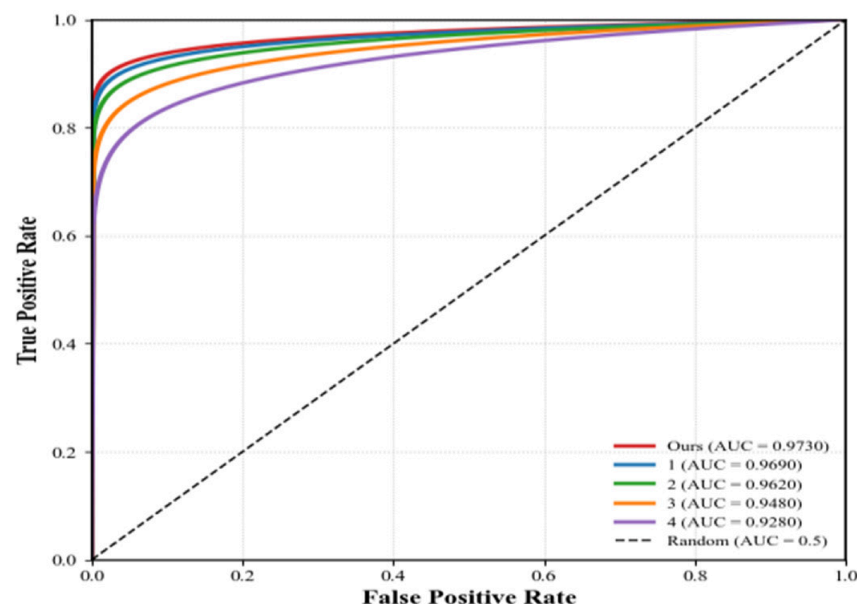


Figure 9. Comparison of ROC curves for different methods.

As shown in Figure 9, the proposed method achieved the highest AUC value of 0.9730. Moreover, its ROC curve consistently remained above those of the competing methods, particularly in the low false positive rate region, indicating superior classification performance and robustness. To further evaluate the stability of the proposed framework, ten random splits of the test set were performed for each method, and the accuracy distributions across different categories were statistically analyzed, with the results presented in Figure 10.

As shown in Figure 10, compared with the other four methods, the proposed method more effectively analyzed EEG signals and accurately extracted discriminative features related to different eye states, which were subsequently input into the DCAM network for training and testing. The experimental results demonstrate that the proposed method achieves the smallest performance fluctuation and the highest median accuracy among all methods, further validating its superior stability and robustness.

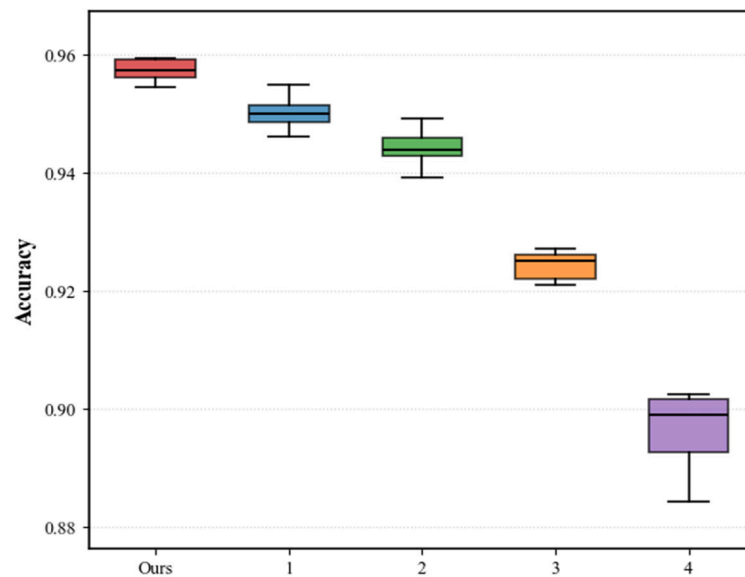


Figure 10. Accuracy comparison of different methods.

5. Conclusions

Severe noise interference in EEG signals degrades recognition performance, while traditional methods often struggle to effectively extract discriminative features across different frequency bands. To address these challenges, this study proposes an eye state recognition method based on VMD-CWT and DCAM. The proposed method combines VMD and CWT to adaptively extract time–frequency features from both high- and low-frequency bands. Furthermore, a frequency-adaptive pooling granularity mechanism is introduced into the traditional context-aware masking network to construct the DCAM network, enabling adaptive attention modeling for EEG state recognition. Finally, the collected datasets are tested and validated, from which the following conclusions were drawn:

- (1) The proposed VMD-CWT method can effectively decompose EEG signals and achieve noise suppression and discriminative feature extraction through adaptive multi-band weighting and feature cascading, thereby generating informative time–frequency feature maps across different frequency bands.
- (2) By integrating VMD-CWT with the proposed DCAM network, the eye state recognition framework improves recognition accuracy while maintaining low computational complexity through a dynamic context-aware network.

