An EEG-Based Depressive Detection Network with Adaptive Feature Learning and Channel Activation

Chenyang Xu[1], Feiyi Fan[2], Jianfei Shen[2], Hanguang Wang[1], Zhongyi Zhang[1], Qinghao Meng[1,✉]

1. School of Electrical and Information Engineering, Tianjin University, Tianjin, China

2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

{xuchenyang, zy_zhang_auto, one_whg, qh_meng}@tju.edu.cn {fanfeiyi, shenjianfei}@ict.ac.cn

Abstract

Electroencephalography (EEG) plays a pivotal role in the diagnosis of various neurological conditions, most notably major depressive disorder (MDD). However, deep learning-based methods currently employed for MDD detection tasks exhibit inadequate generalization capabilities, particularly across different EEG electrode channels, and demonstrate limited feature representation capacity. In this paper, we present a novel approach referred to as adaptive feature learning (AFL), which leverages kernel embedding to facilitate the learning of domain-invariant features across subjects within a reproducing kernel Hilbert space. This method aims to enhance the model's ability to generalize across multiple subjects' EEG signals. Furthermore, our research revealed that batch normalization (BN) layers within the existing MDD detection network frequently result in feature channel suppression, potentially compromising the representation power of the features. To address this issue, we propose channel activation (CA), which employs decorrelation to reactivate suppressed feature maps, thereby enhancing the model's feature representation capability, particularly for subtle EEG changes. The effectiveness of the proposed methods is evaluated using the leave-one-subject-out protocol on MODMA and PRED+CT datasets, yielding detection accuracies of 90.56% (MODMA) and 96.51% (PRED+CT). Our experimental findings exhibit the superior performance of our method compared to state-of-the-art (SOTA) methods in terms of MDD recognition.

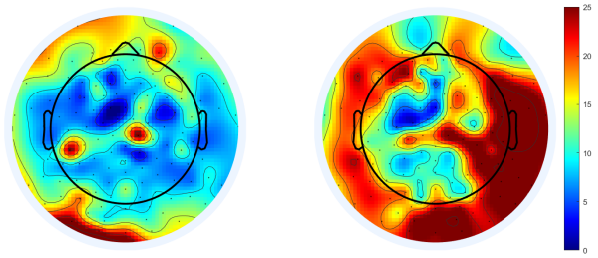
Keywords: Major Depressive Disorder (MDD); Deep Learning; Adaptive Feature Learning (AFL); Channel Activation (CA);

Introduction

MDD is a common neurological illness that causes symptoms such as insomnia, anxiety, and irritability, with severe cases leading to suicidal behavior. EEG is a non-invasive and cost-effective diagnostic tool for neurological disorders such as depression, seizures, Alzheimer's, Parkinson's, and emotion analysis (Saeidi et al., 2021).

Recently, numerous scholars conducted studies on depression using EEG data collected during resting states (Cai et al., 2020; Yang et al., 2018). With the growing popularity of artificial intelligence technology, machine learning (ML) and deep learning (DL)-based approaches to EEG signal recognition are gaining traction.

Recognizing the limitations of traditional machine learning algorithms and their reliance on expert feature engineering and selection, the widespread adoption of deep learning methods has revolutionized the field of MDD recognition using EEG signals. Through the application of deep learning, MDD recognition models are now capable of extracting more complex semantic features and identifying subtle changes in the EEG signals. Deep learning research in the domain of MDD recognition encompasses various approaches, including convolutional neural networks (CNN) (B. Liu, Chang, Peng, & Wang, 2022), graph convolutional networks (GCN) (H.-G. Wang, Meng, Jin, Wang, & Hou, 2023), recurrent neural networks (RNN) (H.-G. Wang, Meng, Jin, & Hou, 2023), Transformer (Qayyum, Razzak, Tanveer, Mazher, & Alhaqbani, 2023), and hybrid architectures that integrate these networks (Sam, Boostani, Hashempour, Taghavi, & Sanei, 2023).



(a) Subject 06 **(b) Subject 15**

Figure 1: Brain topography of EEG value distribution (Subject 06 and Subject 15 are both patients with MDD, MODMA dataset)

The significant variation observed in the EEG signals of depression patients, resulting in distinct data distributions as depicted in Figure 1, poses a challenge for traditional DL methods in addressing these differences. While existing domain adaptation (DA) based methods aim to bridge the distribution gap between the source and target domains (Jiang et al., 2023), they fail to account for the significant inter-subject divergence within the same domain for the MDD detection task. On the other hand, domain generalization (DG) based methods are applicable to a similar problem setting by treating different subjects as different domains (Ma et al., 2023), but they require manual assignment of domain labels.

Additionally, previous EEG-based MDD detection networks required the BN layer to learn additional independent features. However, as demonstrated in Eq. (1), previous research indicates γ_c (re-scale factor of BN layer) can be extremely small, which may suppress the EEG feature map $\tilde{x}_{wcij}$ (Huang, Yang, Lang, & Deng, 2018). Since MDD patients' EEG signal changes are inherently weak, suppressing the EEG feature maps can lead to poorer model representation capability. Existing methods for feature suppression in-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

volve adding attention mechanisms to enable the network to learn and reweight different feature channels (Q. Wang et al., 2020), but the suppressed feature values continue to play a role in downstream classification tasks, which degrades the model's performance (Shao et al., 2020).

$$\begin{aligned} \bar{x}_{wcij} &= (x_{wcij} - \mu) / \sigma, \\ \tilde{x}_{wcij} &= \gamma_c \bar{x}_{wcij} + \beta_c \end{aligned} \tag{1}$$

Where $\bar{x}_{wcij}$ denotes the standardized feature, $\tilde{x}_{wcij}$ represents the normalized feature. σ and μ respectively denotes the standard deviation and mean, respectively, as well as β_c and γ_c are the re-shifting and re-scaling factors for the c^{th} EEG feature channel, respectively.

