MTCA: A Multi-Scale Temporal and Cross-Channel Attention Model for Depression Recognition

1st Kancheng Shen
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
221124120233@zjut.edu.cn

2nd Yan Ling
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
yanling@zjut.edu.cn

3rd Tianyu Zhou
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
211124030099@zjut.edu.cn

4th Ruiji Xu
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
xrj@zjut.edu.cn

5th Zhihu Zhou
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
zhouzhihu@zjut.edu.cn

6th Keji Mao*
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, China
maokeji@zjut.edu.cn

***Abstract*—Spatiotemporal modeling using functional near-infrared spectroscopy (fNIRS) has emerged as a promising approach for objective depression diagnosis, leveraging distinct haemodynamic responses in task-evoked cortical regions between patients with major depressive disorder (MDD) and healthy controls (HC). Prior studies have shown that deep learning models based on fNIRS can effectively assist in the diagnosis of depression. However, mainstream approaches often fail to capture multi-scale temporal dependencies or overlook the importance of interactions across multiple channels. To address this issue, this paper proposes a multi-scale temporal and cross-channel attention (MTCA) model for depression recognition. Our framework integrates: (i) a multi-resolution temporal-aware module that captures short- and long-range dependencies via parallel dilated convolutions; (ii) a channel-wise attention mechanism inspired by iTransformer, which models cross-channel correlations; and (iii) a pseudo-sequence augmentation strategy that synthesizes temporally coherent samples to mitigate data scarcity. Evaluated on a clinically collected fNIRS dataset (N=50, 30 MDD + 20 HC), MTCA achieves 90.8% weighted accuracy and 89.4% unweighted accuracy.These results demonstrate MTCA's effectiveness for depression diagnosis.**

***Index Terms*—Depression, fNIRS, deep learning, multi-scale temporal, cross-channel attention**

I. Introduction

Depression is a widespread condition affecting people globally. During the first year of the COVID-19 pandemic, the rates of common disorders such as depression and anxiety increased by 25% [1]. This further expanded the population affected by these conditions. Additionally, depression has a high recurrence rate, with many patients experiencing multiple relapses. If left untreated, depression may result in suicidal behaviour—a major public-health concern [2].

Currently, the diagnosis of depression relies on self-reports and behavioral assessments from patients, but it lacks clear biomarkers or definitive criteria. Treatment for depression also primarily aims at alleviating symptoms, with limited advancements in effectiveness. Consequently, it is essential to identify biomarkers that demonstrate high sensitivity and specificity to provide more accurate evidence for individualized clinical diagnosis and treatment.

FNIRS is a non-invasive neuroimaging technique that shows great potential for diagnosing depression [3], [4]. This technique measures the brain's haemodynamic response to activity by detecting the absorption of near-infrared light in the 650-950 nm range [5], [6]. By analyzing this absorption, researchers can assess changes in the relative concentrations of oxyhaemoglobin (HbO) and deoxyhaemoglobin (HbR) in the brain, providing valuable information about brain activity.

FNIRS data is a type of time series data. To effectively extract features and learn the complex information embedded within them, deep learning techniques have shown great promise [7]. In recent years, these techniques have advanced rapidly, achieving remarkable success in areas such as time series prediction and medically assisted diagnosis. Methods such as Convolutional Neural Networks (CNN) [8], Recurrent Neural Networks (RNN) [9], and Long Short-Term Memory networks (LSTM) [10] can automatically extract features and learn complex patterns from input data.

