

Original Research

A Multi-Teacher Distilling Framework With Data Privacy for EEG Emotion Recognition

Jiaqi Yang¹, Tianhao Gu^{1,*} , Chong Lin¹, Jiazhen Xu²

¹School of Automation, Qingdao University, 266073 Qingdao, Shandong, China

²Department of Pharmacology, School of Pharmacy, Qingdao University, 266073 Qingdao, Shandong, China

*Correspondence: gutianhao@qdu.edu.cn (Tianhao Gu)

Academic Editor: Bettina Platt

Submitted: 21 June 2025 Revised: 18 September 2025 Accepted: 30 September 2025 Published: 27 November 2025

Abstract

Background: This study addressed three key challenges in subject-independent electroencephalography (EEG) emotion recognition: limited data availability, restricted cross-domain knowledge transfer, and suboptimal feature extraction. The aim is to develop an innovative framework that enhances recognition performance while preserving data privacy. **Methods:** This study introduces a novel multi-teacher knowledge distillation framework that incorporates data privacy considerations. The framework first comprises n subnets, each sequentially trained on distinct EEG datasets without data sharing. The subnets, excluding the initial one, acquire knowledge through the weights and features of all preceding subnets, enabling access to more EEG signals during the training process while maintaining privacy. To enhance cross-domain knowledge transfer, a multi-teacher knowledge distillation strategy was designed, featuring knowledge filters and adaptive multi-teacher knowledge distillation losses. The knowledge filter integrates cross-domain information using a multi-head attention module with a gate mechanism, ensuring effective inheritance of knowledge from all previous subnets. Simultaneously, the adaptive multi-teacher knowledge distillation loss dynamically adjusts the direction of knowledge transfer based on filtered feature similarity, preventing knowledge loss in single-teacher models. Furthermore, a spatio-temporal gate module is proposed to eliminate unnecessary frame-level information from different channels and extract important channels for improved feature representation without requiring expert knowledge. **Results:** Experimental results demonstrate the superiority of the proposed method over the current state of the art, achieving a 2% performance improvement on the DEAP dataset. **Conclusions:** The proposed multi-teacher distillation framework with data privacy addresses the challenges of insufficient data availability, limited cross-domain knowledge transfer, and suboptimal feature extraction in subject-independent EEG emotion recognition, demonstrating strong potential for scalable and privacy-preserving emotion recognition applications.

Keywords: electroencephalography; emotions; data privacy; machine learning

1. Introduction

In recent years, there has been extensive exploration of biological signals to uncover potential relationships between the nervous system and physiological functions in humans [1–3]. This investigation has revealed new avenues in fields such as disease monitoring [4], exercise therapy [5,6], and exercise rehabilitation [7,8]. In the realm of implantable signal processing, electromyography (EMG) signals have successfully controlled bionic hands for amputees [9], demonstrating remarkable success. However, the implantable signal collection mode, which entails open wounds, presents hazards, rendering the non-invasive mode a safer alternative. Electroencephalography (EEG) signals, as a standard non-invasive neural signal, have garnered considerable interest due to their robust temporal resolution and resistance to interference [10]. EEG emotion recognition [11,12], in particular, has emerged as a valuable method for detecting fear and anxiety disorders, contributing to the improvement of mental health. Meanwhile, some publicly available EEG emotion datasets, including SEED (<https://bcmi.sjtu.edu.cn/home/seed/>) [13], DEAP

(<http://www.eecs.qmul.ac.uk/mmv/datasets/deap/>) [14] and DREAMER (<https://zenodo.org/record/546113>) [15], have also provided a foundation and guarantee for the rapid development of this field.

Currently, artificial intelligence (AI)-based approaches, encompassing both traditional machine learning techniques and deep learning (DL) methods [16,17], have become the predominant paradigm in EEG-based emotion recognition. Approaches such as Support Vector Machines (SVM) [18], Long Short-Term Memory (LSTM) [19], and Convolutional Neural Networks (CNN) [20] have achieved notable success, contributing to substantial improvements in overall performance. However, most of these methods operate in a subject-dependent mode, which restricts their generalizability and practical applicability. In contrast, the subject-independent mode offers broader application prospects, but faces three main challenges in EEG emotion recognition: (1) insufficient available data [21]: large volume of cross-subject data is essential for model training; otherwise, it becomes challenging to capture subject-invariant features [22]; however, EEG signals, due to their sensitive and privacy-preserving nature,



are difficult to collect in large quantities, which hinders performance improvement in subject-independent settings; (2) insufficient cross-domain knowledge transfer [23]: the EEG signals of different subjects are from different domains; obtaining universal knowledge from a portion of subjects and achieving cross-domain knowledge transfer to complete the EEG emotion recognition from other subjects are significantly important [24–26]; however, the current limited performance of cross-domain knowledge transfer cannot further optimize recognition precision; (3) insufficient feature extraction: most EEG signal feature extraction methods, such as differential entropy [27], asymmetric spatial pattern [28], and high order zero crossing count [29] rely on manual design, which requires high professional knowledge from researchers and can sometimes cause information loss; besides, existing automatic feature extraction methods generally focus on spatial information extraction [30], ignoring the screening of time segments, which leads to a large amount of unnecessary information interfering with the recognition results.

In response to the above issues, this paper designs the framework from three corresponding aspects. First, to obtain more data for model training, the paper proposes a multi-teacher knowledge distillation (KD) framework, which includes n subnets under serial mode. All subnets, except the initial one, are trained sequentially on distinct EEG datasets, incorporating feature representations from all previously trained subnets to consolidate accumulated knowledge and facilitate cross-domain knowledge transfer [31]. During this process, each subnet relies solely on the pretrained weights of preceding subnets without sharing raw EEG data, thereby preserving data privacy [32] while enhancing performance in EEG-based emotion recognition.

Second, to enhance cross-domain knowledge transfer, the study introduces a knowledge filter (KF) and an adaptive multi-teacher KDloss. The KF integrates features from multiple domains using a gated multi-head attention mechanism, selectively inheriting knowledge relevant to the current domain and thereby improving cross-domain knowledge transfer. Meanwhile, the adaptive multi-teacher KDloss adaptively regulates the direction of knowledge transfer by evaluating feature similarities across all preceding subnets, mitigating potential knowledge loss associated with reliance on a single teacher model.

Third, to improve the effectiveness of automatic feature extraction, the study proposes a spatio-temporal gate (STG) module to handle varying reaction periods of the same emotion across different EEG electrodes. The module identifies salient frame-level segments for each channel by computing similarities between frames and applying a sigmoid function, and subsequently adjusts channel-wise weights through self-attention to optimize frame-level representations. This feature extraction process aligns with the characteristics of EEG signals, effectively filtering out irrelevant information.

A multi-teacher knowledge distillation framework with data privacy (MTKDDP) is suggested for EEG emotion recognition. The framework employs adaptive knowledge distillation, feature filtering, and spatiotemporal gating to tackle challenges related to limited cross-domain knowledge transfer, inadequate data, and suboptimal feature extraction, thereby enhancing generalization, robustness, and automatic feature learning in subject-independent contexts. Experimental results demonstrate a 2% enhancement in performance compared to current methodologies on the DEAP dataset.

2. Related Work

2.1 Knowledge Distillation

KD [33] has emerged as a crucial technique for model compression and transfer learning. By leveraging a loss function that incorporates both soft and hard targets, KD enables the effective transfer of knowledge from a teacher network to a student network. As a widely adopted approach, it facilitates transferring knowledge from one or multiple large teacher models to a more compact student model [33,34]. However, conventional KD approaches generally depend on a single teacher network, which limits knowledge diversity and fails to effectively address cross-domain variability in EEG emotion recognition [23,26]. To overcome this limitation, multi-teacher knowledge distillation (MTKD) has been introduced, allowing a student model to learn from multiple teachers and thereby enhancing the diversity and robustness of knowledge transfer [35–38]. For example, Ye *et al.* [39] propose a multi-teacher feature-ensemble strategy that aggregates representations from multiple teachers to enhance transfer effectiveness. Nevertheless, many existing MTKD methods aggregate teacher outputs using simple averaging or other static rules, which can degrade performance when teacher predictions are of low quality [40,41]. Despite these advances, MTKD still faces limitations in facilitating cross-domain knowledge transfer, as naive feature aggregation may result in information loss and fail to fully capture heterogeneous domain knowledge. To address this issue, this paper proposes an MTKD method that combines KF with an adaptive multi-teacher KDloss, selectively integrating complementary cross-domain information while minimizing redundancy and noise, thereby improving generalization and robustness in subject-independent EEG emotion recognition.

2.2 EEG-Based Emotion Recognition

AI-based technologies for EEG emotion recognition achieve significant progress, encompassing both traditional machine learning and DL approaches. Early methods, such as SVM and K-Nearest Neighbors (KNN), demonstrate feasibility but depend on manual feature engineering and struggle to scale effectively to high-dimensional

EEG data [42,43]. Subsequent variants like the Universal Support Vector Machine (USVM) aim to reduce complexity yet still show generalization limitations [44]. DL frameworks emerge as powerful alternatives, learning complex representations directly from EEG signals. For instance, regional-asymmetric convolution enhances fine-grained pattern capture, while generative modeling supports improved training efficiency [30,45]. Despite these advances, the subject-independent setting continues to face inadequate data availability. Robust, generalizable features require large and diverse datasets, which remain difficult to obtain in practice [11,13]. Data augmentation and transfer learning provide partial alleviation but frequently require cross-subject or cross-institutional sharing of raw EEG data, thereby introducing privacy and compliance concerns [16,17,46]. KD addresses these concerns by enabling knowledge transfer without exposing raw data [32]; nevertheless, most KD approaches in EEG emotion recognition rely on single-teacher frameworks, constraining knowledge diversity and limiting robustness under conditions of data scarcity [47,48]. Recent work on multi-teacher KD introduces adaptive teacher selection and weighting to improve stability and transfer quality, although its application to EEG emotion recognition remains underexplored [36,39].

