

Subject-Calibrated Attention-State Classification during Digital Cognitive Tasks via MTFF: Band-Token Cross-Band Fusion on Single-Channel EEG

Jin Feng

Guilin Normal college

Jing He

Guilin University of Aerospace Technology

Yong Fan

Guilin University of Aerospace Technology

Junwei Huang

Guilin Normal college

Chenliang Jiang

Guilin University of Aerospace Technology

Yunde Li

Guilin University of Aerospace Technology

Guowei Liu

Suqian University

ZhaoLe Yu

1023494842@qq.com

Guilin University of Aerospace Technology

Article

Keywords: single-channel EEG, subject-dependent evaluation, trial-level split, multi-band feature fusion, Transformer, attention-state classification

Posted Date: August 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10533928/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Subject-Calibrated Attention-State Classification during Digital Cognitive Tasks via MTFF:
Band-Token Cross-Band Fusion on Single-Channel EEG

Jin Feng¹, Jing He², Yong Fan³, Junwei Huang⁴, Chenliang Jiang⁵, Yunde Li⁵, Guowei Liu⁶,
ZhaoLe Yu^{3*}

¹School of business , Guilin Normal College, Guilin 541199, China

²School of Management, Guilin University of Aerospace Technology, Guilin 541004, China

³ School of Aeronautics and Astronautics, Guilin University of Aerospace Technology, Guilin 541004, China

⁴Conservatory of Music, Guilin Normal College, Guilin 541199, China

⁵School of Artificial Intelligence, Guilin University of Aerospace Technology, Guilin 541004, China

⁶School of Economics and Management, Suqian University, Suqian223800, China

Corresponding author: ZhaoLe Yu (e-mail:1023494842@qq.com).

Abstract:

In the digital era, learning and work increasingly rely on digital tasks, making objective attention monitoring a practical problem for human-computer interaction and educational technology. Portable consumer-grade single-channel EEG enables low-burden recording, but the signal is vulnerable to ocular, facial EMG, and motion artifacts, and task-derived labels may contain residual uncertainty. To address these constraints, we propose MTFF (Multi-band Temporal Feature Fusion), a classification framework for portable single-channel EEG. MTFF learns intra-band rhythmic patterns in parallel over five canonical bands ($\delta/\theta/\alpha/\beta/\gamma$) within 0.5-40 Hz and compresses each band into a learnable band token. A lightweight Transformer then operates on six tokens (five band tokens plus one class token) to model cross-band complementarity while reducing the cost and overfitting risk of long time-point self-attention. Finally, a capacity-controlled linear decision head with sparse regularization constrains model complexity. Data from 50 healthy university participants performing a digital cognitive task were evaluated using a subject-wise stratified 60/40 trial-level split rather than a leave-one-subject-out protocol. MTFF achieved a subject-wise mean AUC of 0.9214 and accuracy of 0.8640 on held-out test trials; a pooled trial-level ROC analysis across all test trials yielded a test AUC of 0.8928.

Keywords: single-channel EEG; subject-dependent evaluation; trial-level split; multi-band feature fusion; Transformer; attention-state classification

0 Introduction

Digital learning, remote work and human-computer interaction increasingly depend on sustained attention during screen-based cognitive tasks. In these settings, attentional fluctuations can affect learning efficiency, operational accuracy and error risk, yet they are difficult to monitor continuously with subjective reports alone. This has made objective attention-state recognition an important problem for cognitive science, educational technology and human-computer interaction [1-6].

Electroencephalography (EEG) is well suited to this problem because it is non-invasive and captures neural dynamics at millisecond resolution [7]. Attention-related EEG changes are distributed across canonical frequency bands, including delta, theta, alpha, beta and gamma activity, and previous studies have used band-power, time-frequency and machine-learning features to distinguish focused from relaxed or low-attention states [8-17]. However, many existing studies are based on controlled paradigms or laboratory-grade acquisition, leaving a gap between reported classification performance and attention monitoring during everyday digital tasks.

Portable single-channel EEG narrows this gap by reducing preparation burden and supporting low-intrusion recording during computer-based tasks. At the same time, it removes the spatial redundancy available in multi-channel systems. Frontal single-channel signals are more vulnerable to blinks, eye movements, facial electromyography and motion artifacts, and they vary across individuals, sessions and

physiological states. A practical model therefore needs to extract stable information from noisy spectral evidence without relying on spatial filtering.

This requirement makes cross-band fusion a central technical bottleneck. Handcrafted pipelines typically compute band-power or band-limited statistics and then combine them by concatenation or fixed weighting, which can miss transient band-specific dynamics and obscure interactions among bands. Deep architectures and Transformers can learn richer dependencies, but applying high-capacity models to limited single-channel data can amplify subject-specific or noise-related patterns. The unresolved challenge is to represent multi-band evidence explicitly while keeping the model capacity small enough for subject-calibrated data.

To address this challenge, we propose MTFF (Multi-band Temporal Feature Fusion), a band-token fusion framework for portable single-channel EEG. MTFF first decomposes the 0.5-40 Hz signal into five canonical bands and uses parallel temporal-convolution branches to learn local rhythmic patterns within each band. Each branch is compressed into a learnable band token, so the downstream Transformer attends over a six-token sequence consisting of five band tokens and one class token. This design models cross-band complementarity at the token level instead of applying self-attention over long time-point sequences.

MTFF is evaluated on EEG data from 50 healthy university participants completing a digital cognitive task with focused-attention and low-attention baseline phases. The present evidence is intentionally framed as subject-calibrated trial-level

classification under a subject-wise stratified 60/40 split, not as leave-one-subject-out or calibration-free cross-subject validation. Within that boundary, the study tests whether band-token cross-band fusion and a capacity-controlled linear decision head improve balanced attention-state discrimination relative to classical machine-learning baselines.

1 Experimental Design

1.1 Participant Selection Criteria

This study was reviewed and approved by the Academic Research Ethics Review Committee of Guilin Normal College. All experimental protocols were performed in accordance with the relevant guidelines and regulations of the committee and with the Declaration of Helsinki.

A total of 50 university students aged 18 or above were recruited for this study. The inclusion criteria were as follows: right-handedness, good physical health, no history of cardiovascular, neuromuscular, or psychiatric disorders; normal or corrected-to-normal vision, intact hearing and cognitive ability. To minimize the influence of external factors, participants were required to refrain from consuming any stimulant foods or medications for 24 hours prior to the experiment. Before the formal experiment, each participant underwent environmental adaptation training and completed a practice session to ensure full comprehension of the task requirements and correct task execution. All participants provided written informed consent before the formal experiment began.