Cyrus Su Hui Ho et al. [11] utilised a deep learning framework based on end-to-end CNN using a large-scale fNIRS dataset to classify and grade the severity of MDD. This study demonstrated the potential of integrating fNIRS systems with AI algorithms for the classification and staging of MDD in clinical settings using large datasets. Shao Kai et al. [12] proposed a depression recognition model that employs data augmentation and pseudo-sequence methods to address the challenges posed by the difficulty of acquiring and the

limited availability of fNIRS data. Ma Tengfei et al. [13] developed an Attention LSTM model, which incorporates attention mechanisms to focus on long-term connectivity, along with a neural network, InceptionTime, designed to capture short-term feature information. Their proposed model more comprehensively and consistently mines hidden information within fNIRS data. Wang Zenghui et al. [14] introduced a transformer-based fNIRS classification network that investigates both spatial and channel-level representations to enhance data utilization and improve network representation.Nevertheless, two key challenges remain unsolved:

1) Multi-scale temporal modelling: Existing single-scale kernels or fixed-length LSTM windows cannot simultaneously capture both fine-grained and long-term dynamics.
2) Cross-channel interaction: Functional connectivity among the 22–52 fNIRS channels encodes valuable inter-regional coupling information. CNN-based or vanilla Transformer pipelines either process each channel independently or collapse all channels into one token, thus discarding genuine physiological interaction.

To bridge these gaps, we propose the Multi-scale Temporal and Cross-channel Attention Model (MTCA)—a lightweight yet comprehensive framework tailored for small-sample fNIRS depression recognition. Specifically, our contributions are:

- We introduce the first fNIRS model that jointly learns multi-scale temporal dependencies and cross-channel correlations, achieving a weighted accuracy of 90.8% on a clinical dataset of 30 MDD and 20 HC.
- A Temporal-Aware Module employing bidirectional dilated convolutions is designed to cover haemodynamic responses.
- We adapt the recent iTransformer architecture to fNIRS by treating each channel as an independent token.

II. Related works

Early fNIRS depression studies relied on hand-crafted haemodynamic features (mean, peak, slope, etc.) with classical classifiers (SVM, RF, k-NN) [15], [16]. While transparent, these methods are sensitive to feature selection and assume channel independence, failing to capture long-range temporal dependencies or cross-channel functional connectivity.

Deep learning approaches, particularly CNNs, have become prevalent for end-to-end fNIRS classification [11], [12]. However, vanilla convolutions face two limitations: (1) fixed receptive fields cannot simultaneously capture fine-grained and sustained haemodynamic changes; (2) channels are processed independently until late fusion, ignoring physiologically meaningful cross-channel interactions. Recent Transformer-based methods [14], [17] focus primarily on temporal encoding, underutilizing cross-channel dependencies.

Our work addresses these gaps by introducing MTCA, which jointly models multi-scale temporal dynamics via dilated bidirectional convolutions and cross-channel correlations through an iTransformer-inspired attention mechanism.

III. Method

A. Overview

The aim of our work is to develop a novel model for fNIRS data as shown in Fig. 1. (a) Preprocessing of fNIRS data at the data collection stage and converting them to haemoglobin concentration variation data. (b) the data is segmented into multiple pseudo-sequences with a view to increasing the amount of data. (c) Build a network architecture called MTCA. It includes a temporal-aware module using an inflated 1-D CNN and a cross-channel attention module using iTransformer. The proposed model is outlined below:

$$X_1 = Conv2d(Data) \tag{1}$$

$$W = \sigma(TA(X_1) \oplus TA(Rev(X_1))) \tag{2}$$

$$X_2 = W \odot X_1 + X_1 \tag{3}$$

$$Class = Projection(CCA(X_2)) \tag{4}$$

where Data is the fNIRS data after preprocessing, Conv2d denotes 2-D convolution, TA denotes the temporal-aware module, CCA denotes the cross-channel attention module, W denotes the weight matrix, and Projection denotes the fully connected layer. $\oplus$ is the matrix addition, $\odot$ is the matrix Hadamard product, Rev is the data inversion operation, and σ is the sigmoid activation function.