Beyond data limitations, inadequate feature extraction constitutes a second major bottleneck. EEG signals exhibit rich spatiotemporal dynamics, necessitating models that capture spatial-temporal dependencies while effectively suppressing noise. Handcrafted features, such as differential entropy and asymmetric spatial patterns, provide partial solutions but remain susceptible to information loss and rely heavily on expert design [27,28]. DL models improve automatic representation learning, yet many lack dynamic mechanisms to filter redundant or irrelevant frame-level segments. For instance, models designed to learn spatiotemporal dependencies may still omit adaptive cross-channel gating, leading to residual noise and reduced accuracy [49]. Addressing these gaps requires frameworks that simultaneously extract fine-grained spatiotemporal representations and perform dynamic filtering. To this end, the proposed MTKDDP framework sequentially trains multiple subnets on heterogeneous EEG datasets without sharing raw data and distills complementary knowledge from multiple teachers to enhance generalization under conditions of limited data [36,39]. In parallel, a STG module adaptively selects salient frame-level features across channels using attention-based similarity, thereby reducing redundancy and strengthening representation quality [50]. In combination, knowledge filtering and adaptive distillation loss further improve cross-domain generalization while maintaining privacy protection.

3. Method

Here, a comprehensive overview is presented for the four components of the MTKDDP framework, including

EEG signal preprocessing, the MTKDDP architecture, the STG module and the MTKDDP algorithm.

3.1 EEG Signal Preprocessing

The original EEG signals are commonly represented as multi-dimensional tensors with baseline signals. Since neural networks generally require matrix-like input, these tensors do not directly satisfy this requirement. Consequently, preprocessing is necessary to reshape EEG signals into suitable dimensions for network input. Based on the design of the STG module in the paper, the original EEG signal is processed into a signal of size (C, T, N) without the baseline signals, where C represents the number of channels, T is the number of time steps, and N indicates the number of samples. The detailed preprocessing steps are as follows:

First, the baseline in the original EEG signals is removed using signals recorded in a calm state. Specifically, the per-second mean of the calm-state signals is computed as the baseline, which is then subtracted from the emotional signals to complete the preprocessing. Second, similar to Wang's method [48], the processed signal is divided into multiple frame-level signals according to the 3 seconds per segment [50]. Here, the length of each frame level signal is defined as 3 seconds to ensure sufficient emotional information and avoid excessive redundant information. Therefore, if the total length of the original signal is T_{sum} , then $T = T_{sum} / 3$ and the size of the signal is (T, C, N) . Afterwards, the existing signals are transposed as the signal of size (C, T, N) , which meets the input requirements of our framework. Additionally, to reduce model complexity, downsampling is applied, proportionally scaling the data with respect to N to improve computational efficiency.

3.2 MTKDDP Architecture

To optimize performance in subject-independent EEG emotion recognition, this work proposes a multi-teacher KD framework that preserves data privacy [46,51,52], as shown in Fig. 1, for enhancing knowledge transfer and automatic feature extraction. The framework includes n subnets and $(n-2)$ KFs, which are detailed in the subsequent subsection. Except for the initial subnet, each subnet contains only 1.3 M parameters, demonstrating the framework's efficiency. The framework is optimized by sequentially training subnets on different EEG signals to ensure privacy protection. Each subnet comprises two LSTM layers, a gate module, and a multi-dense module. The first LSTM layer extracts temporal features, while the second LSTM layer maps these features to labels. The gate module, positioned between the two LSTM layers and detailed in the following subsection, selectively eliminates unnecessary features. Finally, the multi-dense module, located at the end, performs label prediction. Especially, the first LSTM layer and gate module of the first subnet are different from the subsequent subnets, where the first LSTM layer in the first subnet is

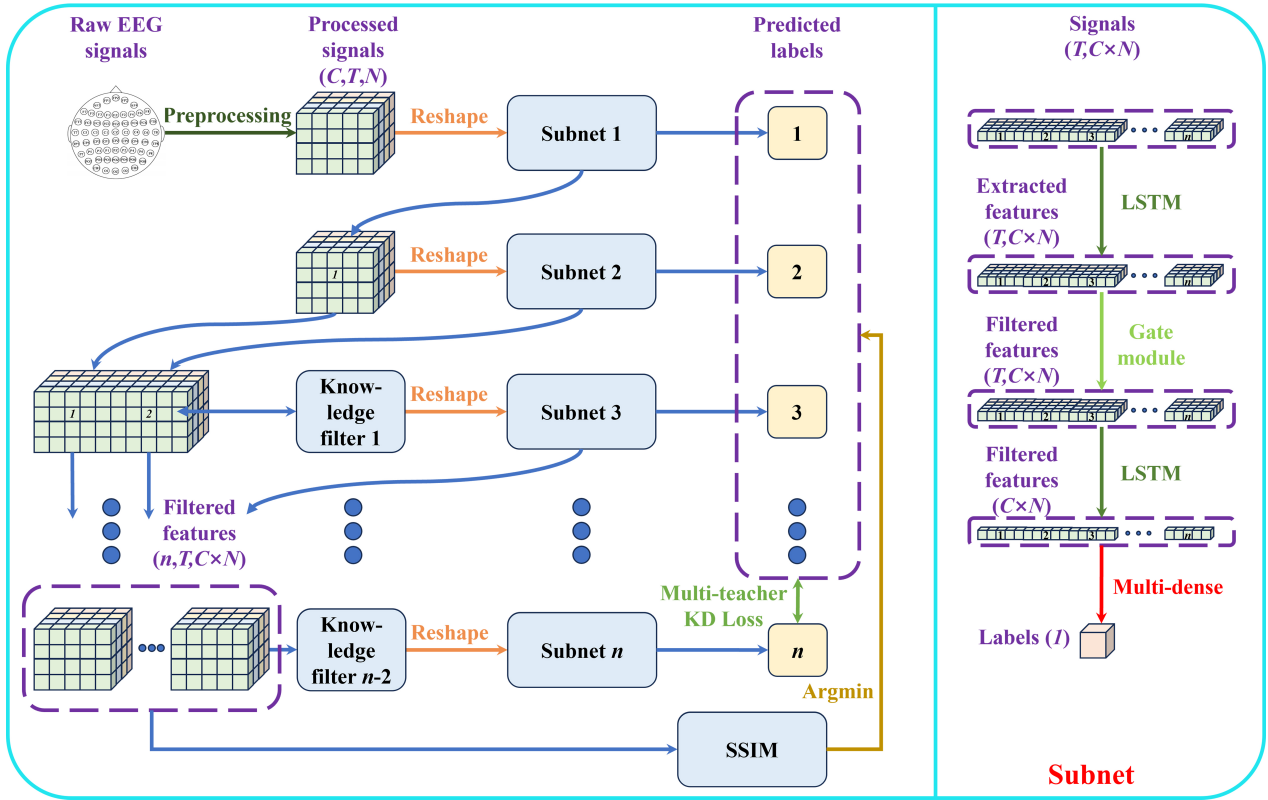


Fig. 1. The architecture of our MTKDDP for EEG emotion recognition. MTKDDP, multi-teacher knowledge distillation framework with data privacy; EEG, electroencephalography; KD, knowledge distillation; SSIM, Structural Similarity; LSTM, Long Short-Term Memory.

trained by EEG signals of single channel together to obtain feature information. The first gate module is the STG module, which screens emotional response periods across different electrode positions and removes irrelevant information, while subsequent gate modules operate based on temporal features. During training, the weights of each trained subnet are frozen, and only the filtered features are provided as input to the following subnet to enhance knowledge transfer. When more than two subnets are employed, the KF selects features generated by all preceding subnets to ensure the effectiveness of knowledge transfer. In addition, except for the first LSTM layer of each subnet using *ReLU* activation function and the output layer using *Sigmoid* activation function, all other activation functions are *Tanh*.

3.3 Spatio-Temporal Gate Module

Capturing emotional information across the entire EEG signal presents challenges, as the response times at different electrode positions are often inconsistent. To overcome the limitations of existing feature extraction methods that focus solely on spatial information, this work introduces the STG module, illustrated in Fig. 2, which extracts the most relevant features between electrode positions and emotional responses. The STG module effectively filters

out unnecessary features from each electrode, thereby enhancing automatic feature extraction. The module comprises three components: a temporal-based gate module set, an LSTM layer, and a spatio-based gate module. In the temporal-based gate module set, gate modules are utilized to filter features C times, with shared weights optimizing the temporal-based gate modules more effectively. Subsequently, after adjusting the feature size, features with dimensions $(T, C \times N)$ undergo feature extraction in the LSTM layer. Following another adjustment to the feature size $(C, T \times N)$, the framework executes feature filtering through the STG module. Throughout this process, the specifics of the gate units, used in temporal-based gate modules and the spatio-based gate modules, are detailed as follows.

