frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 09 January 2025
DOI 10.3389/fnhum.2024.1481493

OPEN ACCESS

EDITED BY
Kashif Saleem,
King Saud University, Saudi Arabia

REVIEWED BY
Ke Liu,
Chongqing University of Posts and Telecommunications, China
Vacius Jusas,
Kaunas University of Technology, Lithuania

*CORRESPONDENCE
Jin Feng
✉ 910950113@qq.com
Chengliang Jiang
✉ jiangcl_0312@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 16 August 2024
ACCEPTED 12 December 2024
PUBLISHED 09 January 2025

CITATION
He J, Huang Z, Li Y, Shi J, Chen Y, Jiang C and Feng J (2025) Single-channel attention classification algorithm based on robust Kalman filtering and norm-constrained ELM. *Front. Hum. Neurosci.* 18:1481493.
doi: 10.3389/fnhum.2024.1481493

Single-channel attention classification algorithm based on robust Kalman filtering and norm-constrained ELM

Jing He[1†], Zijun Huang[2†], Yunde Li[3], Jiangfeng Shi[4], Yehang Chen[3], Chengliang Jiang[3]* and Jin Feng[5]*

[1]School of Management, Guilin University of Aerospace Technology, Guilin, China, [2]School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, China, [3]Biomedical and Artificial Intelligence Laboratory, Guilin University of Aerospace Technology, Guilin, China, [4]School of Automation Science and Engineering, South China University of Technology, Guangzhou, China, [5]Student Affairs Office, Guilin Normal College, Guilin, China

Introduction: Attention classification based on EEG signals is crucial for brain-computer interface (BCI) applications. However, noise interference and real-time signal fluctuations hinder accuracy, especially in portable single-channel devices. This study proposes a robust Kalman filtering method combined with a norm-constrained extreme learning machine (ELM) to address these challenges.

Methods: The proposed method integrates Discrete Wavelet Transformation (DWT) and Independent Component Analysis (ICA) for noise removal, followed by a robust Kalman filter enhanced with convex optimization to preserve critical EEG components. The norm-constrained ELM employs L1/L2 regularization to improve generalization and classification performance. Experimental data were collected using a Schulte Grid paradigm and TGAM sensors, along with publicly available datasets for validation.

Results: The robust Kalman filter demonstrated superior denoising performance, achieving an average AUC of 0.8167 and a maximum AUC of 0.8678 on self-collected datasets, and an average AUC of 0.8344 with a maximum of 0.8950 on public datasets. The method outperformed traditional Kalman filtering, LMS adaptive filtering, and TGAM's eSense algorithm in both noise reduction and attention classification accuracy.

Discussion: The study highlights the effectiveness of combining advanced signal processing and machine learning techniques to improve the robustness and generalization of EEG-based attention classification. Limitations include the small sample size and limited demographic diversity, suggesting future research should expand participant groups and explore broader applications, such as mental health monitoring and neurofeedback.

KEYWORDS

brain-computer interfaces, attentional state, robust Kalman, convex optimization, norm-ELM

Introduction

Attention, characterized by distinct EEG activity patterns, refers to the ability to focus on specific stimuli (Wang et al., 2021). In everyday life, differences in attention levels affect performance in many areas. With advancements in neuroscience, EEG signals are increasingly utilized for analyzing attention states. For instance, Al-Nafjan and Aldayel (2022) applied EEG-based attention assessments in online learning environments, identifying changes in student

attentiveness. In the music domain, EEG signals detect attentiveness to different music, facilitating personalized recommendations (Zhang et al., 2021). In the driving field, attention signals assess driving fatigue. Kashihara (2017) classified brain activity from wavelet signals during attention tasks to identify road signs, developing a system to prevent car accidents. Similarly, Atilla and Alimardani (2021) monitored drivers' attention to detect fatigue. Du et al. (2017) applied physiological signals such as EEG and EOG to driving fatigue detection and proposed a multimodal method combining partial EEG and frontal EEG to enhance driving fatigue detection. In healthcare, EEG-based assessments effectively diagnose and treat attention deficit hyperactivity disorder (ADHD) in children (Miranda et al., 2020). Brain-Computer Interface (BCI) technology enables direct interaction with external devices using electroencephalogram (EEG) signals (Thomas, 2023).

This study focuses on the use of single-channel EEG for attention classification, where the quality of the signal is a key factor in achieving accurate results. Compared to multi-channel EEG systems, single-channel approaches offer a more convenient and faster solution, beneficial for expanding BCI applications.

Related work

Studies have shown that EEG signals vary under different attention tasks. Wang et al. (2021) extracted significant patterns from EEG signals and proposed an attention-based multi-scale convolutional neural network-dynamic graph convolutional network (AMCNN-DGCN) model, tested in fatigue driving environments. Changes in attention states are related to fluctuations in specific EEG frequency bands (Kamiński et al., 2012; Klimesch et al., 2012; Arns et al., 2013; Liu et al., 2013; Pitchford and Arnell, 2019). Increased Beta (β) wave activity indicates heightened attention. High-frequency bands in Delta (δ), Theta (θ), Alpha (α), and Beta (β) rhythms are commonly used as EEG features. Huang et al. (2016) used Granger causality bias coherence analysis for different psychological tasks, such as gaming and resting, to measure local and global network efficiency. They found that local efficiency in the β band was higher during tasks and lower during rest, while global efficiency showed the opposite trend. This distinction can differentiate attention tasks. Common EEG classification features include frequency band energy and power spectral density (Tang and Huang, 2020). These applications demonstrate the potential of BCI technology in studying and monitoring attention.