In this work, firstly, we propose the AFL, which employs kernel embedding to map EEG features into a high-dimensional reproducing kernel Hilbert space (RKHS) to extract high-order moment domain-invariant features. The AFL enhances model generalization by extracting domain-invariant features from the EEG signals of different individuals, thus addressing the challenge of subpar generalization across various subjects. Secondly, we introduce the CA, which activates suppressed EEG feature channels through a decorrelation operation (Huang et al., 2018) following the BN layer. This activation increases the re-scale factor of EEG feature maps, thereby improving their representational ability for downstream MDD detection tasks. Our contributions can be summarized as follows:

- We propose the AFL to solve the varying EEG signal distributions among multiple subjects in the area of MDD detection. The AFL forces the network to learn domain invariant EEG features from multiple subjects, increasing the model's generalization.
- We propose the CA to overcome the suppressive effect on feature maps within BN, which reduces the model's representation ability. The CA utilizes the decorrelation operation to activate all feature maps, enhancing the model's classification performance in MDD detection.
- We experimented with two MDD datasets, comparing them to various SOTA methods. The findings show that our method outperforms SOTA methods in MDD detection performance.

Methodology

This paper proposes an MDD detection model incorporating the AFL and CA. It first processes multi-electrode EEG signals by extracting differential entropy (DE) features, then converting them into signal images, and inputting these into the model for precise MDD detection. The AFL is responsible for increasing the model's generalization, and the CA is responsible for increasing the model's detection performance. The unified architecture is depicted in Figure 2, and comprehensive details are provided in the following sections.

EEG Signals Preprocess

Multi-electrode EEG signals are represented as $S = \{s_1, s_2, \ldots, s_n\}$, where n is the subject number, $S \in \mathbb{R}^{l \times e}$, l is the time series duration, and e is the EEG number of the electrode channels. Following the instructions of existing studies (Jia et al., 2020), we first extract the DE features in five frequency bands: Delta (0.5-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-35 Hz), and Gamma (35-100 Hz) (Abhang, Gawali, & Mehrotra, 2016). Second, to extract EEG features from multiple electrode channels, we convert those one-dimensional time-series signals into multiple signal images using a fixed window(C. Xu, Shen, Fan, Qiu, & Mao, 2023). Finally, the features of the five frequency bands are combined as the input. The preprocessed signal is $X = \{x_1, x_2, \ldots, x_n\}$, where $X \in \mathbb{R}^{n \times f \times e \times l}$, f is the frequency band. We create an adjacent matrix to represent the topological relationship between multiple electrode channels as indicated by (Z.-Y. Zhang, Meng, Jin, Wang, & Hou, 2024).

Adaptive Feature Learning

We propose the AFL to capture high-order, domain-invariant features within EEG data in a high-dimensional RKHS, which can comprehensively incorporate electrode and frequency correlations, temporal dependencies, and subtle signal variations. The AFL first learns multiple subjects' EEG signal distribution features. Then, we employ the MMD loss function to represent the differences between the learned features and the actual EEG signals.

According to kernel embedding technology (Long, Wang, Sun, & Philip, 2014), we can map the low-dimensional EEG DE feature to high-dimension RKHS. In RKHS, we can extract the multiple subjects' domain-invariant features to enhance the model's generalization. The universal approximation theorem (Hornik, Stinchcombe, & White, 1989) allows us to build a network which can extract the domain-invariant features in RKHS. However, our EEG signal image X is a matrix (in this section, we transform the EEG signal to $X \in \mathbb{R}^{b \times e \times l}$, where b is the batch of the EEG signal image); thus, we can write the $\phi_k(\cdot)$ function as follows.

$$f(X_i) = \max_{k \in \mathcal{K}} \frac{1}{l} \sum_{q=1}^{l} \phi_k \left(x_{eq} \right) \tag{2}$$

Where $f(\cdot)$ is a neural network that can automatically select the best kernel from the several distinctive kernels $k \in \mathcal{K}$, and X_i is an EEG signal image.

The kernel embedding idea suggests that the feature mapping function $f(\cdot)$ should be injective. To make the neural network injective, another function (neural network) $f^{-1}(\cdot)$ should make $f^{-1}(f(X_i)) = X_i$ be applicable to all potential X_i. Therefore, we can use an autoencoder to ensure injectivity in the feature mapping. An autoencoder consists of two components: an encoder $f(\cdot)$ and a decoder $f^{-1}(\cdot)$. The encoder maps the input EEG signal image to a vector representing the high-order moments EEG feature in RHKS. The decoder re-