Assuming the input of the gate module is $H = \langle h_1, h_2, \dots, h_T \rangle$, where T is the number of time steps. First, the similarity between extracted features at different time steps as follows:

$$g_{t,t'} = \tanh(W_g h_t + W_g' h_{t'} + b_g) \quad (1)$$

$$e_{t,t'} = \sigma(W_e g_{t,t'} + b_e) \quad (2)$$

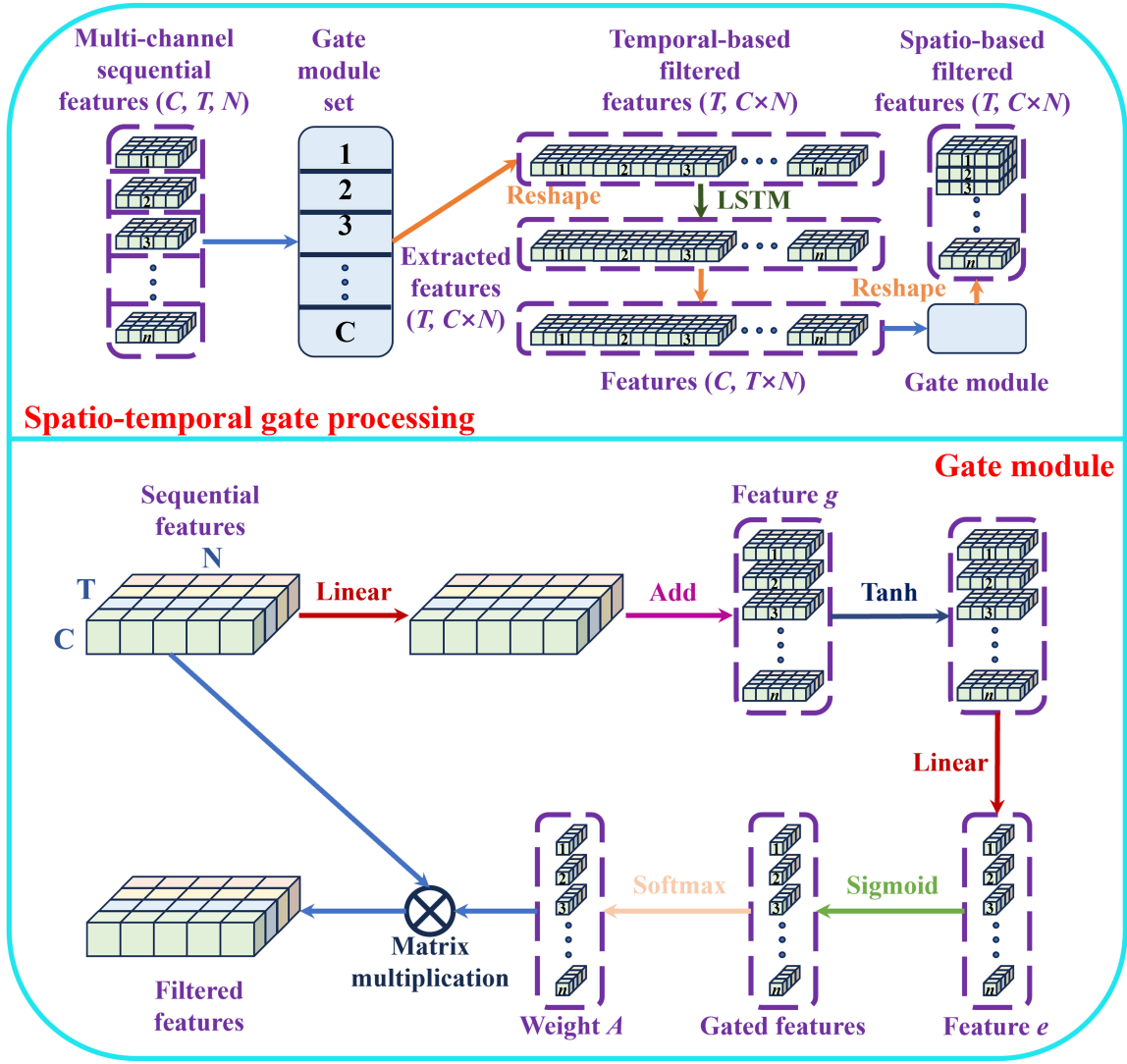


Fig. 2. The details of spatio-temporal gate module.

where W_g , $W_{g'}$ and W_e are the weights of the linear layer, b_g and b_e indicate the bias of the linear layer, t and t' represent the time steps, and σ denotes the *Sigmoid* activation function.

The σ activation function efficiently assigns zero weights to irrelevant features within the correlation ground, effectively eliminating unnecessary elements and optimizing the feature extraction process. According to Eqns. 1,2, $e_t = \langle e_{(t,1)}, e_{(t,2)}, \dots, e_{(t,T)} \rangle$ can be computed. Meanwhile, the weight vector a_t in $A_{T \times T} = \langle a_1, a_2, a_3, \dots, a_T \rangle$ can be obtained as follows:

$$a_t = \text{softmax}(e_t) \quad (3)$$

Hence, the weight matrix $A_{T \times T}$ for filtering features is calculated. Then, the paper can get the filtered feature O as follows:

$$O = A_{T \times T} H'_{T \times 1} \quad (4)$$

Here, the STG module can be used to usefully filter the unnecessary features by Eqns. 1,2,3,4. In addition, all subnets except the first employ gate units as gate modules to enhance feature extraction.

3.4 MTKDDP Algorithm

EEG signals involve personal privacy and are challenging to collect publicly in large quantities, which limits the optimization of models for EEG emotion recognition. In particular, in the subject-independent mode, cross-domain knowledge transfer remains insufficient. Therefore, the paper proposes a multi-teacher KD framework with data privacy, which includes n subnets trained sequentially. The framework enables knowledge transfer to enhance performance without sharing EEG signals, while leveraging multi-teacher KD to facilitate cross-domain knowledge transfer, using the adaptive multi-teacher KDloss to guide the direction of knowledge transfer. The training process is divided into three stages.

First, the preprocessed EEG signals are utilized to train the initial subnet, wherein the network's output is aligned directly with the ground truth distribution through the cross-entropy loss function, expressed as follows:

$$L(p, q) = -(p * \log(q) + (1 - p) * \log(1 - q)) \quad (5)$$

where p denotes the prediction from the first subnet and q represents the ground truth.

Second, the features meticulously filtered by the initial subnet serve as inputs for the subsequent subnet. The KDloss is then applied to guide its outputs toward both the soft targets predicted by the initial subnet and the hard targets derived from the actual labels. This intricate process achieves effective knowledge transfer and successfully concludes the model training phase. The KDloss is represented as follows:

$$\text{KDloss} = \lambda * L(q, \text{softmax}(z)) + (1 - \lambda) * L(\text{softmax}(p/T^*), \text{softmax}(z/T^*)) \quad (6)$$

where z indicates the prediction from the second subnet before the last *Sigmoid* function, T^* indicates the temperature, and λ indicates the trade-off ratio between the real labels and soft targets.

Third, the multi-teacher knowledge KD strategy is employed to train the n -th subnet ($n > 2$), diverging from the preceding two training stages. This phase involves utilizing the KF and the adaptive multi-teacher knowledge KDloss, as illustrated in Fig. 3. Notably, the features filtered from the preceding $(n-1)$ subnets are collectively transferred to the subsequent KF during this process together. The KF, as shown in the Fig. 3, inspired by the multi-head attention [49], can select the important features for optimizing the description of all knowledge from previous subnets. In the process, the filtered feature O_i are concatenated as a feature of size $((n-1) \times C, T, N)$. Then, Q, K and V are used to compute the similarity between the filtered feature and get the weight tensor A . Here, different from the traditional multi-head attention, the *Sigmoid* function, as the gate module, is introduced to eliminate unnecessary features for elevating the knowledge transfer, and the A are turned to A^* through *Sigmoid* function. Afterwards, the features are filtered to select the knowledge from the previous subnets. Finally, the processed features are simplified as the tensors with original dimensions by linear layer or principal component analysis. In addition, the adaptive multi-teacher KDloss is applied into model optimization for avoiding the loss between knowledge transfer, where the similarities between filtered feature O_i are first obtained by Structural Similarity (SSIM) [53].

The adaptive loss is defined as follows:

$$\begin{aligned} \text{KDLoss} = & \lambda * L(q, \text{softmax}(z)) + (1 - \lambda) * (k * \\ & L(\text{softmax}(p_1/T^*), \text{softmax}(z/T^*)) + \\ & (1 - k) * L(\text{softmax}(p_2/T^*), \text{softmax}(z/T^*))) \end{aligned} \quad (7)$$

where k indicates the ratio between the direction of two soft targets. z, p_1 and p_2 represent the prediction from $(n+1)$ -th subnet, n -th subnet and i -th subnet respectively.

Here, to retain maximum knowledge, the subnet exhibiting the lowest similarity between the output and feature O_n is employed as an additional teacher network to optimize the current subnet. Thus, i is solved as follows:

$$i = \text{argmin}(SSIM(O_i, O_n)), i = 1, 2, \dots, n \quad (8)$$

During the training process, k can adjust the balance between the training cost and the final performance. As k increases, the subnet mainly inherits the knowledge of its previous subnet, where the model can converge quickly with weak generalization performance. In contrast, the model achieves improved generalization performance, although convergence is slow. Increasing the number of subnets can enhance both performance and generalization, but this incurs substantial computational cost. To address this issue, a pruning method is employed, in which the model is pruned and training is halted once the feature similarity of three consecutive subnets exceeds a predefined threshold.

4. Experiments

This section presents the implementation details used to demonstrate the performance of the framework. Then, the performance of the MTKDDP framework is analyzed through comparative experiments on the public DEAP dataset [14] and DREAMER dataset [15] in subject-independent mode. Subsequently, the effectiveness of the proposed methods is evaluated sequentially through ablation experiments and visualization analyses.