1.2 Experimental Tasks and Procedures

A digital cognitive task paradigm was designed to induce distinct attentional states. The experiment was conducted in a quiet room to minimize external interference. Participants wore the portable EEG device and completed the tasks on a computer running the experimental software. To control for circadian variations in alertness, all sessions were scheduled in the morning (9:00–12:00).

The experimental session began with a 30-second instructional video explaining the task procedure. Each session consisted of alternating focused-attention and low-attention baseline phases. In the focused-attention phase, participants performed a continuous number-search task: they were presented with a 5 x 5 grid containing numbers 1-25 in random positions and were required to identify and silently count the numbers in ascending order. Upon completing one grid, the next grid appeared immediately. This phase lasted 3 min and was designed to engage sustained attention. In the low-attention baseline phase, participants were instructed to relax, fix their gaze on a blank white wall, and avoid systematic mental activity. This phase lasted 1.5 min. Each focused-baseline cycle was repeated four times per participant to improve the stability of EEG recordings (Fig. 1). The labels used in this study should therefore be interpreted as task-condition labels intended to contrast focused-task engagement with a low-attention baseline, rather than as direct clinical or behavioral measurements of attention.

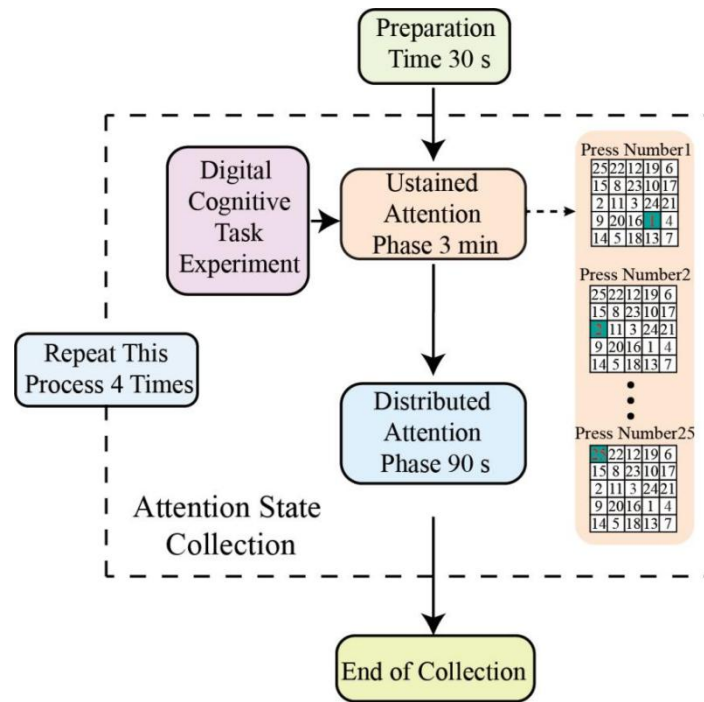


Fig. 1 Experimental paradigm design

2 Data Acquisition and Preprocessing

2.1 EEG Signal Acquisition

A MindWave Mobile 2 portable single-channel EEG device was used for data acquisition, with its core processing unit being the TGAM module. The device uses two dry electrodes: one placed on the forehead to record frontal EEG activity and one clipped to the left earlobe as the reference electrode. The TGAM module samples at 512 Hz and covers a frequency range of 0-100 Hz. It includes built-in noise suppression for environmental interference and 50/60 Hz power-line noise. Because the sensor is a consumer-grade frontal single-channel device, potential contamination from blinks, eye movements, facial EMG, and motion should be considered when interpreting the results.

2.2 Data Preprocessing Pipeline

Raw EEG signals are non-stationary, dynamically varying, and characterized by low amplitude, making them susceptible to contamination from electromagnetic interference and myoelectric artifacts. We therefore apply a conservative preprocessing pipeline to improve robustness before model training.

To further enhance robustness under the single-channel setting, the EEG signal was treated as a mixture. During preprocessing, the single-channel EEG is further denoised via adaptive noise-integrated empirical mode decomposition. For each 5-s window, the EEG is treated as a mixture and adaptively decomposed into a set of mode (sub-band) components, each capturing distinct frequency ranges and transient structures. The identified artifact-like modes are removed via threshold-based suppression, and the remaining modes are recombined to reconstruct a cleaner EEG estimate.

Considering that EEG includes canonical δ (0.5-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz), and γ (30-40 Hz) activity, a 0.5-40 Hz band-pass filter was applied to cover the main frequency range relevant to the present task while suppressing low-frequency drift and high-frequency EMG-dominated components. To reduce label contamination caused by delayed transitions between attentional states, segments within 5 s before and after each state transition were excluded from both focused and baseline periods. The remaining signals were segmented into non-overlapping 5-s trial windows, yielding approximately 200 valid trials per participant. For model evaluation,

trial splitting was performed within each participant as described below; future work should additionally evaluate block-wise or session-wise splits to reduce potential temporal autocorrelation between nearby windows.

3 MTFF Model Framework Based on Multi-band Temporal Feature Fusion

MTFF is designed to bridge the gap between handcrafted spectral features and heavy deep architectures in noisy, small-sample single-channel settings. Compared with handcrafted-feature approaches such as band power or band-power ratios, MTFF directly learns band-specific temporal patterns from the signal and encodes them as a compact set of learnable band tokens, enabling cross-band interaction to be modeled explicitly rather than implicitly mixed by feature concatenation. In contrast to Transformers that tokenize raw time points and operate on long sequences, MTFF performs self-attention on only six tokens (five band tokens plus one class token), which keeps computation lightweight while still modeling cross-band complementarity. To improve robustness under subject-calibrated trial-level evaluation, a linear decision head is adopted to control model capacity by design.

The overall framework is shown in Fig. 2. MTFF consists of (i) temporal-convolution-based multi-band feature extraction, which learns band-specific local temporal patterns, and (ii) Transformer-based feature fusion and classification, which integrates information across frequency bands through self-attention to produce a discriminative global representation.

