Advancing Adolescent Depression Detection through Multi-Task EEG Signals and Biosignal Learning

Aojun Wen[1†], **Zijing Guan**[2,3†], **Zhiqiang Chen**[1], **QuanlinChen**[1], **Jun Xiao**[3], **Jiahui Pan**[1,3], **Haiyun Huang**[1,3*]
[1]School of Artificial Intelligence, South China Normal University, Foshan, 528225, China
[2]School of Automation Science and Engineering, South China University of Technology, Guangzhou, 510640, China
[3]Research Center for Brain–Computer Interface, Pazhou Laboratory, Guangzhou, 510330, China
[†]Equal contribution [*]Corresponding author
{huanghaiyun@m.scnu.edu.cn}

Abstract

The increasing prevalence of depressive disorders in adolescents is becoming a major public health concern. Prompt diagnosis is crucial to mitigate the effects of depression; however, current diagnostic approaches are often burdensome and lack objective biomarkers, which hinders early detection and intervention.This paper introduces a novel multi-tasking framework that integrates attention and resting task EEG signals using Biosignal Learning and Agent Transformer (BLAT) for adolescent depression detection. Our approach involves segment, channel, and positional encoding of EEG data to process signals related to attention and resting tasks. These signals are then analyzed using the Agent Transformer for feature extraction and subsequent depression classification.Our experiments involve 50 adolescents diagnosed with depression and 50 matched controls. By applying BLAT to classify combined attention and resting task EEG signals, we achieved an 85.00% accuracy rate in detecting depression in an independent test set. This outcome underscores the value of our method in providing a viable method for early depression screening in adolescents.

Keywords: Depression detection, EEG, multi-task, attention task, resting task.

Introduction

Depression is a significant global health issue, identified by the World Health Organization as a leading cause of disability, with a profound impact on adolescents due to various pressures (Chen et al., 2024). Adolescents face unique challenges that contribute to the rising prevalence of depression, including family expectations, academic demands, and social interactions (Qi et al., 2025). Addressing these requires comprehensive strategies for early diagnosis and intervention, emphasizing the role of public policies in educational and community settings.

Despite the availability of psychological counseling services, stigma and trust issues prevent many students from seeking help (Yong, Ren, Wang, Yang, & et al., 2024), indicating a need to alter public perceptions of mental health support. Innovative technologies, such as EEG signals and deep learning algorithms, are emerging as effective tools for diagnosing depression (Hao et al., 2021). These methods offer objective data that complement traditional diagnostic frameworks, which often rely on subjective evaluations (Malhi, Mann, & et al., 2018).

Recent advancements in neuroimaging, like fMRI and EEG, have identified biomarkers linked to depression, revealing physiological changes in the brain that enhance diagnostic precision (Hamilton et al., 2012; Olbrich & Arns, 2013). Additionally, attention-based tasks and multimodal networks integrating various physiological data are improving the accuracy of depression recognition (Liu & et al., 2021).Attention performance differs significantly between individuals with depression and non-depressed controls, highlighting key disparities in several dimensions of attentional functioning. Depressed individuals exhibit slower attention allocation, particularly when processing positive information, which may hinder their overall perception and response to such stimuli. They also require greater attentional load during information processing, reflecting increased cognitive effort and reduced processing efficiency. Furthermore, attentional flexibility is impaired in individuals with depression, making it more challenging for them to shift focus between different types of information, especially in scenarios requiring rapid adjustments. Their attention is also less sustained over prolonged tasks, leading to lapses in focus and reduced processing efficiency during extended periods of concentration. In contrast, non-depressed individuals exhibit faster attention allocation, reduced attentional load, greater flexibility, stronger sustained attention, and deeper focus, enabling them to process and comprehend information more efficiently. These differences underscore the unique challenges faced by individuals with depression in attentional performance, suggesting that interventions should prioritize enhancing attention-related capacities to improve cognitive and emotional functioning(Kohei et al., 2022).These developments suggest promising directions for personalized treatment strategies, particularly for adolescents at risk.

This paper seeks to explore these advancements, illuminating current methodologies and identifying future research directions to improve understanding and treatment of depression among vulnerable adolescent populations.