However, existing research faces challenges such as noise interference common in EEG signals, which affects the accuracy of attention detection (Jiang et al., 2019). Traditional methods for removing EEG artifacts, such as wavelet with higher-order statistics (Castellanos and Makarov, 2006), independent component analysis (ICA) (Jung et al., 2001), and principal component analysis (PCA) (Lagerlund et al., 1997), require multiple EEG channels or simultaneous electrooculography (EOG) signal acquisition, which is impractical for portable single-channel EEG devices. Recent advancements have addressed these limitations. Chintala and Thangaraj (2020) recorded horizontal and vertical electrooculogram (EOG) signals as reference signals and employed finite impulse response filters for processing. Somers et al. (2018) proposed an algorithm based on multichannel Wiener filtering, which substituted the pseudo-signal covariance matrix with a low-rank approximation based on generalized eigenvalue decomposition. However, when using filtering methods, the EEG is contaminated by eye signals, and the EOG reference signal is also contaminated by the EEG, leading to a bidirectional contamination issue. To overcome these drawbacks, Ge et al. (2018) proposed a method using higher-order statistical tensors through an Underdetermined Blind Source Separation (UBSS) model to separate artifacts from EEG signals. Saini et al. (2020) proposed a framework based on Variational Mode Decomposition (VMD) and turning-point counting, validated across three standard databases. Research indicates that most eye artifact removal techniques require the collection of multiple EEG channels or simultaneous collection of EOG signals as references, which is not suitable for portable single-channel EEG devices. Despite their effectiveness, these methods often lack real-time processing capabilities, leading to delays in practical applications. Traditional adaptive filtering methods, such as the Least Mean Squares (LMS) algorithm (Mowla et al., 2018) and Kalman filter (Hesar and Mohebbi, 2021), face challenges with real-time, nonlinear, and non-smooth signals. The Kalman filter's adaptive adjustment capabilities optimize filter parameters based on real-time data, making it suitable for dynamic EEG signals, but it may misclassify sparse noise as Gaussian noise.

This study proposes an innovative and robust method for accurately estimating attention levels in EEG signals, focusing on real-time noise reduction and signal stability. To address real-time noise reduction and signal fluctuations, we propose a novel approach combining a robust Kalman filter for noise removal with a norm-constrained Extreme Learning Machine (ELM) for handling dynamic attention changes. The robust Kalman algorithm preserves critical EEG components through signal preprocessing and optimized estimation. The norm-constrained ELM introduces nonlinear mapping for correct estimation, retaining relevant dynamic information and improving attention assessment accuracy and stability.

Methods

Data acquisition

Since attention signals are spontaneous EEG, a specific experimental paradigm is required to elicit EEG signals based on attention mechanisms. In this study, eight subjects participated in the experiment. Data collection was conducted using a 5 × 5 Schulte Grid, as shown in Figure 1.

The Schulte Grid is a classic psychological experimental tool commonly used to measure individual attention, concentration, and visual perception speed. By asking participants to find randomly arranged numbers in turn, the Schulte Grid tests participants' attention concentration, visual field range, and visual search ability. It is widely used in the field of attention assessment and is especially suitable for assessing individual attention fluctuations in a short period of time (Lu et al., 2022). Schulte Grids are based on a visual search task, which has been demonstrated by multiple studies to reliably induce attention-related brain wave changes, particularly attention states associated with alpha and beta waves. It is simple to operate and easy to implement, making it suitable as an experimental paradigm for EEG signal acquisition. Therefore, future research could consider using more complex experimental paradigms, such as the Continuous Performance Test (CPT) or the N-back task, which can more systematically induce sustained attention and test the stability of attention under increased cognitive load.

The experimental paradigm is as follows: To maintain focus, subjects were allowed to rest before the test began. During the test, subjects sequentially identified the positions of Arabic numerals from 1 to 25 on the Schulte Grid. The experiment included a 4-min stimulation period followed by a 1-min interval. Each subject completed three phases, totaling 15 min.

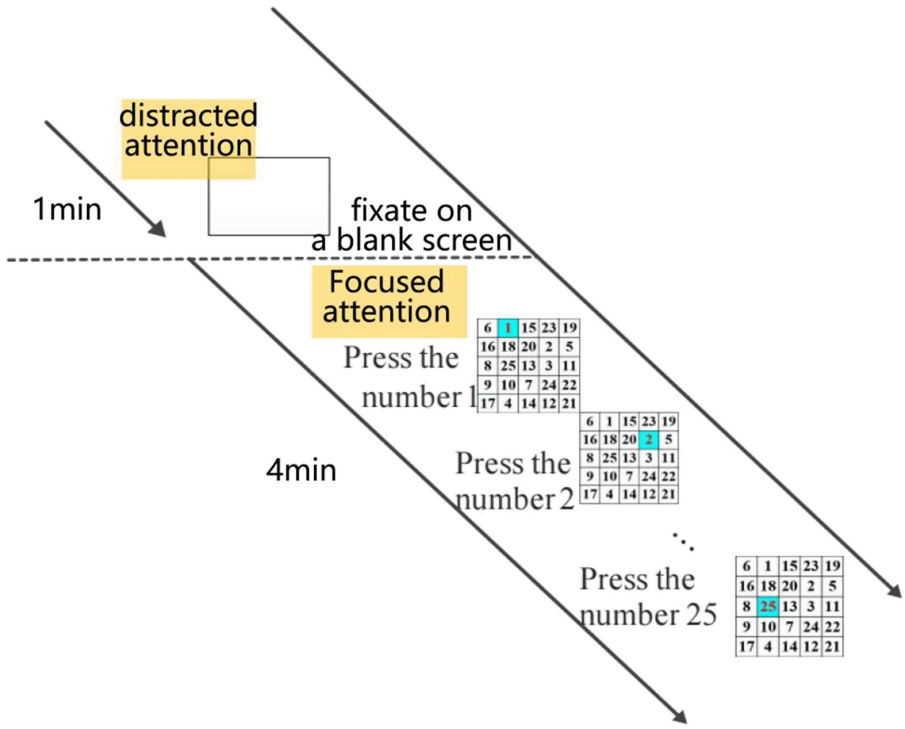


FIGURE 1
Experimental procedure for data collection.

The TGAM sensor was used with three contact points: EEG (EEG acquisition point), REF (reference point), and GND (ground point). The TGAM default port rate is 57,600, with approximately 513 packets per second, and raw EEG data is output at 512 Hz (Wu and Xie, 2020). The data collection process was conducted as follows: A personal computer (PC) served as the data processing center, receiving packets from the EEG module TGAM via a Bluetooth serial communication protocol. The PC paired with the TGAM biosensing chip module via Bluetooth to facilitate data transmission.

This paper also utilized experimental data for monitoring individual attention published by Acı et al. (2019). Each file contained data obtained from the EEG device during the experiment, with raw data sampled at 128 Hz. This study used only the AF3 data channel from 0–10 min (focused) and 10–20 min (unfocused).

