2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

# Complete Multi-Domain Decoupled Fusion Model for EEG-based Person Identification

Zhixun Wang, Jiayu Lu, Tianyang Liu, Ziteng Zhu, Xiaofeng Liu, Bin Wang

*Abstract*— **Electroencephalogram (EEG) signals have unique individual characteristics and have broad application prospects in identity authentication. At present, person identification (PI) based on EEG using the temporal-spatial-spectral feature extraction framework has achieved remarkable success. However, the existing methods suffer from coupled cross-domain feature parameters and insufficient feature fusion during feature extraction, which limits the recognition ability. Moreover, fixed-scale feature extractors can hardly exploit the subject-specific multi-scale information. To address these challenges, we propose CMDFM: a complete multi-domain decoupled fusion model for EEG-based PI. Firstly, we design an independent temporal-spatial-spectral attention mechanism to eliminate cross-domain parameter coupling. Secondly, a full-domain fusion mechanism is designed to comprehensively integrate the features of the temporal domain, spatial domain and spectral domain. Finally, an adaptive multi-scale CNN is designed to adjust the contribution of the multi-scale convolution kernel, thereby making full use of individual-specific multi-scale information. We use four datasets to verify our method. The experimental results show that our method is superior to all the state-of-the-art methods. The code of CMDFM is at https://github.com/2538441690/CMDFM.**

## I. INTRODUCTION

In the field of individual identity authentication, accurate identity identification is the core for ensuring data security. Biometric identification (such as fingerprints [1], irises [2] and faces [3]) has replaced traditional passwords or physical identifiers with its uniqueness, persistence and non-intrusiveness [4], and has become a widely used method of identity authentication. Compared with other biometric features, electroencephalogram (EEG) signals can only be obtained from living organisms and have excellent temporal resolution and rich dynamic characteristics, making it one of the most promising choices in biometric applications [5]. Therefore, person identification (PI) based on EEG has been widely studied in recent years.

In PI studies based on EEG, most previous work have focused on the design of attention mechanisms in a single or partial domain, failing to comprehensively consider the information in the temporal domain, spatial domain and spectral domain [6], [7]. The temporal, spatial and spectral domains respectively represent different characteristics of EEG signals, and these features are complementary, as shown in Fig. 1, focusing only on the feature extraction of a single or part of the domain will lead to the model's inability to fully utilize multi-domain information, thereby limiting its performance ceiling.

Therefore, some recent studies have explored EEG-based PI from the perspective of three-domain representation. However, these three-domain representation methods have the problem of cross-domain feature parameter coupling, resulting in the loss of individual specific information in the three domains [8], [9], [10]. In addition, their fusion mechanisms usually only fuse partial domain information without comprehensive fusion, leading to insufficient cross-domain feature extraction and restricting the further improvement of recognition accuracy [11], [12]. Furthermore, existing methods usually adopt a fixed-scale feature extraction strategy to handle all individuals in the temporal-spatial-spectral feature extraction stage, without considering the heterogeneous requirements of different individuals' features for the extraction scale, resulting in the individual-specific multi-scale temporal-spatial-spectral information not being captured specifically.

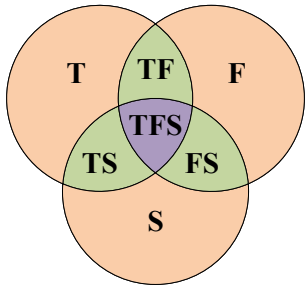


Fig. 1. A complete schematic diagram of the temporal-spatial-spectral feature representation. T, S, F represent temporal domain, spatial domain and frequency domain features. TF, TS, and FS represent temporal-frequency domain, temporal-spatial domain, and frequency-spatial domain features. TFS stands for temporal-frequency-spatial feature.

To address these challenges, we propose a complete multi-domain decoupled fusion model (CMDFM) for EEG. Firstly, an independent temporal-spatial-spectral attention mechanism is proposed. This mechanism aims to extract features independently from the temporal, spatial and spectral domains and minimize the interference between domains to the greatest extent. Meanwhile, a full-domain fusion mech-

This work was supported in part by the National Natural Science Foundation of China under Grant 62176177, in part by the National Natural Science Foundation of China under Grant 62576240, in part by the Science and Technology Cooperation and Exchange Special Projects of Shanxi under grant 202304041101034, in part by the Shanxi Province Higher Education Scientific and Technological Innovation Project under grant Z2025521, in part by the Shanxi Province Application Basic Research Plan under Grant 202303021211055. (Corresponding authors: Bin Wang.)

Zhixun Wang, Jiayu Lu, Tianyang Liu, Ziteng Zhu, Xiaofeng Liu, and Bin Wang are with the College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, 030024, China (e-mail: 2024510437@link.tyut.edu.cn; 2023310111@link.tyut.edu.cn; liutianyang0723@link.tyut.edu.cn; 2024001951@link.tyut.edu.cn; liuxiaofeng03@tyut.edu.cn; wangbin01@tyut.edu.cn).

anism is designed to effectively integrate the features of the three domains into a unified representation. We integrate the independent features in the temporal domain, spatial domain and frequency domain, the interactive features in the temporal-frequency domain, temporal-spatial domain and frequency-spatial domain, as well as the global feature in the temporal-frequency-spatial domain, thereby obtaining the most comprehensive individual characteristics. In addition, an input-dependent weight adaptive mechanism is adopted to dynamically adjust the kernel scale based on individual characteristics, so as to fully utilize the individual-specific multi-scale information. Finally, we introduce label smoothing cross-entropy loss to reduce excessive confidence in PI, thereby reducing overfitting.