In this study, we explore various fusion methods of resting and attention tasks for depression detection. By employing deep learning techniques, we extract effective features from these multi-task datasets, significantly improving overall classification accuracy. The model's efficacy is rigorously validated through extensive testing on an independent test set. The primary contributions of this work are:

- By extracting and integrating features from resting and attention task data and utilizing Biosignal Learning and Agent Transformer (BLAT), we have successfully pre-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

served the rich information present in EEG signals.

- Our self-collected dataset includes data from 100 adolescent participants, providing a large sample that encompasses both resting and attention task data.
- Through the use of a comprehensive EEG dataset for model training and subsequent validation on an independent test set with different participants, we ensure a thorough evaluation of the model's performance under real-world conditions.

Proposed Model

Physiological signals provide critical insights into health conditions, but their complexity poses challenges for analysis. Compared to images (RGB or grayscale), audio, or natural language, physiological electrical signals are more complex, primarily due to their multiple channels. Converting these diverse physiological signals into a unified 'sentence' structure is not a straightforward task. This paper employs the Biosignal Learning and Agent Transformer (BLAT) to address this challenge. BLAT utilizes a novel biosignal tokenization module (Yang, Westover, & Sun, 2023) that individually segments and flattens each channel into tokens to form a consistent biosignal 'sentence'. Through innovative parsing techniques, BLAT effectively extracts information and identifies multi-channel EEG signals from both depressive subjects and healthy controls.

Biosignal Tokenization

Tokenization To accommodate varying lengths, we tokenize each channel's record into t-second tokens, with adjacent tokens overlapping by p seconds (where p $<$ t, e.g., t = 1 and p = 0.5). Consequently, the k-th token in the i-th channel (k = 1, 2, 3, ...) can be represented using slice notation as $S[i,(t-p)(k-1):(t-p)(k-1)+t]$. The number of tokens in each channel is constrained by the inequality: $(t-p)(k-1)+t \leq J$ (where J is the signal length). Here, the overlap p is crucial for preserving temporal information in shorter signals.

Flattening We flatten tokens from all channels into a consistent 'sentence'.

Segment embedding We derive the segment embedding from a spectral perspective. Initially, we extract an energy vector for each token $S[i,(t-p)(k-1):(t-p)(k-1)+t]$ across all frequency bands, facilitated by the Fast Fourier Transform (FFT). Subsequently, a fully connected network (FCN) processes this energy vector to generate the segment embedding.

Channel embedding (spatial) We learn an embedding table for all different channels and add the corresponding channel embedding to the token. Each color represents one channel in Figure 1.

Positional encoding (temporal) In the context of biosignals, the sequential arrangement of segments within a channel encapsulates temporal information. Consequently, we incorporate relative positional encoding into the ultimate token embedding through the utilization of sinusoidal and cosine functions, thereby obviating the necessity for trainable parameters.

We denote the final tokenized biosignal 'sentence' as $\mathbf{X} \in \mathbb{R}^{N\times l}$, where N represents the number of tokens and l denotes the dimension of the token embedding. In Figure 1, the highlighted orange region signifies the spatial or temporal-relevant tokens with respect to the current token (i.e., tokens at the same time step or within the same channel). Our token embeddings are adept at capturing both segment-specific features and spatio-temporal characteristics.

Agent transformer

We propose the utilization of the Transformer architecture to facilitate the learning of 'sentence' embeddings for biosignals. Recognizing the inherent quadratic complexity in both time and space of the conventional Transformer model, we integrate the Agent attention mechanism to address the computational challenges associated with long biosignal sequences. This adaptation ensures a more efficient processing capability, aligning with the demands of biosignal learning applications.