B. Data Acquisition

Due to the absence of a publicly available fNIRS dataset for depression research, we prospectively collected data from a psychiatric hospital. In total, 30 MDD and 20 HC were enrolled. The dataset contains sensitive personal and clinical information. In accordance with the ethics approval obtained from a psychiatric hospital and to protect participant privacy, the data cannot be made publicly available. However, anonymized data may be shared upon reasonable request and with appropriate ethical clearance, subject to compliance with relevant data protection regulations.

fNIRS mainly exploits the difference in absorbance of near-infrared light at different wavelengths of 600-900 nm between HbO and HbR in brain tissues to detect the haemodynamic activities in the cerebral cortex in real time and directly [18], [19].

We used the Hitachi ETG-4000 functional near-infrared imaging system (695 nm and 830 nm; 22 channels; 10 Hz sampling rate).The locations of emitters, detectors, and the corresponding channel distribution are shown in Fig. 2.

The experiment employed a verbal fluency task (VFT) [20], a standard paradigm for assessing prefrontal activation in depression. Participants were instructed to verbally generate as many words as possible beginning with a given letter. Throughout the 125-second recording session, fNIRS data were acquired at a sampling rate of 10 Hz, resulting in 1,250

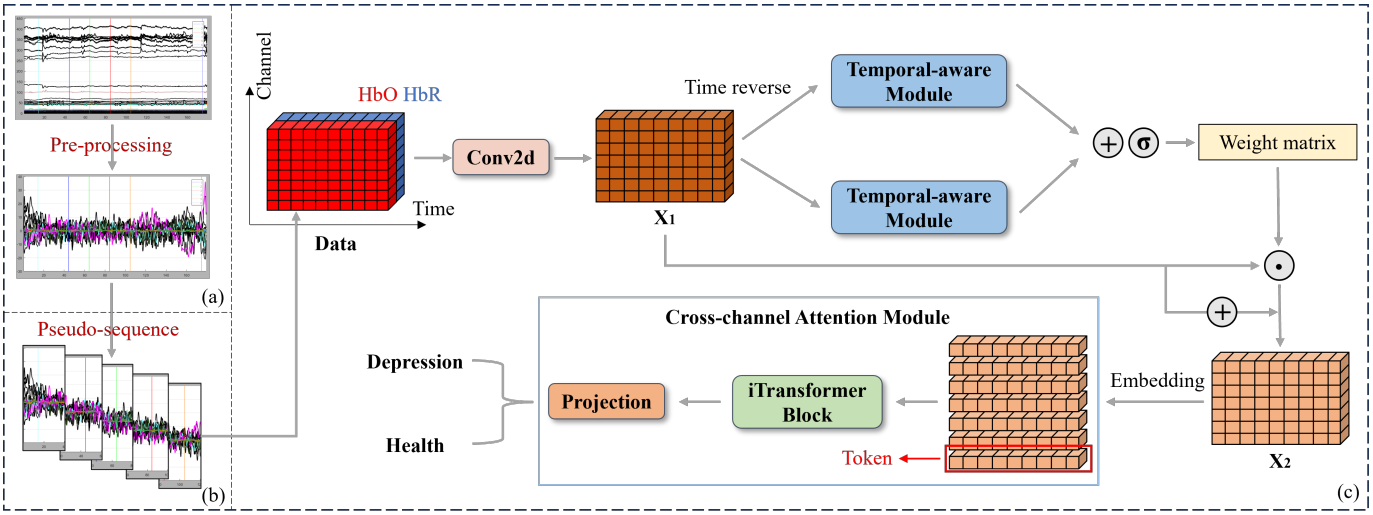

Fig. 1: Flow chart of the proposed model. (a) Pre-processing: conversion of raw optical density to ΔHbO/ΔHbR. (b) Pseudo-sequence augmentation: slices the data according to certain rules. (c) Network architecture: temporal-aware and cross-channel attention modules.