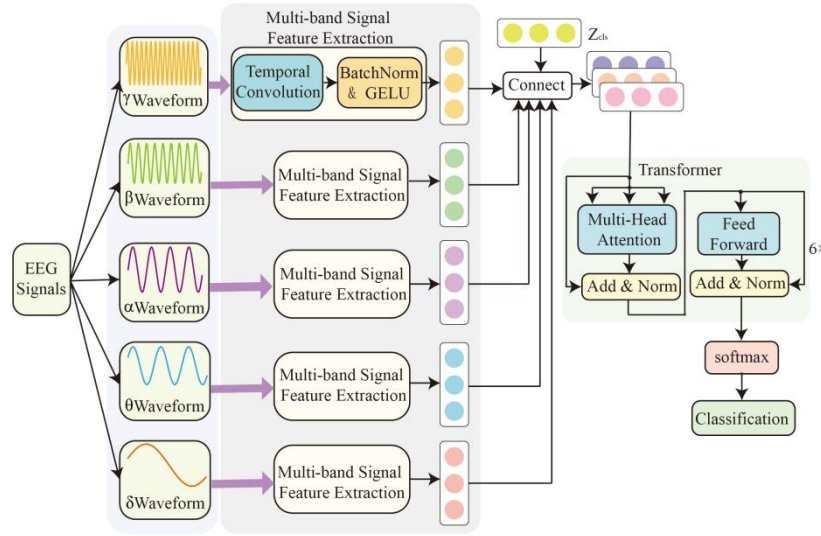


Fig. 2 Overall algorithm framework

3.1 Multi-band Feature Extraction Using Temporal Convolution

Raw EEG contains mixed activity from multiple frequency bands, each reflecting different neurocognitive processes. Directly feeding raw signals to a classifier may lead to inter-band interference. We therefore extract band-specific local temporal features using parallel temporal-convolution branches. Guided by electrophysiological studies [23–26], the 0.5–40 Hz range is divided into five canonical bands—delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ)—as summarized in Table 1.

Table 1 Mental state of 5 kinds of brain waves and corresponding frequency bands

Waveform	Frequency Band/Hz	Mental State
Delta(δ)	0.5~4.0	Deep sleep, unconsciousness, etc.
Theta(θ)	4.0~8.0	Light sleep, meditation, etc.
Alpha(α)	8.0~12.0	Eyes-closed relaxation, quiet sitting, etc
Beta(β)	12.0~30.0	Working, studying, intense thinking, etc.
Gamma(γ)	30.0~40.0	Highly focused thinking, etc.

For each band, we design a long temporal convolution whose receptive field covers 2–3 cycles of the lowest frequency of that band, enabling the network to better capture discriminative rhythmic patterns. The resulting band-wise features are projected to a unified embedding dimension ($C = 64$) to form the five band tokens. The complete training and inference pipeline is summarized in Table 2.

Table 2 MTFF training and inference procedure.

Input: raw EEG $x(t)$, sampling rate f_s , labels y (focused vs. low-attention baseline)

1) Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is applied to each window to obtain $IMF(x_k)$. Artifact-like modes are identified by a lightweight score $g(\cdot)$ and are removed before recombining the remaining modes to reconstruct a cleaner EEG $\tilde{x}(t)$. (Eq. 1 – 3)

2) For each window, compute five band-wise feature sequences via temporal convolution branches (Eq. 4 – 5)

3) Concatenate band tokens with a trainable class token; add positional embedding (Eq. 6 – 7)

4) Apply L Transformer encoder layers with H attention heads to fuse cross-band information (Eq. 8 – 10)

5) Use the class token as the fused representation; apply a sparse classifier for prediction (Eq. 11-12)

The original EEG signal can be expressed as:

$$x(t) = \sum_{k=1}^K x_k(t) + \varepsilon(t) \quad (1)$$

For each 5-second window of discrete EEG data $x(t)$, the CEEMDAN algorithm extracts K mode components (IMFs) and residuals. Here, $x_k(t)$ denotes the k-th IMF, and $\varepsilon(t)$ represents the final residual.

Calculate a pseudo-score s_k for each IMF

$$s_k = g(x_k(t)) \quad (2)$$

$g(\cdot)$ combines component energy, kurtosis, and high-frequency energy ratio.

$$\tilde{x}(t) = x(t) - \sum_{k \in \{k | s_k > \tau\}} x_k(t) \quad (3)$$

CEEMDAN is applied to each window to obtain $IMF(x_k)$. Artifact-like modes are identified by a lightweight score $g(\cdot)$ and are removed before recombining the remaining modes to reconstruct a cleaner EEG $\tilde{x}(t)$. τ is the false positive discrimination threshold.

$\tilde{x}(t)$ corresponding to δ 、 θ 、 α 、 β 、 γ , five waveform frequency bands. The signal extraction of each frequency band is completed by designing the corresponding time convolution:

$$h_k(t) = \sigma(\text{BN}((x \cdot \omega^{(0)} \cdot \omega^{(k)})(t))) \quad (4)$$

Of which, BN means BatchNorm, σ Indicates GELU activation function, $\omega^{(0)}$ is the initial one-dimensional convolution kernel (kernel=3, padding=1), used to suppress noise and enhance local information; $\omega^{(k)}$ for frequency band k one dimensional long convolution kernel, its length is set to $\text{kernel_size}_k \approx 1.5 \cdot \frac{F_s}{f_{\min,k}}$, to ensure that the receptive field covers 2-3 complete cycles, and its stripe is set to $\frac{1}{4} \text{kernel_size}_k$, that is, the adjacent windows overlap by 75%, taking into account the efficiency of low-frequency signal capture and calculation.

Finally, in order to map the signal features of five frequency bands to a unified spatial dimension and obtain a compact feature representation, the characteristics of each frequency band are $h_k(t)$ according to the following formula.

$$z_k = \text{Proj}(h_k(t)) \in \mathbb{R}^C \quad (5)$$

Of which, $k \in \{\delta, \theta, \alpha, \beta, \gamma\}$, $C=64$ is a unified spatial dimension. Finally, five frequency band feature representations are obtained $\{z_\delta, z_\theta, z_\alpha, z_\beta, z_\gamma\}$.