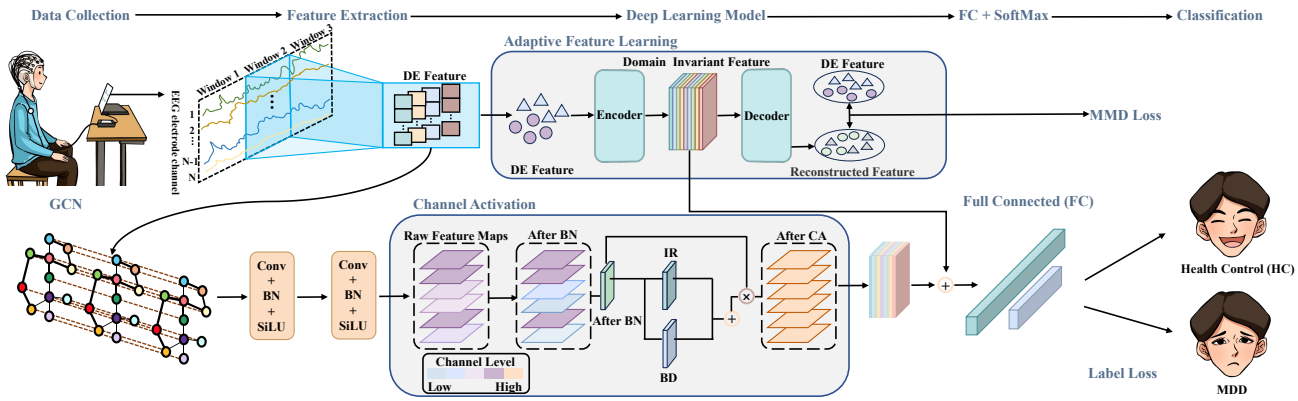


Figure 2: Overview of GCN-CNN Network Based on The AFL And CA (i.e., the AFL-CA model).

constructs this vector into an output of the same size as the input signal image.

We use the MMD loss function to compare the DE feature of the EEG signal to the reconstructed feature. The MMD function's general form is shown in Eq. (3), where X_m and X_n are two EEG signal images, and l is the time windows of two EEG signals.

$$\mathrm{MMD}\left(X_m, X_n\right) = \left\| \frac{1}{l}\sum_{i=1}^{l}\left(\phi_k\left(x_{ei}\right)\right) - \frac{1}{l}\sum_{j=1}^{l}\left(\phi_k\left(x_{ej}\right)\right) \right\|_2 \tag{3}$$

The MMD function instructs the autoencoder to learn high-order domain invariant feature representations of the EEG signal in RKHS. The MMD function can be described as Eq. (4) in our task.

$$\mathrm{Loss}_{\mathrm{MMD}}\left(X_m, f^{-1}\left(f\left(X_m\right)\right)\right) = \frac{1}{l}\left\| \sum_{j=1}^{l} x_{ej} - \left(f^{-1}\left(f\left(x_{ej}\right)\right)\right) \right\|_2 \tag{4}$$

Based on the operations above, we can compel the network to learn domain invariant features of multiple subjects in RKHS, thereby enlarging the feature space. Consequently, the AFL enhances the model's generalize ability and improves the performance of cross-subject MDD detection.

Channel Activation

We find that BN reduces the re-scale factor, which results in the suppression of EEG feature maps. Drawing inspiration from the decorrelation technique described in (Huang et al., 2018; Shao et al., 2020), which indicates decorrelation can increase the re-scale factor magnitude using the inverse square root of the covariance matrix Σ (i.e., $\Sigma^{-\frac{1}{2}}$), based on this, we propose a method to reinvigorate suppressed feature maps, thereby enhancing their utility for MDD classification. Following the BN layer, our CA method reactivates feature maps by executing channel decorrelation, as delineated in Eq. (5).

$$p_{wij} = D_w^{-\frac{1}{2}}\left(\mathrm{Diag}(\gamma)\bar{x}_{wij} + \beta\right) \tag{5}$$

Here, p_{wij} is the c^{th} EEG feature channel output of the CA, w is the batch of the EEG feature map, and i and j are the pixels in the EEG signal image. To generate a column vector $\bar{x}_{wij}$, we stack the items across all $\bar{x}_{wcij}$ channels. Similarly, stacking β_c and γ_c yields β and γ. $\mathrm{Diag}(\gamma)$ denotes a diagonal matrix. The decorrelation operation, $D_w^{-\frac{1}{2}}$, requires all EEG feature maps to contribute more or less to signal representation.

The CA needs to not only activate a batch of feature maps but also consider the channel dependency statistics within each feature map. We can decorrelate the EEG feature channel dependency statistics for the batch feature maps and embed each signal image in the matrix $D_w^{-\frac{1}{2}}$. D_w can be written as follows.

$$D_w = \lambda\Sigma + (1-\lambda)\,\mathrm{Diag}\left(g\left(\tilde{\sigma}_w^2\right)\right) \tag{6}$$

Where the covariance matrix Σ is generated after normalization over an entire batch of feature maps $\{\bar{x}_w\}_{w=1}^{W}$. $\tilde{\sigma}_w^2$ is a variance vector measured across all channels. g is responsible for modeling feature channel dependencies and returns an adaptive instance variance. The Jensen inequality (Pečarić, 1996) allows us to relax $D_w^{-\frac{1}{2}}$ as Eq. (7).

$$\begin{aligned} D_w^{-\frac{1}{2}} &= \left[\lambda\Sigma + (1-\lambda)\,\mathrm{Diag}\left(\mathrm{Diag}\left(f\left(\tilde{\sigma}_w^2\right)\right)\right)\right]^{-\frac{1}{2}} \\ &\preceq \lambda \underbrace{\Sigma^{-\frac{1}{2}}}_{\text{Batch Decorrelation (BD)}} + (1-\lambda) \underbrace{\left[\mathrm{Diag}\left(g\left(\tilde{\sigma}_w^2\right)\right)\right]^{-\frac{1}{2}}}_{\text{Instance Reweighting (IR)}} \end{aligned} \tag{7}$$

Where $\preceq$ stands for matrix comparison symbols.