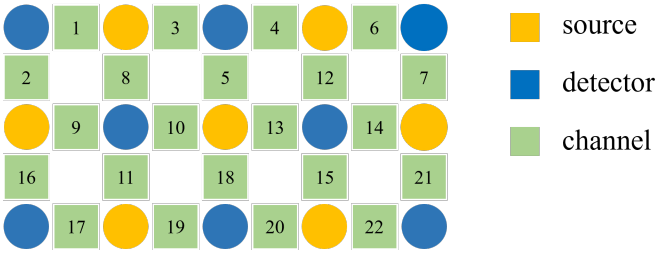

Fig. 2: Channel distribution of the near-infrared device.The orange circles denote near-infrared light source probes, the blue circles denote detector probes, and the green squares denote channels.

time points per subject. The signal comprises measurements from two near-infrared wavelengths, which are subsequently processed to derive hemodynamic responses.

C. Data Pre-processing

Preprocessing of fNIRS is well established. Because HbO and HbR are reliable indicators of cerebral blood-flow changes related to depression, we restrict subsequent analyses to these two chromophores.

The modified Beer-Lambert law is applied to convert the optical density ΔOD to the concentration changes of ΔHbO and ΔHbR caused by near-infrared light absorption, with the following equations:

$$\begin{bmatrix}\Delta HbO\\ \Delta HbR\end{bmatrix} = d \cdot DPF^{-1} \cdot \begin{bmatrix}\epsilon_{HbO,\lambda_1} & \epsilon_{HbO,\lambda_2}\\ \epsilon_{HbR,\lambda_1} & \epsilon_{HbR,\lambda_2}\end{bmatrix}^{-1} \cdot \begin{bmatrix}\Delta OD_{\lambda_1}\\ \Delta OD_{\lambda_2}\end{bmatrix} \quad (5)$$

d is the distance between the light source and the detector (in cm), DPF is the differential path factor, which is used to correct the scattering path of light in the tissue, λ_1 and λ_2 are the different illumination wavelengths, and ϵ is the molar extinction coefficient matrix of haemoglobin (in $\mathrm{cm}^{-1} \cdot \mathrm{M}^{-1}$).

Fig. **??** displays the temporal dynamics of ΔHbO and ΔHbR for one MDD and one HC following pre-processing.

D. Data Pseudo-sequence

To mitigate the challenge of limited clinical samples while preserving subject-wise independence, we adopt a a pseudo-sequential approach during training. Specifically, each subject's fNIRS recording (sampled at 10 Hz) is partitioned into segments of 50 seconds (500 time points). The step size is set to half of the window size. Specifically, given that fNIRS measures two hemodynamic components-HbO and HbR—each recording is structured as a 3D tensor of shape $2 \times C \times T = 2 \times 22 \times 500$, where $C = 22$ denotes the number of fNIRS channels and $T = 500$ represents the temporal duration. Critically, all segments from the same subject are assigned to the same data split.

E. Temporal-Aware Module

A 2-D convolution layer with a single kernel is first applied to extract coarse-grained representations and to project the input into a two-dimensional latent space.

To extract finer-grained temporal information, a bi-directional temporal-aware module based on 1D dilated convolution is added after the 2-D convolution, as shown in Fig. 3. Dependencies are learnt from the forward and backward directions, and complementary information about the past and future of the fNIRS data is integrated. Specifically, the input is first divided into two parts X_1 and $\mathrm{Rev}(X_1)$ in both forward and reverse order, and then entered into two identical temporal-aware modules, followed by summing the two parts and assigning the weight property, and finally multiplying the weights with the input data and using residual concatenation, to obtain the output. Each temporal-aware module mainly

consists of a 1D dilated convolution, a batch normalisation, a ReLU function, and a spatial dropout Composition. Specifically, the dilated convolution adopts a kernel size of 2 and a dilation rate of 1.

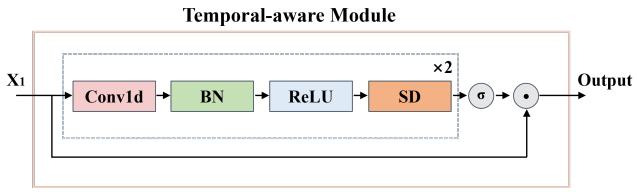

Fig. 3: Overview of temporal-aware module.