After obtaining the band level features, this paper further introduces the transformer module for global cross band information integration. The core of transformer is the self attention mechanism. Its advantage is that it can use the correlation information between sequences to model, so as to break through the bottleneck of traditional convolution limited to local receptive fields. Specifically, the five frequency band features obtained from the time convolution branch are represented as $\{z_\delta, z_\theta, z_\alpha, z_\beta, z_\gamma\} \in \mathbb{R}^{64}$ obtained from the time convolution branch with a trainable category vector $z_{cls} \in \mathbb{R}^{64}$ to obtain the input sequence:

$$Z = [z_{cls}; z_\delta; z_\theta; z_\alpha; z_\beta; z_\gamma] \in \mathbb{R}^{6 \times 64} \quad (6)$$

In order to distinguish the semantic roles with other frequency bands of z_{cls} and make the model learn the relative position between frequency bands, the position information of the Department of science is added and embedded in $E \in \mathbb{R}^{6 \times 64}$:

$$X^{(0)} = Z + E \quad (7)$$

For layer l and n attention heads:

$$Q_n = X^{(l)} W_n^Q, K_n = X^{(l)} W_n^K, V_n = X^{(l)} W_n^V \quad (8)$$

Where, $W_h^{Q,K,V} \in \mathbb{R}^{64 \times 64}$ is the corresponding learnable parameter matrix, and $Q_n, K_n, V_n \in \mathbb{R}^{6 \times 64}$ is the query matrix, key matrix and value matrix of the n attention header, respectively. Then the attention weight output is:

$$A_n = \text{soft max}(\frac{Q_n K_n^\top}{\sqrt{d_n}}), O_n = A_n V_n \quad (9)$$

Where, $A_n \in \mathbb{R}^{6 \times 6}$ is the attention weight matrix, which reflects the correlation between vectors; $O_n \in \mathbb{R}^{6 \times 64}$ represents the output representation of the n attention head, then the output mapping of multiple attention heads:

$$\begin{cases} \text{MHSA}(X^{(l)}) = \text{Concat}_n(O_n)W^O \\ \widetilde{X}^{(l)} = \text{LN}(X^{(l)} + \text{MHSA}(X^{(l)})) \quad \text{LN} \\ X^{(l+1)} = \text{LN}(\widetilde{X}^{(l)} + \text{MLP}(\widetilde{X}^{(l)})) \end{cases} \quad (10)$$

Where, $W^O \in \mathbb{R}^{(N \cdot 64) \times 64}$ is the learnable parameter matrix after multi head splicing, which is used to map the multi head output back to dimension 64, and N is the number of attention heads; LN is layernorm normalization; $\text{MHSA}(X^{(l)})$ is the overall output of multi head sub attention operation on $X^{(l)}$ input; MLP is a two-layer feedforward network including Gelu activation function and dropout; $\widetilde{X}^{(l)}$ is the intermediate representation in layer l after residual connection and LN normalization. $X^{(l+1)}$ is the output of the $l+1$ layer of transformer; $X^{(L)}$ is obtained after passing through layer $L = 6$.

Through the above transformer's multi-layer attention interaction, Class token z_{cls} can effectively integrate the context representation of frequency band features. Therefore, the final output can be used as the global expression of the characteristics of each frequency band, that is, $z_{global} = X_0^{(L)} \in \mathbb{R}^{64}$.

Finally, because z_{global} has passed through multi-layer attention interaction and has good discriminant ability, in order to avoid the risk of over fitting the model on

small sample EEG data, this paper only uses a linear classifier to classify the current subjects' attention state

$$\hat{y} = \text{softmax}(W(z_{global}) + b) \quad (11)$$

Where, is W hidden layer weight, and b is offset.

To prevent overfitting, this study incorporates a L^1 -norm regularization term into the objective function $E(\omega)$, ensuring the model produces sparse solutions.

$$\arg \min E(\omega) = \|y - X\omega\|^2 + \lambda \sum_{i=1}^K \|\omega_i\|_{L^1} \quad (12)$$

The weight is ω , the real label of the data sample is y , the regularization intensity is $\lambda > 0$, the total number of neurons is K , and the output of the layer is X .

4 Experimental results and analysis

To evaluate classification performance, we adopted a subject-wise stratified 60/40 trial-level split. For each participant, trials from the focused-attention and low-attention baseline conditions were divided into training and test sets at a ratio of 6:4 while preserving class proportions. The training set therefore contained trials from all participants, and the test set consisted of held-out trials from the same participants. This protocol evaluates subject-dependent trial-level generalization. Hyperparameters and early stopping were selected using only the training subset, and all baseline models (DT, MLP, RF, and SVM) were evaluated under the same split for fair comparison. Two complementary AUC summaries were reported: pooled trial-level

AUC, computed after concatenating all test trials, and subject-wise mean AUC, computed by first calculating AUC separately for each participant and then averaging across participants.

4.1 classification performance evaluation index

The classification performance metrics adopted in this study can all be derived from the confusion matrix [27], which comprises four essential elements: True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). Here, TP denotes the number of samples that are truly in a state of attentional focus and are correctly identified by the model; FN refers to the number of samples that are truly attentive but misclassified as distracted; FP represents the number of samples that are actually distracted but incorrectly predicted as attentive; and TN indicates the number of samples accurately identified as distracted.

Based on the above four core elements, a series of commonly used classification performance indicators can be further calculated, including area under the receiver operating characteristic curve (AUC), accuracy, F1 score, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). These indicators can quantitatively evaluate the classification performance of the model from multiple dimensions. The specific calculation formula is as follows:

$$\left\{ \begin{array}{l} \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \\ \text{F1} = \frac{2TP}{2TP + FP + FN} \\ \text{Sensitivity} = \frac{TP}{TP + FN} \\ \text{Specificity} = \frac{TN}{TN + FP} \\ \text{PPV} = \frac{TP}{TP + FP} \\ \text{NPV} = \frac{TN}{TN + FN} \end{array} \right. \quad (13)$$

4.2 Accuracy under the Subject-wise 60/40 Trial-level Split

To evaluate MTFF in the attention-state classification task, validation was conducted across all participants under the subject-wise 60/40 trial-level split. Fig. 3 reports training and test accuracy aggregated across participants. MTFF achieved average accuracy rates of 91.9% on the training trials and 86.4% on held-out test trials, respectively. These results indicate good trial-level discrimination within a subject-calibrated setting, but they should not be interpreted as evidence of calibration-free generalization to unseen participants.

