Leveraging Semi-Supervised Learning and Self-Attention for EEG-Based Emotion Recognition

Arun Kumar S
BM24RESCH12001
Indian Institute of Technology Hyderabad
bm24resch12001@iith.ac.in

Bapaladoddi Vamshi Krishna
BM21BTECH11006
Indian Institute of Technology Hyderabad
bm21btech11006@iith.ac.in

Burra Vishal Mathews
CS21BTECH11010
Indian Institute of Technology Hyderabad
cs21btech11010@iith.ac.in

Potta Vennala
CS21BTECH11046
Indian Institute of Technology Hyderabad
cs21btech11046@iith.ac.in

Pundi Bindu Sree
CS21BTECH11048
Indian Institute of Technology Hyderabad
cs21btech11048@iith.ac.in

Abstract

EEG emotion recognition through semi-supervised learning has gained attention for its applications in emotion classification. Our model utilizes data augmentation, label guessing, and convex combinations of unlabeled and labeled datasets, along with pairwise representation alignment to harmonize their distributions. Incorporating self-attention mechanisms further enhances the model's capability to effectively capture long-range EEG contextual information. To address the mismatch between large amounts of unlabeled data and a limited set of labeled data, we utilize pairwise representation alignment to effectively bridge this gap. Initially, our model conducts data augmentation, followed by pseudo-labeling, sharpening of the pseudo-labels, and finally applies MixUp to combine labeled and unlabeled data. We then perform representation alignment and classify the emotions with the self - attention mechanism based on the enhanced data. We plan to compare the performance of this model with existing semi-supervised models to evaluate its effectiveness in emotion classification. We will utilize publicly available EEG-based emotion classification datasets, such as SEED, SEED IV, and SEED V.

1. Introduction

Human emotions play a crucial role in daily life, influencing behaviors, interactions, and perceptions. Understanding emotions through algorithms can improve how computers meet human needs. This has led to the rise of affective computing, a field dedicated to creating computational models that can accurately identify human emotional states [10]. Emotion classification is particularly important in mental health monitoring, as it can help detect emotional changes, track progress, and provide timely interventions for those in need.

Humans express emotions through facial expressions [2] and speech [8], which can be quantified with physiological measures such as electrocardiogram (ECG) [11], electrodermal activity (EDA) [1], photoplethysmogram (PPG) [9], and electroencephalogram (EEG) [15]. EEG is often preferred in affective computing due to its strong link to brain activity, making it a valuable tool for recognizing emotions.

Deep learning has shown great promise in various applications, but it also faces challenges, particularly in the context of electroencephalogram (EEG) data for emotion recognition. One significant issue is performance degradation, which can occur due to the limited availability of high-quality labeled data. The process of annotating emotions in EEG data presents additional complications, such as the need for pre-stimulation protocols, self-assessment

by participants, and evaluations from multiple experts, all of which can lead to inconsistencies and inaccuracies. To address these challenges, we can utilize semi-supervised learning, which leverages unlabeled data alongside a smaller set of labeled examples. This approach can help improve model performance by capturing more comprehensive patterns in the data, ultimately enhancing emotion recognition in EEG analysis.

In this study, we use a pairwise representation semi-supervised framework called PARSE, as introduced by Guangyi Zhang *et al*. [16], to enhance emotion recognition in EEG data. This framework employs a Semi-Supervised Learning (SSL) approach that leverages large amounts of unlabeled EEG data alongside a small labeled subset. As a novel contribution, we plan to incorporate a self-attention mechanism within PARSE. This integration aims to improve accuracy by allowing the model to focus on salient features and capture long-range dependencies in the preprocessed EEG data encoder. By leveraging self-attention, the framework can better understand complex relationships between different segments of the EEG signals, which is crucial for effective emotion classification. This addition will help the framework better manage long-range dependencies and variations in the data, ultimately leading to more robust and reliable outcomes in emotion classification.

1.1. Problem Statement

The use of semi-supervised learning (SSL) in EEG representation learning faces several critical challenges that can hinder its effectiveness. One major issue is the distribution of unlabeled samples, which often varies significantly from the labeled subset. This discrepancy can lead to low-confidence pseudo labels being generated during the training process. When the model encounters a diverse set of unlabeled data, the confidence threshold for these pseudo labels may not accurately reflect the true emotional states, resulting in unreliable training signals. Consequently, the quality of the pseudo labels becomes a significant concern, as low-quality annotations can further degrade the model's performance and learning efficiency.