4.1 Implementation Details

In comparative experiments, the selected frameworks include classical models such as Convolutional Long Short-Term Memory (CLSTM) [54] and Attention-LSTM [55], as well as state-of-the-art frameworks including Recurrent Attention Convolutional Neural Network (RACNN) [30], Attention-based Temporal Dense Dual (ATDD)-LSTM [56], Frame-Level Teacher-Student Learning With Data Privacy (FLTSDP) [47], Frame-Level Distilling Neural Network (FLDnet) [48], Cascaded Gated Recurrent Unit-Multi-Channel Dynamic Graph Network (CGRU-MDGN) [42], and Multi-Task Learning Fusion Network (MTL-FuseNet) [57]. A dual strategy is employed for selecting comparison algorithms, incorporating both classical and state-of-the-art (SOTA) approaches to ensure a com-

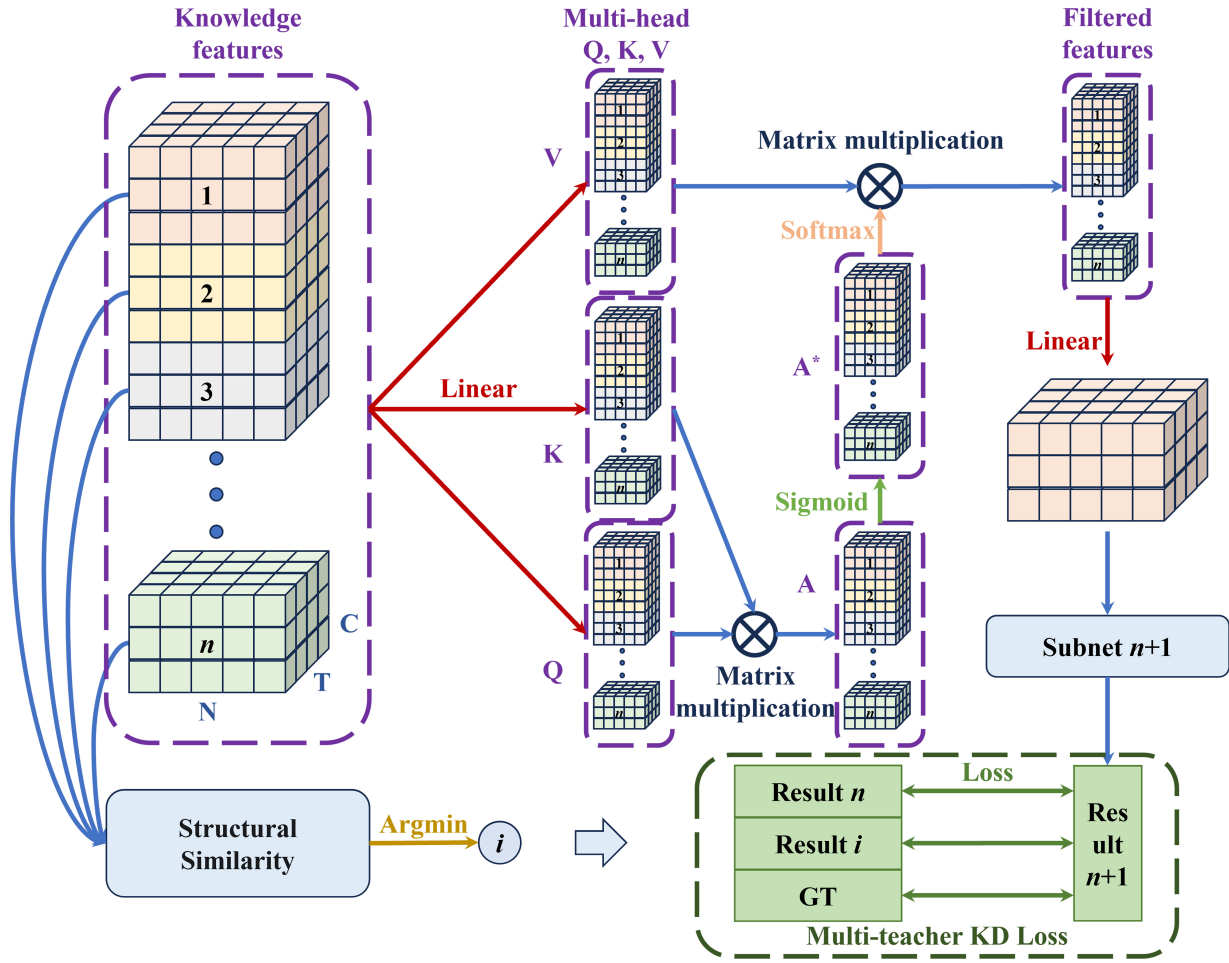


Fig. 3. The details of multi-teacher knowledge distillation strategy. GT, ground truth.

prehensive and fair evaluation. Classical models such as CLSTM and Attention-LSTM serve as widely recognized baselines in EEG-based emotion recognition due to their established capability in feature extraction and temporal sequence modeling, providing reliable benchmarks for assessing the overall effectiveness of the proposed method. In addition, recent SOTA methods—including FLTSDP, which integrates data privacy protection mechanisms, as well as other advanced approaches based on knowledge distillation or transfer learning—are incorporated to evaluate performance against the latest developments in data privacy, cross-domain knowledge transfer, and feature extraction. This combination ensures that the evaluation reflects both competitiveness with established baselines and superiority in cutting-edge scenarios.

Here, the EEG signals are first preprocessed into $32 \times 20 \times 60$ tensors and $14 \times 20 \times 48$ tensors respectively for training. Second, considering the number of participants in the public datasets, where the DEAP dataset [14] and DREAMER dataset [15] have 32 and 23 subjects respectively, the number of subnets in the framework is defined as 3 to ensure that each subnet can have certain generaliza-

tion performance in 120 epochs during training. In addition, the multi-teacher KD parameters k , T^* and learning rate of this experiment are set to 0.9, 0.9 and 0.001, respectively. The emotional dimensions—valence, arousal, dominance, and liking—were annotated based on participants' self-reported ratings in the DEAP and DREAMER datasets. For each stimulus, participants provide evaluations using a standardized 9-point Likert scale. For binary classification, a threshold of 5 is applied, where scores equal to or greater than 5 are labeled as “high” and scores below 5 as “low”, in accordance with widely adopted practices in EEG-based emotion recognition [58]. Although these labels are self-reported and conventional inter-rater reliability metrics are not applicable, both datasets serve as well-established benchmarks with validated rating procedures. This labeling protocol ensures a clear separation between high and low emotional states while maintaining a balanced distribution of samples for model training, thereby providing a reliable foundation for subsequent comparative and ablation experiments.

To statistically validate the differences among multiple algorithms, the Friedman test is employed as a nonpara-

Table 1. Comparisons of accuracies and standard deviations on the DEAP dataset.

Prediction target	Attention-LSTM	RACNN	ATDD-LSTM	FLDNet	FLTSDP	CGRU-MDGN	MTLFuseNet	MTKDDP
Valence	76.77% \pm 10.71	80.55% \pm 12.50	74.73% \pm 13.07	83.85% \pm 11.34	92.40% \pm 5.20	69.92% \pm 9.42	71.33% \pm 5.24	94.24% \pm 4.50
Arousal	70.71% \pm 8.88	74.64% \pm 8.72	67.44% \pm 8.03	78.22% \pm 10.14	82.51% \pm 7.28	70.08% \pm 10.28	73.28% \pm 7.74	92.58% \pm 4.98
Dominance	72.06% \pm 9.79	74.64% \pm 11.01	68.41% \pm 9.98	77.52% \pm 10.30	89.95% \pm 5.63	71.01% \pm 13.43	-	93.27% \pm 4.00
Liking	76.48% \pm 9.06	79.30% \pm 11.80	75.47% \pm 9.38	82.42% \pm 9.56	91.20% \pm 4.40	-	-	94.64% \pm 3.51
Valence - dp	65.31% \pm 11.98	80.88% \pm 12.19	56.62% \pm 31.69	-	91.45% \pm 6.04	-	-	94.89% \pm 4.46
Arousal - dp	62.67% \pm 7.76	67.69% \pm 10.23	58.94% \pm 19.06	-	85.03% \pm 7.90	-	-	91.92% \pm 4.27
Dominance - dp	66.07% \pm 6.93	69.07% \pm 11.99	62.22% \pm 19.61	-	87.49% \pm 4.50	-	-	91.90% \pm 4.65
Liking - dp	66.61% \pm 10.43	77.59% \pm 11.28	66.61% \pm 10.43	-	89.14% \pm 6.60	-	-	93.87% \pm 5.07
Trainingtime (s/epoch)	18.21	85.06	123.21	33.67	36.01	-	-	38.10

Note: dp indicates that only a few subjects in the training dataset are used to train with data privacy. Attention-LSTM, Attention-based Long Short-Term Memory; RACNN, Recurrent Attention Convolutional Neural Network; ATDD-LSTM, Attention-based Temporal Dense Dual-Long Short-Term Memory; FLDNet, Frame-Level Distilling Neural Network; FLTSDP, Frame-Level Teacher-Student Learning With Data Privacy; CGRU-MDGN, Cascaded Gated Recurrent Unit - Multi-Channel Dynamic Graph Network; MTLFuseNet, Multi-Task Learning Fusion Network; MTKDDP, Multi-Teacher Knowledge Distillation Framework with data privacy.

metric approach suitable for comparing the rankings of several models across multiple datasets or experimental conditions [59]. This test first ranks the results of different algorithms across multiple datasets or experimental conditions and then computes the Friedman statistic τ_F based on the average ranks. The statistic is subsequently compared with the critical value from the chi-square distribution at a given confidence level (e.g., $\alpha = 0.05$) to assess whether the differences among algorithms are statistically significant. The key parameters include the number of algorithms k , the number of datasets or experimental cases N , and the corresponding critical value. When the null hypothesis—that all algorithms perform equally—is rejected, post hoc procedures such as the Nemenyi test are applied to conduct pairwise comparisons and identify which algorithms differ significantly.

4.2 Experiments on DEAP Dataset

First, to effectively demonstrate the performance of the framework, the larger DEAP dataset is employed for comparative experiments without data privacy in the subject-independent mode. Experiments utilize 5-fold cross-validation to evaluate the effectiveness and robustness of all methods. Consequently, 32 subjects are randomly divided into five groups for training and testing. The results, presented in Table 1, indicate that the framework outperforms all comparison methods across the four

emotional dimensions, achieving an improvement exceeding 2%. These results further demonstrate the effectiveness of the framework in automatic feature extraction and cross-domain knowledge transfer. Compared with the FLTSDP [47] and FLDnet [48] frameworks, which also perform frame-level feature filtering, the proposed method demonstrates substantially enhanced performance. This improvement indicates that the STG module accurately captures distinct reaction times across electrode positions, thereby more effectively avoiding interference from unnecessary features. Furthermore, MTKDDP outperforms these cross-domain integrated knowledge methods in terms of overall generalization and knowledge transfer. The CGRU-MDGN proposed in Guo and Wang [49] effectively processes both the temporal dynamics and spatial features of EEG signals, including their non-Euclidean relationships, thereby enhancing the accuracy of emotion recognition. However, this model has limitations in cross-domain knowledge transfer, while MTKDDP adopts multi-teacher KD strategy to integrate knowledge of different subnets more effectively and enhance the effect of cross-domain knowledge transfer. In addition, MTLFuseNet is an emotion recognition model based on EEG deep latent feature fusion and multi-task learning [57]. MTKDDP addresses challenges such as multi-task feature interference, significant knowledge forgetting, and low computational efficiency. It employs multi-teacher hierarchical distillation to decou-

ple task-specific features, dynamic knowledge solidification to mitigate forgetting, and combines parameter sharing with asynchronous updates to improve efficiency, thereby substantially enhancing accuracy and robustness. Furthermore, whereas conventional multi-teacher approaches often incur prohibitive computational overhead, MTKDDP achieves comparable time complexity to SOTA single-teacher frameworks through optimized parameter sharing and asynchronous gradient updates. This dual advantage of increased robustness against forgetting and preserved operational efficiency highlights the distinctive applicability of the framework in real-world scenarios.