The contributions of the paper are summarized as follows.

- We design an independent multi-domain attention mechanism to completely decouple feature learning in the temporal domain, spatial domain and spectral domain.
- We propose a full-domain fusion mechanism to comprehensively integrate temporal, spatial and spectral information.
- We propose an input-dependent adaptive multi-scale CNN to better capture the unique features at different scales.

## II. RELATED WORK

### A. Domain Disentanglement in EEG Representations

The existing studies generally combine the temporal-domain, spatial-domain and frequency-domain features for analysis. Most methods directly extract temporal, spatial and spectral mixed features from the original EEG signals through deep learning. However, this approach fails to focus on the unique individual characteristics of each domain [8], [9], [13], [14], [15], [16]. Cui et al. design a universal deep multiview network (DMV), which extracts different types of discriminatory features from multiple angles through various convolutional branches [17]. Miao et al. propose the multi-loss domain adapter (MLDA), which solves the domain shift problem through edge distribution adaptation and conditional distribution adaptation [18]. There are also some methods to convert the original EEG signals into two-dimensional image processing [5], [19], [20] or to reconstruct the original EEG signal by constructing a functional connection matrix [21], [22] and then perform feature extraction through deep learning methods.

Although these studies have made remarkable progress, the lack of consideration for the unique characteristics of the temporal domain, spatial domain and spectral domain has limited the recognition performance. Therefore, some recent studies aim to conduct independent feature extraction on the three domains of EEG signals to take into account the unique characteristics of each domain. Jin et al. propose MD-CAT, designing spatial attention to handle channel relationships separately and simultaneously designing temporal-spectral attention to jointly extract temporal-spectral domain features [23]. Chen et al. propose JTFT, which firstly extracts spatial domain features through transformer layer, and then jointly extracts temporal-spectral representations through low-rank attention layer [24]. The joint representation of the temporal-spectral domain by these two methods will cause the coupling of temporal-spectral domain features. Gao et al. propose SFT-Net, which extracts spatial-spectral features through depth-separable convolution and captures temporal dependencies through LSTM [10]. The joint extraction of spatial-spectral features by this method will cause information coupling between the spectral domain and the spatial domain.

Overall, although some methods have taken into account the characteristics of the temporal domain, spatial domain, or spectral domain, they still have not achieved complete feature decoupling.

### B. Domain Feature Fusion

In the field of EEG representations, some previous studies only focus on information in one of the temporal domain, spectral domain and spatial domain, which will cause the loss of a large amount of individual specific information [6], [7]. Therefore, considering the fusion of information from different domains is of vital importance for PI.

Some studies have taken into account the fusion of single domain features in the temporal domain, spectral domain and spatial domain [11], [12]. Cai et al. propose AITST, which utilizes the attention mechanism of the serial architecture to fuse spatial and temporal features [25]. Ding et al. propose TSception, which extracts temporal and spatial features in a parallel manner and finally fuses temporal-spatial features through a fusion layer [26]. Shao et al. propose CLT, which sequentially pass the processed EEG signals through electrode attention and spectral attention in the feature extraction stage to fuse spatial and spectral domain features [27].

All the above-mentioned methods only focus on the fusion of single domain features and fail to take into account the interaction features of two domains and the global features of three domains, which restricts the performance of PI. Recently, Miao et al. propose TSFF-Net, a method to extract temporal-spectral features and temporal-spatial features of EEG signals respectively through parallel CNN architecture, and finally fuse temporal-spectral features and temporal-spatial features in a weighted way [28]. However, this method ignores the independent features of a single domain, which may lead to the loss of individual-specific information.

Therefore, some methods have begun to focus on the fusion of single domain independent features and two domain interactive features [23], [29]. Jin et al. propose the TSFAN model. Firstly, the temporal domain features are extracted by the temporal domain feature extractor, and then the spatial-spectral features are extracted by the spatial-spectral feature extractor, and fusion is achieved through series [30].

However, the above-mentioned methods still cannot fully pay attention to the information in the temporal domain, spectral domain and spatial domain. Therefore, a comprehensive fusion framework is needed to fully explore the

complementary advantages of the information in the three domains, thereby improving the recognition performance.

### C. EEG-Based Feature Extraction Using CNN

Traditional convolutional neural networks use fixed-sized convolution kernels for feature extraction, which makes it difficult for them to capture the multi-scale features specific to different individuals [31]. Therefore, some studies have attempted to expand the receptive field of the network during the feature extraction process. Wang et al. utilize a stacked multi-scale convolutional neural network (MS-CNN) to extract multi-scale temporal-spatial features of EEG signals [32]. Yuan et al. propose Strip R-CNN, which utilizes large strip convolution to capture spatial information [33]. Recently, some people have weighted fused local features with global context by combining CNN and attention mechanisms, including MBSC [34], Sctnet [35] and RepViT [36].

However, all the above-mentioned methods are still limited by predefined fixed-scale combinations. In PI tasks, due to the significant differences in characteristics among different individuals, the above fixed-scale strategy cannot meet the heterogeneous needs of individuals. Therefore, an adaptive method that can dynamically adjust the feature extraction scale based on the input EEG features is needed to specifically extract individual-specific EEG information.

## III. METHODS

### A. Overall Structure

This model comprehensively integrates the features of the temporal domain, spatial domain and spectral domain to achieve accurate PI.