To simplify, we abbreviate Softmax and linear attention as:

$$O^{\mathrm{S}} = \sigma\left(QK^T\right)V \triangleq \mathrm{Attn}^{\mathrm{S}}(Q,K,V), \tag{1}$$

$$O^{\phi} = \phi(Q)\phi(K)^T V \triangleq \mathrm{Attn}^{\phi}(Q,K,V). \tag{2}$$

Where $Q,K,V \in \mathbb{R}^{N\times C}$ denote query, key and value matrices, and $\sigma(\cdot)$ represents the Softmax function. Then our agent attention can be written as:

$$O^{\mathrm{A}} = \underbrace{\mathrm{Attn}^{\mathrm{S}}(Q,A,\underbrace{\mathrm{Attn}^{\mathrm{S}}(A,K,V)}_{\text{Agent Aggregation}})}_{\text{Agent Broadcast}}. \tag{3}$$

Experiment

Dataset

A cohort of 100 middle school students was enrolled in the present study, stratified into two groups: the Depression (DEP) group and the Healthy Control (HC) group, with each group consisting of 50 participants, using HAMD scores to distinguish between healthy and depressed subjects.As depicted in Figure 2,each subject underwent both attention and resting task assessments. For the non-attention condition, participants engaged in a one-minute period of tranquil imagery (stimulus-free). The EEG acquisition apparatus used in this study featured 30 channels with a sampling rate of 250 Hz. Each participant will have four tasks at the end: eyes-open resting task**(EO)**, eyes-closed resting task**(EC)**, attention task**(ATT)**, and non-attention task**(Non-ATT)**. Approval for this research was obtained from the Ethics Committee of

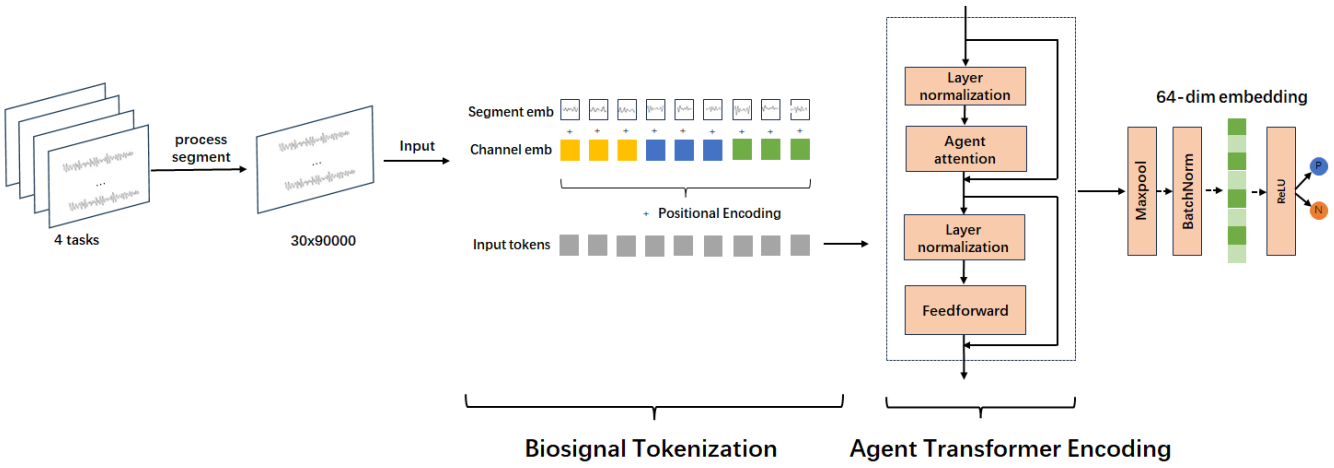


Figure 1: Biosignal Learning and Agent Transformer(BLAT). Upon acquisition of a novel data sample, an initial phase of data preprocessing is conducted, encompassing (Filtering, data cleaning, Notch filtering, re-referencing, removal of empty channels, normalization, and flattening), culminating in the formulation of a biosignal 'sentence' via the biosignal tokenization module. Subsequently, the intricate interrelations within the 'sentence' are elucidated through the application of the Agent transformer module.

the Affiliated Brain Hospital of Guangzhou Medical University, and all participants and their guardians signed informed consent forms.