Algorithm overview

Conventional EEG processing algorithms often overlook the dynamic interdependencies within EEG signal changes, leading to inaccurate recognition of attention states. Therefore, our study introduces a robust Kalman filter method for dynamic attention detection. The Kalman filter's ability to update signal estimates in real-time allows for the dynamic capture of changes in attentional mechanisms, addressing the shortcomings of traditional methods in terms of real-time performance and adaptability. The algorithm designed in our study is depicted in Figure 2.

The proposed algorithm includes three main improvements. First, the preprocessing stage combines Discrete Wavelet Transformation (DWT) and Independent Component Analysis (ICA) to better remove noise and retain dynamic useful information from EEG signals. Second, an improved Kalman filter algorithm integrates convex optimization techniques and Bayesian theory to enhance signal estimation, utilizing an augmented reality model and incorporating sparse noise control. Third, L1/L2 norm regularization in the Extreme Learning Machine (ELM) algorithm is chosen for its benefits in simplifying the model and improving generalization capability. Specifically, L1 regularization aids in feature selection, while L2 regularization helps in learning more effective features. By incorporating these regularization constraints, the ELM algorithm's generalization performance is significantly enhanced.

Denoising

This study employs the Independent Component Analysis (ICA) algorithm to decompose and evaluate EEG signals. The Sample Entropy algorithm (Liu et al., 2023) calculates the entropy of the decomposed signal components to determine the entropy information of each component. Removing lower-entropy components effectively eliminates noise and artifacts.

Robust Kalman target tracking model

In Equation 1, the system was first expressed as a convex optimization problem by constructing an attention state tracking model for single-channel EEG, introducing tuning parameters to balance data fit and noise sparsity. The optimal estimate is obtained by optimizing the objective function, dynamically updating the state estimate of the EEG by considering both prior and measurement information to improve the system's adaptive ability. Specifically, the attention level tracking model is expressed as:

$$\begin{aligned} x_t &= Ax_{t-1} + w_t \\ y_t &= Cx_t + v_t + z_t \end{aligned} \tag{1}$$

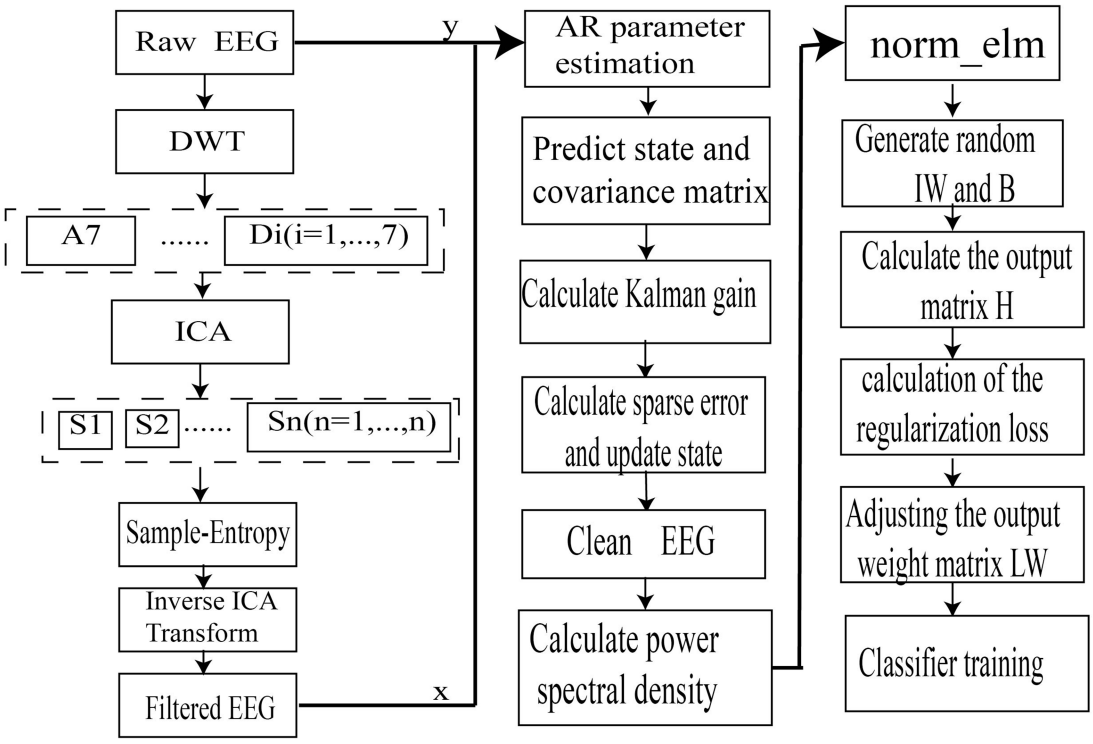


FIGURE 2
System flowchart.

where $x_t \in R^N$ denotes the clean EEG at time t, $y_t \in R^M$ is the original EEG signals at time t, A is the dynamic change matrix of the brain's attentional state, C maps state vectors to observation vectors, $w_t \in R^N$ denotes the noisy EEG, $v_t \in R^M$ represents measurement noise, and $z_t \in R^M$ denotes sparse noise, which can cause signal distortion, introduce artifacts, and reduce the signal-to-noise ratio.

The estimation of attention level involves predicting the next time point's attention level based on known original EEG signals x_{t+1}. To obtain the optimal probability estimate of $p(x_{t+1}|y_t)$, the maximum likelihood estimation theory (Sur and Candès, 2019) constructs the likelihood function $L(y_t|x_{t+1})$:

$$L(y_t|x_{t+1}) = \frac{p(x_{t+1})p(v_t)}{p(x_{t+1})} = \frac{1}{\left(\sqrt{2\pi}\right)^{\mathrm{m}}|V|^{1/2}}\exp\left(-\frac{1}{2}v_t^{\mathrm{T}}V^{-1}v_t\right) \tag{2}$$

where V denotes the covariance matrix of the measurement noise v_t, V^{-1} reflecting the statistical properties of the observation noise. According to the full probability formula, the maximum likelihood function $L(x_{t+1})$ is expressed as:

$$L(x_{t+1}) = \frac{1}{\left(\sqrt{2\pi}\right)^{\mathrm{n}}|\mathrm{V}|^{1/2}} \exp\left(\left(x-\hat{x}_{t|t-1}\right)^T V^{-1}\left(x-\hat{x}_{t|t-1}\right)\right) \tag{3}$$