Fig. 2-(a) shows the overall structure of the complete multi-domain decoupled fusion model (CMDFM) we propose. This network first inputs the EEG signals into an independent temporal-spatial-spectral attention encoder, in which three dedicated branches independently extract discriminative features from the temporal, spatial and spectral domains respectively. Then, these domain-specific representations are integrated through the all-domain fusion mechanism. Then, the fused features are decomposed and encoded through patch and position embedding module. Then, the transformer architecture is adopted to capture long-range dependencies. Finally, the prediction results are obtained through the fully connected layer.

### B. Independent Multidomain Attention Mechanism

To address the issue of individual information loss caused by feature coupling in traditional methods, this model designs independent attention mechanisms for the temporal domain, spatial domain, and spectral domain of EEG signals. It is a feature extraction component designed for a single dimension of EEG signals. The core logic is to enhance its feature expression by learning the weights of the key regions of this dimension. At the same time, avoid coupling with information from other dimensions and retain the individual's specificity in that dimension. Fig. 2-(b) is a schematic diagram of the temporal domain attention mechanism (the spatial domain and the spectral domain follow exactly the same logic), and its working process is described in detail in combination with the components in the figure.

The input of the temporal domain attention mechanism is the feature tensor of the EEG signal in the temporal dimension, denoted as $X \in \mathbb{R}^{T\times C\times F}$, where $T$ is the time step, $C$ is the number of channels, and $F$ is the number of spectral bands. To calculate the weight of the temporal dimension, the input needs to be projected onto two parallel branches to generate local features and global reference features of the temporal dimension respectively, which are used to measure the importance of local features of the temporal dimension.

For the feature projection in the temporal dimension (denoted as $W_v$), a 1×1 convolutional layer is used to process $X$. Subsequently, the spatial and spectral dimensions are merged through the reshape operation to obtain the dimension feature matrix. This matrix retains the local features of the temporal dimension.

For the global reference feature projection (denoted as $W_q$), a 1×1 convolutional layer is used to process $X$. The objective is to generate global reference features, compress the time step to 1, and output the global reference feature tensor. Also, the spatial and spectral dimensions are merged through the reshape operation to obtain the global reference vector. This vector is a global summary of the temporal dimension.

Finally, perform matrix multiplication on the global reference vector and the dimension feature matrix to obtain the correlation score $W_z$ between each time point and the global information. The higher the score, the greater the importance. Finally, the original temporal step is restored through a 1×1 convolutional layer, and the temporal dimension weights are obtained through normalization and activation functions. The temporal domain attention weighted feature tensor $Z_t \in \mathbb{R}^{T\times C\times F}$ is obtained by taking the dot product of these weights with the original input.

For the temporal attention mechanism, its calculation formula is as follows:

$$Z_t = x \odot \sigma[C_z(R(C_v(x)) \otimes S(R(C_q(x))))] \qquad (1)$$

In the formula, $x$ represents the input of the EEG signal and $Z_t$ represents the output of the temporal attention mechanism, $C_q(\cdot)$, $C_v(\cdot)$ and $C_z(\cdot)$ are adaptive multi-scale convolution layer, and $C_z(\cdot)$ is followed by a normalization layer. Here $R(\cdot)$ stands for reshaping. $\sigma(\cdot)$ indicates the sigmoid function and $S(\cdot)$ indicates the LogSoftmax function. The operators $\odot$ and $\otimes$ correspond to the dot-product and multiplication of matrices.

### C. Global Fusion Mechanism

The global fusion mechanism aims to address the issue of insufficient inter-domain information integration in traditional methods. This mechanism takes the output of the independent domain attention mechanism (features of the temporal domain $Z_t$, frequency domain $Z_f$, and spatial