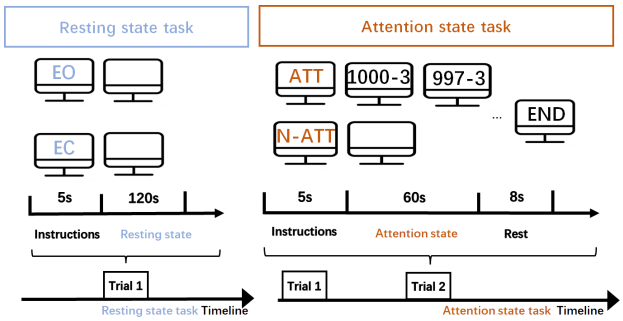


Figure 2: The experimental design consists of three distinct segments: two 2-minute resting EEG segments and one 1-minute attention task segment. The first segment involves a resting task with eyes open, followed by a resting task with eyes closed. The graphical user interface (GUI) for the attention task is shown on the right side of the image. During the attention task, participants perform mental arithmetic subtraction.

Pre-process

The preprocessing of raw EEG signals was performed using the EEGLAB toolbox. The specific steps are as follows:

- **Signal Filtering and Cleaning:** Firstly, a band-pass filter with a range of 1 to 50 Hz was applied to the raw EEG signals. Subsequently, the filtered data was cleaned by filling in any missing values. To further remove power line noise, a notch filter at 50 Hz was employed.
- **Signal Re-referencing:** The filtered EEG signals were then re-referenced to the common average reference. This step helps to reduce the influence of the reference electrode and improve the stability and reliability of the signals.
- **Data Normalization:** Finally, Z-score normalization was applied to the preprocessed EEG data. This step helps to eliminate the effects of magnitude differences between channels, improving the comparability and accuracy of the data for subsequent analysis. Due to varying experimental durations, the resting task was maintained for 2 minutes, while both the attention and non-attention tasks were each sustained for 1 minute. The sampling frequency for all tasks was consistently set at 250 Hz. Consequently, the experimental data yielded a total of 120 seconds * 250 Hz sampling points for the resting task, whereas the attention and non-attention tasks each generated 60 seconds * 250 Hz sampling points.

Through this preprocessing pipeline, the quality of the EEG data was significantly enhanced, providing a solid foundation for further analysis and application.

Baseline

We consider the following representative models:

- **SPaRCNet (Jing et al., 2023):** This model is based on 1D-CNN with dense residual connections, offering more advanced performance compared to well-known models such as ConvNet(Hartmann, Schirrmeister, & Ball, 2017) and CSCM(Sakhavi, Guan, Yan, & et al., 2018).
- **CNN-Transformer (Peh, Yao, & Dauwels, 2022):** This model outperforms the CNN-LSTM models (Zhang, Zong, Dou, Zhao, & et al., 2019) by integrating convolutional neural networks with transformer architectures.

Table 1: Classification Performance for Different Task Combinations

Task Combinations	F1-Score±SD(%)	AUROC±SD(%)	Acc±SD(%)	TPR±SD(%)	TNR±SD(%)
Non-ATT	61.38±5.36	62.80±6.13	62.00±6.78	68.33±7.78	57.27±13.97
EC	69.11±6.45	66.78±7.87	67.00±7.48	71.62±10.95	61.93±9.91
ATT	62.22±15.32	66.77±9.13	68.00±10.30	59.59±21.95	73.95±16.07
EO	67.63±15.40	69.00±11.78	69.00±11.14	62.37±14.19	75.63±12.85
Non-ATT,ATT	70.35±9.40	68.84±8.64	69.00±8.00	67.18±9.19	70.51±13.49
EO,ATT	71.15±4.46	71.50±4.24	71.00±3.74	72.22±5.62	70.77±9.50
Non-ATT,EO	70.54±13.24	72.00±10.58	72.00±10.77	71.79±8.23	72.21±13.39
EC,EO	72.89±9.91	73.51±6.31	72.00±7.48	85.11±12.29	61.91±12.93
EC,ATT	73.37±1.81	75.78±4.01	74.00±2.00	75.88±12.21	75.69±8.81
Non-ATT,EC	76.79±10.91	73.66±9.75	74.00±10.68	85.90±10.72	61.41±11.81
Non-ATT,EC,EO	68.89±6.18	71.81±5.65	72.00±6.00	70.28±6.31	73.33±11.08
Non-ATT,EO,ATT	74.69±5.29	75.41±3.55	75.00±3.16	70.26±6.18	**80.56±6.81**
Non-ATT,EC,ATT	76.41±11.02	76.23±7.42	76.00±8.60	81.13±6.28	71.32±9.36
EC,EO,ATT	77.28±6.35	79.49±3.49	79.00±3.74	79.28±15.05	79.70±11.67
Non-ATT,EC,EO,ATT	**87.27±2.77**	**84.16±3.49**	**85.00±3.16**	**88.25±6.82**	80.06±6.24