F. Cross-Channel Attention Module

In the field of temporal data modelling, Transformer-based architectures usually encode multiple variables at the same time step into a unified multidimensional temporal token, and based on the temporal token, model the temporal correlation between different moments using the attention mechanism [21], [22]. However, this architecture has the following limitations:

- Data points at the same time step may have different physical meanings, be collected at unaligned times and have scale differences. Forcing them to be encoded as a uniform temporal token and no longer distinguishing between different channels leads to the following problems: correlations between multiple variables are eliminated and efficient variable-based representations cannot be learnt. Due to the time lag of variables, the amount of information contained in the temporal token at one point in time is more limited, and starting from the temporal token at each moment in time may not be conducive to modelling global temporal correlations.
- When modelling the long-term correlation of data along the temporal direction, Transformer will face the challenges of performance degradation as well as sharp increase in computation as the length of the history window keeps increasing.

To address the above issues, Yong Liu et al. [23] proposed the iTransformer model. The model considers different variables separately, each of which is encoded as an independent token, and uses the attention mechanism to model the correlation between different variables, and has achieved remarkable results. Fig. 4 illustrates a comparison of how tokens are encoded in Transformer and iTransformer.

Inspired by the effectiveness of iTransformer in modeling multivariate time series, we adapt its core mechanism to design a cross-channel attention module tailored for fNIRS-based depression recognition. As illustrated in fig. 5, Each fNIRS channel is treated as an independent token, and Multivariate Attention is computed across these channel tokens to explicitly capture functional dependencies between cortical region. Layer normalization is applied per channel token to mitigate scale discrepancies arising from heterogeneous hemodynamic

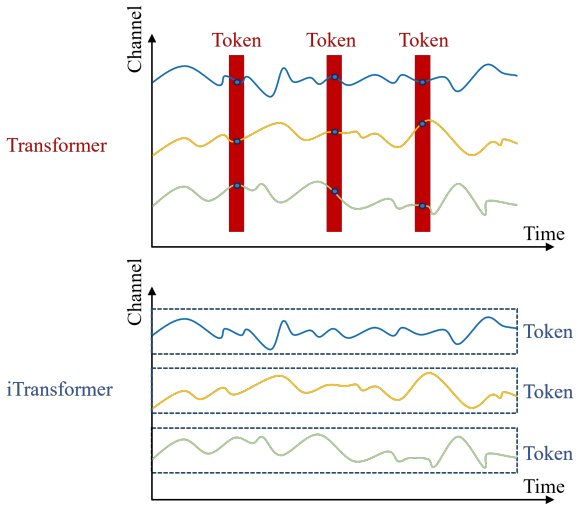

Fig. 4: Differences in token encoding methods between Transformer and iTransformer.

responses. Subsequently, Feed-forward networks process each token independently, enhancing its representation with global contextual information.

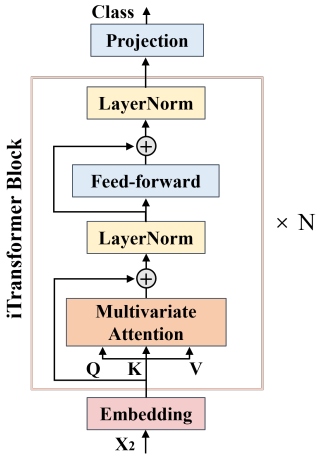

Fig. 5: Overview of cross-channel attention module.

IV. EXPERIMENTS AND RESULT

A. Training Strategy

To mitigate overfitting on small-sample fNIRS data, we employ label smoothing [24] with $\epsilon = 0.1$, converting hard labels into soft distributions. The cross-entropy loss with label smoothing is: $\mathcal{L}_{\text{smooth}} = -\sum_{i=1}^{K} \tilde{y}_i \log(p_i)$, where $\tilde{y}_i$ assigns $(1-\epsilon) + \epsilon/K$ to the correct class and ϵ/K to others.