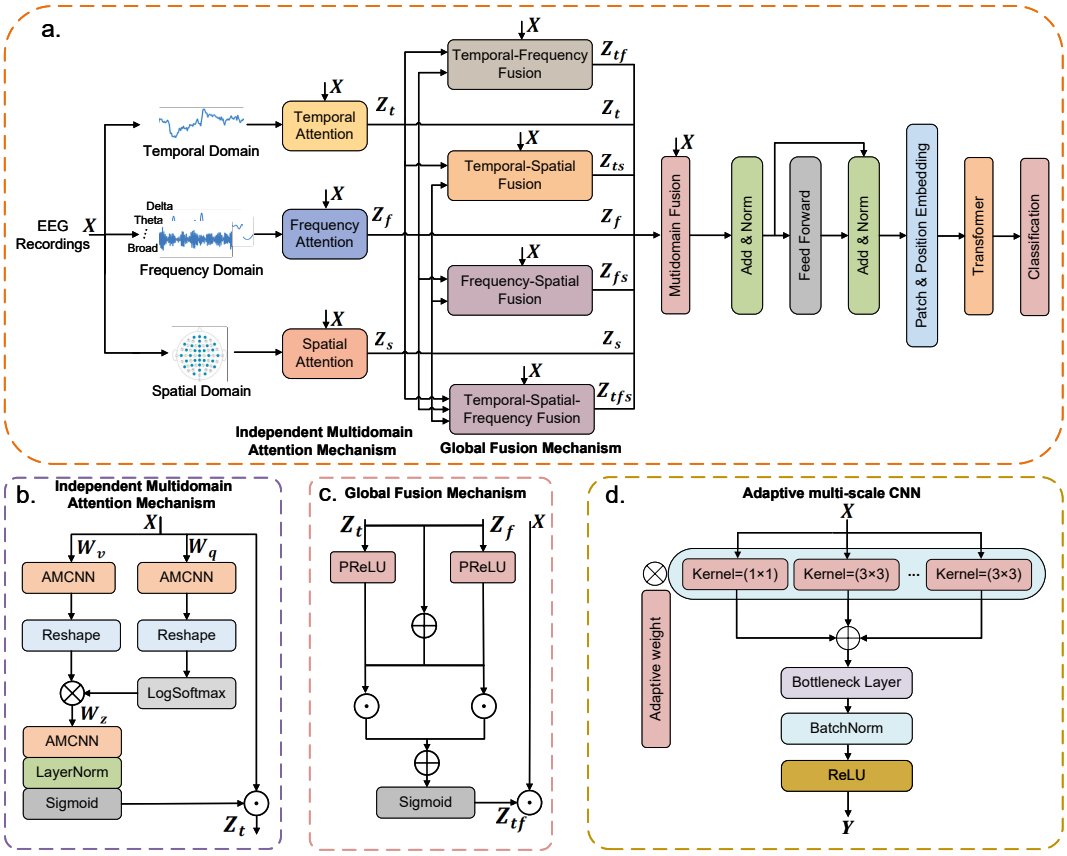


Fig. 2. Overall structure of our CMDFM model: (a) The overall framework of the complete multi-domain decoupled fusion model; (b) Schematic diagram of the temporal-domain attention mechanism; (c) Schematic diagram of the temporal-frequency fusion mechanism; (d) Schematic diagram of adaptive multi-scale CNN.

domain $Z_s$) as input. By integrating single domain independent features, two domain complementary features and three domain global features, multi-domain fusion features that fully retain individual specificity are generated.

Fig. 2-(c). is a schematic diagram of the temporal-spectral fusion mechanism (temporal-spatial fusion, spectral-spatial fusion, temporal-spatial-spectral fusion and the final seven-domain fusion all follow exactly the same logic), and its working process is described in detail in combination with the components in the figure.

The temporal-spectral fusion mechanism takes the temporal-domain ($Z_t$) and frequency-domain ($Z_f$) features as input. Firstly, the features from these two domains are input into the PReLU activation function for nonlinear enhancement. Then, the intermediate fusion features obtained by adding the temporal-domain and spectral-domain features are dot-product with the activated features to obtain the weighted features in the temporal-domain and spectral-domain. Finally, the weighted features are added. After passing through the sigmoid function, the dot product is performed with the original input ($X$) through residual connection to obtain the temporal-frequency fusion feature ($Z_{tf}$).

For the temporal-spectral fusion mechanism, its calculation formula is as follows:

$$Z_{tf} = x \odot \sigma((P(Z_t) \oplus P(Z_f)) \odot (Z_t + Z_f)) \quad (2)$$

Here, $Z_t$ and $Z_f$ are the outputs of the temporal-domain attention mechanism and the frequency-domain attention mechanism respectively. $P(\cdot)$ stands for the PReLU activation function. The operators $\odot$ and $\oplus$ correspond to the dot-product and addition of matrices.

### D. Adaptive Multi-Scale CNN

This paper proposes an adaptive multi-scale convolutional neural network (AMCNN) module and integrates it into a complete multi-domain decoupled fusion model. This module constructs a multi-branch structure by using multiple convolutional kernels of different scales, and introduces a dynamic weight generation mechanism. According to the distribution of input features, the weight coefficients of convolutional kernels of each scale are automatically learned and assigned, achieving the input adaptive adjustment of the contribution degree of convolutional kernels. Subsequently, the computational complexity is significantly reduced through the bottleneck compression layer, and the nonlinear modeling capability is introduced through the ReLU activation function. This design significantly enhances the ability to capture

individual-specific features of EEG signals. The structure is shown in Fig. 2-(d).

For each convolution kernel size k, define the convolution operation:

$$F = \textit{Conv2D}_{k\times k}(x) \tag{3}$$

Then the branch weights are generated by the adaptive branch weight generator:

$$W = S(\textit{Conv2D}_{1\times 1}(G(x))) \tag{4}$$

Here, $S(\cdot)$ indicates the LogSoftmax function and $G(\cdot)$ is global average pooling. Next, fuse each branch output by weight:

$$Z = \sum_{i=1}^{C} W_i \odot F_i \tag{5}$$

Here, $C$ represents the number of branches. Finally, the final output is obtained through the bottleneck layer, Batch Norm layer, and ReLU activation function:

$$Y = \textit{ReLU}(\textit{BN}(\textit{Bneck}(Z))) \tag{6}$$

Here, $\textit{Bneck}(\cdot)$ refers to the bottleneck layer, $\textit{BN}(\cdot)$ refers to the batch normalization layer, and $\textit{ReLU}(\cdot)$ is the ReLU activation function.

### E. Patch and Position Embedding

To better capture the global features of EEG signal, the features output by the global fusion mechanism are divided into fixed-size patches [37]. At the same time, we introduce the position embedding mechanism, which enhances the model's perception of the sequence structure by defining a position embedding matrix. Finally, the position embedding matrix is added to the patch sequence to obtain the output of this module.

### F. Transformer Encoder

The transformer module is designed to further capture the complex nonlinear information in the EEG signals. It integrates global key information through the scaled dot-product attention mechanism. In addition, the multi-head attention mechanism is employed to enable the model to capture the dependencies between different patches. Finally, The output of the transformer is sent to the linear layer for classification.

### G. Label Smoothing Loss

In order to reduce the common overfitting phenomenon in PI, we use the label smoothing cross-entropy loss instead of the traditional cross-entropy [38].

The traditional cross entropy loss function is:

$$L = -\sum_{i=1}^{c} y_i \log(\hat{y}_i) \tag{7}$$

Here, $L$ is cross-entropy loss. $C$ is the number of classes. $y_i$ is the one-hot encoding of the true label. $\hat{y}_i$ is the predicted probability distribution.