Table 2: Classification Performance for Different Models

Model	Acc±SD(%)	AUROC±SD(%)	TPR±SD(%)	TNR±SD(%)
STTransformer	65.00±8.20	66.78±9.45	67.34±8.52	62.89±8.12
FFCL	72.00±5.92	74.56±6.78	71.24±6.65	73.12±5.89
CNNTransformer	77.00±5.64	79.34±6.25	76.89±5.94	76.45±5.67
SPaRCNet	81.00±3.72	**85.78±3.91**	87.12±4.24	79.56±3.64
BLAT	**85.00±3.16**	84.16±3.49	**88.25±6.82**	**80.06±6.24**

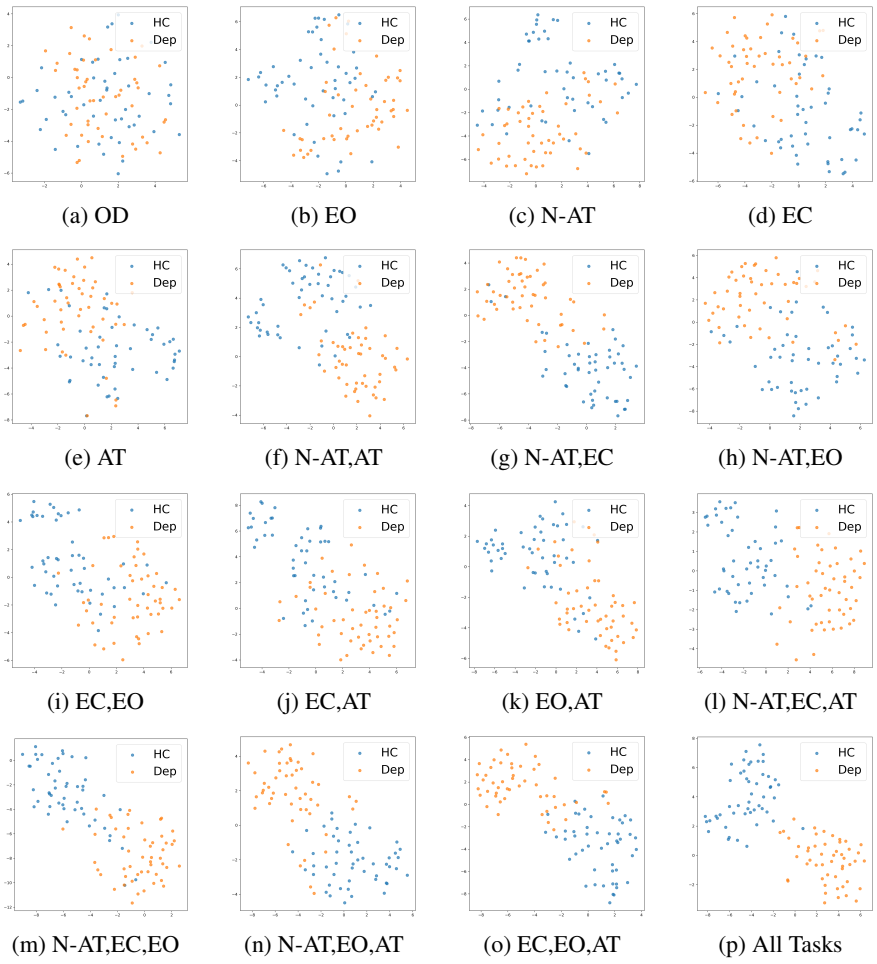
- **FFCL (Li, Ding, Zhang, Xiu, & et al., 2022):** This approach combines embeddings from both CNN and LSTM encoders to achieve feature fusion, enhancing the overall model capability.
- **ST-Transformer (Song, Jia, Yang, Xie, & et al., 2021):** This model introduces a multi-level EEG transformer designed to simultaneously learn spatial and temporal features, demonstrating empirical superiority over EEGNet.