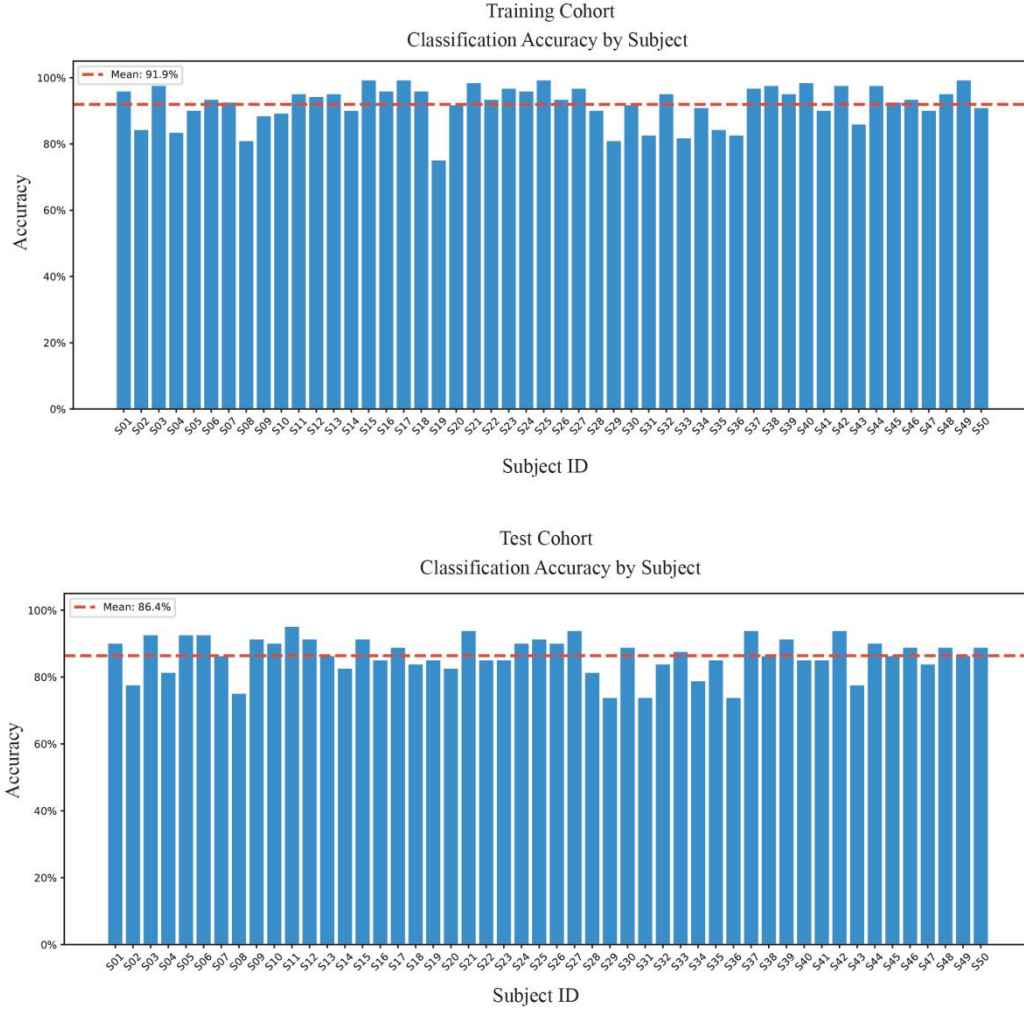


Fig. 3 Subject-wise training and test accuracy under the 60/40 trial-level split

4.3 Comparative Experimental Analysis

To further validate the advantages of MTFF over conventional methods, several commonly used classification models in EEG analysis were selected for comparison, including Decision Tree (DT) [28], Multilayer Perceptron (MLP) [29], Random Forest (RF) [30], and Support Vector Machine (SVM) [31]. Previous studies have shown that these models perform effectively in EEG signal processing tasks. For instance, Guan et al. [32] used a Decision Tree to classify motor imagery tasks and improved the Kappa value on Dataset IV of a BCI competition from 0.57 to 0.607. Experimental

results from Wong et al. [33] demonstrated that an MLP-based learning method outperformed mean-based or optimal filtering methods in EEG classification accuracy. Wang Bingbing et al. [34] divided attention into five categories and proposed an attention detection method based on Random Forest, improving the detection accuracy to 76.17%. Meanwhile, Jana et al. [35] enhanced epilepsy classification accuracy by 1%–4% compared to a single-view SVM model by introducing a multi-view SVM approach. In summary, these classical algorithms have proven effective across various EEG classification tasks, providing a reasonable baseline for comparative evaluation in this study.

Fig.4 presents violin plots illustrating the distribution of prediction scores at the trial level for each model, along with the corresponding independent-samples t-test statistics (t-values) and significance levels (p-values). Visual inspection of the figure reveals a marked separation between the score distributions of trials predicted as class 0 and class 1 by the proposed MTFF method. Moreover, the MTFF model yields a t-value of -45.22, which is the smallest (i.e., largest in absolute magnitude) among all models, with a corresponding p-value below 0.05. This indicates that the difference in predictions between the two classes is most statistically significant for the MTFF model, suggesting stronger statistical support for its class separability. Here, class 1 denotes the focused-attention phase and class 0 denotes the low-attention baseline phase. The x-axis indicates the predicted class and the y-axis indicates the model score/probability for class 1.