4.3 Learning Mechanism With Data Privacy

Second, to evaluate the knowledge transfer capability of the framework, comparative experiments with private data are conducted on the DEAP dataset. In these experiments, the training set under 5-fold cross-validation is divided into three groups, each containing non-overlapping EEG data from 9, 8, and 8 subjects, respectively. The results, presented in Table 1, demonstrate the optimal performance achieved by the framework. Compared with classical methods, the framework exhibits a clear performance advantage, particularly in the stability and accuracy of predictions, which also reflects the effectiveness of the privacy protection mechanisms. In comparison with the FLTSDP framework [47], which similarly incorporates privacy protection mechanisms, the superior performance of the framework highlights the contribution of the multi-teacher KD strategy in facilitating cross-domain knowledge transfer.

4.4 Ablation Study

Third, to assess the effectiveness of the proposed modules, two types of ablation studies are conducted on the DEAP dataset. In the first type of ablation experiment, the STG module, teacher-student (TS) framework, KF module, and adaptive multi-teacher knowledge distillation loss (AMTKDL) are sequentially integrated into the base model, which corresponds to the Attention-based Long Short-Term Memory (attLSTM) model from Wang's framework [48]. In the second type of ablation experiment, individual removal of the STG or KF modules is tested, further confirming the essential contribution of each module to feature selection and knowledge integration. The results of these ablation experiments are presented in Table 2. As shown, the proposed modules successfully enhance the performance of EEG emotion recognition, particularly the STG module, indicating its key role in eliminating unnecessary features and effectively improving overall performance. Furthermore, compared with a single network, multi-classifiers in the TS framework effectively integrate knowledge from previous subnets, enhancing network robustness. The KF module optimizes knowledge composition to further improve feature extraction, as reflected in the experimental results. Fi-

nally, the AMTKDL refines the robustness of knowledge transfer and further enhances the precision of the framework.

To quantitatively assess the findings from the ablation experiments, the Friedman test is conducted, followed by the Nemenyi post-hoc test. The Friedman statistic τ_F was 21.00, exceeding the critical value of 2.488 at a significance level of $\alpha = 0.05$, indicating that the performance differences among the ablation configurations are statistically significant. Subsequently, the Nemenyi post-hoc test was applied, with the critical distance (CD) calculated as 5.2498 based on $q_{0.05}$, allowing for a precise distinction of pairwise differences between the algorithms. As illustrated in the Friedman test (Fig. 4), the complete model (A5+DP) achieves the highest performance, whereas removal of either the STG or KF module leads to notable performance degradation. These results underscore the critical role of the STG and knowledge fusion components in ensuring the robustness and effectiveness of the proposed framework.

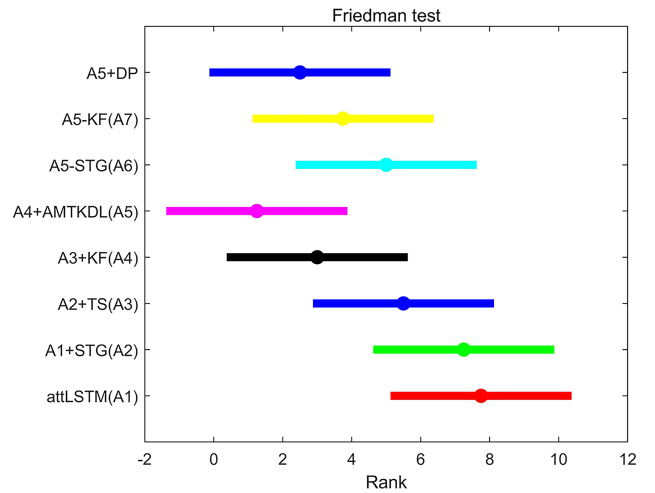


Fig. 4. Friedman test of ablation study. DP, data privacy; KF, knowledge filter; STG, spatiotemporal gate; AMTKDL, adaptive multi-teacher knowledge distillation loss; TS, teacher-student; attLSTM, Attention-based Long Short-Term Memory.

4.5 Ablation Study in Knowledge Transfer and Visualization of KF

Then, to evaluate the capability of knowledge transfer among multiple subnets, ablation experiments on knowledge transfer methods and feature screening visualization experiments are conducted. First, several knowledge transfer methods—including the TS framework with Student-1 and Student-2, integrated networks, KF, and AMTKDL—are presented in Table 3. Here, the base model corresponds to the attLSTM model from Wang's framework [48] with the STG module. From the results in Table 3, the TS network clearly enhances EEG emotion recognition; however, deeper student networks do not substantially im-

Table 2. Ablation study results on the DEAP dataset.

Prediction target	Valence	Arousal	Dominance	Liking
attLSTM (A1)	87.68% \pm 7.94	80.83% \pm 6.45	80.70% \pm 8.29	85.52% \pm 7.26
A1+STG (A2)	87.21% \pm 8.40	87.26% \pm 14.60	87.98% \pm 6.31	87.53% \pm 6.12
A2+TS (A3)	91.05% \pm 8.40	89.26% \pm 5.12	90.27% \pm 4.35	89.70% \pm 3.62
A3+KF (A4)	93.45% \pm 4.85	90.47% \pm 4.58	92.64% \pm 4.24	93.93% \pm 3.96
A4+AMTKDL (A5)	94.24% \pm 4.51	92.58% \pm 4.98	93.27% \pm 4.00	94.64% \pm 3.51
A5-STG (A6)	93.82% \pm 6.35	87.35% \pm 6.83	88.67% \pm 8.44	92.88% \pm 4.39
A5-KF (A7)	93.20% \pm 7.45	91.64% \pm 7.43	92.06% \pm 5.88	93.21% \pm 7.33
A5+DP	94.89% \pm 4.46	91.92% \pm 4.27	91.90% \pm 4.65	93.87% \pm 5.07

Note: attLSTM is the base model.

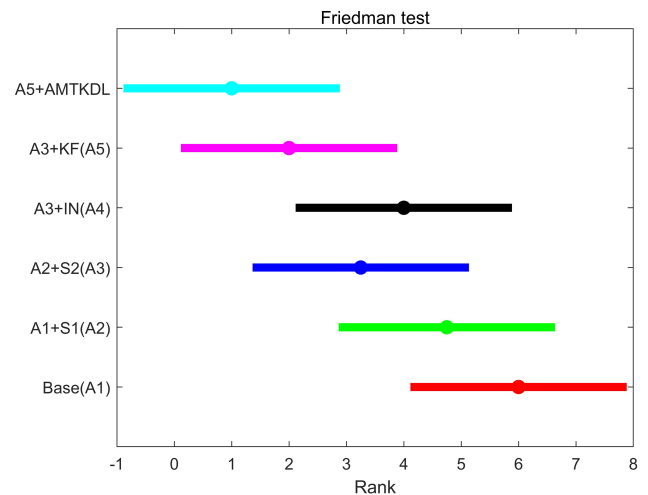
Table 3. Ablation study results of knowledge transfer on the DEAP dataset.

Prediction target	Valence	Arousal	Dominance	Liking
Base (A1)	87.21% \pm 8.40	87.26% \pm 5.17	87.98% \pm 6.31	87.53% \pm 6.12
A1+S1 (A2)	90.09% \pm 6.26	87.89% \pm 5.59	89.18% \pm 6.03	89.04% \pm 6.10
A2+S2 (A3)	91.05% \pm 8.40	89.26% \pm 5.12	90.27% \pm 4.35	89.70% \pm 3.62
A3+IN (A4)	91.08% \pm 5.33	88.01% \pm 5.65	88.98% \pm 5.27	89.45% \pm 6.75
A3+KF (A5)	93.45% \pm 4.85	90.47% \pm 4.58	92.64% \pm 4.24	93.93% \pm 3.96
A5+AMTKDL	94.24% \pm 4.51	92.58% \pm 4.98	93.27% \pm 4.00	94.64% \pm 3.51

Note: attLSTM with STG module is the base model. S1, S2, and IN are defined as student-1, student-2, and integrated network, respectively. IN, integrated networks.

prove recognition accuracy. Similarly, the integrated network from FLTSDP [47] does not further improve performance and even underperforms compared with the deeper student networks. The KF mechanism effectively reorganizes knowledge, capturing information more aligned with the current domain, better integrating knowledge from previous subnets, and thereby facilitating EEG emotion recognition. To further assess the contributions of different components in the knowledge transfer process, the Friedman test accompanied by the Nemenyi post-hoc analysis was conducted on the results reported in Table 3. The computed Friedman statistic τ_F reached 57.0, which is above the critical threshold of 2.488 at a significance level of $\alpha = 0.05$, confirming that the performance differences across the ablation settings are statistically meaningful. The subsequent Nemenyi test was performed to identify specific pairwise differences, with the critical distance (CD) calculated as 3.7702 based on $q_{0.05}$. As depicted in the Friedman test (Fig. 5), the configuration including all knowledge transfer components (A5+AMTKDL) achieves the top ranking, whereas omitting student networks or the KF module leads to noticeable reductions in performance. These results highlight the importance of sequential knowledge integration and feature refinement for enhancing the effectiveness and stability of the proposed framework.