A tuning parameter λ is introduced to balance data fitting and noise sparsity z_t. The conditional probability $p(x_{t+1}|y_t)$ is obtained by minimizing the power exponent in Equations 2, 3 to calculate the optimal estimate:

$$\min_{x,v_t,y_t} v_t^T V^{-1} v_t + \left(x-\hat{x}_{t|t-1}\right)^T \sum^{-1}\left(x-\hat{x}_{t|t-1}\right) + \lambda \| z_t \|_1 \quad s.t. \quad y_t = Cx + v_t + z_t \tag{4}$$

In Equation 4, from $v_t = y_t - Cx - z_t$, define the optimization objective function $F(x,z_t)$ as:

$$F(x,z_t) = \left(y_t - Cx - z_t\right)^T V^{-1}\left(y_t - Cx - z_t\right) + \left(x-\hat{x}_{t|t-1}\right)^T V^{-1} \left(x-\hat{x}_{t|t-1}\right) + \lambda \| z_t \|_1 \tag{5}$$

Applying the partial derivative to x in Equation 5, one obtains.

$$\frac{\partial F(x,z_t)}{\partial x} = 2C^T V^{-1} Cx - 2C^T V^{-1} y_t + 2C^T V^{-1} z_t + 2\sum^{-1} x - 2\sum^{-1}\hat{x}_t \tag{6}$$

The optimal state estimate is obtained by setting Equation 6 to zero:

$$x = \hat{x}_t + \sum C^T \left(V + C\sum C^T\right)^{-1} (e_t - z_t) \tag{7}$$

In Equation 7, $e_T = Z_T - c\hat{x}_t$, c is a constant; by updating the real-time estimate of non-stationary noise at each iteration, the optimal estimate $\hat{x}$ of attention level is obtained by weighting the observed and predicted values. The improved convex optimization Kalman expression is:

$$\begin{aligned} &\min(\mathrm{e_t} - z_t)\left(I - C\sum C^T\left(V + C\sum C^T\right)^{-1}\right)^T \\ &V^{-1}\left(I - C\sum C^T\left(V + C\sum C^T\right)^{-1}\right) \\ &+\sum C^T\left(V + C\sum C^T\right)^{-1T}\sum^{-1}\sum C^T \\ &\left(V + C\sum C^T\right)^{-1}(e_t - z_t) + \lambda \| z_t \|_1 \end{aligned} \tag{8}$$

In Equation 8, the sparse noise z_t is the only variable. Controlling z_t aiding in the detection and analysis of attention-related signal features while reducing noise effects.

Norm-constrained ELM classifier

In the second step, the ELM algorithm with L1/L2 norm constraints is used to correct the EEG estimates. The EEG state estimates serve as inputs to construct an ELM model with T hidden layer nodes, introducing regularization constraints to improve generalization. The specific steps are:

$$\begin{aligned} &\Phi(w_1, w_2, \ldots, w_T, b_1, b_2, \ldots, b_T, a_1, a_2, \ldots, a_T)\beta = \\ &\begin{bmatrix} g(w_1 a_1^{\ T} + b_1) & \ldots & g(w_T a_1^{\ T} + b_T) \\ \vdots & \ddots & \vdots \\ g(w_1 a_N^{\ T} + b_1) & \ldots & g(w_T a_N^{\ T} + b_T) \end{bmatrix}_{N\times T} * \beta = t \\ &\beta = [\beta_1, \beta_2, \ldots, \beta_T]^T, t = [t_1, t_2, \ldots, t_N]^T \end{aligned} \tag{9}$$

In Equation 9, where Φ is the hidden layer output matrix, β is the hidden layer weight matrix, and t is the training set target matrix. The output weights connecting the hidden and output layers β_j are solved by minimizing the squared difference algorithm with the objective function (Yıldırım and Özkale, 2023):

$$\hat{\beta} = \arg\min \frac{1}{2} \| \Phi\beta - t \|_2^2 = \Phi^{\dagger} t \tag{10}$$

In Equation 10, where: $\Phi^{\dagger}$ is the pseudo-inverse; $\hat{\beta}$ is the estimated value of the target output;

$$\hat{\beta} = \arg\min \frac{1}{2} \| \Phi\beta - t \|_2^2 + \lambda_1 \| \beta \|_1 + \lambda_2 \| \beta \|_2 \tag{11}$$

The values of λ_1 and λ_2, which are regularization hyperparameters constrained by L1 and L2 norms, are primarily used to enhance the generalization capability and stability of the model. Formula 11 is defined in the following matrix form:

$$\Phi^* = \sqrt{1+\lambda_2}\begin{pmatrix} \Phi \\ \sqrt{\lambda_2} I \end{pmatrix}, t^* = \begin{pmatrix} t \\ 0 \end{pmatrix} \tag{12}$$

Let $\gamma = \lambda_1 / \sqrt{\lambda_1 + \lambda_2}, \beta^* = \sqrt{1+\lambda_2}\beta$, and combined with formula 12, then formula 11 can be expressed as shown below.

$$\begin{aligned} \hat{\beta} &= \frac{1}{\sqrt{1+\lambda_2}}\hat{\beta}^* \\ &= \frac{1}{\sqrt{1+\lambda_2}} \arg\min \frac{1}{2} \| \Phi^*\beta^* - t^* \|_2^2 + \gamma \| \beta^* \|_1 \end{aligned} \tag{13}$$

Since λ_2 is a set constant, Equation 17 can be further expanded to get.

$$\begin{aligned} \hat{\beta} &= \frac{1}{2}\Phi^*\beta^* - t^{*T}\Phi^*\beta^* - t^* + \lambda\beta^{*T}\beta^* \\ &= \frac{1}{2}\beta^{*T}\Phi^{*T}\Phi^*\beta^* - \beta^{*T}\Phi^{*T}t^* - t^{*T}\Phi^*\beta^* \\ &+ t^{*T}t^* + \lambda\beta^{*T}\beta^* \end{aligned} \tag{14}$$

In Equation 14, where the γ, β^* parameter is a constraint function that mainly corrects the estimate of the attentional state.

