# BrainTune: An Artifact-Aware Meta-Contrastive Framework for Fast and Lasting Personalization of EEG Emotion Models

**Wei Gu**[1], **Wenxin Chen**[2], **Bao-Liang Lu**[1] **& Wei-Long Zheng**[1*]
[1]School of Computer Science, Shanghai Jiao Tong University
[2]Global College, Shanghai Jiao Tong University
[*]Corresponding author (weilong@sjtu.edu.cn)

## Abstract

The practical application of EEG emotion recognition is hindered by two critical challenges: reliance on offline processing pipelines that assume access to complete session data, and high inter-subject variability that complicates personalization. To address these, we propose BrainTune, an Artifact-Aware Meta-Contrastive Framework. BrainTune features a causal stream-compatible pipeline with an artifact-aware graph neural network to bypass offline dependencies, as well as a meta-contrastive strategy using supervised multi-level contrastive learning for the backbone to handle session/subject/task domains and meta-learning for the classifier to enable rapid personalization from brief calibration. Experiments on SEED, DEAP, and DREAMER demonstrate that BrainTune achieves state-of-the-art performance and exhibits "lasting personalization", where early calibration progressively benefits future sessions. Code is available at https://github.com/iewug/BrainTune.

**Keywords:** Emotion Recognition; EEG; Personalization; Meta-Learning; Contrastive Learning

## Introduction

Electroencephalography (EEG) holds significant promise for emotion recognition (Wu et al., 2023), yet its transition from laboratory settings to real-world applications is impeded by two critical challenges: operational infeasibility due to offline dependencies and a severe personalization gap.

The first challenge, operational infeasibility, stems from a deep-rooted reliance on offline EEG processing methods. Many established pipelines depend on session-wide knowledge for tasks like time-consuming manual artifact removal (Jiang, Liu, et al., 2023; Koelstra et al., 2011), trial-wide signal smoothing (Zheng et al., 2017), or feature normalization requiring full-session statistics. The second challenge, the personalization gap, is driven by large domain shifts in EEG signals. As illustrated in Figure 1, the feature distribution shifts substantially not only between subjects but also across different sessions for the same individual. While domain adaptation (DA) seems a natural fit, it is often unsuitable as it requires access to unlabeled target data during training—an unrealistic assumption for new users. Our work therefore pursues a more practical paradigm: personalizing a pre-trained model with only a brief, one-time calibration.

To ensure operational feasibility, BrainTune abandons session-wide statistics in favor of a strictly causal stream-compatible pipeline, which processes data strictly sequentially without accessing future information. It handles noise

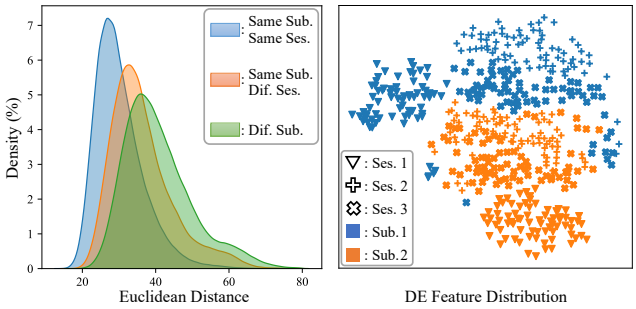


Figure 1: Left: Kernel density of Euclidean distances between differential entropy (DE) features for all subjects in SEED. Right: DE feature distribution for two subjects across three sessions. The domain shift is smallest within a session, larger across a subject's sessions, and largest between subjects.

via an Artifact-Aware Graph Neural Network (GNN) that dynamically aggregates clean neighbors to compensate for corrupted channels, eliminating the need for offline cleaning. For personalization, BrainTune combines Meta-Learning (Finn et al., 2017) with Supervised Multi-Level Contrastive Learning (Khosla et al., 2020). This structure guides the model to progressively handle intra-session, inter-session and inter-subject variants and enables the classifier to rapidly adapt to a new user with one brief calibration.

Our main contributions are summarized as follows:

- We propose BrainTune, a novel framework that uniquely combines an Artifact-Aware GNN with a Meta-Contrastive Learning mechanism to achieve rapid and lasting EEG model personalization.
- Our framework removes critical dependencies on offline EEG processing (e.g., manual artifact cleaning, trial-wide smoothing and session-wide normalization), enabling a stream-compatible pipeline for new user deployment.
- We demonstrate through extensive experiments on three benchmark datasets, including SEED (Zheng & Lu, 2015), DEAP (Koelstra et al., 2011), and DREAMER (Katsigiannis & Ramzan, 2017), that BrainTune significantly outperforms existing methods.
- We validate the concept of lasting personalization, showing on the multi-session SEED dataset that calibration data from earlier sessions progressively improves performance in future, unseen sessions.