Figure 2 and Eq. (7) illustrate that D_w consists of two parts: $\Sigma^{-\frac{1}{2}}$ for the covariance matrix across the batch of EEG feature maps, which can activate the batch of feature channels (BD), and $\left[\mathrm{Diag}\left(f\left(\tilde{\sigma}_w^2\right)\right)\right]^{-\frac{1}{2}}$ can adjust the correlation between EEG feature channels of each signal image (IR). λ is a learnable ratio balancing BD and IR.

Batch Decorrelation The BD computes the covariance matrix Σ from a batch of EEG feature maps to activate the batch of EEG feature channels. We assume that $\bar{X} \in \mathbb{R}^{W\times Z}$ represents the EEG feature maps following the BN layer, where $Z = C \times E \times L$, W is the batch size, C is the feature channels,

E and L are the height and width of the signal image. The BD can be computed as follows:

$$\Sigma = \gamma\gamma^T \odot \frac{1}{Z}\bar{X}\bar{X}^T \tag{8}$$

where Σ stands for the EEG feature channel correlation matrix. For example, Σ_{ij} is the dependency between the i^{th} and j^{th} EEG feature channels, scaled by $\gamma_i\gamma_j$ after normalization. $\odot$ indicates an elementwise multiplication.

In addition, we utilize Newton's iterative method to compute the inverse square root, avoiding the computationally intensive decorrelation operation (Bini, Higham, & Meini, 2005).

Instance Reweighting The IR can estabilsh channel dependencies within each feature map (Shao et al., 2020). We can calculate the IR part input $\tilde{\sigma}_w^2$ by Eq. (9).

$$\tilde{\sigma}_w^2 = \operatorname{diag}\left(\gamma\gamma^T\right) \odot \frac{\left(\sigma_p^2\right)_w}{\sigma_q^2} \tag{9}$$

Where the diagonal of a given matrix is extracted using $\operatorname{diag}\left(\gamma\gamma^T\right)$. The variances estimated by σ_q^2 and $\left(\sigma_p^2\right)_w$ are shown by BN and each feature map normalization, respectively. The vector division is done element-wise in Eq. (9). The IR input is scaled using γ_c^2 for the c^{th} channel.

$$\left[\operatorname{Diag}\left(g\left(\tilde{\sigma}_w^2\right)\right)\right]^{-\frac{1}{2}} = \operatorname{Diag}\left(\tilde{s}\left(\tilde{\sigma}_w^2;\theta\right)\right) \cdot \underbrace{\frac{1}{WC}\sum_{w,c}^{W,C}\left(\tilde{\sigma}_w^2\right)_c^{-\frac{1}{2}}}_{\textbf{Part A}} \tag{10}$$

In the IR, as depicted in Eq. (10), $\left[\operatorname{Diag}\left(g\left(\tilde{\sigma}_w^2\right)\right)\right]^{-\frac{1}{2}}$ adjusts the correlations between EEG feature maps. **Part A** in Eq. (10) denotes the inverse square root of variances computed across the batch feature channels and each feature map. According to (Hu, Shen, & Sun, 2018), channel dependencies can be established using a sub-network parameterized by θ. The $\tilde{s}$ represents the sigmoid function that generates weights, which control the strength of the inverse square root of variance for each EEG feature map, ensuring the output maintains the same magnitude as the BD.

Loss Function and Baseline Network

The proposed AFL-CA model's loss function consists of bi-classification cross-entropy (CE) and MMD loss. The model's overall loss function can be expressed in the following way:

$$\text{Loss}_{\text{ALL}} = \text{Loss}_{\text{CE}} + \alpha\,\text{Loss}_{\text{MMD}} \tag{11}$$

where α is the coefficient of Loss_{MMD}.

The baseline network includes 1 layer GCN and 2 layers CNN, as shown in Figure 2. The design of GCN network and adjacency matrix follows the settings of (Z.-Y. Zhang et al., 2024).

Experiments

The benchmark datasets, experimental platform, experimental settings and evaluation metrics are each presented in this section.

Datasets Description

We evaluate our model's efficacy with two widely used MDD datasets (MODMA (Cai et al., 2020) and PRED+CT (Cavanagh, Bismark, Frank, & Allen, 2019)). The MODMA dataset includes 24 subjects with MDD and 29 HC subjects. We used resting state data containing 128 channels in MODMA dataset. The PRED+CT dataset includes 75 HC subjects and 46 MDD subjects, and we chose 43 HC and 43 MDD patients for the experimental data.

Platform, Hyperparameters and Evaluation Metrics

Table 1: Hyperparameters Setting.

Hyperparameters	MODMA	PRED+CT
Fully Connected Layer of Encoder	4	4
Fully Connected Layer of Decoder	4	4
GCN Number	1	1
CNN Number	2	2
Learning Rate	0.001	0.001
Batch_Size	5×150	10×150
GCN Dropout	0.3	0.3
CNN-1 Dropout	0.5	0.3
CNN-2 Dropout	0.5	0.5
Optimizer	Adam	Adam
L2 Regularization Coefficient	0.2	0.2
Epoch	50	70

All the models were trained/tested on one NVIDIA RTX 4090 24 GB GPU, Intel E5-2686 CPU, and 64 GB memory. The deep learning framework Pytorch was used to implement the experiments. The hyperparameters settings are shown in Table 1. The proposed method was evaluated using four metrics: accuracy (Acc), F1-Score (F1), polygon area metric (PAM) (Aydemir, 2021) and Kappa coefficient (Cohen, 1960).