We adopt 5-fold cross-validation [25], ensuring subject-level splitting to prevent data leakage before applying pseudo-sequence augmentation. The Adam optimizer is used with learning rate 1×10^{-3}, weight decay 1×10^{-4}, cosine annealing over 100 epochs, batch size 32, and early stopping

patience of 20 epochs. All experiments run on an NVIDIA RTX 3090 GPU with PyTorch 2.0.

B. Depression Classification Results

We compare MTCA against MCNet [11] and fNIRS-T [14] using identical training schemes. Table I reports mean ± std % for unweighted accuracy (Ua, standard classification accuracy) and weighted accuracy (Wa, recall-weighted by class size to handle imbalance): Wa $= \frac{w_1 Recall_1 + w_0 Recall_0}{w_1 + w_0}$, where $w_i = N/N_i$ and N_1, N_0 are class sizes.

It can be observed that MTCA outperforms other models in both Ua and Wa. Specifically, the proposed MTCA achieves an average Ua of 89.44%, an average Wa of 90.80%. MCNet models each channel independently and merges them only at the final fully-connected layer. This hampers cross-channel interaction modeling and yields a Wa of only 50%. In contrast, FNIRS-T explicitly incorporates cross-channel relationships, leading to significantly improved performance over MCNet. Our proposed MTCA model addresses these aspects more comprehensively: it employs a bidirectional timing-aware module for enhanced temporal modeling and utilizes an advanced cross-channel attention module (iTransformer) to effectively capture channel-level interactions. This dual-level modeling strategy enables MTCA to achieve state-of-the-art results.

TABLE I: Comparison of different models

Model	Ua (%)	Wa (%)
MCNet [11]	71.95 ± 16.21	50.00 ± 0.00
fNIRS-T [14]	87.57 ± 7.26	85.53 ± 9.93
UniRepLKNet [26]	85.90 ± 6.63	87.01 ± 7.27
MTCA*	**89.44 ± 5.94**	**90.80 ±6.94**

C. Training and Test Curves

Fig. 6 and 7 show the five-fold cross-validation accuracy of MTCA. Although individual folds exhibit intermittent fluctuations—expected in small-sample fNIRS studies—the median trend steadily rises and saturates after approximately 25 epochs. The inter-fold standard deviation remains below 5% during the last 10 epochs, indicating statistical stability despite local variance.

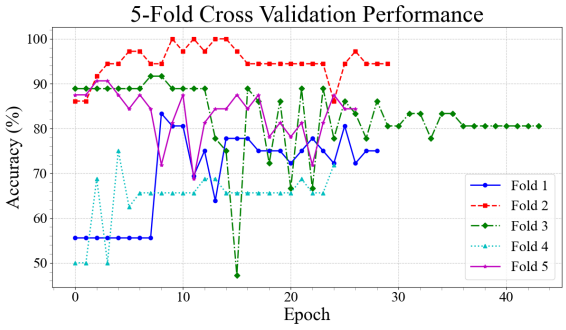

Fig. 6: Unweighted accuracy curves for each fold under 5-fold cross-validation.

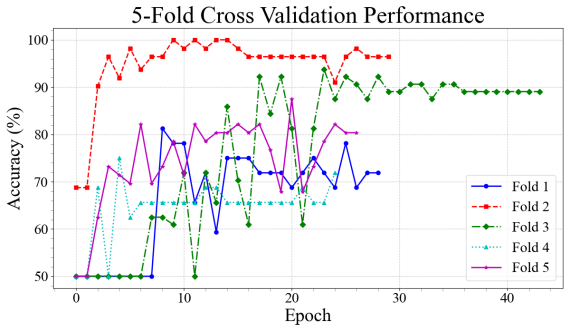

Fig. 7: Weighted accuracy curves for each fold under 5-fold cross-validation.