Equation 13 takes the partial derivative of β^* and makes the derivative zero, resulting in a weight estimate that minimizes the number of targets as in Equation 15.

$$\hat{\beta}^* = \left(\Phi^{*T}\Phi^* + \lambda I\right)^{-1}\Phi^{*T}t^* \tag{15}$$

The final robust Kalman is corrected to obtain an expression for the estimate as in Equation 16:

$$output = \hat{\beta}^{*T}\Phi^{*T} \tag{16}$$

To evaluate the proposed algorithm's performance, we conducted experiments on both self-collected and public EEG datasets, focusing on key evaluation metrics such as AUC (Area Under the Curve) and ROC curve (Receiver Operating Characteristic Curve) to assess classification accuracy (Nahm, 2022). Below, we provide a detailed comparison of the results across different algorithms.

By definition, the AUC can be obtained by summing the area of each part under the ROC curve. The ROC curve is assumed to form $(\mathrm{x}_1 = 0, x_m = 1)$ by connecting points in sequence with coordinates $\{(\mathrm{x}_1, \mathrm{y}_1), (\mathrm{x}_2, \mathrm{y}_2), \ldots, (x_m, y_m)\}$, the AUC is calculated as:

$$\mathrm{AUC} = \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{m}-1}(\mathrm{x}_{\mathrm{i}+1} - \mathrm{x}_{\mathrm{i}})(\mathrm{y}_{\mathrm{i}} + \mathrm{y}_{\mathrm{i}+1}) \tag{17}$$

Result

Denoising experiment

In the experiment, a small amount of sparse noise was added to the original signal. The denoising capabilities of different algorithms were then compared to validate the effectiveness of the proposed algorithm. The experimental results showed that the robust Kalman model proposed in this study had the highest similarity with the original signal, indicating that it was least affected by sparse noise. The results of the denoising experiment are shown in Figure 3A. By calculating the mean square error (MSE) values (Papaioannou et al., 2021) for various comparison algorithms, it is evident that the proposed algorithm can accurately track the signal and provides superior filtering performance, as shown in the MSE results in Figure 3B. The mean squared error (MSE) values after applying the Kalman filter fluctuate around an average of 0.2. This is because, when sparse noise is present in the signal data, the traditional Kalman algorithm mistakenly identifies other noise signals as Gaussian noise, thereby affecting the denoising accuracy. LMS adaptive filtering exhibited larger MSE fluctuations due to over-filtering. In contrast, the MSE values after using the robust Kalman filter are lower than those of the traditional Kalman filter. This further indicates that the robust Kalman model proposed in this study is capable of removing noise while retaining as much effective information from the original signal as possible.

Assessing effectiveness

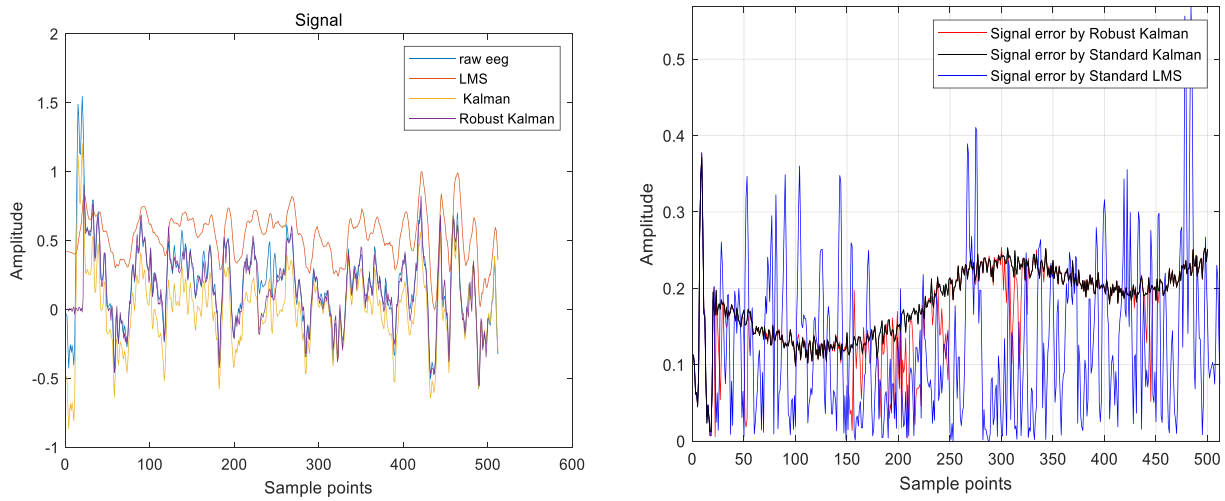
In the experiment, considering that Extreme Learning Machine (ELM) uses random weights as initial model weights, ELM may overfit these noisy points due to the presence of outliers in some data, thereby lacking an accurate fit to the true data situation. To further prevent overfitting in ELM, which can decrease the model's generalization performance, regularization terms (L1 or L2) were added to constrain the model parameters, reducing model complexity. Figure 4 compares the experimental results of the conventional ELM and the ELM algorithm based on L1/L2 regularization terms.

As shown in the experimental results in Figure 5, the data processed by the robust Kalman algorithm achieved an AUC of 0.8678 after applying the ELM algorithm with L1/L2 regularization. In comparison, the traditional Kalman filter algorithm had an AUC of 0.8168, the traditional adaptive filtering algorithm had an AUC of 0.6971, and the eSense algorithm in the TGAM module with the robust Kalman algorithm had an AUC of 0.7379.

If the ROC curve of one learner is completely "enclosed" by the curve of another learner, the performance of the latter is superior. If the ROC curve of the two learners crosses, it is difficult to claim the superiority of the two; in that case, the reasonable criterion is to compare the area under the ROC curve, AUC.

In this study, EEG signals were collected from eight participants, and different processing methods were applied to each participant's data. The experimental results are shown in Table 1. The results indicate that the proposed robust Kalman algorithm achieved an average AUC of 0.8167 on the test sets of the 8 participants, with the difference between the training and test set AUCs ranging from 0.02 to 0.04. In contrast, the traditional Kalman filter algorithm achieved an average AUC of 0.7417 on the test sets, the adaptive LMS algorithm achieved an average test set AUC of 0.6214, and the eSense algorithm (López-Ahumada et al., 2023) integrated into the TGAM module achieved an average test set AUC of 0.7011. This further demonstrates that the proposed robust Kalman algorithm has strong classification performance and effectively suppresses overfitting.