Comparison with Other Methods

We compared some results from other literature. Moreover, the PRED+CT dataset is less used; we reproduced some classical networks on the PRED+CT dataset. The classical networks including ResNet-3 (3 layers ResNet) (C. Xu et al., 2023, 2022), ShuffleNet (X. Zhang, Zhou, Lin, & Sun, 2018), Vision Transformer (ViT) (Dosovitskiy et al., 2020), Time-Series DCN (S. Xu, Zhang, Huang, Wu, & Song, 2022) and Swin Transformer (Z. Liu et al., 2021). The experimental results are summarized in Table 2 (MODMA) and Table 3 (PRED+CT). For the MODMA dataset, our method outperforms SOTA methods on the accuracy, F1-Score, Kappa, and PAM metrics by 0.93%, 1.29%, 8.09%, and 7.26%, respectively. For the PRED+CT dataset, our method outperforms SOTA methods on the accuracy, F1-Score, Kappa, and PAM metrics by 0.51%, 7.04%, 13.92%, and 18.06%, respectively.

Table 2: Model Results on The MODMA Dataset (%).

Method	Acc	F1	Kappa	PAM
(Y. Wang, Liu, & Yang, 2021)	86.67	90.51	-	-
(Chen, Guo, Hao, & Hong, 2022)	84.91	84.00	-	-
(Chen, Hong, Guo, Hao, & Hu, 2022)	86.49	84.85	-	-
(Su, Zhang, Cai, Zhang, & Li, 2023)	82.27	-	-	-
(Tasci et al., 2023)	83.96	81.10	-	-
(W. Liu, Jia, Wang, & Ma, 2022)	89.63	90.19	-	-
EEG Transformer (Qayyum et al., 2023)	72.03	62.30	-	-
EEGNet (B. Liu et al., 2022)	78.46	77.91	56.00	-
SENet (Qayyum et al., 2023)	82.15	78.58	-	-
DANN (Jiang et al., 2023)	85.08	84.09	-	-
DAN (Wu, Ma, Lian, Cai, & Zhao, 2022)	87.40	-	-	-
DCANN (Jiang et al., 2023)	86.85	85.97	-	-
ViT*	66.04	75.67	27.28	40.80
Swin-Transformer*	58.49	71.05	10.44	31.56
Time Series DCN*	86.79	89.23	72.66	72.82
ShuffleNet*	84.90	87.87	68.63	69.44
ResNet-3*	75.47	79.36	49.59	54.56
Ours	**90.56**	**91.80**	**80.75**	**80.08**
ΔSOTA	0.93↑	1.29↑	8.09↑	7.26↑

Table 3: Model Results on The PRED+CT Dataset (%).

Method	Acc	F1	Kappa	PAM
(Z.-Y. Zhang et al., 2024)	83.17	82.93	-	65.69
(Sam et al., 2023)	96.00	-	-	-
(H.-G. Wang, Meng, Jin, & Hou, 2023)	90.38	89.51	79.10	74.07
ViT*	80.23	75.36	60.46	56.64
Swin-Transformer*	63.95	45.61	27.90	26.87
Time Series DCN*	68.60	63.01	37.20	39.87
ShuffleNet*	74.41	70.27	48.83	48.67
ResNet-3*	63.95	56.33	27.90	33.09
Ours	**96.51**	**96.55**	**90.02**	**92.13**
ΔSOTA	0.51↑	7.04↑	13.92↑	18.06↑

Ablation Study

Table 4: Ablation Study Analysis. (Black: MODMA, Blue: PRED+CT)

Method	Acc	F1	Kappa	PAM
Baseline	81.13&86.04	85.18&86.67	62.19&72.09	63.09&71.49
+AFL	88.67&95.34	90.32&95.23	76.82&90.69	76.48&73.11
+CA	86.79&88.37	87.71&88.89	73.44&76.74	73.11&75.80
Ours	**90.56&96.51**	**91.80&96.55**	**80.75&93.02**	**80.08&92.13**

According to existing research (Zeng, Chen, Xu, & Zhang, 2023) indicated, we conducted ablation study on the AFL-CA model. Table 4 presents the results of ablation studies, demonstrating enhancements in the performance of MDD detection. Integrating the AFL and CA results in the highest performance improvements compared to the baseline: 7.54% (MODMA, accuracy), 9.3% (PRED+CT, accuracy), 13.39% (MODMA, PAM), and 17.95% (PRED+CT, PAM). Adding only the AFL improves accuracy by 7.54% (MODMA) and 9.3% (PRED+CT), while PAM metrics rise by 13.39% (MODMA) and 17.95% (PRED+CT) compared to the baseline. Integrating the CA alone improves accuracy by 5.66% (MODMA) and 2.33% (PRED+CT), while increasing PAM metrics by 10.02% (MODMA) and 4.31% (PRED+CT). Overall, the AFL-CA model achieves a remarkable accuracy improvement of 9.43% (MODMA) and 10.47% (PRED+CT), along with a significant increase in PAM metrics by 19.99% (MODMA) and 20.64% (PRED+CT).