Moreover, the internal distribution of the unlabeled dataset poses additional complications, as the lack of generalizability across different EEG datasets can limit the framework's applicability. Vast differences in data distribution among various EEG due to factors such as equipment, different participants, and experimental conditions which can adversely affect the training process. This lack of consistency makes it challenging for the model to learn robust representations that are applicable across different contexts. Therefore, improving the quality and reliability of pseudo labels, while also addressing the distributional differences in EEG data, is essential for enhancing the overall performance and generalizability of semi-supervised learning frameworks like PARSE.

2. Related Work

2.1. Emotion Recognition With EEG

Emotion recognition using EEG typically involves preprocessing the EEG recordings, extracting relevant features, and then applying a classifier to learn from these features[15]. In the preprocessing we do down sampling, artifact removal and noise filtering to clean the EEG data. After preprocessing, features such as Differential Entropy (DE) and logarithmic Power Spectral Density (PSD) are extracted from key EEG frequency bands (e.g, alpha, beta, and gamma). Once features are extracted, different classifiers such as K-nearest neighbor [10], support vector machines [7], [14], [18], [**?**], logistic regression , random forest, and naive Bayes [12] are employed to learn and predict emotional states from the nonlinear output information.

2.2. Semi supervised learning

Deep learning networks struggle with overfitting and slow convergence when trained on limited labeled data. Unsupervised pre-training methods like Stacked Auto-Encoders (SAE) and Deep Belief Networks (DBN)[5][6] help regularize models and improve performance. Recent techniques, such as contrastive learning and pseudo-labeling, further enhance results by leveraging unlabeled data. Consistency regularization ensures stable predictions despite data augmentations, and methods like the P-model and Mean Teacher enforce this stability. Holistic approaches like MixMatch and FixMatch combine these ideas to boost semi-supervised learning, improving performance with small labeled datasets.

2.3. Parse Framework

The proposed solution for EEG-based emotion recognition leverages semi-supervised learning (SSL) to handle scenarios with limited labeled data and large amounts of unlabeled data. The architecture uses two essential **augmentations-weak and strong**-on labeled and unlabeled EEG data, followed by a label-guessing process to generate pseudo-labels for the unlabeled data. The MixUp technique combines labeled and unlabeled data via convex combinations, improving generalization by addressing noisy perturbations. A pairwise alignment mechanism helps align the distributions of labeled and unlabeled data, minimizing domain shifts. The architecture consists of a **1-D convolutional** encoder for feature extraction, a classifier for emotion recognition, and a discriminator to distinguish between labeled

and unlabeled data. The total loss function combines supervised, unsupervised, classification, and discriminator losses to optimize the model's performance on emotion recognition while ensuring the alignment of labeled and unlabeled data representations.

3. Dataset Description

In the SEED-IV dataset[17] 15 participants completed 3 separate sessions, each consisting of 24 trials, resulting in a total of 72 trials per participant. During each trial, participants watched carefully selected film clips while EEG data was recorded using a 62-channel ESI NeuroScan System. The recorded data from each session was segmented into 4-second non-overlapping intervals, with each segment treated as an independent data sample for model training. This approach allowed for efficient feature extraction from the EEG data corresponding to the participants' responses to the film clips.
Link to the Dataset (SEED-IV) :
https://bcmi.sjtu.edu.cn/home/seed/seed-iv.html

4. Preprocessing:

EEG preprocessing is often dataset-dependent due to the various collection equipment, as well as experimental conditions and protocols. It often contains downsampling, artefact removal, and noise filtering. In our experiments, we use the preprocessed data that are made public by the original papers which introduced the datasets

4.1. Downsampling

In the SEED-IV dataset, EEG signals were down-sampled from 1000 Hz to 200 Hz.

4.2. Artefact Removal and Noise Filtering

EEG signals are often contaminated by Electrooculogram(EOG) and Electromyogram (EMG) . EOG signals,which reflect the ocular activities such as eye movements and eye blinks, are most active below 4Hz . EMG signals reflect the muscular activities around the face and are domi-nant above 30Hz . Both EOG and EMG have a high overlap with EEG, which is dominant in 0.3 - 50Hz. Consequently, different artefact removal and noise filtering methods were used in the original papers that published the datasets.