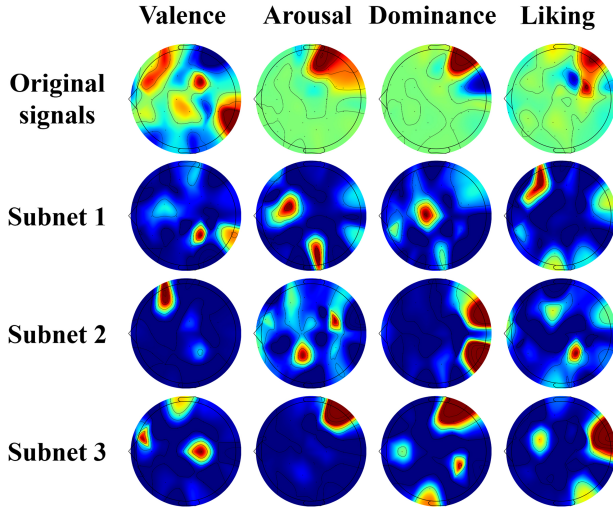
Additionally, the visualization results of the KF mechanism are presented in Fig. 6, illustrating feature extraction outcomes for the four emotional dimensions derived from EEG signals across 32 channels. The “Original signals” row visualizes raw EEG activity, with widespread red regions indicating noisy and unfocused signals with no

**Fig. 5. Friedman test of ablation study in knowledge transfer.**

clear alignment to emotion-specific responses. Rows corresponding to Subnet 1–3 illustrate features extracted by successive subnets. Subnet 1 shows scattered and inaccurate activation, capturing noisy or irrelevant activity. Subnet 2 displays more targeted activations, partially aligning with known emotion-related regions, but still inconsistent. By Subnet 3, activations are highly precise, with red regions sharply highlighting channels critical for each emotional dimension, as the KF module fuses outputs from previous subnets, filtering out noise and domain-irrelevant features while retaining emotion-relevant ones. This visualization demonstrates that early subnets do not fully capture target-domain emotion features; however, as sub-

Table 4. Accuracy comparison under varying privacy budgets.

δ	0.01	0.001	0.0001	0.00001
Valence	93.49% \pm 6.57	92.76% \pm 6.77	92.50% \pm 6.70	91.98% \pm 6.57
Arousal	89.29% \pm 8.62	89.00% \pm 8.39	88.47% \pm 8.69	88.34% \pm 8.79
Dominance	89.76% \pm 8.20	89.74% \pm 7.84	89.39% \pm 7.91	89.45% \pm 8.28
Liking	92.16% \pm 7.32	91.50% \pm 7.44	90.64% \pm 6.86	90.06% \pm 7.64

**Fig. 6. Visualization results of KF.**

nets progress under KF guidance, features become increasingly domain-specific and accurately aligned with known neural correlates of emotion. Overall, Fig. 6 illustrates how KF and adaptive subnets collaboratively refine raw, noisy signals into precise, domain-relevant, and neurophysiologically plausible features, validating the effectiveness of adaptive knowledge transfer and the framework's capability to enhance EEG emotion recognition.

4.6 Visualization of Temporal-Based Gates

To evaluate the adaptive feature extraction capability of the STG module, temporal gate weights from multiple EEG channels are visualized, as shown in Fig. 7. The vertical axis represents electrode channels from AF4 to O2, covering frontal, central, parietal, and occipital regions, while the horizontal axis denotes 20 discrete time steps. Color intensity corresponds to gating weights, with red regions indicating higher weights. The visualization reveals that high-weight time intervals vary across channels. Frontal channels exhibit pronounced activation during early time steps, whereas central and parietal channels show sustained or delayed activation in later periods. This distribution indicates that frame-level features are both channel-specific and time-specific, with different electrodes carrying emotion-relevant information within distinct temporal windows.

The STG mechanism applies gating across both temporal and spatial dimensions. For each channel, a temporal gate assigns weights to frame-level features across time

steps, followed by spatial gating that integrates contributions from all channels. This process prioritizes informative intervals and suppresses irrelevant or noisy segments, generating spatio-temporal representations optimized for emotion recognition. The strategy enhances signal quality by emphasizing high-weight intervals and reducing noise. Activation patterns in frontal and central-parietal regions align with established stages of emotional processing, improving physiological interpretability. By focusing on emotion-relevant channel-time features, the mechanism strengthens discriminative capability. Overall, the STG module functions as a spatio-temporal filter that automatically identifies critical EEG features for emotion recognition, ensuring effective and interpretable feature extraction.

4.7 Privacy-Preserving Performance

To evaluate the privacy-preserving capability of the proposed framework, differential privacy [60] is incorporated using the Differentially Private Stochastic Gradient Descent (DP-SGD) mechanism, which limits the sensitivity of individual samples through gradient clipping (threshold $C = 1.0$) and Gaussian noise injection. During training, the privacy budget is set to ϵ , and four different δ values (10^{-2} , 10^{-3} , 10^{-4} , 10^{-5}) are tested to investigate the trade-off between privacy and accuracy, with smaller δ values providing stronger privacy at the cost of increased noise.

As shown in Table 4, accuracy decreases slightly as δ decreases, for example from approximately 91.4% to 89.96%, with performance reduction within 2% and standard deviations remaining relatively stable. These results indicate that the framework maintains robust performance while providing effective privacy protection, confirming the successful integration of DP-SGD without substantial loss of model utility.

4.8 Experiments on DREAMER Dataset

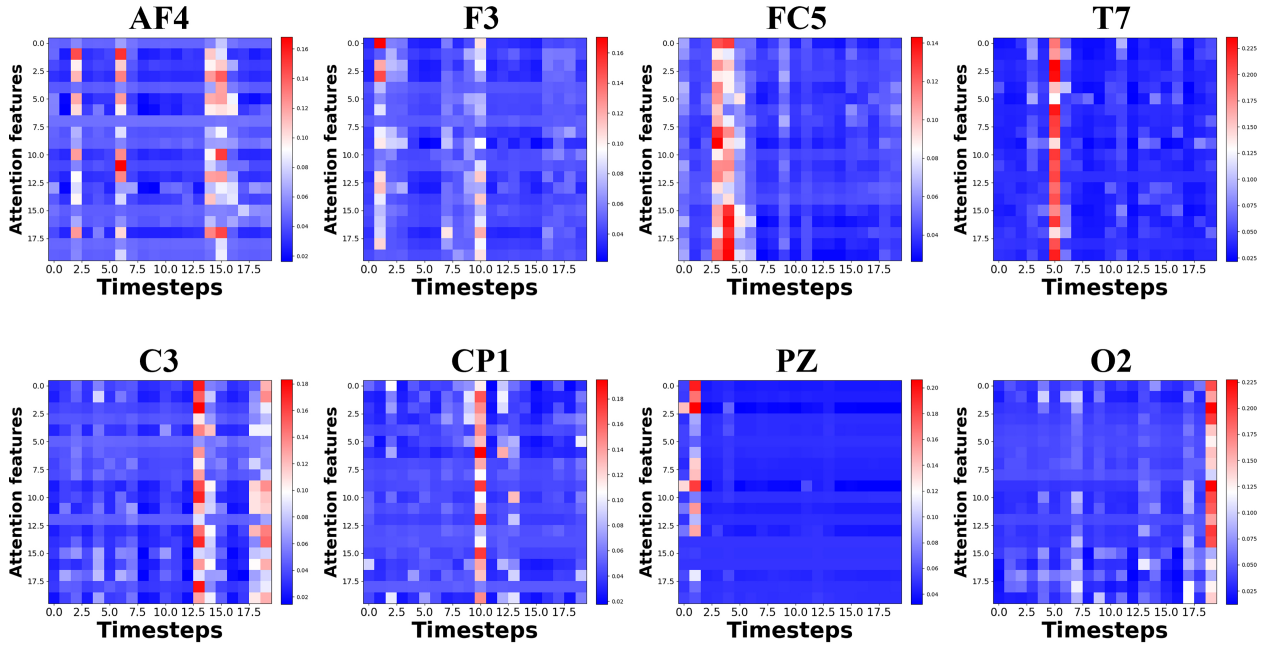
Experiments on the DREAMER dataset (Table 5) demonstrate that the MTKDDP framework consistently outperforms SOTA baselines across all three emotional dimensions. The framework achieves marked improvements over single-model approaches, illustrating the effectiveness of leveraging complementary knowledge from multiple teachers to mitigate dataset bias and enhance robustness. Reduced variability further reflects superior stability across subjects and sessions.

This strong performance stems from the synergistic integration of the STG module's noise-invariant feature ex-

Table 5. Comparisons of accuracies on the DREAMER dataset.

Prediction target	Attention-LSTM	CLSTM	ATDD-LSTM	FLDNet	FLTSDP	RACNN	MTLFuseNet	MTKDDP
Valence	83.19% \pm 14.97	83.72% \pm 14.60	86.00% \pm 13.00	89.91% \pm 12.51	91.54% \pm 8.14	83.69% \pm 14.41	80.03% \pm 8.01	96.61% \pm 3.73
Arousal	82.17% \pm 11.45	81.41% \pm 12.92	82.09% \pm 11.03	87.67% \pm 10.02	90.61% \pm 8.13	83.04% \pm 12.05	83.33% \pm 11.24	96.39% \pm 3.32
Dominance	86.67% \pm 11.29	85.32% \pm 9.58	86.37% \pm 9.69	90.28% \pm 6.06	91.00% \pm 8.02	83.93% \pm 12.55	-	97.11% \pm 2.92
Valence - dp	81.74% \pm 15.46	80.66% \pm 18.04	85.37% \pm 10.61	-	93.17% \pm 7.72	83.87% \pm 14.89	-	96.56% \pm 2.96
Arousal - dp	81.43% \pm 10.7	81.13% \pm 12.47	82.95% \pm 8.27	-	91.43% \pm 7.69	80.82% \pm 12.85	-	96.78% \pm 3.26
Dominance - dp	84.15% \pm 9.97	83.17% \pm 11.96	83.54% \pm 12.57	-	92.13% \pm 6.71	81.59% \pm 13.99	-	97.39% \pm 2.59

Note: dp indicates that only a few subjects in the training dataset are used to train with data privacy. CLSTM, Convolutional LongShort Term Memory.