Visualization Analysis

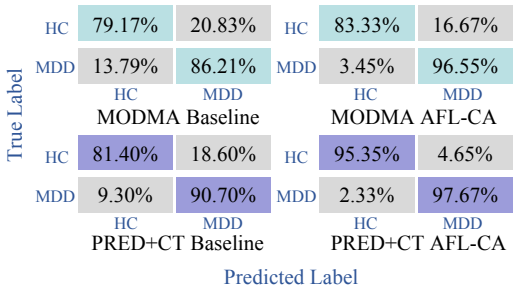


Figure 3: Confusion Matrices.

Figure 3 compares the confusion matrices of the baseline and proposed models for MODMA and PRED+CT datasets, providing a quantitative evaluation. In MODMA, the error rate decreases from 20.83% to 16.67% in HC and from 13.39% to 3.45% in MDD. In PRED+CT, the error rate decreases from 18.60% to 4.65% in HC and from 9.30% to 2.33% in MDD. Notably, the PRED+CT dataset exhibits a more significant improvement, benefiting from its larger EEG data size. The AFL improves the model's generalization by learning various EEG data distributions measured using MMD, forcing the network to learn in the direction of the actual data. We attribute the AFL-CA model's success to extracting diverse EEG signal characteristics and activating suppressed feature maps, significantly improving MDD detection performance.

Different Weighting λ Coefficients Analysis for the BD and IR

Eq. (7) shows that λ adjusts the strength of the two parts in the CA. We establish an ablation study on the CA to assess their impact on the final performance, as shown in Figure 4. The results show that BD and IR on the MODMA dataset can improve accuracy by 1.88% and 1.89%, respectively, compared to the baseline networks. These findings show that the CA performs best in the two scenarios. In this case, the CA must train a learnable λ to acquire a tunable ratio; from this, we can attain the best performance.

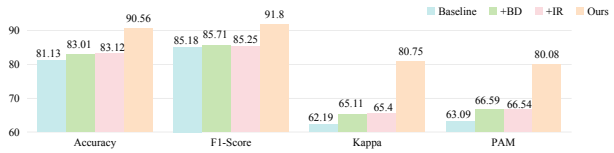


Figure 4: The BD Part and IR Part Analysis.

EEG Single Frequency Band Analysis

Following the existing method (Qi, Xu, & Li, 2023), we conducted a single-band analysis experiment, as shown in Table 5, revealing a significant impact of the Alpha and Beta frequency bands on MDD detection. Using only the Beta band on the MODMA dataset yields 88.67% accuracy and 76.59%

Table 5: Single Frequency Band Analytical Experiment Results.

Method	Acc	F1	Kappa	PAM
Delta	77.36&88.37	80.00&89.36	54.00&76.74	57.30&76.11
Gamma	77.35&90.69	79.99&91.30	53.97&81.39	57.34&80.41
Beta	88.67&94.18	90.00&91.30	76.98&88.37	76.59&87.01
Alpha	83.01&93.03	86.56&93.33	64.58&86.05	66.17&84.91
Theta	81.13&93.02	84.37&92.85	61.08&86.04	63.34&84.47
Ours	**90.56&96.51**	**91.80&96.55**	**80.75&93.02**	**80.08&92.13**

PAM metrics, with a slight decrease of 1.89% in accuracy and 3.49% in PAM compared to all five bands. Similarly, using only the Beta band on the PRED+CT dataset yields 94.18% accuracy and 87.01% PAM metrics, with a slight decrease of 2.33% in accuracy and 5.12% in PAM compared to all bands. While acceptable, simultaneous use of all five bands is preferable, emphasizing the multi-frequency band significance of depression EEG signals.

Parameter Sensitivity Analysis

We also look into the weights of the additional loss function MMD used in the AFL. We experimented with different weights for the loss function in Eq. (11). Table 6 shows that the best performance occurs when $\alpha = 0.005$, resulting in 90.56% (MODMA) and 96.51% (PRED+CT). However, performance drops sharply as the coefficient weight values α increase. The accuracy metrics are worst when $\alpha = 50$, resulting in 84.9% (MODMA) and 86.04% (PRED+CT). The observed phenomenon could be attributed to an increase in α, in which network training forces the model to learn features that closely resemble the data distribution of the original signal. While the AFL can effectively capture more data distribution features, the AFL's encoder features and certain original signals may be too similar, resulting in overfitting.

Table 6: Performance variation when the loss function is weighted α differently in Eq. (11).

α	Acc	F1	Kappa	PAM
$\alpha = 50$	84.90&86.04	86.20&86.04	69.54&72.09	69.75&71.01
$\alpha = 5$	86.80&83.72	88.13&83.72	73.25&67.44	73.15&66.82
$\alpha = 0.5$	86.79&88.37	89.23&88.63	72.66&76.74	72.82&75.60
$\alpha = 0.05$	88.67&91.86	90.00&91.56	76.98&83.72	76.59&81.96
$\alpha = 0.005$	**90.56&96.51**	**91.80&96.55**	**80.75&93.02**	**80.08&92.13**

Discussion

What is the Difference among the AFL, DA, and DG Methods?