Label smoothing prevents the model from over-confident predictions by softening the hard one-hot labels. Specifically, it assigns a reduced probability $1-\epsilon$ to the true class and distributes the remaining probability $\epsilon$ uniformly over all classes. The label smoothing loss is defined as:

$$L = -\sum_{i=1}^{c} \left[(1-\epsilon)y_i + \frac{\epsilon}{c}\right] \log(\hat{y}_i) \tag{8}$$

where $L$ is label smoothing cross-entropy loss. $\epsilon$ is the smoothing coefficient.

## IV. EXPERIMENTS AND RESULTS

### A. Experimental Settings

*1) Datasets and Preprocessing:* We evaluate our model on four different datasets: THU-EP [39], SEED [40], SEED-IV [41], and SEED-V [42]. The THU-EP dataset consists of EEG data captured by a 32-channel wireless EEG system using Neuracle at a sampling rate of 250 Hz. In contrast, the SEED, SEED-IV and SEED-V datasets use 62-channel EEG data recorded by the ESI neural scanning system, following the international 10-20 system with a sampling rate of 1000Hz. All data are filtered within the frequency range of 0.05-47Hz, and artifacts such as eye movement are removed using independent component analysis (ICA). In this study, to avoid the influence of emotions, we only use neutral state fragments from four datasets to focus on the PI task and exclude all other emotional states. Detailed information about the dataset can be found in Tab. 1.

TABLE I
DETAILS OF THE DATASETS.

| Item | THU-EP | SEED | SEED-IV | SEED-V |
|---|---|---|---|---|
| Subjects | 80 | 15 | 15 | 16 |
| Channels | 32 | 62 | 62 | 62 |
| Emotions | 9 | 3 | 4 | 5 |
| Neutral Clips | 4 | 5 | 6 | 3 |

*2) Baselines:* We compare the CMDFM with two of the latest baselines. The first type is the baseline with coupling problems, including Tensorials, MLDA, SFT-Net, JTFT, MD-CAT, and DMV-MAT. The second type is baselines with insufficient fusion, including TSFAN, TSception, AITST, TSFF-NET, MD-CAT, CLT, etc. For specific details, please refer to the relevant work.

*3) Evaluation Index:* This study employed 5-fold cross-validation to assess the average performance of the model and took the overall average accuracy (OAA) and standard deviation (SD) as the evaluation metrics.

*4) Implementation Details:* For each experiment, we set the learning rate to 0.01 and use Stochastic Gradient Descent Momentum (SGDM) as the optimizer [43]. Set the batch size to 32 and the weight attenuation to 0.0005. The training process is carried out over 100 epochs. In addition, the smoothing factor of the model is set to 0.1. The number of network layers is set to 3. The entire model is implemented using PyTorch and experimented on a machine equipped with an NVIDIA GeForce GTX 4090 GPU.

TABLE II
PERFORMANCE COMPARISON WITH BASELINES.

| Baselines | THU-EP OAA ± SD (%) | SEED OAA ± SD (%) | SEED-IV OAA ± SD (%) | SEED-V OAA ± SD (%) |
|---|---|---|---|---|
| Tensorials[9] | - | 98.80±1.35 | - | - |
| MLDA[18] | - | 99.90±0.10 | 96.70±0.28 | 99.80±0.10 |
| TSFAN[30] | - | 94.11±4.40 | 94.10±3.65 | - |
| TSception[26] | 68.91±17.94 | 99.79±0.13 | 89.63±1.07 | 99.83±0.05 |
| AITST[25] | 97.67±0.63 | 98.53±0.25 | 98.89±0.39 | 99.08±0.12 |
| SFT-Net[10] | 97.95±0.11 | 99.62±0.09 | 99.71±0.12 | 99.46±0.14 |
| TSFF-Net[28] | 97.93±0.40 | 99.88±0.05 | 76.40±8.01 | 99.84±0.08 |
| JTFT[24] | 75.65±5.78 | 99.28±0.23 | 94.71±4.07 | 99.54±0.14 |
| MD-CAT[23] | 97.37±1.46 | 99.66±0.09 | 99.33±1.49 | 99.78±0.11 |
| DMV-MAT[17] | 97.90±0.45 | 99.79±0.06 | 99.85±0.12 | 99.30±0.18 |
| CLT[27] | 98.23±0.80 | 99.83±0.06 | 99.89±0.12 | 99.42±0.19 |
| **ours** | **99.68±0.21** | **99.99±0.01** | **99.99±0.01** | **99.92±0.12** |

### *B. Comparison Experiments*

We compare our CMDFM with the recent SOTA algorithm of EEG-based PI. For methods without formal code, the results are cited from the original paper. For the remaining baselines, we conduct a fair comparison using the code they published.

As shown in Tab. 2, compared with the two types of baseline models, CMDFM consistently maintains the best performance on the four benchmark datasets. It is worth noting that its PI accuracy on the complex dataset THU-EP has increased by 1.45%. The recognition accuracy on the SEED and SEED-IV datasets has reached 99.99%, approaching the theoretical limit. The accuracy of PI reaches 99.92% on the SEED-V dataset, which also showed the best performance.

The experimental results show that CMDFM effectively solves the common problems of cross-domain feature parameter coupling and insufficient multi-domain feature fusion in the baseline model. Adaptive multi-scale CNN overcomes the defect of insufficient capture of individual-specific information caused by the fixed-scale strategy commonly found in baseline models. The experimental results confirm that this framework provides a better solution for EEG-based PI.