**Fig. 7. Visualization results of temporal-based gates.**

traction and the MTKD framework's ability to transfer robust, generalized representations. These results highlight the framework's potential for practical EEG-based emotion recognition applications under substantial inter-subject variability and diverse recording conditions.

4.9 Friedman Test and Nemenyi Test

To demonstrate the superiority, effectiveness, and robustness of the proposed MTKDDP framework, the Friedman test and Nemenyi post-hoc test were conducted. The basic assumption of the Friedman test is that all algorithms perform equally well. Comparisons are conducted across four publicly available EEG datasets, involving the MTKDDP framework and seven other represen-

tative algorithms—including Attention-LSTM, RACNN, CLSTM, ATDD-LSTM, MTLFuseNet, and three baseline methods—comprising a total of eight algorithms. The Friedman statistic τ_F is computed as 4.5224, exceeding the critical value of 2.488 at a significance level of $\alpha = 0.05$ indicating significant differences among the algorithms. Subsequently, the Nemenyi post-hoc test is conducted, and the Friedman ranking diagram is shown in Fig. 8. Specifically, Fig. 8 is constructed based on the results of the DEAP and DREAMER datasets, where the statistical analyses are conducted on the Valence and Arousal dimensions, and the outcomes from these two dimensions are combined to generate the comprehensive ranking diagram. In this figure, the critical distance is calculated as 5.2498, and the cen-

tral point represents the average ranking. The ranking diagram indicates that the MTKDDP framework consistently outperformed all other methods with statistically significant improvements. These results demonstrate that integrating multiple subnets and extracting robust spatio-temporal features is crucial for enhancing EEG-based emotion recognition performance.

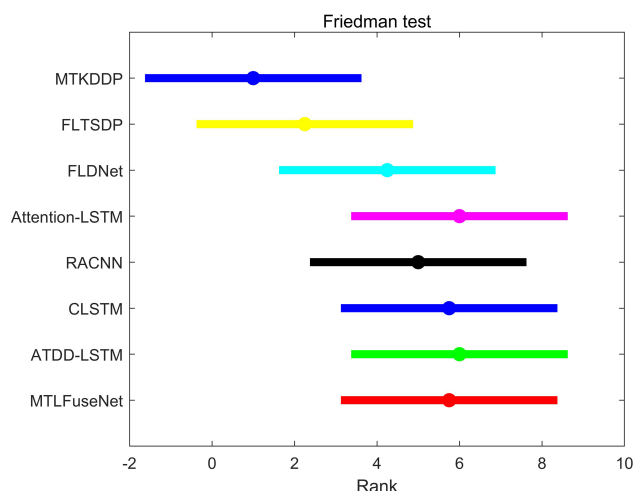


Fig. 8. Friedman on DEAP and DREAMER datasets.

5. Conclusions

For EEG emotion recognition in subject-independent mode, the paper proposes a multi-teacher KD framework with data privacy. The framework sequentially trains multiple subnets with different EEG signals and employs a multi-teacher KD strategy that integrates KF and adaptive KDloss to enhance knowledge transfer. Moreover, the STG module alleviates response time variations across electrodes, thereby improving feature extraction and eliminating irrelevant information. Experimental results demonstrate that the MTKDDP framework consistently outperforms SOTA methods, achieving higher accuracy on the DEAP and DREAMER datasets. While the current work remains at the experimental stage and has not yet been clinically deployed, these findings indicate strong potential for future applications in real-world clinical scenarios.

However, MTKDDP still requires further verification on diverse cross-domain EEG datasets. As more EEG signals become available, the framework can add subnets to improve performance and generalization, though with higher computational cost. To address this, future work applies a pruning strategy, stopping training when feature similarity among three consecutive subnets exceeds a threshold. Despite more subnets, each maintains a small number of parameters, keeping the overall model size manageable for practical deployment.

Availability of Data and Materials

The datasets used and analyzed in the present study are available upon request from the corresponding author.

Author Contributions

JQY, THG, and CL designed the research study. JQY and THG performed the research. CL and JZX provided help and advice on the experiments, JZX analyzed the data, and JQY drafted the manuscript. THG and CL guided the experimental design and revised the article. All authors contributed to critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This work is supported by National Natural Science Foundation of China under Grant No.62403264, China Postdoctoral Science Foundation under Grant No.2024M761556, Qing-dao Natural Science Foundation under Grant No.24-4-4-zrjj-94-jch, Postdoctoral Innovation Project of Shandong Province under Grant No.SDCX-ZG 202400312, Qingdao Postdoctoral Applied Foundation under Grant No.QDBSH20240102029, Natural Science Foundation of Shandong Province under Grant No.ZR2025ZD15 and Systems Science Plus Joint Research Program of Qingdao University under Grant No.XT2024202.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Ni G, Xu Z, Bai Y, Zheng Q, Zhao R, Wu Y, *et al.* EEG-based assessment of temporal fine structure and envelope effect in mandarin syllable and tone perception. *Cerebral Cortex*. 2023; 33: 11287–11299. <https://doi.org/10.1093/cercor/bhad366>.
- [2] Zhang R, Lee S, Hwang M, Hiranaka A, Wang C, Ai W, *et al.* Noir: Neural signal operated intelligent robots for everyday activities. *arXiv*. 2023. <https://doi.org/10.48550/arXiv.2311.01454>. (preprint)
- [3] Mao X, Shan W, Yu J. Automatic Diagnosis and Subtyping of Ischemic Stroke Based on a Multi-Dimensional Deep Learning System. *IEEE Transactions on Instrumentation and Measurement*. 2024; 73: 1–11. <https://doi.org/10.1109/TIM.2024.3458059>.
- [4] Dou Y, Xia J, Fu M, Cai Y, Meng X, Zhan Y. Identification of epileptic networks with graph convolutional network incorporating oscillatory activities and evoked synaptic responses.

- NeuroImage. 2023; 284: 120439. <https://doi.org/10.1016/j.neuroimage.2023.120439>.
- [5] Wang Y, Wang F, Kong Y, Gao T, Zhu Q, Han L, *et al.* High definition transcranial direct current stimulation of the Cz improves social dysfunction in children with autism spectrum disorder: A randomized, sham, controlled study. *Autism Research*. 2023; 16: 2035–2048. <https://doi.org/10.1002/aur.3018>.
 - [6] Utsumi K, Takano K, Okahara Y, Komori T, Onodera O, Kansaku K. Operation of a P300-based brain-computer interface in patients with Duchenne muscular dystrophy. *Scientific Reports*. 2018; 8: 1753. <https://doi.org/10.1038/s41598-018-20125-6>.
 - [7] Orban M, Elsamanty M, Guo K, Zhang S, Yang H. A Review of Brain Activity and EEG-Based Brain-Computer Interfaces for Rehabilitation Application. *Bioengineering*. 2022; 9: 768. <https://doi.org/10.3390/bioengineering9120768>.
 - [8] Xu R, Zhang H, Shi X, Liang J, Wan C, Ming D. Lower-Limb Motor Assessment With Corticomuscular Coherence of Multiple Muscles During Ankle Dorsiflexion After Stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2023; 31: 160–168. <https://doi.org/10.1109/TNSRE.2022.3217571>.
 - [9] Ortiz-Catalan M, Zbinden J, Millenaar J, D'Accolti D, Controzzi M, Clemente F, *et al.* A highly integrated bionic hand with neural control and feedback for use in daily life. *Science Robotics*. 2023; 8: eadf7360. <https://doi.org/10.1126/scirobotics.adf7360>.
 - [10] Abenna S, Nahid M, Bouyghf H, Ouacha B. EEG-based BCI: A novel improvement for EEG signals classification based on real-time preprocessing. *Computers in Biology and Medicine*. 2022; 148: 105931. <https://doi.org/10.1016/j.compbiomed.2022.105931>.
 - [11] Li X, Zhang Y, Tiwari P, Song D, Hu B, Yang M, Zhao Z, *et al.* EEG based emotion recognition: A tutorial and review. *ACM Computing Surveys*. 2022; 55: 1–57. <https://doi.org/10.1145/3524499>.
 - [12] Li S, Li Y, Li R, Li Y, Wu J, Yang W, *et al.* The Modulatory Effect of Exogenous Orienting on Audiovisual Emotional Integration: An ERP Study. *Journal of integrative neuroscience*. 2024; 23: 210. <https://doi.org/10.31083/j.jin2311210>.
 - [13] Zheng WL, Lu BL. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*. 2015; 7: 162–175. <https://doi.org/10.1109/TAMD.2015.2431497>.
 - [14] Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, *et al.* DEAP: A database for emotion analysis; Using physiological signals. *IEEE Transactions on Affective Computing*. 2011; 3: 18–31. <https://doi.org/10.1109/T-AFCC.2011.15>.
 - [15] Katsigiannis S, Ramzan N. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics*. 2018; 22: 98–107. <https://doi.org/10.1109/JBHI.2017.2688239>.
 - [16] Chen X, Li C, Liu A, McKeown MJ, Qian R, Wang ZJ. Toward open-world electroencephalogram decoding via deep learning: A comprehensive survey. *IEEE Signal Processing Magazine*. 2022; 39: 117–134. <https://doi.org/10.1109/MSP.2021.3134629>.
 - [17] Zhao H, Ye N, Wang R. A survey on automatic emotion recognition using audio big data and deep learning architectures. In *2018 IEEE 4th International Conference on Big Data Security on Cloud* (pp. 139–142). Omaha, NE, USA. 2018.
 - [18] Alarcao SM, Fonseca MJ. Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing*. 2017; 10: 374–393. <https://doi.org/10.1109/TAFFC.2017.2714671>.
 - [19] Kawakami K. Supervised sequence labelling with recurrent neural networks [Ph. D. thesis]. Technische Universität München. 2008.
 - [20] Tripathi S, Acharya S, Sharma R, Mittal S, Bhattacharya S. Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2017; 31: 4746–4752. <https://doi.org/10.1609/aaai.v31i2.19105>.
 - [21] Deng Y, Yu H, Peng F, Yan F, Wu Y, Yan L. EEG signal classification based on neural network with depthwise convolution. *Journal of Physics: Conference Series*. 2022; 2219: 012056. <https://doi.org/10.1088/1742-6596/2219/1/012056>.
 - [22] Gao Z, Wang X, Yang Y, Li Y, Ma K, Chen G. A channel-fused dense convolutional network for EEG-based emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*. 2020; 13: 945–954. <https://doi.org/10.1109/TCDS.2020.2976112>.
 - [23] Wan Z, Yang R, Huang M, Zeng N, Liu X. A review on transfer learning in EEG signal analysis. *Neurocomputing*. 2021; 421: 1–14. <https://doi.org/10.1016/j.neucom.2020.09.017>.
 - [24] Chen H, Sun S, Li J, Yu R, Li N, Li X, *et al.* Personal-zscore: Eliminating individual difference for EEG-based cross-subject emotion recognition. *IEEE Transactions on Affective Computing*. 2023; 14: 2077–2088. <https://doi.org/10.1109/TAFFC.2021.3137857>.
 - [25] Li J, Qiu S, Shen YY, Liu CL, He H. Multisource Transfer Learning for Cross-Subject EEG Emotion Recognition. *IEEE Transactions on Cybernetics*. 2020; 50: 3281–3293. <https://doi.org/10.1109/TCYB.2019.2904052>.
 - [26] Li W, Huan W, Hou B, Tian Y, Zhang Z, Song A. Can emotion be transferred?—A review on transfer learning for EEG-based emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*. 2022; 14: 833–846. <https://doi.org/10.1109/TCDS.2021.3098842>.
 - [27] Duan RN, Zhu JY, Lu BL. Differential entropy feature for EEG-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering* (pp. 81–84). Diego, CA, USA. IEEE. 2013.
 - [28] Huang D, Guan C, Ang KK, Zhang H, Pan Y. Asymmetric spatial pattern for EEG-based emotion detection. In *The 2012 International Joint Conference on Neural Networks* (pp. 1–7). Brisbane, QLD, Australia. IEEE. 2012.
 - [29] Petrantonakis PC, Hadjileontiadis LJ. Emotion recognition from EEG using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine*. 2010; 14: 186–197. <https://doi.org/10.1109/TITB.2009.2034649>.
 - [30] Cui H, Liu A, Zhang X, Chen X, Wang K, Chen X. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowledge-Based Systems*. 2020; 205: 106243. <https://doi.org/10.1016/j.knosys.2020.106243>.
 - [31] Ye F, Bors AG. Dynamic Self-Supervised Teacher-Student Network Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2023; 45: 5731–5748. <https://doi.org/10.1109/TPAMI.2022.3220928>.
 - [32] Griggs WS, Norman SL, Deffieux T, Segura F, Osmani BF, Chau G, *et al.* Decoding motor plans using a closed-loop ultrasonic brain-machine interface. *Nature Neuroscience*. 2024; 27: 196–207. <https://doi.org/10.1038/s41593-023-01500-7>.
 - [33] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv*. 2015. <https://doi.org/10.48550/arXiv.1503.02531>. (preprint)
 - [34] Yuan F, Shou L, Pei J, Lin W, Gong M, Fu Y, *et al.* Reinforced multi-teacher selection for knowledge distillation. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021; 35: 14284–14291. <https://doi.org/10.1609/aaai.v35i16.17680>.