Compared to the AFL approach, DA and DG have disadvantages. The DA-based method needs source domain data to pretrain the model. Then, the pretrained model transfers to the target domain using the target domain data to guide the finetune model. However, we cannot obtain patients' EEG data (target domain data) in actual application. Although DG-based methods do not require target domain data, they require the manual creation of many domain labels for each subject, which increases the manual workload. In contrast, the AFL improves cross-subject classification accuracy by converting EEG signals' data to RKHS, learning the multiple subjects' domain invariant feature. The AFL eliminates the need for manual domain label creation and target domain data. The results in Table 2 and Table 3 demonstrate our method's significant improvement over DG-based approaches (Z.-Y. Zhang et al., 2024) and DA-based approaches (Jiang et al., 2023), emphasizing the advantage of the AFL.

Why the CA Outperforms Attentional Mechanisms in MDD Detection Area?

Table 4 shows that the MODMA dataset, including the CA, achieves 86.79% accuracy, surpassing the SENet and EEG Transformer model (Qayyum et al., 2023). In SENet, accuracy decline is caused by the SE attention mechanism compressing spatiotemporal features, resulting in a loss of temporal dependence and interdependence among EEG channels. In contrast to image data, the Transformer model's self-attention focuses on local-global signal correlations. EEG signals involve multiple channels and temporal points, with changes in MDD patients occurring over contextual signal segments rather than isolated points (C. Xu et al., 2022). The self-attention mechanism frequently ignores significant temporal and multi-channel correlations, resulting in lower classification accuracy. Unlike attention mechanisms, the CA activates feature maps via decorrelation, resulting in diverse contributions to the MDD classification task and improved overall network performance.

Why Can the CA Activate the Feature Channel?

The CA activates EEG feature channels through BD and IR. Eq. (1) links suppressed channels to γ_c with small values. As previous work indicates (Huang et al., 2018), decorrelation operation boosts the re-scale factor. Hence, BD can utilize decorrelation to activate suppressed EEG feature maps, expressed as $p_{wij}^{\mathrm{BD}} = \mathrm{Diag}\left(\Sigma^{-\frac{1}{2}}\gamma\right)\bar{x}_{wij} + \Sigma^{-\frac{1}{2}}\beta$. Compared with Eq. (1), an equivalent γ for the BD part can be defined as $\hat{\gamma} = \Sigma^{-\frac{1}{2}}\gamma$. Due to the covariance of re-scale, it can increase the re-scale factor magnitude (Shao et al., 2020). BD activates suppressed channels by increasing the re-scale factor magnitude. We match the IR's output magnitude to BD's, ensuring activation of every EEG signal image and establishing channel dependencies. The combined BD and IR enhance EEG feature maps' representation ability, thereby increasing MDD detection in the downstream network.

Conclusions

This paper proposes the AFL-CA model for MDD detection, which combines the AFL and CA. The AFL employs kernel embedding to learn domain-invariant features in RKHS, which solves the MDD detection model's poor generalization ability problem. The CA reactivates suppressed EEG feature maps to ensure that they all contribute to representation, which increases the MDD detection model's representation ability. Ablation studies on two MDD datasets demonstrate the effectiveness of the AFL and CA. Comparison experi-

ments confirm the AFL-CA model's superiority. These advantages address poor MDD detection model generalization and classification performance, which bodes well for future research.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62203321), China Postdoctoral Science Foundation (2021M692390) and Tianjin Research Innovation Project for Postgraduate Students (2022BKY090).

References

Abhang, P. A., Gawali, B. W., & Mehrotra, S. C. (2016). Technological basics of eeg recording and operation of apparatus. *Introduction to EEG-and Speech-Based Emotion Recognition*, 19–50.

Aydemir, O. (2021). A new performance evaluation metric for classifiers: polygon area metric. *Journal of Classification*, *38*, 16–26.

Bini, D. A., Higham, N. J., & Meini, B. (2005). Algorithms for the matrix pth root. *Numerical Algorithms*, *39*(4), 349–378.

Cai, H., Gao, Y., Sun, S., Li, N., Tian, F., Xiao, H., ... others (2020). Modma dataset: a multi-modal open dataset for mental-disorder analysis. *arXiv preprint arXiv:2002.09283*.

Cavanagh, J. F., Bismark, A. W., Frank, M. J., & Allen, J. J. (2019). Multiple dissociations between comorbid depression and anxiety on reward and punishment processing: Evidence from computationally informed eeg. *Computational Psychiatry (Cambridge, Mass.)*, *3*, 1.

Chen, T., Guo, Y., Hao, S., & Hong, R. (2022). Exploring self-attention graph pooling with eeg-based topological structure and soft label for depression detection. *IEEE Transactions on Affective Computing*, *13*(4), 2106–2118.

Chen, T., Hong, R., Guo, Y., Hao, S., & Hu, B. (2022). Ms ˆ{2}-gnn: Exploring gnn-based multimodal fusion network for depression detection. *IEEE Transactions on Cybernetics*.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Ppsychological Measurement*, *20*(1), 37–46.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... others (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359–366.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7132–7141).

Huang, L., Yang, D., Lang, B., & Deng, J. (2018). Decorrelated batch normalization. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 791–800).

Jia, Z., Lin, Y., Wang, J., Zhou, R., Ning, X., He, Y., & Zhao, Y. (2020). Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification. In *Ijcai* (Vol. 2021, pp. 1324–1330).

Jiang, W., Su, N., Pan, T., Miao, Y., Lv, X., Jiang, T., & Zuo, N. (2023). Eeg-based subject-independent depression detection using dynamic convolution and feature adaptation. In Y. Tan, Y. Shi, & W. Luo (Eds.), *Advances in swarm intelligence.* Springer Nature Switzerland.