D. Ablation Experiments

We use the following comparative experiments to verify the validity of the different modules in the model. TA + CCA is our proposed model. We performe the experiment after removing the TA module and CCA module respectively. The comparison of the experimental results is shown in Table II. The experimental results show that removing either the TA module or the CCA module leads to a significant decrease in model accuracy. This validates the critical role of the proposed TA module in effectively modeling temporal information, and the CCA module in effectively capturing cross-channel dependencies.

TABLE II: Comparison of experimental results

Model	Ua (%)	Wa (%)
MTCA*	**89.44**	**90.80**
MTCA without TA	79.03	79.64
MTCA without CCA	74.86	74.50

V. CONCLUSION

In this paper, we propose a multi-scale temporal and cross-channel model for depression recognition. To the best of our knowledge, MTCA is the first framework to jointly model multi-scale temporal dynamics (via parallel dilated convolutions) and cross-channel functional connectivity (through an iTransformer-inspired attention mechanism) within a unified architecture. Experimental results on a clinically collected dataset (N=50) demonstrate that MTCA achieves 90.8% weighted accuracy, significantly outperforming existing methods.Critically, this performance enables practical clinical deployment: MTCA can be integrated into portable fNIRS systems to deliver objective depression screening.

Currently we have only used fNIRS data, whereas the diagnosis and identification of depression can be informed by multimodal data, such as electroencephalograms (EEGs) [27], magnetic resonance imaging (MRIs), and behavioural data (e.g. questionnaires, recordings of daily activities, etc.). If multimodal data are not combined, important complementary information will be missed. Therefore, we will fuse fNIRS with EEG, MRI, and behavioural data in future work. By

establishing a multimodal data fusion model, we will make full use of the advantages of different modal data to improve the accuracy and reliability of depression identification. For example, combining the spatial information of fNIRS data and the temporal information of EEG data provides a more comprehensive understanding of the brain activity characteristics of patients with controls.

ACKNOWLEDGMENT

This research is supported by "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province under No. 2026C02A1019, 2026C02A1222; Zhejiang Provincial Natural Science Foundation of China under Grant No: LTGG23F020002; Zhejiang University of Technology School-enterprise Cooperation Project under Grant No: KYY-HX-20250798,KYY-HX-20230493, KYY-HX-20250381, KYY-HX-20250423; Medical and Health Science Program of Zhejiang Province, Grant No: WKJ-ZJ-26092.

REFERENCES

[1] W. H. Organization, *World mental health report: Transforming mental health for all*. Geneva: World Health Organization, 2022.

[2] H. Herrman, V. Patel, C. Kieling, M. Berk, C. Buchweitz, P. Cuijpers, T. A. Furukawa, R. C. Kessler, B. A. Kohrt, M. Maj *et al.*, "Time for united action on depression: a lancet–world psychiatric association commission," *The Lancet*, vol. 399, no. 10328, pp. 957–1022, 2022.

[3] S. Zheng, C. Lei, T. Wang, C. Wu, J. Sun, and H. Peng, "Feature-level fusion for depression recognition based on fnirs data," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 2906–2913.

[4] J. Zhong, G. Ma, L. Zhang, Q. Wang, S. Qiao, H. Peng, and B. Hu, "Spatio-temporal scale information fusion of functional near-infrared spectroscopy signal for depression detection," *Knowledge-Based Systems*, vol. 283, p. 111165, 2024.

[5] A. Villringer and U. Dirnagl, "Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging," *Cerebrovascular and brain metabolism reviews*, vol. 7, no. 3, pp. 240–276, 1995.

[6] E. J. Doherty, C. A. Spencer, J. Burnison, M. Čeko, J. Chin, L. Eloy, K. Haring, P. Kim, D. Pittman, S. Powers *et al.*, "Interdisciplinary views of fnirs: Current advancements, equity challenges, and an agenda for future needs of a diverse fnirs research community," *Frontiers in integrative neuroscience*, vol. 17, p. 1059679, 2023.