- [35] Ma C, Zhang Y, Tu M, Zhao Y, Zhou Y, Zong C. Knowledge Multi-teacher knowledge distillation for end-to-end text image machine translation. In International Conference on Document Analysis and Recognition (pp. 484–501). Springer Nature: Cham Switzerland. 2023.
- [36] Liu Y, Zhang W, Wang J. Adaptive Multi-Teacher Multi-level Knowledge Distillation. *Neurocomputing*. 2020; 415: 106–113. <https://doi.org/10.1016/j.neucom.2020.07.048>.
- [37] Son W, Na J, Choi J, Hwang W. Densely Guided Knowledge Distillation using Multiple Teacher Assistants. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9395–9404). 2021.
- [38] Gui S, Wang Z, Chen J, Zhou X, Zhang C, Cao Y. MT4MTL-KD: A Multi-Teacher Knowledge Distillation Framework for Triplet Recognition. *IEEE Transactions on Medical Imaging*. 2024; 43: 1628–1639. <https://doi.org/10.1109/TMI.2023.3345736>.
- [39] Ye X, Jiang R, Tian X, Zhang R, Chen Y. Knowledge Distillation via Multi-Teacher Feature Ensemble. *IEEE Signal Processing Letters*. 2024; 31: 566–570. <https://doi.org/10.1109/LSP.2024.3359573>.
- [40] Shang R, Li W, Zhu S, Jiao L, Li Y. Multi-teacher knowledge distillation based on joint Guidance of Probe and Adaptive Corrector. *Neural Networks*. 2023; 164: 345–356. <https://doi.org/10.1016/j.neunet.2023.04.015>.
- [41] Zhang H, Chen D, Wang C. Confidence-aware multi-teacher knowledge distillation. In ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4498–4502). Singapore. IEEE. 2022.
- [42] Liu YJ, Yu M, Zhao G, Song J, Ge Y, Shi Y. Real-time movie-induced discrete emotion recognition from EEG signals. *IEEE Transactions on Affective Computing*. 2017; 9: 550–562. <https://doi.org/10.1109/TAFFC.2017.2660485>.
- [43] Pihl L, Tjahjedi T. A mutual information based adaptive windowing of informative EEG for emotion recognition. *IEEE Transactions on Affective Computing*. 2018; 11: 722–735. <https://doi.org/10.1109/TAFFC.2018.2840973>.
- [44] Richhariya B, Sharma A, Tanveer M. Improved universum twin support vector machine. In 2018 IEEE Symposium Series on Computational Intelligence (pp. 2045–2052). Bangalore, India. IEEE. 2018.
- [45] Gu Y, Zhong X, Qu C, Liu C, Chen B. A Domain Generative Graph Network for EEG-Based Emotion Recognition. *IEEE Journal of Biomedical and Health Informatics*. 2023; 27: 2377–2386. <https://doi.org/10.1109/JBHI.2023.3242090>.
- [46] Song M, Wang Z, Zhang Z, Song Y, Wang Q, Ren J, *et al.* Analyzing user-level privacy attack against federated learning. *IEEE Journal on Selected Areas in Communications*. 2020; 38: 2430–2444. <https://doi.org/10.1109/JSAC.2020.3000372>.
- [47] Gu T, Wang Z, Xu X, Li D, Yang H, Du W. Frame-Level Teacher-Student Learning With Data Privacy for EEG Emotion Recognition. *IEEE Transactions on Neural Networks and Learning Systems*. 2023; 34: 11021–11028. <https://doi.org/10.1109/TNNLS.2022.3168935>.
- [48] Wang Z, Gu T, Zhu Y, Li D, Yang H, Du W. FLDNet: Frame-Level Distilling Neural Network for EEG Emotion Recognition. *IEEE Journal of Biomedical and Health Informatics*. 2021; 25: 2533–2544. <https://doi.org/10.1109/JBHI.2021.3049119>.
- [49] Guo W, Wang Y. Convolutional gated recurrent unit-driven multidimensional dynamic graph neural network for subject-independent emotion recognition. *Expert Systems with Applications*. 2024; 238: 121889. <https://doi.org/10.1016/j.eswa.2023.121889>.
- [50] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, *et al.* Attention is all you need. In *Neural Information Processing Systems* (pp. 5998–6008). 2017.
- [51] Duan Y, Zhou J, Wang Z, Wang YK, Lin CT. Dewave: Discrete EEG waves encoding for brain dynamics to text translation. *arXiv*. 2024. <https://doi.org/10.48550/arXiv.2309.14030>. (preprint)
- [52] You L, Liu S, Chang Y, Yuen C. A triple-step asynchronous federated learning mechanism for client activation, interaction optimization, and aggregation enhancement. *IEEE Internet of Things Journal*. 2022; 9: 24199–24211. <https://doi.org/10.1109/JIOT.2022.3188556>.
- [53] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004; 13: 600–612. <https://doi.org/10.1109/tip.2003.819861>.
- [54] Li X, Song D, Zhang P, Yu G, Hou Y, Hu B. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In 2016 IEEE International Conference on Bioinformatics and Biomedicine (pp. 352–359). Shenzhen, China. IEEE. 2016.
- [55] Du X, Ma C, Zhang G, Li J, Lai YK, Zhao G, *et al.* An efficient lstm network for emotion recognition from multichannel EEG signals. *IEEE Transactions on Affective Computing*. 2020; 13: 1528–1540. <https://doi.org/10.1109/TAFFC.2020.3013711>.
- [56] Yin Y, Zheng X, Hu B, Zhang Y, Cui X. EEG emotion recognition using fusion model of graph convolutional neural networks and lstm. *Applied Soft Computing*. 2021; 100: 106954. <https://doi.org/10.1016/j.asoc.2020.106954>.
- [57] Li R, Ren C, Ge Y, Zhao Q, Yang Y, Shi Y, *et al.* MTLFuseNet: a novel emotion recognition model based on deep latent feature fusion of EEG signals and multi-task learning. *Knowledge-Based Systems*. 2023; 276: 110756. <https://doi.org/10.1016/j.knsys.2023.110756>.
- [58] Zhang X, Cheng X, Liu H. TPRO-NET: an EEG-based emotion recognition method reflecting subtle changes in emotion. *Scientific Reports*. 2024; 14: 13491. <https://doi.org/10.1038/s41598-024-62990-4>.
- [59] Demšar J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*. 2006; 7: 1–30.
- [60] Boenisch F, Mühl C, Dziedzic A, Rinberg R, Papernot N. Have it your way: Individualized Privacy Assignment for DP-SGD. *Advances in Neural Information Processing Systems*. 2023; 36: 19073–19103.