In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* 

## Related Work

### Practicality of EEG Processing Pipelines

A significant but often overlooked barrier to deploying EEG models is the reliance on processing steps that assume access to complete, offline datasets. Most academic research operates on data that has undergone extensive preprocessing. For instance, feature normalization is commonly performed using statistics (mean and standard deviation) derived from an entire experimental session for a single subject. Similarly, popular feature smoothing techniques like Linear Dynamic Systems (LDS) require access to the entire data of a trial to function (Zheng et al., 2017). Manual or semi-automated artifact rejection (Jiang, Liu, et al., 2023; Koelstra et al., 2011) also depends on offline inspection. While early works explored real-time classification (Lan et al., 2016; Y. Liu et al., 2011), this challenge has been largely sidestepped in recent high-impact research.

### Personalization of EEG Emotion Models

Handling inter-subject variability is a central challenge in EEG emotion recognition, primarily addressed through Domain Generalization and Domain Adaptation.

**Domain Generalization (DG)** trains a single model on multiple source subjects to generalize to an unseen target without fine-tuning. Methods include learning domain-invariant features through disentangled representations (Ma et al., 2019), using a mixture of experts to capture diverse patterns (X.-H. Liu et al., 2024), or employing contrastive learning to pull same-class samples together across domains (Cai & Pan, 2023). However, the "one-size-fits-all" nature of DG often leads to suboptimal performance, as it struggles to capture the unique nuances of a new individual.

**Domain Adaptation (DA)**, in contrast, seeks to adapt a model trained on source subjects using data from the target subject. Adversarial training, typically using a Gradient Reversal Layer (GRL), is a common technique to align feature distributions between source and target domains (Ganin et al., 2016; Luo & Lu, 2021; Zhou et al., 2024). The fundamental limitation of DA for real-world applications is its impractical assumption that a large amount of unlabeled target data is available during training.

**Rapid Personalization with Calibration Data**. A more practical paradigm, which our work adopts, involves personalizing a pre-trained model for a new user through a brief calibration phase using a small amount of their own data. This approach strikes a balance between the performance of user-specific models and the convenience of a general model, making it highly suitable for real-world deployment. One notable effort in this direction is Plug-and-Play Domain Adaptation (PPDA) (L.-M. Zhao et al., 2021), which adapts a model using minimal unsupervised data from a new subject. However, its complex framework, involving five different kinds of loss functions, makes it difficult to train and scale. Crucially, neither traditional DA/DG nor PPDA explicitly address the offline processing pipeline issues.

## Methodology

### Overview

The framework of proposed BrainTune is depicted in Figure 2, which contains three stages. In Stage 1, the collected raw EEG data can be represented as $X' \in \mathbb{R}^{T' \times C}$, where $T'$ is the number of sample points in one sample and $C$ is the number of channels. Then the raw EEG data are sent into a Differential Entropy (DE) extractor and an artifact processor, which output $X \in \mathbb{R}^{T \times C \times F}$ for the DE features and boolean $M \in \mathbb{R}^{T \times C}$ for masks. Here $T$ is the number of seconds corresponding to the sampling points which is equal to $T'$ divided by sample frequency, and $F$ is the number of the DE frequency bands. Then the DE features $X$ and masks $M$ will be used to train the backbone of BrainTune by Supervised Multi-Level Contrastive Learning (Khosla et al., 2020). The backbone contains an Artifact-Aware GNN, Transformer I and Transformer II as a shallow, intermediate and deep layer, respectively. The outputs of the three layers are represented as $H_1, H_2, H_3 \in \mathbb{R}^{T \times D}$ where $D$ is the dimension of these hidden features. The three contrastive learning losses ($L_{\text{session}}$, $L_{\text{subject}}$, and $L_{\text{task}}$) enable the model to perform classification within the same session, within the same subject, and across all subjects, respectively. After the backbone is trained, the classifier is trained by meta-learning in a Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017) style.

In Stage 2, the classifier is fine-tuned on a new subject's data with a fixed backbone. In Stage 3, the personalized model is frozen and tested on the subject's this or other sessions.

### Stream-Compatible Preprocessing

DE features (Duan et al., 2013) are spectral energy features containing five frequency bands (namely $\delta$: 1-3 Hz, $\theta$: 4-7 Hz, $\alpha$: 8-13 Hz, $\beta$: 14-30 Hz, and $\gamma$: 31-50 Hz). They are calculated through Fast Fourier Transformations and satisfy real-time processing requirements. A uniform mean MEAN $\in \mathbb{R}^F$ and standard deviation STD $\in \mathbb{R}^F$ calculated by non-masked DE features of training dataset are used to normalized all subjects' DE features.