Additionally, to further validate the performance of the proposed algorithm on other datasets, this study used a publicly available datasets published by Acı et al. (2019) on Kaggle. This dataset contains data related to two states (focused and unfocused) of individuals. For this study, only the AF3 channel data from the public dataset was used. As shown in Table 2, the robust Kalman algorithm achieved an average AUC of 0.8344 on the test sets of three groups of subjects. This indicates that the proposed algorithm also performs well on other datasets.



(a) Signal Plot (b) Mean Squared Error

FIGURE 3
Results of denoising experiments. **(A)** Signal plot. **(B)** Mean squared error.

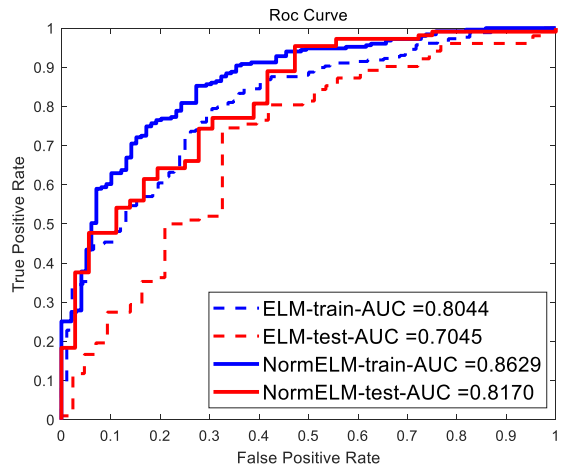


FIGURE 4
Comparison results of different ELM algorithms.

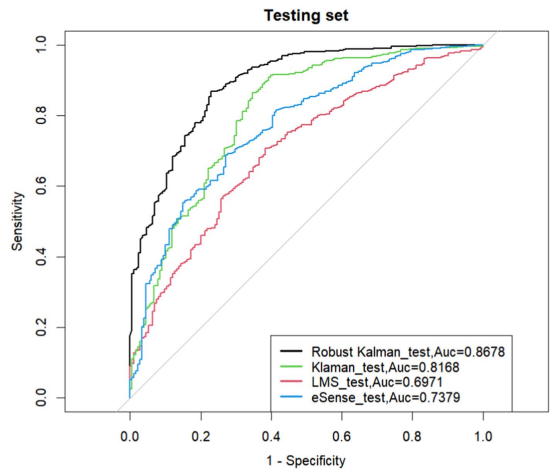


FIGURE 5
ROC curves of different algorithms.

The experimental results show that the robust Kalman algorithm produces an average AUC of 0.8167 and a maximum AUC of 0.8678 on the in-house collected datasets. It also showed good performance on the public datasets, with an average AUC of 0.8344 and a maximum AUC of 0.8950. These results highlight the effectiveness of the algorithm in accurately estimating levels of attention, significantly surpassing conventional methods such as LMS adaptive filtering and the conventional Kalman filter. The study demonstrates that the robust Kalman algorithm proposed significantly improves the accuracy of attention level estimation, showing superior classification accuracy and robust model generalization capability. These results offer strong evidence for the practical implementation of this novel approach in tasks related to monitoring attention states.

Discussion

This study addresses a critical gap in the current field of attention tracking, specifically in evaluating and monitoring attention levels using electroencephalogram (EEG) signals. We present an effective and practical method for estimating attention levels from EEG signals. Although various methods exist to assess attention, the issue of noise and artifacts in EEG signals limits the practical application of these techniques, especially in portable single-channel devices (Grosselin et al., 2019). This study proposes an enhanced Extreme Learning Machine (ELM) algorithm that integrates L1/L2 norm regularization with DWT and employs ICA for preprocessing. Additionally, convex optimization techniques are utilized to enhance the Kalman filter, thereby improving signal estimation and model generalization. By combining advanced signal processing and machine learning methods, this research aims to enhance the accuracy and robustness of attention level estimation.

Experimental results indicate that the proposed model not only improves signal integrity but also achieves real-time adaptability, making it highly suitable for practical Brain-Computer Interface (BCI) applications. Compared to previous studies, the proposed algorithm demonstrates substantial superiority in real-time processing and noise reduction. Traditional EEG artifact (Jiang et al., 2019) removal methods typically require multi-channel EEG signals or synchronous acquisition of Electrooculography (EOG) signals as references and are predominantly offline batch processing methods, leading to delays and a lack of real-time application

TABLE 1 AUC results for different algorithms on 8 subjects.

AUC	Robust Kalman train	Robust Kalman test	Kalman train	Kalman test	LMS train	LMS test	eSense train	eSense test
1	0.8629	0.8170	0.7528	0.7482	0.7054	0.6971	0.7289	0.6913
2	0.8127	0.8055	0.7825	0.7782	0.6606	0.6203	0.7816	0.7454
3	0.8832	0.8658	0.7789	0.7523	0.6100	0.5548	0.8139	0.7772
4	0.7455	0.7265	0.6517	0.6191	0.5336	0.5188	0.7506	0.6640
5	0.8767	0.8514	0.8172	0.7852	0.7204	0.7034	0.7293	0.6715
6	0.8172	0.8058	0.7667	0.7416	0.6236	0.6096	0.7796	0.6774
7	0.9065	0.8678	0.8425	0.8168	0.7241	0.6971	0.7492	0.7379
8	0.8110	0.7937	0.7062	0.6922	0.5891	0.5703	0.6666	0.6442
ave	0.8395	0.8167	0.7623	0.7417	0.6459	0.6214	0.7500	0.7011

TABLE 2 AUC results of different algorithms on public data sets.