### *C. Ablation Study*

*1) Ablation Study on CMDFM:* We conduct a comprehensive ablation study on four datasets to evaluate the contribution of the core components in CMDFM. Due to space limitations, the experimental results regarding SEED, SEED-IV and SEED-V are presented in the *github repository*.

Specifically, we evaluate the performance of the model by respectively removing the temporal-spatial-frequency decoupling fusion model (SFT) and the adaptive multi-scale CNN (AMCNN). Finally, we dissolve the label smoothing loss (LS) to explore its effect on inhibiting overfitting.

As shown in Tab. 3, excluding any one of these modules will lead to a decline in performance, thereby emphasizing the necessity of these three parts. It is worth noting that the accuracy of PI decreases significantly by 0.71% compared with the optimal model by ablating the complete multi-domain decoupled fusion model. In addition, the ablation of the adaptive multi-scale CNN also has a significant impact, decreasing by 0.61% compared to the optimal model. Finally, after ablating the label smoothing loss, it decreases by 0.33% compared with the optimal model. These results indicate that the complete multi-domain decoupled fusion model and the adaptive multi-scale CNN significantly enhance the feature extraction capability, while the label smoothing loss reduces overfitting in PI. Their complementary contributions are utilized to improve PI performance.

TABLE III
ABLATION STUDY ON CMDFM.

| Modules | OAA ± SD (%) |
|---|---|
| transformer | 95.18±1.02 |
| w/o SFT | 98.97±0.42 |
| w/o AMCNN | 99.07±0.55 |
| w/o LS | 99.35±0.16 |
| **ours** | **99.68±0.21** |

*2) Ablation Study of Global Fusion Mechanism:* To study the contribution of each domain in the global integration mechanism, we conduct ablation study based on the optimal model. We respectively dissolve the single-domain features of the temporal domain, spatial domain and spectral domain, the dual-domain features of the temporal-spectral domain, the temporal-spatial domain and the spatial-spectral domain, as well as the three-domain features of temporal-spatial-spectral domain, to explore the influence of each domain on the model performance.

As shown in Tab. 4, ablating any domain will lead to a decline in model performance. These results indicate that the comprehensive fusion of features from the three dimensions of temporal, spatial and spectral can enhance the accuracy of PI, highlighting the advantages of our fusion method.

TABLE IV
ABLATION STUDY OF GLOBAL FUSION MECHANISM.

| Modules | OAA ± SD (%) |
|---|---|
| w/o spectral attention | 99.18±0.35 |
| w/o temporal attention | 99.18±0.43 |
| w/o spatial attention | 99.25±0.39 |
| w/o temporal-spectral fusion | 99.13±0.35 |
| w/o spatial-spectral fusion | 99.13±0.41 |
| w/o temporal-spatial fusion | 99.25±0.35 |
| w/o temporal-spatial-spectral fusion | 99.05±0.23 |
| **ours** | **99.68±0.21** |

TABLE V
EFFECTIVENESS OF ADAPTIVE MULTI-SCALE CNN.

| Modules | Year | OAA ± SD (%) |
|---|---|---|
| CNN | 2023 | 99.32±0.39 |
| MS-CNN | 2024 | 99.55±0.14 |
| MBSC | 2024 | 98.80±0.70 |
| Sctnet | 2024 | 99.22±0.06 |
| RepViT | 2024 | 99.05±0.67 |
| Strip R-CNN | 2025 | 99.35±0.19 |
| **ours** | **2025** | **99.68±0.21** |

*3) Effectiveness of Adaptive Multi-Scale CNN:* To verify the contribution of the adaptive multi-scale CNN module, we compare the optimal model of the adaptive multi-scale CNN with the models using other CNN modules. When making comparisons, the complete multi-domain decoupled fusion model remains unchanged, and the label smoothing cross-entropy loss is also used.

As shown in Tab. 5, we compare the adaptive multi-scale CNN with two types of baselines. The first category is CNN and its variants, including CNN, MS-CNN and Strip R-CNN. The second type is CNN that combines attention mechanisms, including MBSC, Sctnet and RepViT. The comparison results show that our adaptive multi-scale CNN has always outperformed all our competitors. The results show that the adaptive weight convolution kernel can effectively and dynamically adjust the contribution of each convolution kernel according to the input, thereby accurately capturing individual specific information.

*4) Effectiveness of Label Smoothing Loss:* Due to the low signal-to-noise ratio characteristic of EEG signals, the PI model suffers from overfitting due to noise interference. Therefore, this paper introduces label smoothing loss to enhance the generalization ability of PI task.

To quantitatively assess its contribution, we compare the optimal model using label smoothing loss with models using other losses, including Kullback Leibler divergence loss [44], focus loss [45], center loss [46] and categorical cross-entropy loss [23], as shown in Tab. 6. When making comparisons, keep the complete multi-domain decoupled fusion model and the adaptive multi-scale CNN unchanged. The experimental results show that the model adopting label smoothing loss significantly outperforms all baseline methods in recognition performance. This confirms that this loss function can effectively alleviate the overfitting problem that is widespread in PI task.

TABLE VI
EFFECTIVENESS OF LABEL SMOOTHING LOSS.

| Losses | Year | OAA ± SD (%) |
|---|---|---|
| Kullback Leibler Divergence Loss | 2023 | 98.95±0.53 |
| Focus Loss | 2023 | 98.23±0.61 |
| Center Loss | 2023 | 99.38±0.40 |
| Categorical Cross-entropy Loss | 2024 | 98.87±0.54 |
| **ours** | **2025** | **99.68±0.21** |

### *D. Hyperparameter Analysis*

We conduct a hyperparameter analysis to investigate the impact of different hyperparameters on the PI performance of our proposed CMDFM. These hyperparameters include the smoothing factor in the loss function, the number of model layers, the batch size, the learning rate, and the weight decay coefficient. For more detailed information, please refer to the hyperparameter analysis section in the *github repository*.