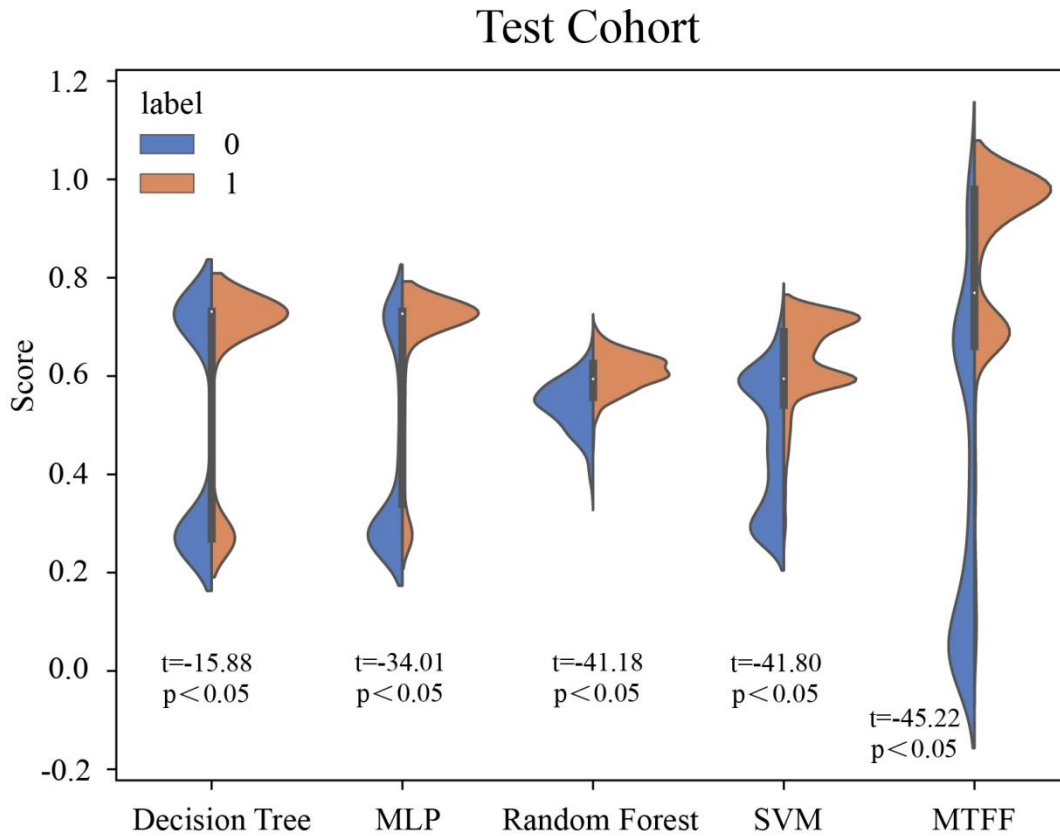


Fig. 4 Violin plots of different models

Fig. 5 displays pooled trial-level ROC curves and corresponding AUC values for different models on the training and test trials. In this analysis, trials from all participants were concatenated before ROC calculation. The training AUC values for Decision Tree, MLP, Random Forest, SVM, and MTFF were 0.9973, 1.0000, 1.0000, 0.8953, and 0.9478, respectively. The pooled test-trial AUC values were 0.6339, 0.7903, 0.8576, 0.8556, and 0.8928, respectively.

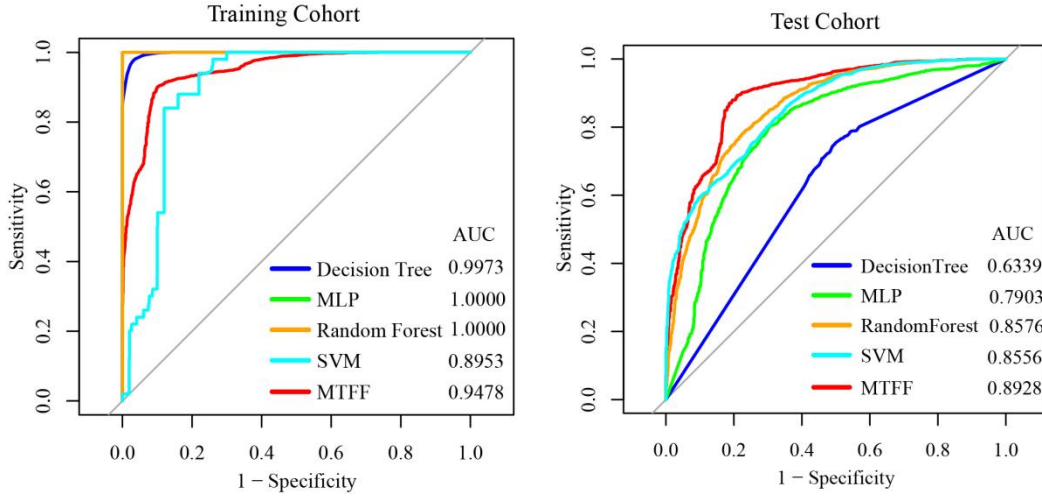


Fig. 5 Pooled trial-level ROC curves of all models across training and test trials

A comparison between training and test AUC values reveals marked overfitting in the Decision Tree, MLP, and Random Forest models. These models achieved extremely high AUC values on the training trials, reaching or approaching 1.0000, while their performance declined on held-out test trials. The SVM model demonstrated relatively stable pooled AUC values across the training and test sets.

In contrast, MTFF maintained better consistency between the training trials (AUC = 0.9478) and pooled test trials (AUC = 0.8928). MTFF also achieved the highest pooled test-trial AUC among the compared models, suggesting stronger trial-level classification performance under the subject-wise split.

The observed train-test gap of conventional baselines suggests overfitting to training trials: several models reached near-ceiling training AUC yet degraded on held-out test trials, indicating that they may exploit noise-specific or task-segment-specific cues rather than stable EEG signatures. In contrast, MTFF maintained a smaller train-test gap and achieved the best pooled test-trial AUC, which aligns with its design goals for portable single-channel EEG. Specifically, band-aware

tokenization preserves transient band-specific dynamics beyond handcrafted summaries, token-level cross-band attention enables explicit yet lightweight inter-band fusion, and the capacity-controlled linear head constrains the decision boundary to reduce spurious correlations in small-sample noisy settings.

To evaluate classification performance more comprehensively, we further computed subject-wise metrics. For each participant, AUC, accuracy, F1 score, sensitivity, specificity, PPV, and NPV were calculated on held-out test trials; the resulting values were then averaged across participants, as shown in Table 3. Fig. 6 provides a radar chart to visualize the multidimensional comparison among methods.

Table 3 Subject-wise averaged performance metrics on held-out test trials

classification models	AUC	Accuracy	Precision	Sensitivity	Specificity	PPV	NPV
Decision Tree	0.6278	0.6685	0.7494	0.7455	0.5084	0.7586	0.5046
MLP	0.8140	0.7375	0.8313	0.9548	0.2861	0.7398	0.7898
Random Forest	0.8402	0.6825	0.8096	0.9996	0.0238	0.6804	0.9285
SVM	0.7887	0.6190	0.8059	0.8400	0.1600	0.6750	0.3250
MTFF	0.9214	0.8640	0.8988	0.8966	0.7961	0.9045	0.7974