Author Contributions: Conceptualization, M.-T.G.; Methodology, J.-J.W.; Software, X.L.; Validation, Y.-H.G.; Formal analysis, Y.-H.G., X.L. and L.-H.Z.; Resources, J.-J.W. and L.-H.Z.; Data curation, Y.-H.G. and X.L.; Writing—original draft, M.-T.G., Y.-H.G., X.L. and W.-Y.Z.; Writing—review & editing, M.-T.G., Y.-H.G., W.-Y.Z., J.-J.W. and L.-H.Z.; Visualization, W.-Y.Z.; Supervision, W.-Y.Z.; Project administration, Y.-H.G. and L.-H.Z.; Funding acquisition, M.-T.G. and Y.-H.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Henan Province Science and Technology Research Project (Grant No. 242102241061 and No. 262102210077), Doctoral Research Start-up Fund of Pagdanganan University (Grant No. PXY-BSQD-2023048 and No. PXY-BSQD-2023018), and the Key Scientific Research Project in Colleges and Universities of Henan Province in China (No. 24A520033).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Li, X.; Zhang, Y.; Tiwari, P.; Song, D.; Hu, B.; Yang, M.; Zhao, Z.; Kumar, N.; Marttinen, P. EEG based emotion recognition: A tutorial and review. *ACM Comput. Surv.* **2022**, *55*, 1–57. [\[CrossRef\]](#)
- Nilashi, M.; Abumalloh, R.A.; Ahmadi, H.; Samad, S.; Alghamdi, A.; Alrizq, M.; Alyami, S.; Nayer, F.K. Electroencephalography (EEG) eye state classification using learning vector quantization and bagged trees. *Heliyon* **2023**, *9*, e15258. [\[CrossRef\]](#) [\[PubMed\]](#)
- Riaz, F.; Hassan, A.; Rehman, S.; Niazi, I.K.; Dremstrup, K. EMD-based temporal and spectral features for the classification of EEG signals using supervised learning. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2015**, *24*, 28–35. [\[CrossRef\]](#) [\[PubMed\]](#)
- Abo-Zahhad, M.; Ahmed, S.M.; Abbas, S.N. A new multi-level approach to EEG based human authentication using eye blinking. *Pattern Recognit. Lett.* **2016**, *82*, 216–225. [\[CrossRef\]](#)
- Elshewey, A.M.; Alhussan, A.A.; Khafaga, D.S.; Elkenawy, E.-S.M.; Tarek, Z. EEG-based optimization of eye state classification using modified-BER metaheuristic algorithm. *Sci. Rep.* **2024**, *14*, 24489. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mondal, S.; Nag, A. A computational eye state classification model using EEG signal based on data mining techniques: Comparative analysis. *Phys. Eng. Sci. Med.* **2025**, *48*, 1761–1774. [\[CrossRef\]](#) [\[PubMed\]](#)
- Oh, S.L.; Hagiwara, Y.; Raghavendra, U.; Yuvaraj, R.; Arunkumar, N.; Murugappan, M.; Acharya, U.R. A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Comput. Appl.* **2020**, *32*, 10927–10933.
- Wu, W.; Zhang, Y.; Jiang, J.; Lucas, M.V.; Fonzo, G.A.; Rolle, C.E.; Cooper, C.; Chin-Fatt, C.; Krepel, N.; Cornelissen, C.A. An electroencephalographic signature predicts antidepressant response in major depression. *Nat. Biotechnol.* **2020**, *38*, 439–447. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lu, N.; Li, T.; Ren, X.; Miao, H. A deep learning scheme for motor imagery classification based on restricted Boltzmann machines. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *25*, 566–576. [\[CrossRef\]](#) [\[PubMed\]](#)
- Peng, P.; Song, Y.; Yang, L.; Wei, H. Seizure prediction in EEG signals using STFT and domain adaptation. *Front. Neurosci.* **2022**, *15*, 1880. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gao, Q.; Omran, A.H.; Baghersad, Y.; Mohammadi, O.; Alkhafaji, M.A.; Al-Azzawi, A.K.J.; Al-Khafaji, S.H.; Emami, N.; Toghraie, D.; Golkar, M.J. Electroencephalogram signal classification based on Fourier transform and Pattern Recognition Network for epilepsy diagnosis. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106479. [\[CrossRef\]](#)
- Manshouri, N.; Maleki, M.; Kayikcioglu, T. EEG-based stereoscopic research of the PSD differences in pre and post 2D&3D movies watching. *Biomed. Signal Process. Control* **2020**, *55*, 101642. [\[CrossRef\]](#)
- Ding, Z.; Ouyang, G.; Chen, H.; Li, X. Closed-loop transcranial magnetic stimulation of real-time EEG based on the AR mode method. *Biomed. Phys. Eng. Express* **2020**, *6*, 035010. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sun, J.; Wei, M.; Luo, N.; Li, Z.; Wang, H. Euler common spatial patterns for EEG classification. *Med. Biol. Eng. Comput.* **2022**, *60*, 753–767. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rajwal, S.; Aggarwal, S. Convolutional neural network-based EEG signal analysis: A systematic review. *Arch. Comput. Methods Eng.* **2023**, *30*, 3585–3615. [\[CrossRef\]](#)
- Samal, P.; Hashmi, M.F. Role of machine learning and deep learning techniques in EEG-based BCI emotion recognition system: A review. *Artif. Intell. Rev.* **2024**, *57*, 50. [\[CrossRef\]](#)
- Subasi, A. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst. Appl.* **2007**, *32*, 1084–1093. [\[CrossRef\]](#)
- He, H.; Liu, Q.; Xu, C.; Lu, Y.; Huai, Y.; Li, Y. GVICA: A Multi-Channel EEG Hierarchical Noise Reduction Framework Based on GWO Dynamically Optimized VMD-ICA Fusion. *IEEE Trans. Biomed. Eng.* **2026**. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fraïwan, L.; Lweesy, K.; Khasawneh, N.; Wenz, H.; Dickhaus, H. Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. *Comput. Methods Programs Biomed.* **2012**, *108*, 10–19. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fu, R.; Hou, Q.; Wang, S.; Wang, L.; Chen, J.; Wen, G. Effectiveness of adaptive attention-based network for situation awareness recognition. *IEEE Sens. J.* **2024**, *24*, 20092–20102. [\[CrossRef\]](#)
- Han, C.-H.; Choi, G.-Y.; Hwang, H.-J. Deep Convolutional Neural Network Based Eye States Classification Using Ear-EEG. *Expert Syst. Appl.* **2022**, *192*, 116443. [\[CrossRef\]](#)
- Pandey, P.; Seeja, K.R. Subject independent emotion recognition from EEG using VMD and deep learning. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 1730–1738. [\[CrossRef\]](#)
- Du, J.; Chen, W.; Vong, C.M.; Liu, P.; Wang, T. Context-CAM: Context-level weight-based CAM with sequential denoising to generate high-quality class activation maps. *IEEE Trans. Image Process.* **2025**, *34*, 3431–3446. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gurve, D.; Krishnan, S. Deep learning of EEG time-frequency representations for identifying eye states. *Adv. Data Sci. Adapt. Anal.* **2018**, *10*, 1840006. [\[CrossRef\]](#)
- Wang, T.; Guan, S.-U.; Man, K.L.; Ting, T. EEG eye state identification using incremental attribute learning with time-series classification. *Math. Probl. Eng.* **2014**, *2014*, 365101. [\[CrossRef\]](#)