## V. CONCLUSION

This study proposes an innovative EEG feature learning framework (CMDFM). We design an independent temporal-spatial-spectral attention mechanism, successfully achieving the effective decoupling and extraction of EEG signal features in the temporal domain, spatial domain and spectral domain, and significantly reducing cross-domain interference. Meanwhile, a full-domain fusion mechanism is proposed, which efficiently integrates the decoupled multi-domain features into a unified and information-rich individual representation. Finally, we propose an input-based weight adaptive mechanism to accurately extract individual-specific information. Extensive experiments have shown that our model consistently outperforms the most advanced methods.

## REFERENCES


[1] X. Guan, Z. Pan, J. Feng, and J. Zhou, "Joint identity verification and pose alignment for partial fingerprints," *IEEE Transactions on Information Forensics and Security*, 2024.

[2] S. Kadhim, J. K. S. Paw, Y. C. Tak, S. Ameen, and A. Alkhayyat, "Deep learning for robust iris recognition: Introducing synchronized spatiotemporal linear discriminant model-iris," *Advances in Artificial Intelligence and Machine Learning*, 2025.

[3] F. Zhang, T. Yang, L. Cao, K. Du, Y. Guo, P. Song, and C. Shao, "Deep supervised anomaly detection for generalized face forgery detection," *Pattern Recognition*, vol. 169, p. 111976, 2026.

[4] H.-L. Chan, P.-C. Kuo, C.-Y. Cheng, and Y.-S. Chen, "Challenges and future perspectives on electroencephalogram-based biometrics in person recognition," *Frontiers in neuroinformatics*, vol. 12, p. 66, 2018.

[5] A. Seyfizadeh, R. L. Peach, P. Tovote, I. U. Isaias, J. Volkmann, and M. Muthuraman, "Enhancing security in brain–computer interface applications with deep learning: Electroencephalogram-based user identification," *Expert Systems with Applications*, vol. 253, p. 124218, 2024.

[6] P. Arnau-González, M. Arevalillo-Herráez, S. Katsigiannis, and N. Ramzan, "On the influence of affect in eeg-based subject identification," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 391–401, 2018.

[7] Y. Zhang, X. Li, C. Zhao, W. Zheng, M. Wang, Y. Zhang, H. Ma, and D. Gao, "Affective eeg-based person identification using channel attention convolutional neural dense connection network," *Security and Communication Networks*, vol. 2021, no. 1, p. 7568460, 2021.

[8] L. Alshehri and M. Hussain, "A lightweight gct-eegnet for eeg-based individual recognition under diverse brain conditions," *Mathematics*, vol. 12, no. 20, p. 3286, 2024.

[9] W. Li, Y. Yi, M. Wang, B. Peng, J. Zhu, and A. Song, "A novel tensorial scheme for eeg-based person identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–17, 2022.

[10] D. Gao, K. Wang, M. Wang, J. Zhou, and Y. Zhang, "Sft-net: a network for detecting fatigue from eeg signals by combining 4d feature flow and attention mechanism," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 8, pp. 4444–4455, 2023.

[11] Y. Du, Y. Xu, X. Wang, L. Liu, and P. Ma, "Eeg temporal–spatial transformer for person identification," *Scientific Reports*, vol. 12, no. 1, p. 14378, 2022.

[12] Y. Hu, L. Sun, X. Mao, and S. Zhang, "Eeg data augmentation method for identity recognition based on spatial–temporal generating adversarial network," *Electronics*, vol. 13, no. 21, p. 4310, 2024.

[13] R. Jin, Y. Wang, R. Bai, H. Xie, and G. Wang, "Epi-cgan: Robust eeg-based person identification using conditional generative adversarial network," in *2022 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2022, pp. 1–9.

[14] W. Alsumari, M. Hussain, L. Alshehri, and H. A. Aboalsamh, "Eeg-based person identification and authentication using deep convolutional neural network," *Axioms*, vol. 12, no. 1, p. 74, 2023.

[15] B. Bandana Das, S. Kumar Ram, K. Sathya Babu, R. K. Mohapatra, and S. P. Mohanty, "Person identification using autoencoder-cnn approach with multitask-based eeg biometric," *Multimedia Tools and Applications*, vol. 83, no. 35, pp. 83 205–83 225, 2024.

[16] X. Lu, A. Wen, L. Sun, H. Wang, Y. Guo, and Y. Ren, "An epileptic seizure prediction method based on cbam-3d cnn-lstm model," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp. 417–423, 2023.

[17] W. Cui, Y. Xiang, Y. Wang, T. Yu, X.-F. Liao, B. Hu, and Y. Li, "Deep multiview module adaption transfer network for subject-specific eeg recognition," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[18] Y. Miao, W. Jiang, N. Su, J. Shan, T. Jiang, and N. Zuo, "Mlda: Multi-loss domain adaptor for cross-session and cross-emotion eeg-based individual identification," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 12, pp. 5767–5778, 2023.

[19] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, "Affective eeg-based person identification using the deep learning approach," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 3, pp. 486–496, 2019.

[20] S. K. Khare and V. Bajaj, "Time–frequency representation and convolutional neural network-based emotion recognition," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 7, pp. 2901–2909, 2020.