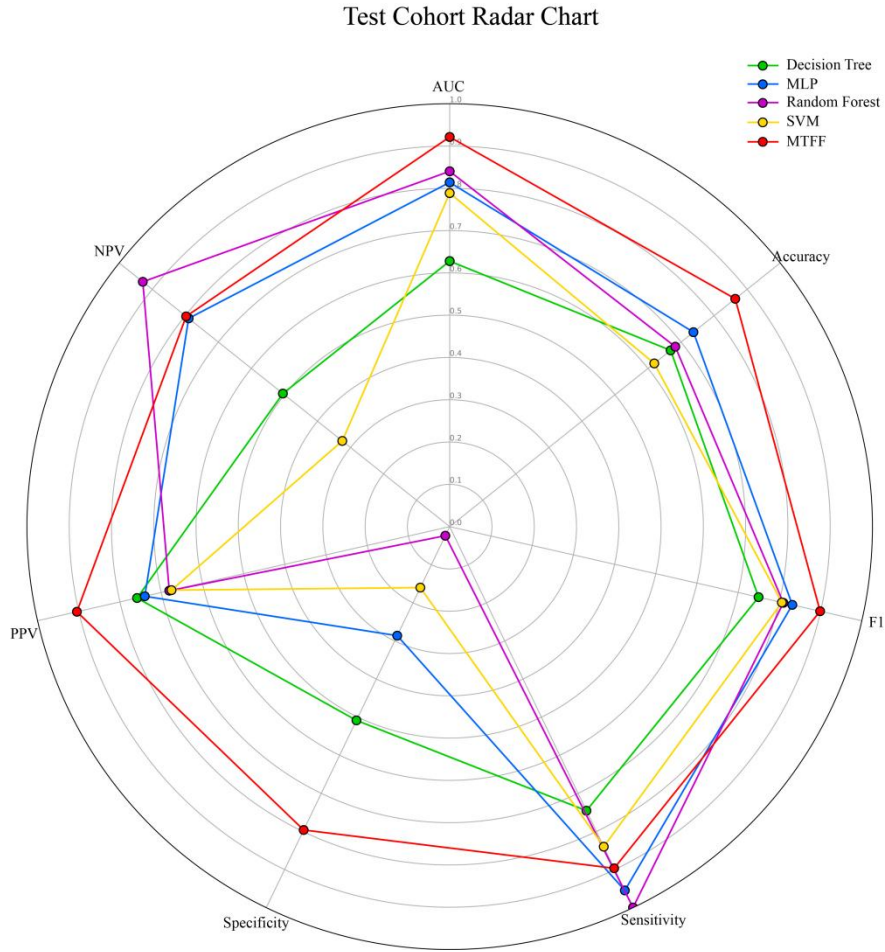


Fig. 6 Radar chart illustrating subject-wise averaged performance metrics on held-out test trials

The results in Table 3 and Fig. 6 show that MTFF achieved more balanced performance across key indicators. Decision Tree and SVM were lower than MTFF on most metrics, suggesting limited ability to capture discriminative EEG patterns in this setting. Although MLP and Random Forest showed very high sensitivity, their specificity values were only 0.2861 and 0.0238, respectively, indicating a tendency to misclassify baseline trials as focused-attention trials. In contrast, MTFF maintained high sensitivity while substantially improving specificity, yielding a more balanced discrimination between focused-task and baseline conditions.

Overall, MTFF demonstrated strong trial-level classification performance and balanced multidimensional metrics under the subject-wise 60/40 split. These findings support its potential for subject-calibrated portable EEG attention-state monitoring, but further validation under block-wise, session-wise, and true subject-independent protocols is required before claims of cross-subject deployment can be made.

5 Conclusion

This work studies portable single-channel attention-state classification through the lens of cross-band fusion. Under a subject-wise stratified 60/40 trial-level evaluation, MTFF achieved the best subject-wise averaged performance among the compared models (AUC = 0.9214, Accuracy = 0.8640) and showed a more balanced sensitivity-specificity trade-off than classical baselines. In particular, some baselines exhibited very high sensitivity but extremely low specificity, indicating excessive false positives for baseline trials. In contrast, MTFF maintained high sensitivity while substantially improving specificity, yielding more reliable subject-calibrated trial-level discrimination.

The training-test AUC comparison further highlights the challenge of overfitting in portable single-channel EEG. Several baselines reached near-saturated training AUC, whereas their pooled test-trial AUC decreased substantially. MTFF showed a smaller train-test gap and achieved the highest pooled test-trial AUC among the compared models, indicating improved stability under the present subject-wise trial-level split.

These observations are consistent with the fusion-centric design of MTFF. First, band-wise temporal modeling provides a structured pathway to preserve transient rhythmic dynamics within each canonical band, instead of collapsing them into handcrafted summaries. Second, representing each band as a compact token enables interaction modeling at the band level, which offers an explicit fusion interface under the single-channel constraint. Third, performing self-attention on a fixed and very small token set (six tokens) concentrates modeling capacity on cross-band interactions while avoiding the cost and potential overfitting risk of applying self-attention over long time-point sequences. Finally, the capacity-controlled linear decision head with sparse regularization constrains decision boundary complexity, which may help reduce the reliance on subject- or noise-specific cues under distribution shift.

This work has limitations. First, attention labels were derived from task phases and may contain residual label noise, especially near state transitions; behavioral or subjective attention measures should be incorporated in future studies. Second, although conservative preprocessing and denoising were applied, consumer single-channel frontal EEG remains susceptible to eye-movement, blink, motion, and facial EMG contamination. Third, the current evaluation uses a subject-wise 60/40 trial-level split and therefore does not establish leave-one-subject-out or calibration-free cross-subject generalization. Fourth, random trial-level splitting of continuous recordings may retain temporal autocorrelation between training and test windows; block-wise or session-wise validation should be added. Fifth, we do not isolate the contribution of each component, such as band tokenization, token-level fusion, and capacity control,

via ablation analyses, and we do not provide direct interpretability evidence, such as attention maps or band-occlusion analyses, for the learned cross-band interactions. Future work should address these limitations by incorporating richer behavioral ground truth, stronger artifact-handling and calibration strategies, component-wise ablation, interpretable band-contribution analyses, and true subject-independent validation.

Data Availability

The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request. Due to participant privacy restrictions, the data are not publicly available.

Funding Declaration

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by Guangxi Natural Science Foundation under Grant (No.2025JJH180042, No.2025JJH170032), National Science Foundation of China (72264007), Research Project of Guilin Normal University (KYA202301).

References

- [1] Li Chengyu & Zhao Weijie. (2020). Progress in the brain-computer interface: an interview with Bin He. *National Science Review*, 7 (2), 480-483.
- [2] Lodge, J. M., & Harrison, W. J. (2019). The role of attention in learning in the digital age. *The Yale journal of biology and medicine*, 92(1), 21.
- [3] Mengyao Zhai & Zhaohui Yuan. (2024). Research on Attention of Students in Junior High School Art Classroom Based on Flow Theory. *Innovation Humanities and Social Sciences Research*, 20,
- [4] Kaveti Pavan, Ankit Singh, Digvijay S Pawar & Nagarajan Ganapathy. (2025). Assessment of Driver Inattention State Using Multimodal Wearable Signals and Cross-Attention-Driven Hierarchical Fusion.. *Studies in health technology and informatics*, 323, 121-125.
- [5] Quynh Anh Nguyen, Nam Anh Dao & Long Nguyen. (2025). Enhanced Pilot Attention Monitoring: A Time-Frequency EEG Analysis Using CNN–LSTM Networks for Aviation Safety. *Information*, 16 (6), 503-503.
- [6] Lee Seungkeon, Lee Meyoung, Lee Hakjin, Jeong Daesik & Lee Eui Chul. (2025). Automated Detection of Construction Worker Inattention Using IMU Sensors and Visual Focus of Attention. *Tehnički vjesnik*, 32 (2), 610-619.