Implement Details

In the configuration of the BLAT model, we employ 8 attention heads and 4 Transformer layers, with the temperature parameter set to $T = 2$. The Adam optimizer is utilized, featuring a learning rate of 1×10^{-3} and an L2 regularization term of 1×10^{-5}. The training, validation, and testing workflows are orchestrated using the PyTorch Lightning framework, with a maximum of 100 epochs while monitoring the AUROC metric for binary classification performance. The GPU model employed is 3660, and the PyTorch version is 1.12.1. The dataset, comprising 100 subjects, is divided into training, validation, and test sets in a 60:20:20 ratio. An independent test set is reserved for the final autonomous evaluation of the model's predictive performance.The dataset partitioning is performed using a random seed to ensure consistency across five experiments, thereby eliminating any risk of data leakage. The time series features are directly constructed by concatenating the time series from four tasks. Experimental results suggest that the order of concatenation has minimal impact on the direct identification of depressed subjects from raw data.

Results and Discussion

The Table 1 demonstrates that the task combination Non-ATT, EC, EO, and ATT achieves the best performance across all metrics: the F1-Score reaches 87.27%, significantly higher than other combinations; AUROC is 84.16%, reflecting strong capability in distinguishing positive and negative samples; accuracy is 85.00%, the highest value in the table; TPR is 88.25%, indicating the model's effectiveness in identifying positive cases; meanwhile, TNR is 80.06%, which, despite not being the highest, remains at a high level. This indicates that integrating all task features (Non-ATT, EC, EO, ATT) can significantly enhance the overall performance of the model. In comparison, mid-level combinations such as EC, EO, ATT perform relatively well with an F1-Score of 77.28% and AUROC of 79.49%, suggesting balanced performance even with reduced features. However, single-task features (e.g., Non-ATT or EC) show poor performance, particularly Non-ATT with an F1-Score of only 61.38%. These findings highlight the complementary nature of task features, where integrating multiple features significantly improves classification performance, offering insights for future optimization. Future research can focus on understanding the individual contribu-



(a) OD (b) EO (c) N-AT (d) EC

(e) AT (f) N-AT,AT (g) N-AT,EC (h) N-AT,EO

(i) EC,EO (j) EC,AT (k) EO,AT (l) N-AT,EC,AT

(m) N-AT,EC,EO (n) N-AT,EO,AT (o) EC,EO,AT (p) All Tasks

Figure 3: The t-SNE method was employed to conduct scatter plot analysis on the dimensionality-reduced features of subjects in the dataset. After training on the training set division across all task combinations, t-SNE was applied to all subjects. Subplots are represented as follows: OD = Original Data, EO = Eyes-open, N-AT = Non-ATT, EC = Eyes-closd, AT = ATT.

tions of each task and exploring additional combinations to further optimize the model.

Table 2 presents the classification performance of various models. Notably, the BLAT model outperforms other models across several key performance metrics. Specifically, BLAT achieves a test accuracy (Acc) of 85.00%, which is significantly higher than that of SPaRCNet (81.00%), CNNTransformer (77.00%), and FFCL (72.00%). This substantial performance improvement indicates that the BLAT model possesses a stronger classification capability for specific tasks. Furthermore, the area under the receiver operating characteristic curve (AUROC) is 84.16%, highlighting its exceptional performance in distinguishing between positive and negative class samples. The BLAT model also excels in true positive rate (TPR), reaching 88.25%, indicating its effectiveness in capturing positive class samples. The true negative rate (TNR) is 80.06%, further enhancing the model's overall capability in multi-class classification tasks. In conclusion, the BLAT model demonstrates superior performance across multiple metrics, particularly in accuracy, showcasing its unique advantages and stronger predictive capability. These results suggest that the BLAT model holds significant promise for tasks related to electroencephalogram (EEG) signal analysis, warranting further exploration in future research.