[21] R. Ashenaei, A. A. Beheshti, and T. Y. Rezaii, "Stable eeg-based biometric system using functional connectivity based on time-frequency features with optimal channels," *Biomedical Signal Processing and Control*, vol. 77, p. 103790, 2022.

[22] T. Behrouzi and D. Hatzinakos, "Graph variational auto-encoder for deriving eeg-based graph embedding," *Pattern Recognition*, vol. 121, p. 108202, 2022.

[23] J. Jin, Z. Chen, H. Cai, and J. Pan, "Affective eeg-based person identification with continual learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–16, 2024.

[24] Y. Chen, S. Liu, J. Yang, H. Jing, W. Zhao, and G. Yang, "A joint time-frequency domain transformer for multivariate time series forecasting," *Neural Networks*, vol. 176, p. 106334, 2024.

[25] H. Cai, J. Jin, H. Wang, L. Li, Y. Huang, and J. Pan, "Aitst—affective eeg-based person identification via interrelated temporal–spatial transformer," *Pattern Recognition Letters*, vol. 174, pp. 32–38, 2023.

[26] Y. Ding, N. Robinson, S. Zhang, Q. Zeng, and C. Guan, "Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 2238–2250, 2022.

[27] X. Shao, C. Chang, J. Q. Gan, and H. Wang, "An interpretable contrastive learning transformer for eeg-based person identification," *IEEE Transactions on Information Forensics and Security*, 2025.

[28] Z. Miao and M. Zhao, "Time–space–frequency feature fusion for 3-channel motor imagery classification," *Biomedical Signal Processing and Control*, vol. 90, p. 105867, 2024.

[29] Y. Zhou, X. Wen, and D. Zhang, "Simultaneously cognitive workload assessment and personal identification using deep cnn," in *2024 IEEE 9th International Conference on Computational Intelligence and Applications (ICCIA)*. IEEE, 2024, pp. 176–181.

[30] X. Jin, X. Yang, W. Kong, L. Zhu, J. Tang, Y. Peng, Y. Ding, and Q. Zhao, "Tsfan: tensorized spatial-frequency attention network with domain adaptation for cross-session eeg-based biometric recognition," *Journal of Neural Engineering*, vol. 21, no. 4, p. 046005, 2024.

[31] A. Mahmoud, K. Amin, M. M. Al Rahhal, W. S. Elkilani, M. L. Mekhalfi, and M. Ibrahim, "A cnn approach for emotion recognition via eeg," *Symmetry*, vol. 15, no. 10, p. 1822, 2023.

[32] X. Wang, M. Dang, K. Yang, X. Cui, D. Zhang, and C. Chen, "The ensemble multi-scale convolution neural network for visual target detection eeg-based brain-computer interfaces," *Biomedical Signal Processing and Control*, vol. 96, p. 106583, 2024.

[33] X. Yuan, Z. Zheng, Y. Li, X. Liu, L. Liu, X. Li, Q. Hou, and M. Cheng, "Strip r-cnn: Large strip convolution for remote sensing object detection. arxiv 2025," *arXiv preprint arXiv:2501.03775*, 2025.

[34] F. Bougourzi, F. Dornaika, A. Nakib, and A. Taleb-Ahmed, "Embtrattunet: a novel edge loss function and transformer-cnn architecture for multi-classes pneumonia infection segmentation in low annotation regimes," *Artificial Intelligence Review*, vol. 57, no. 4, p. 90, 2024.

[35] Z. Xu, D. Wu, C. Yu, X. Chu, N. Sang, and C. Gao, "Sctnet: Single-branch cnn with transformer semantic information for real-time segmentation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 38, no. 6, 2024, pp. 6378–6386.

[36] A. Wang, H. Chen, Z. Lin, J. Han, and G. Ding, "Repvit: Revisiting mobile cnn from vit perspective," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 15 909–15 920.

[37] A. Panzino, G. Orrù, G. L. Marcialis, and F. Roli, "Eeg personal recognition based on 'qualified majority'over signal patches," *IET Biometrics*, vol. 11, no. 1, pp. 63–78, 2022.

[38] G. Xia, O. Laurent, G. Franchi, and C.-S. Bouganis, "Towards understanding why label smoothing degrades selective classification and how to fix it," *arXiv preprint arXiv:2403.14715*, 2024.

[39] X. Hu, F. Wang, and D. Zhang, "Similar brains blend emotion in similar ways: Neural representations of individual difference in emotion profiles," *Neuroimage*, vol. 247, p. 118819, 2022.

[40] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on autonomous mental development*, vol. 7, no. 3, pp. 162–175, 2015.

[41] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.

[42] W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, "Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 715–729, 2021.

[43] Y. Liu, Y. Gao, and W. Yin, "An improved analysis of stochastic gradient descent with momentum," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18 261–18 271, 2020.

[44] T. Xu, W. Dang, J. Wang, and Y. Zhou, "Dagam: a domain adversarial graph attention model for subject-independent eeg-based emotion recognition," *Journal of Neural Engineering*, vol. 20, no. 1, p. 016022, 2023.

[45] J. Shen, Y. Zhang, H. Liang, Z. Zhao, K. Zhu, K. Qian, Q. Dong, X. Zhang, and B. Hu, "Depression recognition from eeg signals using an adaptive channel fusion method via improved focal loss," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3234–3245, 2023.

[46] H. Zhi, Z. Yu, T. Yu, Z. Gu, and J. Yang, "A multi-domain convolutional neural network for eeg-based motor imagery decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3988–3998, 2023.