AUC	Robust Kalman train	Robust Kalman test	Kalman train	Kalman test	LMS train	LMS test
1	0.8143	0.7777	0.7620	0.7460	0.6537	0.6283
2	0.9196	0.8950	07155	0.6791	0.6768	0.5845
3	0.8565	0.8305	0.6581	0.6389	0.6270	0.6136
ave	0.8635	0.8344	0.85473	0.6880	0.6525	0.6088

capabilities. Adaptive filtering methods and the Kalman filter face challenges in handling real-time, nonlinear, and non-smooth signals. The Kalman filter tends to treat sparse noise as Gaussian noise (Roth et al., 2017), whereas adaptive filtering methods have the problem of overfiltering in the filtering wave (Wang et al., 2023). In contrast, the robust Kalman algorithm introduced in this study can effectively mitigate noise interference through real-time adjustment of filter parameters while preserving the integrity of the original signal. These findings have profound implications for the application of BCI technology in fields such as education and healthcare, providing valuable insights for both theoretical advancement and practical application (Bonci et al., 2021). For example, the adaptive characteristics of the robust Kalman filter make it particularly suitable for real-time applications, ensuring minimal delay and high responsiveness, while effectively handling sparse noise to enhance signal integrity and retain essential signal components necessary for accurate attention assessment. Furthermore, the observed robust generalization capabilities across various datasets underscore its broad potential in a range of BCI applications, from educational tools to medical diagnostics.

The study acknowledges certain limitations, particularly the reliance on a relatively small participant group, which may affect the generalizability of the results. The small sample size, which does not encompass individuals from different age groups and health statuses, may limit the applicability of the results to a broader population (Vasileiou et al., 2018). Additionally, despite the high accuracy achieved by the proposed algorithm in signal processing, the inherent variability of EEG signals in real-world environments may significantly impact its performance (Souza and Naves, 2021). Future research should aim to validate the robustness and applicability of the algorithm using larger and more diverse datasets. Furthermore, future research can explore the findings from multiple perspectives, such as increasing the sample size and incorporating more diverse populations to enhance the robustness of the results. Additionally, applying the robust Kalman filter in other domains, such as mental health monitoring, cognitive training, and neurofeedback, could provide a deeper understanding of the method's capabilities. Through these additional studies, we hope to enrich the understanding of this field and offer practical recommendations.

Conclusion

The robust Kalman processing algorithm proposed in this paper demonstrates excellent performance in tracking and assessing attention levels. This algorithm utilizes convex optimization techniques on top of the Kalman filter to eliminate noise and obtain optimal estimates, followed by secondary corrections using extreme learning machines based on L1/L2 norms to improve system robustness and generalization. Experimental results show that the model achieves a maximum Test_AUC of 0.8167 and an average Test_AUC of 0.8678 on self-collected single-channel datasets, as well as a maximum Test_AUC of 0.8950 and an average Test_AUC of 0.8344 on the AF3 channel of public datasets. These results outperform traditional comparative algorithms, laying the foundation for EEG feature extraction and classification research.

The success of attention classification using single-channel EEG in this study highlights the importance of signal quality. Single-channel EEG provides a more efficient and portable solution without compromising accuracy. This makes it especially useful for real-time and practical applications in education, healthcare, and other fields.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://www.kaggle.com/datasets/inancigdem/eeg-data-for-mental-attention-state-detection/data.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

JH: Writing – review & editing, Data curation, Resources. ZH: Writing – original draft, Conceptualization, Software. YL: Writing – review & editing, Methodology, Formal analysis. JS: Writing – review & editing. YC: Writing – review & editing, Methodology, Validation. CJ: Writing – review & editing, Investigation, Visualization. JF: Writing – review & editing, Conceptualization, Project administration.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The study on postoperative progression prediction of non-small cell

lung cancer based on robust transfer learning in a federated learning framework (82460361), Science Foundation of Guilin University of Aerospace Technology (XJ22KT28), and GUAT Special Research Projection the Strategic Development of Distinctive Interdisciplinary Fields (TS2024231).

Acknowledgments

GPT-4o, the fourth-generation general-purpose pre-trained model released by OpenAI, was used for partial translation during the preparation of this manuscript.

Conflict of interest

The 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.

Publisher's note

References

Acı, C. I., Kaya, M., and Mishchenko, Y. (2019). Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods. *Expert Syst. Appl.* 134, 153–166. doi: 10.1016/j.eswa.2019.05.057

Al-Nafjan, A., and Aldayel, M. (2022). Predict students' attention in online learning using EEG data. *Sustain. For.* 14:655. doi: 10.3390/su14116553

Arns, M., Conners, C. K., and Kraemer, H. C. (2013). A decade of EEG theta/beta ratio research in ADHD: a meta-analysis. *J. Atten. Disord.* 17, 374–383. doi: 10.1177/1087054712460087

Atilla, F., and Alimardani, M. (2021). EEG-based classification of driver attention using convolutional neural network. 2021 IEEE 2nd international conference on human-machine systems, Magdeburg: IEEE.

Bonci, A., Fiori, S., Higashi, H., Tanaka, T., and Verdini, F. (2021). An introductory tutorial on brain–computer interfaces and their applications. *Electronics* 10:560. doi: 10.3390/electronics10050560

Castellanos, N. P., and Makarov, V. A. (2006). Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. *J. Neurosci. Methods* 158, 300–312. doi: 10.1016/j.jneumeth.2006.05.033

Chintala, S., and Thangaraj, J. (2020). Ocular artifact elimination from EEG signals using RVFF-RLS adaptive algorithm. 2020 National Conference on Communications (NCC), pp. 1–6.

Du, L.-H., Liu, W., Zheng, W.-L., and Lu, B.-L. (2017). "Detecting driving fatigue with multimodal deep learning," in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER), Shanghai, China*. 74–77.

Ge, S., Yang, Y., Ni, W., and Zhang, R. (2018). "Multiple ocular artifacts removal based on high-order statistical tensor algorithm," in *2018 3rd International Conference on Intelligent Green Building and Smart Grid (IGBSG), Yilan, Taiwan, 2018*. 1–4.

Grosselin, F., Navarro-Sune, X., Vozzi, A., Pandremmenou, K., De Vico Fallani, F., Attal, Y., et al. (2019). Quality assessment of Single-Channel EEG for wearable devices. *Sensors (Basel, Switzerland)* 19:601. doi: 10.3390/s19030601

Hesar, H. D., and Mohebbi, M. (2021). An adaptive Kalman filter bank for ECG denoising. *IEEE J. Biomed. Health Inform.* 25, 13–21. doi: 10.1109/JBHI.2020.2982935

Huang, D., Ren, A., Shang, J., Lei, Q., Zhang, Y., Yin, Z., et al. (2016). Combining partial directed coherence and graph theory to analyse effective brain networks of different mental tasks. *Front. Hum. Neurosci.* 10:235. doi: 10.3389/fnhum.2016.00235

Jiang, X., Bian, G. B., and Tian, Z. (2019). Removal of artifacts from EEG signals: a review. *Sensors (Basel, Switzerland)* 19:987. doi: 10.3390/s19050987

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp.* 14, 166–185. doi: 10.1002/hbm.1050

Kamiński, J., Brzezicka, A., Gola, M., and Wróbel, A. (2012). β band oscillations engagement in human alertness process. *Int. J. Psychophysiol.* 85, 125–128. doi: 10.1016/j.ijpsycho.2011.11.006

Kashihara, K. (2017) Automatic discrimination of attention levels estimated by frontal EEG activity in drivers. IEEE Life Sciences Conference (LSC), 2017, pp. 194–197.