The artifact processor implemented in this paper is adopted from L. Zhao et al. (2023) and the PREP pipeline (Bigdely-Shamlo et al., 2015), which can also support real-time processing. For each raw EEG window, the metrics contain:

- **Constant signals and outliers**: The median absolute deviations from the median (MAD) or the standard deviation (STD) is smaller than $T_{flat} = 1$. The data value is out of range of $T_{range} = [-120, 150]\ \mu V$.
- **Abnormal amplitude**. The z-score of robust standard deviation (rSD) is larger than $T_{ampzsd} = 5.0$.
- **High frequency noise**. The z-score of high frequency component's rSD is larger than $T_{hfzsd} = 8.0$.
- **Vertical electrooculography**. The signal from the prefrontal channel is monotonically increasing (little vibration is allowed) for more than $T_{veog1} = 0.08\ s$, and the cumulative increase exceeds the threshold $T_{veog2} = 60\ \mu V$.

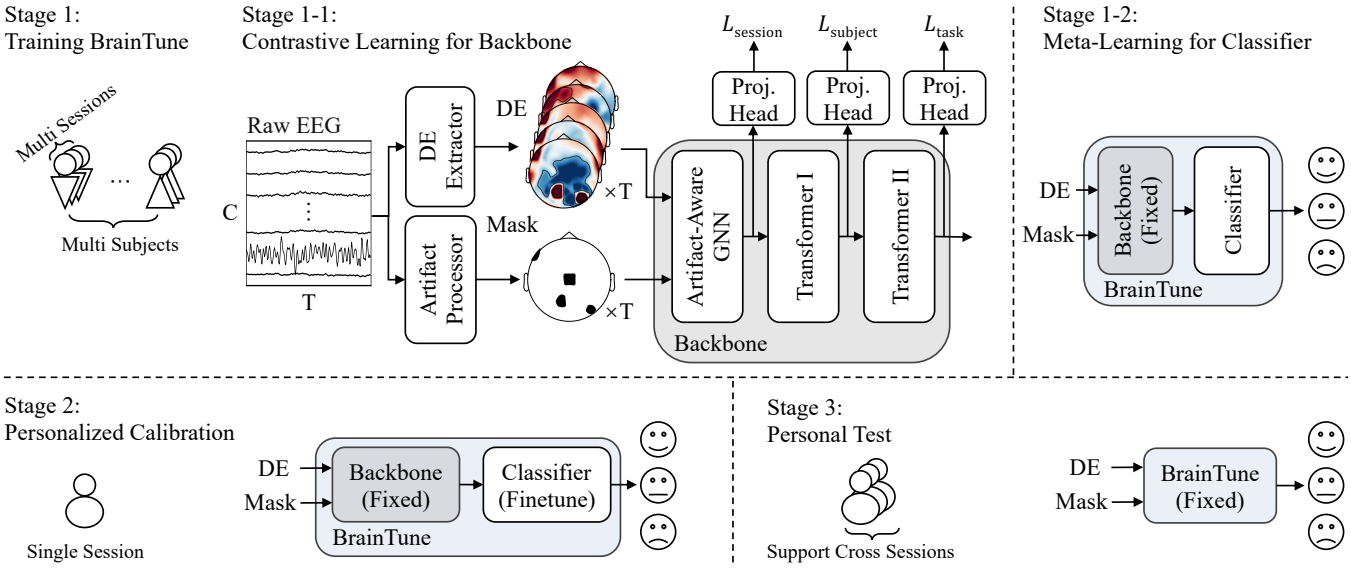


Figure 2: The framework of BrainTune. In Stage 1, the model is trained on multi subjects with multi sessions. The supervised multi-level contrastive learning is used for training the backbone, which learns session domain, subject domain and task domain progressively. The meta-learning is used for training the classifier to enhance generalizability and quick start. Stage 2 fine-tunes the model using a short period of data from a new subject. In Stage 3, BrainTune is tested on the subject's this or other sessions.

## Artifact-Aware Graph Neural Network

Standard GNNs treat all nodes equally, which is detrimental when artifacts occur. Our Artifact-Aware GNN (Figure 3) dynamically adjusts the graph topology based on the quality mask $M_t \in \mathbb{R}^C$ of the current time $t$. The network uses a hierarchical structure: channel nodes connect to region nodes, which aggregate to a master node. Aggregation is achieved by concatenating features and passing them through a linear layer. The key innovation lies in the conditional aggregation mechanism:

- If a node is clean, it aggregates features from its clean neighbors.
- If a node is noisy, its feature is discarded and replaced by the average of its clean neighbors.
- If a node and all neighbors are noisy, the node remains masked.

This logic allows the network to "self-heal" by leveraging spatial correlations in EEG signals, bypassing the need for offline artifact removal algorithms.