26. Saghafi, A.; Tsokos, C.P.; Goudarzi, M.; Farhidzadeh, H. Random eye state change detection in real-time using EEG signals. *Expert Syst. Appl.* **2017**, *72*, 42–48. [[CrossRef](#)]
27. Saghafi, A.; Tsokos, C.P.; Farhidzadeh, H. Common spatial pattern method for real-time eye state identification by using electroencephalogram signals. *IET Signal Process.* **2017**, *11*, 936–941. [[CrossRef](#)]
28. Khosla, A.; Khandnor, P.; Chand, T. A novel method for EEG based automated eyes state classification using recurrence plots and machine learning approach. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6912. [[CrossRef](#)]
29. Xiao, Q.; Li, Y.; Zhai, X.; Jiang, W.; Jin, Y.; Yuan, K.; Shi, W. A novel noise-robust and lightweight underwater acoustic target recognition method based on BSCQT and DCAM. *Signal Process.* **2025**, *241*, 110386. [[CrossRef](#)]
30. Mohanto, R.; Shamim-Al-Mamun, M.; Sufi, F.B.; Ali, S.; Alahmadi, T.J.; Moni, M.A.; Islam, M.B. CWT-based transfer learning model with optimal channel selection for the detection of MDD using EEG signals. *Knowl.-Based Syst.* **2026**, *336*, 115331. [[CrossRef](#)]
31. Maddirala, A.K.; Veluvolu, K.C. ICA with CWT and k-means for eye-blink artifact removal from fewer channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 1361–1373. [[CrossRef](#)] [[PubMed](#)]
32. Feng, W.; Wang, X.; Xie, J.; Liu, W.; Qiao, Y.; Liu, G. Real-time eeg-based driver drowsiness detection based on convolutional neural network with gumbel-softmax trick. *IEEE Sens. J.* **2024**, *25*, 1860–1871.
33. Chen, J.; Cui, Y.; Wei, C.; Polat, K.; Alenezi, F. Driver fatigue detection using EEG-based graph attention convolutional neural networks: An end-to-end learning approach with mutual information-driven connectivity. *Appl. Soft Comput.* **2025**, *186*, 114097.
34. Li, Q.; Zhang, T.; Chen, C.L.P.; Zhang, X.; Hu, B. DGC-Link: Dual-Gate Chebyshev Linkage Network on EEG Emotion Recognition. *IEEE Trans. Affect. Comput.* **2025**, *16*, 3499–3511. [[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.