[7] S. Ding, K. Wang, W. Jiang, C. Xu, H. Bo, L. Ma, and H. Li, "fnirs emotion recognition method based on multi-scale fusion features," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2024, pp. 5202–5209.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, pp. 84–90, 2012.

[9] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," *arXiv preprint arXiv:1409.2329*, 2014.

[10] A. Graves, "Long short-term memory," *Supervised sequence labelling with recurrent neural networks*, pp. 37–45, 2012.

[11] C. S. H. Ho, J. Wang, G. W. N. Tay, R. Ho, S. F. Husain, S. K. Chiang, H. Lin, X. Cheng, Z. Li, and N. Chen, "Interpretable deep learning model for major depressive disorder assessment based on functional near-infrared spectroscopy," *Asian Journal of Psychiatry*, vol. 92, p. 103901, 2024.

[12] K. Shao, Y. Hao, L. Hu, X. Zong, and M. Chen, "Data augmentation and pseudo-sequence of fnirs for depression recognition," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 2223–2226.

[13] T. Ma, H. Lyu, J. Liu, Y. Xia, C. Qian, J. Evans, W. Xu, J. Hu, S. Hu, and S. He, "Distinguishing bipolar depression from major depressive disorder using fnirs and deep neural network," *Progress In Electromagnetics Research*, vol. 169, pp. 73–86, 2020.

[14] Z. Wang, J. Zhang, X. Zhang, P. Chen, and B. Wang, "Transformer model for functional near-infrared spectroscopy classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2559–2569, 2022.

[15] J. Chao, S. Zheng, H. Wu, D. Wang, X. Zhang, H. Peng, and B. Hu, "fnirs evidence for distinguishing patients with major depression and healthy controls," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2211–2221, 2021.

[16] G. Wang, N. Wu, Y. Tao, W. H. Lee, Z. Cao, X. Yan, and G. Wang, "The diagnosis of major depressive disorder through wearable fnirs by using wavelet transform and parallel-cnn feature fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023.

[17] X. Si, H. Huang, J. Yu, and D. Ming, "The fnirs-based emotion recognition by spatial transformer and wgan data augmentation towards developing a novel affective bci," *IEEE Transactions on Affective Computing*, 2024.

[18] S. C. Wriessnegger, D. Kirchmeyr, G. Bauernfeind, and G. R. Müller-Putz, "Force related hemodynamic responses during execution and imagery of a hand grip task: A functional near infrared spectroscopy study," *Brain and Cognition*, vol. 117, pp. 108–116, 2017.

[19] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application," *Neuroimage*, vol. 63, no. 2, pp. 921–935, 2012.

[20] V. Quaresima, P. Giosuè, R. Roncone, M. Casacchia, and M. Ferrari, "Exploring prefrontal cortex oxygenation in schizophrenia by functional near-infrared spectroscopy," in *Oxygen Transport to Tissue XXVII*. Bari: Springer, 2006, pp. 229–235.

[21] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in neural information processing systems*, vol. 34, pp. 22 419–22 430, 2021.

[22] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," *arXiv preprint arXiv:2211.14730*, 2022.

[23] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "itransformer: Inverted transformers are effective for time series forecasting," *arXiv preprint arXiv:2310.06625*, 2023.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[25] T. Fushiki, "Estimation of prediction error by using k-fold cross-validation," *Statistics and Computing*, vol. 21, no. 2, pp. 137–146, 2011.

[26] X. Ding, Y. Zhang, Y. Ge, S. Zhao, L. Song, X. Yue, and Y. Shan, "Unireplknet: A universal perception large-kernel convnet for audio video point cloud time-series and image recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 5513–5524.

[27] H. Chen, I. Y. Chan, Z. Dong, and T.-A. Samuel, "Unraveling trust in collaborative human-machine intelligence from neurophysiological perspective: a review of eeg and fnirs features," *Advanced Engineering Informatics*, vol. 67, p. 103555, 2025.