## Supervised Multi-Level Contrastive Learning

Supervised Multi-Level Contrastive Learning is employed to train the backbone of BrainTune. In the shallow layer, data from the same subject and session are treated as positive pairs, while data from different subjects are regarded as negative pairs. In the intermediate layer, data from the same subject across different sessions are considered positive pairs, while data from different subjects are treated as negative pairs. Finally, in the deep layer, data with the same labels are treated as positive pairs, while data with different labels are regarded

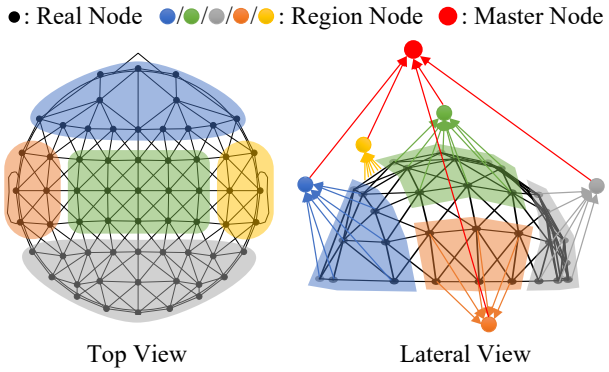


Figure 3: The hierarchical network of Artifact-Aware GNN aggregates features from real nodes to the master node. Left: Real nodes in the bottom layer are connected with undirected edges. Right: Region nodes and master nodes are connected with directed edges.

as negative pairs. The training loss is therefore a weighted sum of $L_{\text{session}}$, $L_{\text{subject}}$ and $L_{\text{task}}$:

$$Loss = \alpha \cdot L_{\text{session}} + \beta \cdot L_{\text{subject}} + \gamma \cdot L_{\text{task}}, \tag{1}$$

and each supervised contrastive loss (Khosla et al., 2020) is calculated as:

$$L_i^{sup} = -\frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p/\tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a/\tau)}, \tag{2}$$

where $z_i$ is a feature in the batch, $P(i)$ is $z_i$'s positive pairs, $A(i)$ is the other features in the batch except $z_i$, and $\tau$ is the temperature hyperparameter.

This hierarchical loss structure is designed to mirror the scale of EEG data variance. As shown in Figure 1, the domain shift is smallest within a session, larger across a subject's sessions, and largest between subjects. Due to the nature of back propagation, the shallow layer learns emotion classification on a per-session basis, the intermediate layer learns it on a per-subject basis, and the deep layer learns it by considering all subjects as a whole. This strategy resolves smaller domain shifts before tackling larger ones.

### Meta-Learning

To further enhance the ability to do domain generalization and enable quick adaption to a new subject with minimal data, Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017) is used to train the classifier, which is composed of two linear layers. When training the classifier, the backbone is keeping fixed. Each subject's session is treated as an individual task. For SEED, the support set is derived from a subject's one session, and the query set comes from his or her other two sessions, while for datasets with single session both the support and query set come from the same session.

## Experiments

### Datasets

We evaluate BrainTune on three widely-used affective BCI datasets: the SEED series (Zheng & Lu, 2015), DEAP (Koelstra et al., 2011), and DREAMER (Katsigiannis & Ramzan, 2017).

For the SEED series, which includes SEED-III, SEED-IV (Zheng et al., 2018), and SEED-V (W. Liu et al., 2021), we perform a three-class classification task (happy, neutral, sad), using only data corresponding to these emotions. Under this setup, SEED-III, SEED-IV, and SEED-V contain data from 15, 15, and 16 subjects viewing 15, 18, and 9 video clips, respectively. A key feature of the SEED series is that each subject participated in three separate experimental sessions.

For DEAP and DREAMER, we conduct a four-class classification task based on the arousal-valence model: Low Arousal-Low Valence (LALV), Low Arousal-High Valence (LAHV), High Arousal-Low Valence (HALV), and High Arousal-High Valence (HAHV). DEAP consists of data from 32 subjects (40 trials each), while DREAMER includes 23 subjects (18 trials each). In contrast to SEED, subjects in these two datasets completed only a single session.

Our evaluation employs a rigorous Leave-One-Subject-Out based protocol to simulate a real-world scenario where the model encounters a new user. For each fold, one subject is designated as the target subject, while the rest form the training and validation pool. The target subject's data is then partitioned into a calibration set and a test set. For the SEED evaluation, a subject from SEED-III is selected as the target. The training and validation pool comprises all subjects from SEED-IV and SEED-V, plus the remaining 14 subjects from SEED-III. The target subject's test set consists of the last 12 trials from their third session. To investigate calibration data size, we incrementally form the calibration set starting from the first three trials of their first session (674 s). For DEAP, the first 20 trials (20 min) are for calibration and the next 20 for testing; for DREAMER, the first 4 trials (844 s) are for calibration and the remaining 14 for testing. The relatively long calibration time for DEAP is intended to ensure that each label can appear in the calibration data