In the ablation study detailed in Table 3, the significance of each component in the performance of the BLAT model is evident. The complete model, integrating position encoding, channel embedding, and the Transformer, achieves optimal results with a test accuracy of 85.00% and an AUROC of 84.16%. Removing position encoding causes a decline in performance, reducing test accuracy to 81.00% and AUROC to 81.50%, underscoring its essential role in providing relative positional information necessary for capturing temporal dependencies crucial for EEG signal analysis. Further degradation is noted with the exclusion of the Transformer, where test accuracy diminishes to 79.00% and AUROC to 78.03%, highlighting its importance in modeling complex patterns and long-range dependencies. Additionally, the absence of channel embedding results in a test accuracy drop to 78.00% and AUROC to 77.66%, indicating its role in preserving feature

Table 3: Results of Ablation Study

Position Encoding	Channel Embedding	Transformer	Acc±SD(%)	AUROC±SD(%)	TPR±SD(%)	TNR±SD(%)
✓		✓	78.00±11.66	77.66±12.33	79.90±11.27	75.43±14.88
✓	✓		79.00±8.60	78.03±9.06	81.81±12.33	74.25±14.31
	✓	✓	81.00±8.00	81.50±7.75	81.64±9.50	**81.37±10.89**
✓	✓	✓	**85.00±3.16**	**84.16±3.49**	**88.25±6.82**	80.06±6.24

channel information critical for maintaining model accuracy and robustness. Collectively, position encoding, the Transformer, and channel embedding are vital for the model's optimal performance, each contributing uniquely to capturing and leveraging the diverse features and dependencies inherent in EEG signals.

Figure 3 presents the results of t-SNE dimensionality reduction and scatter plot analysis for the all subjects. The visualization results of the t-SNE scatter plot are highly consistent with the model's classification accuracy, further validating the importance of integrating multi-task features. Individual task-specific features, such as Non-ATT or EC, demonstrate limitations in effectively distinguishing between the depression group and the healthy control group, which aligns with their relatively low classification accuracy. Notably, when all task features are combined, the model achieves its highest classification performance at 85.00%, highlighting the complementary nature of features derived from different tasks and their ability to provide a more comprehensive representation of participants' conditions. Therefore, the t-SNE scatter plot not only offers an intuitive visualization of feature distributions but also provides strong evidence to support the reasons behind the model's improved performance.

Conclusion

This study presents a novel multi-task framework for detecting adolescent depression by integrating attention and resting task EEG signals through the Biosignal Learning and Agent Transformer (BLAT). Our method segments, channels, and encodes EEG data, effectively processing both attention and resting task signals. By applying BLAT to classify these combined signals, we achieved an 85% accuracy rate in an independent test set, highlighting the approach's promise for early depression screening and the potential of multi-task EEG data fusion to enhance classification performance. The integration of features from multiple tasks, supported by t-SNE analysis, significantly enhances the model's ability to differentiate between depressed and healthy subjects. This suggests that tasks provide complementary information, leading to a more comprehensive depression assessment. Our method's high classification accuracy with a sample of 100 adolescents underscores its robustness and generalizability.

The current study presents promising results and outlines key future research directions to improve EEG-based depression detection. Firstly, a deeper examination of task contributions could refine feature extraction and integration. Secondly, exploring new task combinations and EEG tasks, such as emotion induction or cognitive load, may enhance feature diversity and performance. Thirdly, fine-tuning the BLAT model through hyperparameter optimization or advanced architectures could increase accuracy. Fourthly, cross-dataset validation is essential to assess generalizability across diverse populations. Lastly, translating research into clinical practice through large-scale trials and collaboration with healthcare professionals is crucial for real-world implementation.In conclusion, integrating multi-task EEG data with the BLAT model offers a promising method for early adolescent depression screening. This study aids in developing objective, efficient diagnostic tools vital for addressing adolescent depression. Future research will refine these findings to increase the clinical applicability of EEG-based detection methods.