Liu, B., Chang, H., Peng, K., & Wang, X. (2022). An end-to-end depression recognition method based on eegnet. *Frontiers in Psychiatry*, *13*, 864393.

Liu, W., Jia, K., Wang, Z., & Ma, Z. (2022). A depression prediction algorithm based on spatiotemporal feature of eeg signal. *Brain Sciences*, *12*(5), 630.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 10012–10022).

Long, M., Wang, J., Sun, J., & Philip, S. Y. (2014). Domain invariant transfer kernel learning. *IEEE Transactions on Knowledge and Data Engineering*, *27*(6), 1519–1532.

Ma, S., Zhang, Y., Chen, Y., Xie, T., Song, S., & Jia, Z. (2023). Exploring structure incentive domain adversarial learning for generalizable sleep stage classification. *ACM Transactions on Intelligent Systems and Technology*.

Pečarić, J. (1996). Power matrix means and related inequalities. *Mathematical Communications*, *1*(2), 91–110.

Qayyum, A., Razzak, I., Tanveer, M., Mazher, M., & Alhaqbani, B. (2023). High-density electroencephalography and speech signal based deep framework for clinical depression diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Qi, X., Xu, W., & Li, G. (2023). Neuroimaging study of brain functional differences in generalized anxiety disorder and depressive disorder. *Brain Sciences*, *13*(9), 1282.

Saeidi, M., Karwowski, W., Farahani, F. V., Fiok, K., Taiar, R., Hancock, P., & Al-Juaid, A. (2021). Neural decoding of eeg signals with machine learning: A systematic review. *Brain Sciences*, *11*(11), 1525.

Sam, A., Boostani, R., Hashempour, S., Taghavi, M., & Sanei, S. (2023). Depression identification using eeg signals via a hybrid of lstm and spiking neural networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 4725–4737.

Shao, W., Tang, S., Pan, X., Tan, P., Wang, X., & Luo, P. (2020). Channel equilibrium networks for learning deep representation. In *International conference on machine learning* (pp. 8645–8654).

Su, Y., Zhang, Z., Cai, Q., Zhang, B., & Li, X. (2023). 3dmkdr: 3d multiscale kernels cnn model for depression recognition based on eeg. *Journal of Beijing Institute of Technology*, *32*(2), 230–241.

Tasci, G., Loh, H. W., Barua, P. D., Baygin, M., Tasci, B.,

Dogan, S., ... Acharya, U. R. (2023). Automated accurate detection of depression using twin pascal's triangles lattice pattern with eeg signals. *Knowledge-Based Systems*, *260*, 110190.

Wang, H.-G., Meng, Q.-H., Jin, L.-C., & Hou, H.-R. (2023, oct). Amgcn-l: an adaptive multi-time-window graph convolutional network with long-short-term memory for depression detection. *Journal of Neural Engineering*, *20*(5), 056038.

Wang, H.-G., Meng, Q.-H., Jin, L.-C., Wang, J.-B., & Hou, H.-R. (2023). Amg: A depression detection model with autoencoder and multi-head graph convolutional network. In *2023 42nd chinese control conference (ccc)* (p. 8551-8556).

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). Eca-net: Efficient channel attention for deep convolutional neural networks. In *2020 IEEE/CVF conference on computer vision and pattern recognition, CVPR 2020, seattle, wa, usa, june 13-19, 2020* (pp. 11531–11539). Computer Vision Foundation / IEEE. doi: 10.1109/CVPR42600.2020.01155

Wang, Y., Liu, F., & Yang, L. (2021). Eeg-based depression recognition using intrinsic time-scale decomposition and temporal convolution network. In *The fifth international conference on biological information and biomedical engineering* (pp. 1–6).

Wu, W., Ma, L., Lian, B., Cai, W., & Zhao, X. (2022). Few-electrode eeg from the wearable devices using domain adaptation for depression detection. *Biosensors*, *12*(12), 1087.

Xu, C., Mao, Z., Fan, F., Qiu, T., Shen, J., & Gu, Y. (2022). A shallow convolution network based contextual attention for human activity recognition. In *International conference on mobile and ubiquitous systems: Computing, networking, and services* (pp. 155–171).

Xu, C., Shen, J., Fan, F., Qiu, T., & Mao, Z. (2023). An enhanced human activity recognition algorithm with positional attention. In *Asian conference on machine learning* (pp. 1181–1196).

Xu, S., Zhang, L., Huang, W., Wu, H., & Song, A. (2022). Deformable convolutional networks for multimodal human activity recognition using wearable sensors. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–14.

Yang, J., Niu, J., Zeng, S., Wang, Y., La, R., Mao, W., & Cai, H. (2018). Resting state eeg based depression recognition research using voting strategy method. In *2018 ieee international conference on bioinformatics and biomedicine (bibm)* (pp. 2666–2673).

Zeng, X., Chen, Y., Xu, B., & Zhang, T. (2023). Modaldrop: Modality-aware regularization for temporal-spectral fusion in human activity recognition. In *Icassp 2023-2023 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 1–5).

Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 6848–6856).

Zhang, Z.-Y., Meng, Q.-H., Jin, L.-C., Wang, H.-G., & Hou, H.-R. (2024). A novel EEG-based graph convolution network for depression detection: Incorporating secondary subject partitioning and attention mechanism. *Expert Systems with Applications*, *239*, 122356.