### Implementation Details

DE features are extracted from non-overlapping 1s windows ($T = 10, F = 5$) with global standardization and without any smoothing methods or artifact removal. Channel counts $C$ are 62 (SEED), 32 (DEAP), and 14 (DREAMER). For the backbone (3-layer Artifact-Aware GNN, 1-layer Transformer I and 1-layer Transformer II), we use a contrastive batch size of 4096 (SEED/DEAP) or 512 (DREAMER, which has fewer samples) and AdamW optimizer (learning rate = 1e-3). Feature dimension $D$ is 128. Loss weights $(\alpha, \beta, \gamma)$ are (1, 0.1, 1) for SEED and (1, 0, 2) for other two single-session datasets. For the classifier, meta-learning support/query batch sizes are (32, 32) for SEED, (16, 32) for DEAP, and (8, 16) for DREAMER. We use SGD (learning rate = 1e-2) for the inner loop and AdamW (learning rate = 1e-3) for the outer. Calibration uses a batch size of 32 for 3 epochs (learning rate = 1e-4).

### Experiment Results

We evaluate BrainTune in two settings. Setting I tests lasting personalization, where calibration data from early sessions improves performance on future sessions on the SEED dataset. Setting II compares no-calibration, within-session, and cross-session calibration scenarios. We report the average results across all test subjects.

Baselines include a subject-specific SOTA, EmoGT (Jiang, Yan, et al., 2023), a classical DA model, DANN (Ganin et al., 2016), an unsupervised pre-training SOTA, MV-SSTMA (R. Li et al., 2022), and DG SOTAs, MoGE (X.-H. Liu et al., 2024), DMMR (Wang et al., 2024), and MAET (Jiang, Liu, et al., 2023). To ensure a fair comparison, all baselines were trained and tested using the exact same stream-compatible DE features as BrainTune. For calibration, we froze the backbone and fine-tuned the classifier for 3 epochs (learning rate = 1e-4), except for MV-SSTMA, which has its own strategy. Note that DANN requires access to all unlabeled target data during training, a condition not permitted for other methods, giving it an unfair advantage.

Figure 4 shows the results for Setting I. The calibration data starts with the first three video clips from the first session, incrementally adding three clips at a time, up to the first three clips from the third session. BrainTune achieves significantly higher accuracy, with performance jumping as data from a new session is added for calibration, reaching 70.82 ± 9.82%. In contrast, EmoGT and MAET exhibit negative transfer, where more calibration data hurts performance. Even with its unfair advantage, DANN underperforms BrainTune, likely due to its difficulty with multi-domain source data and the large source-

Table 1: Setting II. “M.-S." is short for “MV-SSTMA", “S1/2/3" is short for “Session 1/2/3", “w/o Cal" means “without calibration" and “w/ Cal" means “with calibration". “S1/2/3" means the first three video clips from which session are used for calibration. Accuracy is reported for SEED, while balanced accuracy is reported for DEAP and DREAMER due to their imbalanced labels. BrainTune has the best performance.

| Method | SEED Series | | | | | DEAP | | DREAMER | |
|---|---|---|---|---|---|---|---|---|---|
| | w/o Cal | S1 | S1+S2 | S1+S2+S3 | S3 | w/o Cal | w/ Cal | w/o Cal | w/ Cal |
| MAET | 55.07±11.88 | 48.87±10.02 | 51.38±9.98 | 53.29±10.41 | 54.43±9.57 | 37.79±11.05 | 38.43±10.87 | 38.86±10.21 | 39.09±8.47 |
| MoGE | 52.45±14.21 | 53.02±14.11 | 53.76±14.22 | 54.45±14.32 | 53.21±14.24 | 34.17±10.14 | 35.83±12.38 | 35.74±10.15 | 36.44±11.23 |
| M.-S. | / | 51.68±14.27 | 54.02±13.73 | 60.00±12.65 | 55.77±13.82 | / | 36.77±11.12 | / | 37.96±9.67 |
| EmoGT | 53.55±13.98 | 51.93±11.36 | 53.25±12.03 | 55.01±12.19 | 57.43±14.11 | 36.06±12.40 | 41.00±14.69 | 37.28±11.54 | 41.54±10.79 |
| DMMR | 54.32±13.39 | 55.13±13.15 | 57.74±14.54 | 58.75±14.65 | 61.76±14.20 | 34.31±9.48 | 38.24±13.32 | 38.10±10.32 | 43.79±9.78 |
| DANN | 56.02±10.77 | 56.87±11.19 | 57.24±12.40 | 57.97±14.01 | 57.32±11.82 | 35.90±10.52 | 38.83±10.96 | 39.63±9.47 | 40.84±9.05 |
| **Ours** | **59.33**±10.61 | **65.42**±10.42 | **66.39**±10.45 | **68.63**±11.87 | **68.03**±9.88 | **42.57**±12.29 | **51.71**±10.68 | **41.66**±9.30 | **52.31**±7.19 |