Klimesch, W., et al. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007

Lagerlund, T. D., Sharbrough, F. W., and Busacker, N. E. (1997). Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *J. Clin. Neurophysiol.* 14, 73–82. doi: 10.1097/00004691-199701000-00007

Liu, N. H., Chiang, C. Y., and Chu, H. C. (2013). Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensors* 13, 10273–10286. doi: 10.3390/s130810273

Liu, J., Lu, H., Zhang, X., Li, X., Wang, L., Yin, S., et al. (2023). Which multivariate multi-scale entropy algorithm is more suitable for analyzing the EEG characteristics of mild cognitive impairment? *Entropy (Basel)* 25:396. doi: 10.3390/e25030396

López-Ahumada, R., Jiménez-Naharro, R., and Gómez-Bravo, F. (2023). A hardware-based configurable algorithm for eye blink signal detection using a single-channel BCI headset. *Sensors (Basel, Switzerland)* 23:5339. doi: 10.3390/s23115339

Lu, A., Wang, D., He, S., Zhongcheng, Q., Zhang, W., and Li, Z. (2022). Attention mechanisms underlying dual-color digital visual search based on Schulte grid: an event-related potential study. *Brain Behav.* 12:e2471. doi: 10.1002/brb3.2471

Miranda, P., Cox, C. D., Alexander, M., Danev, S., and Lakey, J. R. T. (2020). In quest of pathognomonic/endophenotypic markers of attention deficit hyperactivity disorder (ADHD): potential of EEG-based frequency analysis and ERPs to better detect, prevent and manage ADHD. *Med. Devices* 13, 115–137. doi: 10.2147/MDER.S241205

Mowla, M. R., Huggins, J. E., Natarajan, B., and Thompson, D. E. (2018). P300 latency estimation using least mean squares filter. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2018, 1976–1979. doi: 10.1109/EMBC.2018.8512644

Nahm, F. S. (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J. Anesthesiol.* 75, 25–36. doi: 10.4097/kja.21209

Papaioannou, A. G., Kalantzi, E., Papageorgiou, C. C., Korombili, K., Bokou, A., Pehlivanidis, A., et al. (2021). Complexity analysis of the brain activity in autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) due to cognitive loads/demands induced by Aristotle's type of syllogism/reasoning. A power spectral density and multiscale entropy (MSE) analysis. *Heliyon* 7:e07984. doi: 10.1016/j.heliyon.2021.e07984

Pitchford, B., and Arnell, K. M. (2019). Resting EEG in alpha and beta bands predicts individual differences in attentional breadth. *Conscious. Cogn.* 75:102803. doi: 10.1016/j.concog.2019.102803

Roth, M., Hendeby, G., Fritsche, C., and Gustafsson, F. (2017). The ensemble Kalman filter: a signal processing perspective. *EURASIP J. Adv. Signal Process.* 2017, 1–16. doi: 10.1186/s13634-017-0492-x

Saini, M., Payal, S. U., and Satija, U. (2020). An effective and robust framework for ocular artifact removal from single-channel EEG signal based on variational mode decomposition. *IEEE Sensors J.* 20, 369–376. doi: 10.1109/JSEN.2019.2942153

Somers, B., Francart, T., and Bertrand, A. (2018). A generic EEG artifact removal algorithm based on the multi-channel wiener filter. *J. Neural Eng.* 15:036007. doi: 10.1088/1741-2552/aaac92

Souza, R. H. C., and Naves, E. L. M. (2021). Attention detection in virtual environments using EEG signals: a scoping review. *Front. Physiol.* 12:727840. doi: 10.3389/fphys.2021.727840

Sur, P., and Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* 116, 14516–14525. doi: 10.1073/pnas.1810420116

Tang, Y., and Huang, H. (2020). An EEG-based brain-computer interface for attention state recognition. 2020 International Symposium on Autonomous Systems (ISAS), pp. 100–104.

Thomas, S. (2023). Building a brain-computer interface to restore communication for people with paralysis. *Nat. Electron.* 6, 924–925. doi: 10.1038/s41928-023-01100-0

Vasileiou, K., Barnett, J., Thorpe, S., and Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Med. Res. Methodol.* 18, 1–18. doi: 10.1186/s12874-018-0594-7

Wang, L., Pu, Y. F., Liu, P., and Hao, Y. (2023). Multiscale hybrid method for speckle reduction of medical ultrasound images. *Multimed. Tools Appl.* 83, 55219–55234. doi: 10.1007/s11042-023-17335-0

Wang, H., Xu, L., Bezerianos, A., Chen, C., and Zhang, Z. (2021). "Linking attention-based multiscale CNN with dynamical GCN for driving fatigue detection," in *IEEE Transactions on Instrumentation and Measurement, vol. 70.* 1–11.

Wang, B., Xu, Z., Luo, T., and Pan, J. (2021). EEG-based closed-loop neurofeedback for attention monitoring and training in young adults. *J. Healthc. Eng.* 2021, 1–13. doi: 10.1155/2021/5535810

Wu, Y., and Xie, N. (2020). Attention optimization method for EEG via the TGAM. *Comput. Math. Methods Med.* 2020, 6427305–6427311. doi: 10.1155/2020/6427305

Yıldırım, H., and Özkale, M. R. (2023). A combination of ridge and Liu regressions for extreme learning machine. *Soft* 27, 2493–2508. doi: 10.1007/s00500-022-07745-x

Zhang, T. Z., Chang, T., and Wu, M. H. (2021) A brainwave-based attention diagnosis and music recommendation system for reading improvement. 2021 IEEE International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC), pp. 1–4.