- [7] Rahman, S. A., Khalil, M. I., Zhou, H., Guo, Y., Ding, Z., Gao, X., & Zhang, D. (2025). Advancement in Graph Neural Networks for EEG Signal Analysis and Application: A Review. IEEE Access.
- [8] Kästle, J. L., Anvari, B., Krol, J., & Wurdemann, H. A. (2021). Correlation between Situational Awareness and EEG signals. *Neurocomputing*, 432, 70-79.
- [9] Al-Nafjan, A., & Aldayel, M. (2022). Predict students' attention in online learning using EEG data. *Sustainability*, 14(11), 6553.
- [10] Li, D., **e, L., Chai, B., Wang, Z., & Yang, H. (2022). Spatial-frequency convolutional self-attention network for EEG emotion recognition. *Applied Soft Computing*, 122, 108740.
- [11] Xu, X., Nie, X., Zhang, J., & Xu, T. (2023). Multi-level attention recognition of EEG based on feature selection. *International Journal of Environmental Research and Public Health*, 20(4), 3487
- [12] Kaushik, P., Moye, A., Vugt, M. V., & Roy, P. P. (2022). Decoding the cognitive states of attention and distraction in a real-life setting using EEG. *Scientific Reports*, 12(1), 20649.
- [13] Alirezaei, M., & Sardouie, S. H. (2017, November). Detection of human attention using EEG signals. In 2017 24th National and 2nd International Iranian Conference on biomedical engineering (ICBME) (pp. 1-5). IEEE.

- [14] Barry, R. J., Clarke, A. R., & Johnstone, S. J. (2003). A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography. *Clinical neurophysiology*, 114(2), 171-183.
- [15] You, S. D. (2021). Classification of relaxation and concentration mental states with eeg. *Information*, 12(5), 187.
- [16] Hassan, R., Hasan, S., Hasan, M. J., Jamader, M. R., Eisenberg, D., & Pias, T. (2020, May). Human attention recognition with machine learning from brain-EEG signals. In 2020 IEEE 2nd Eurasia conference on biomedical engineering, healthcare and sustainability (ECBIOS) (pp. 16-19). IEEE.
- [17] Xue, H., Wang, D., **, M., Gao, H., Wang, X., **a, L., ... & Lin, J. (2023). Hydrogel electrodes with conductive and substrate-adhesive layers for noninvasive long-term EEG acquisition. *Microsystems & Nanoengineering*, 9(1), 79.
- [18] Yan Huang, Yao Deng, Lingda Kong, et al. Vigilant attention mediates the association between resting EEG alpha oscillations and word learning ability, *NeuroImage*, Volume 281, 2023, 120369, ISSN 1053-8119, <https://doi.org/10.1016/j.neuroimage.2023.120369>.
- [19] Verma, D., Bhalla, S., Santosh, S. S., Yadav, S., Parnami, A., & Shukla, J. (2023, September). Attentionet: Monitoring student attention type in learning with eeg-based measurement system. In 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 1-8). IEEE.

- [20] Chaddad, A., Wu, Y., Kateb, R., & Bouridane, A. (2023). Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques. *Sensors*, 23(14), 6434.
- [21] Sharma, R., & Meena, H. K. (2024). Emerging trends in EEG signal processing: A systematic review. *SN Computer Science*, 5(4), 415.
- [22] Sun, Q., Zhou, Y., Gong, P., & Zhang, D. (2025). Attention detection using EEG signals and machine learning: a review. *Machine Intelligence Research*, 22(2), 219-238.
- [23] Chattu, P., & Prasad, C. V. P. R. (2025). Feature Extraction of EEG Signal Using Convolutional Neural Networks by Removing Artifacts. *Informatica*, 49(23).
- [24] Tangyongchao (2021) Research on attention analysis method based on brain computer interface (Master's thesis, South China University of Technology) Master <https://doi.org/10.27151/d.cnki.ghnlu.2021.003232>.
- [25] Insana, S. (2020). Quantitative EEG Including the Five Common Bandwidths (Delta, Theta, Alpha, Sigma, and Beta). In *Encyclopedia of Behavioral Medicine* (pp. 1827-1830). Cham: Springer International Publishing.
- [26] Cai, W., Chen, L., Dai, Y., Chen, B., Zheng, D., & Li, Y. (2024). Association between EEG power during sleep and attention levels in patients with major depressive disorder. *Nature and Science of Sleep*, 855-864.

- [27] Li Chen (2023). Research on Attention State Classification Based on EEG Signals (Master's Thesis, Central China Normal University) Master's degree <https://doi.org/10.27159/d.cnki.ghzsu.2023.001311>.
- [28] Alghamdi, A. M., Ashraf, M. U., Bahaddad, A. A., Almarhabi, K. A., Al Shehri, W. A., & Daraz, A. (2025). Cross-subject EEG signals-based emotion recognition using contrastive learning. *Scientific Reports*, 15(1), 28295.
- [29] Saraya, G. K. (2025). Predicting Consumer Behavior: A Critical Review of EEG-Based Neuromarketing and the Decision Tree Model. *NeuroRegulation*, 12(2), 132-132.
- [30] Li, W., Tian, Y., Hou, B., Dong, J., Shao, S., & Song, A. (2023). A Bi-Stream hybrid model with MLPBlocks and self-attention mechanism for EEG-based emotion recognition. *Biomedical Signal Processing and Control*, 86, 105223.
- [31] Bedolla-Ibarra, M. G., Cabrera-Hernandez, M. D. C., Aceves-Fernández, M. A., & Tovar-Arriaga, S. (2022). Classification of attention levels using a Random Forest algorithm optimized with Particle Swarm Optimization. *Evolving Systems*, 13(5), 687-702.
- [32] Peng, C. J., Chen, Y. C., Chen, C. C., Chen, S. J., Cagneau, B., & Chassagne, L. (2020). An EEG-based attentiveness recognition system using Hilbert–Huang transform and support vector machine. *Journal of Medical and Biological Engineering*, 40(2), 230-238.

- [33] Wong, D. K., Uy, E. T., Guimaraes, M. P., Yang, W., & Suppes, P. (2006). Interpretation of perceptron weights as constructed time series for EEG classification. *Neurocomputing*, 70(1-3), 373-383.
- [34] Wang Bingbing, Xu zeju, Luo Tong&pan Jiahui (2021) Adolescent attention detection and training system based on EEG signal Computer system applications, 30 (10), 76-85 <https://doi.org/10.15888/j.cnki.csa.008107>.
- [35] Jana, G. C., Praneeth, M. S., & Agrawal, A. (2023). A multi-view SVM approach for seizure detection from single channel EEG signals. *IETE Journal of Research*, 69(6), 3120-3131.