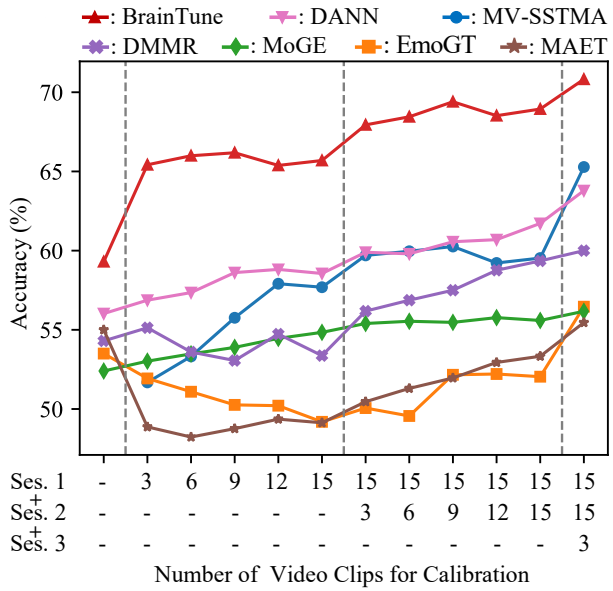


Figure 4: Setting I, the average curve of each subject in SEED-III as the amount of calibration data progressively increases. BrainTune consistently outperforms all baselines.

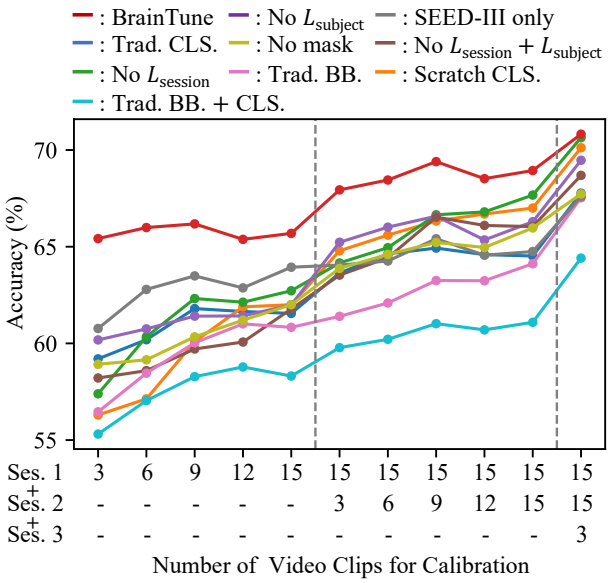


Figure 5: Ablation study for BrainTune on SEED. “BB.", “CLS." and “Trad." are short for “backbone", “classifier" and “traditional". Any ablation is detrimental.

target data imbalance. Lastly, MV-SSTMA, lacking emotion-specific pre-training, is suboptimal with limited calibration data, and it cannot function without calibration.

Table 1 presents the results of Setting II. For the SEED series, generally speaking, S3 (within-session calibration) is higher than S1 (cross-session calibration), and S1 is higher than the result without fine-tuning. Similarly, for the DEAP and DREAMER datasets, we can also observe that calibration can improve the model's accuracy. Crucially, BrainTune demonstrates superior “fine-tunability", achieving the most significant performance gains from calibration on all datasets. In contrast, some baselines like MAET even exhibit negative transfer under the within-session and cross-session calibration scenario (i.e., S1 and S3 on SEED), struggling to leverage calibration data effectively.

To demonstrate stream compatibility and fast-adaptation capabilities, the time from raw EEG input to inference output and the time for fine-tuning with three preprocessed calibration videos were measured, which were 325±4 ms and 716±53 ms, respectively. These are the average results of ten tests conducted in an environment with an i3-10105 CPU and an RTX 3060 GPU. The reason for not comparing latency with other methods is that those methods are originally designed to require offline preprocessing of EEG data. Besides, the efficient model design ensures that each subject's complete training and testing procedure requires under 20 minutes.

## Ablation Study

In this section, the term “traditional" refers to training using conventional supervised learning approaches and the cross-entropy loss function exclusively. The ablation studies of