Acknowledgment

This work was supported in part by the STI 2030-Major Projects, under Grant 2022ZD0208900; in part by the National Key R&D Program of China, under Grant 2023YFE0207800; in part by the National Natural Science Foundation of China under Grant 62306120 and Grant U22A20293.

References

Chen, Tianyangli, Liuchuanjian, Caohuiyin, Lanjinyin, Yangjun, ... et al. (2024). Related factors of psychological therapy fear in adolescent depression patients. , *26*(12), 1315-1321.

Hamilton, J. P., Etkin, A., Furman, D. J., Lemus, M. G., Johnson, R. F., Gotlib, I. H., & et al. (2012). Functional neuroimaging of major depressive disorder: a meta-analysis and new integration of baseline activation and neural response data. *American Journal of Psychiatry*, *169*, 693–703.

Hao, S., Jiaqing, L., Shurong, C., Zhaolin, Q., Lanfen, L., Xinyin, H., ... et al. (2021). Multi-modal adaptive fusion transformer network for the estimation of depression level. *Sensors*, *21*(14), 4764-4764.

Hartmann, K. G., Schirrmeister, R. T., & Ball, T. (2017). Hierarchical internal representation of spectral features in deep convolutional networks trained for eeg decoding. *2018 6th International Conference on Brain-Computer Interface (BCI)*, 1-6.

Jing, J., Ge, W., Hong, S., Fernandes, M. B., Lin, Z., Yang, C., ... et al. (2023). Development of expert-level classification of seizures and rhythmic and periodic patterns during eeg interpretation. *Neurology*, *100*(17), e1750–e1762. (Epub 2023 Mar 6)

Kohei, F., Hiroki, W., Atsushi, M., Junpei, S., Yasushi, N., & S., I. A. (2022). Impact of depressed state on attention and language processing during news broadcasts: Eeg analysis and machine learning approach. *Scientific Reports*, *12*(1).

Li, H., Ding, M., Zhang, R., Xiu, C., & et al. (2022). Motor imagery eeg classification algorithm based on cnn-lstm feature fusion network. *Biomedical Signal Processing and Control*, *72*, 103342.

Liu, Y., & et al. (2021). Multi-modal attention network for eeg-based emotion and depression recognition. *IEEE Journal of Biomedical and Health Informatics*.

Malhi, G. S., Mann, J. J., & et al. (2018). Depression. *The Lancet*, *392*, 2299–2312.

Olbrich, S., & Arns, M. (2013). Eeg biomarkers in major depressive disorder: discriminative power and prediction of treatment response. *International Review of Psychiatry*, *25*, 604–618.

Peh, W. Y., Yao, Y., & Dauwels, J. (2022). Transformer convolutional neural networks for automated artifact detection in scalp eeg. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 3599-3602.

Qi, K., Hua, L., ni Tong, J., jie Xiong, J., yin Pan, Z., Li, N., ... long Jin, Y. (2025). A study on sleep quality under the shadow of school bullying: The interwoven effects of depressed mood, low self-esteem, and negative parenting practices. *Acta Psychologica*, *253*, 104717-104717.

Sakhavi, S., Guan, C., Yan, S., & et al. (2018). Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(11), 5619–5629. (Epub 2018 Mar 9)

Song, Y., Jia, X., Yang, L., Xie, L., & et al. (2021). Transformer-based spatial-temporal feature learning for eeg decoding. *arXiv preprint arXiv:2106.11170*.

Yang, C., Westover, M. B., & Sun, J. (2023). Biot: Cross-data biosignal learning in the wild. *ArXiv*, *abs/2305.10351*.

Yong, N., Ren, Y. L., Wang, C. Y., Yang, C., & et al. (2024). Investigation on the attitude of middle school students with depression in the city toward professional psychological help-seeking. *Patient preference and adherence*, *18*, 11-13.

Zhang, R., Zong, Q., Dou, L., Zhao, X., & et al. (2019). A novel hybrid deep learning scheme for four-class motor imagery classification. *Journal of Neural Engineering*, *16*(6), 066004.