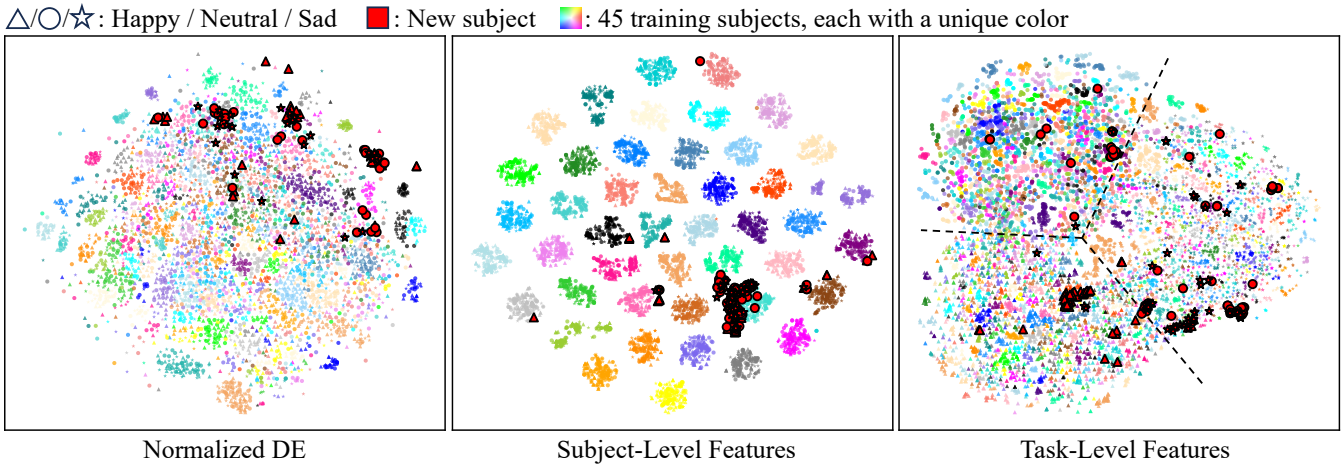


Figure 6: Visualization of the input DE features, subject-level features, and task-level features, using SEED as an example. The background displays data from three sessions conducted with 45 subjects. Each training subject is represented by a unique color, while the new subject is marked in red with a black border. BrainTune automatically maps the new subject into existing domains, realizing domain generalization.

BrainTune are conducted on SEED series, which cover four aspects: ablation of classifier training methods, ablation of backbone training methods, data volume ablation, and artifact module ablation. The ablation of classifier training methods includes two scenarios: using a traditional classification head and training a classification head from scratch during the calibration phase, with the backbone still trained using supervised multi-level contrastive learning. The ablation of backbone training methods involves the removal of the shallow or intermediate loss function, as well as training with traditional methods, while the classifier continues to be trained using meta-learning. The data volume ablation focuses on training solely with the SEED-III, excluding SEED-IV and SEED-V. Artifact module ablation sets all artifact masks to zero, effectively disabling the artifact module. Additionally, we also present the result with the traditional approach, that is training both backbone and classifier in a traditional method.

Figure 5 presents the results of all ablation experiments on BrainTune, demonstrating that any type of ablation reduces its performance. Several key conclusions can be drawn: 1) Traditional training methods, whether applied to the backbone network or the classification head, negatively impact the model's generalization capability; 2) A classification head without prior knowledge performs poorly during small-scale data calibration, though it can achieve satisfactory results with sufficient calibration data; 3) Data volume ablation achieves higher accuracy under small-scale data calibration compared to other structural ablations, underscoring the rationality of BrainTune's structural design. These comprehensive ablation studies demonstrate the validity of BrainTune.

### Visualization

To better explain the design of the backbone network, we visualized the input DE features, the intermediate layer output (subject-level features), and the final layer output (task-level features), as shown in Figure 6. Visualizations were generated by randomly sampling a fixed number of instances for each emotion from each subject during each session in SEED series. The background comprises data from three sessions conducted on 45 training subjects, with each subject represented by a unique color. The test subject's three sessions are red icons with black borders.

It can be observed that the DE features of each subject in each session tend to form a distinct domain. Through the shallow and intermediate layers, the model progressively clarifies each subject's domain, automatically mapping the new subject into existing domains. The deep layer further separates the feature space into three categories. However, for the new subject, neutral and sad emotions are often confused, which aligns with findings from prior studies (R. Li et al., 2022; Y. Li et al., 2018; Zheng & Lu, 2016).

## Conclusion

This paper proposes BrainTune to overcome two critical barriers in EEG emotion recognition: the reliance on offline processing and the personalization gap. The efficacy of BrainTune stems from three synergistic designs. First, the unified standardization methods and Artifact-Aware GNN enable the model's input to operate without seeing future EEG data and avoid the impact of artifacts, thus achieving stream compatibility. Second, the hierarchical contrastive backbone progressively disentangles intra-session, inter-session and inter-subject variations, enabling "lasting personalization" where the model continuously benefits from historical data. Third, the meta-learned classifier enables rapid personalization from brief calibration. Together, these designs provide a blueprint for transitioning affective BCIs from laboratory settings to real-world use.

## Acknowledgments

This work was supported in part by grants from Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project (2025ZD0218900), National Key Research and Development Program of China (2024YFC3606800), National Natural Science Foundation of China (62376158), STI 2030-Major Projects+2022ZD0208500, Medical-Engineering Interdisciplinary Research Foundation of Shanghai Jiao Tong University "Jiao Tong Star" Program (YG2023ZD25, YG2024ZD25 and YG2024QNA03), Shanghai Jiao Tong University 2030 Initiative, the Lingang Laboratory (Grant No. LGL-1987), GuangCi Professorship Program of RuiJin Hospital Shanghai Jiao Tong University School of Medicine, and Shanghai Jiao Tong University SCS-Shanghai Emotionhelper Technology Co., Ltd Joint Laboratory of Affective Brain-Computer Interfaces.

## References

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*, 16.

Cai, H., & Pan, J. (2023). Two-phase prototypical contrastive domain generalization for cross-subject EEG-based emotion recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1–5.

Duan, R.-N., Zhu, J.-Y., & Lu, B.-L. (2013). Differential entropy feature for EEG-based emotion classification. *International IEEE/EMBS Conference on Neural Engineering*, 81–84.

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *International Conference on Machine Learning*, 1126–1135.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of machine learning research*, *17*(59), 1–35.

Jiang, W.-B., Liu, X.-H., Zheng, W.-L., & Lu, B.-L. (2023). Multimodal adaptive emotion transformer with flexible modality inputs on a novel dataset with continuous labels. *Proceedings of the 31st ACM International Conference on Multimedia*, 5975–5984.

Jiang, W.-B., Yan, X., Zheng, W.-L., & Lu, B.-L. (2023). Elastic graph transformer networks for EEG-based emotion recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1–5.

Katsigiannis, S., & Ramzan, N. (2017). 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*, *22*(1), 98–107.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., & Krishnan, D. (2020). Supervised contrastive learning. *Advances in Neural Information Processing Systems*, *33*, 18661–18673.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31.

Lan, Z., Sourina, O., Wang, L., & Liu, Y. (2016). Real-time EEG-based emotion monitoring using stable features. *The Visual Computer*, *32*, 347–358.

Li, R., Wang, Y., Zheng, W.-L., & Lu, B.-L. (2022). A multi-view spectral-spatial-temporal masked autoencoder for decoding emotions with self-supervised learning. *Proceedings of the 30th ACM International Conference on Multimedia*, 6–14.

Li, Y., Zheng, W., Cui, Z., Zhang, T., & Zong, Y. (2018). A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition. *International Joint Conference on Artificial Intelligence*, 1561–1567.

Liu, W., Qiu, J.-L., Zheng, W.-L., & Lu, B.-L. (2021). Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, *14*(2), 715–729.

Liu, X.-H., Jiang, W.-B., Zheng, W.-L., & Lu, B.-L. (2024). MoGE: Mixture of graph experts for cross-subject emotion recognition via decomposing EEG. *2024 IEEE International Conference on Bioinformatics and Biomedicine*, 3515–3520.

Liu, Y., Sourina, O., & Nguyen, M. K. (2011). Real-time EEG-based emotion recognition and its applications. *Transactions on Computational Science XII: Special Issue on Cyberworlds*, 256–277.

Luo, Y., & Lu, B.-L. (2021). Wasserstein-distance-based multi-source adversarial domain adaptation for emotion recognition and vigilance estimation. *IEEE International Conference on Bioinformatics and Biomedicine*, 1424–1428.

Ma, B.-Q., Li, H., Zheng, W.-L., & Lu, B.-L. (2019). Reducing the subject variability of EEG signals with adversarial domain generalization. *International Conference on Neural Information Processing*, 30–42.

Wang, Y., Zhang, B., & Tang, Y. (2024). DMMR: Cross-subject domain generalization for EEG-based emotion recognition via denoising mixed mutual reconstruction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*, 628–636.

Wu, D., Lu, B.-L., Hu, B., & Zeng, Z. (2023). Affective brain–computer interfaces (aBCIs): A tutorial. *Proceedings of the IEEE*, *111*(10), 1314–1332.

Zhao, L.-M., Yan, X., & Lu, B.-L. (2021). Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*, 863–870.

Zhao, L., Zhang, Y., Yu, X., Wu, H., Wang, L., Li, F., Duan, M., Lai, Y., Liu, T., Dong, L., et al. (2023). Quantitative signal quality assessment for large-scale continuous scalp electroencephalography from a big data perspective. *Physiological Measurement*, *44*(3), 035009.

Zheng, W.-L., Liu, W., Lu, Y., Lu, B.-L., & Cichocki, A. (2018). Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, *49*(3), 1110–1122.

Zheng, W.-L., & Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175.

Zheng, W.-L., & Lu, B.-L. (2016). Personalizing EEG-based affective models with transfer learning. *International Joint Conference on Artificial Intelligence*, 2732–2738.

Zheng, W.-L., Zhu, J.-Y., & Lu, B.-L. (2017). Identifying stable patterns over time for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, *10*(3), 417–429.

Zhou, R., Ye, W., Zhang, Z., Luo, Y., Zhang, L., Li, L., Huang, G., Dong, Y., Zhang, Y.-T., & Liang, Z. (2024). EEGMatch: Learning with incomplete labels for semisupervised EEG-based cross-subject emotion recognition. *IEEE Transactions on Neural Networks and Learning Systems*.