At last, a band-pass filter was applied to filter the noise and artefacts outside the dominant EEG frequency range. Specifically, the frequency ranges of the bandpass filter are 1- 75Hz for SEED-IV Dataset.

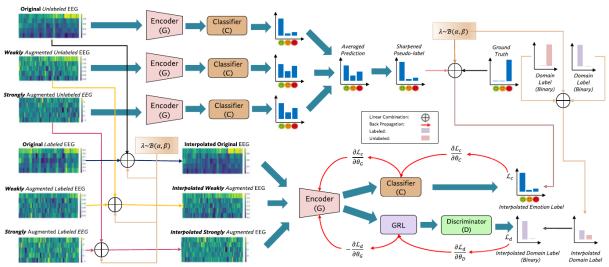

Figure 1. The architecture of our proposed semi-supervised framework with pairwise representation alignment

4.3. Feature Extraction

In the SEED-IV dataset[17], EEG data was divided into 4-second non-overlapping segments for feature extraction. The extracted features include differential entropy (DE) values from five EEG frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-31 Hz), and gamma (31-50 Hz). These features were calculated for each of the 62 EEG channels, resulting in 310 features per segment (62 channels × 5 frequency bands).The extracted features were reshaped from a 2D matrix [62 channels, 5 features per channel] into a 1D feature vector of length 310. The features were further scaled using min-max normalization, which ensures they range between 0 and 1 before being used in the analysis. The process begins by extracting frequency band features from each EEG channel in sequence.

4.4. Evaluation Protocol

For the SEED-IV dataset[17], the evaluation protocol involves using the first 16 trials (with 4 trials per emotion class) as the training set, and the remaining 8 trials as the testing set. This approach ensures a consistent split between training and testing data for emotion recognition tasks. Since the class distributions in SEED-IV are almost balanced, accuracy is used as the primary evaluation metric for assessing model performance.

5. Implemented Architecture

5.1. Data Augmentation

Data augmentation involves applying **weak** and **strong augmentations** to both labeled and unlabeled EEG data:

- **Weak Augmentation**: Aims to preserve the core features of the EEG signals, ensuring the retention of identifiable patterns essential for effective training.
- **Strong Augmentation**: Introduces distortions that simulate challenging conditions, enabling the model to become robust to domain shifts. Strong augmentations are

pivotal for generating pseudo-labels.

5.1.1 Need for Strong and Weak Augmentations

Weak augmentations maintain consistency in training by preserving essential patterns, while strong augmentations provide robustness to variations in the data [4]. The combination ensures:

1. **Robust pseudo-label generation**: Strong augmentations introduce noise to simulate domain shifts.
2. **Model generalization**: Weak augmentations help prevent overfitting to noisy pseudo-labels.

5.2. Prediction Averaging and Sharpening

To stabilize pseudo-labels, predictions are averaged across original and augmented unlabeled data[3]. Averaging reduces noise and provides more generalizable labels:

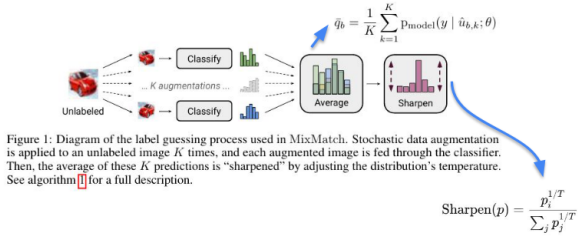

Figure 1: Diagram of the label guessing process used in MixMatch. Stochastic data augmentation is applied to an unlabeled image K times, and each augmented image is fed through the classifier. Then, the average of these K predictions is "sharpened" by adjusting the distribution's temperature. See algorithm 1 for a full description.

- **Prediction Sharpening**: Sharpens the averaged predictions by reducing their entropy, transforming soft probabilities into more confident labels. This is achieved by adjusting the temperature parameter T, with $T < 1$ making probabilities closer to 0 or 1. Sharpened pseudo-labels mimic true labels more effectively, ensuring better training stability.

5.3. Pseudo-Label Generation

Pseudo-labels are derived from the sharpened predictions of the unlabeled data. They act as approximations of true labels, enabling the inclusion of unlabeled data in supervised training. However, the quality of pseudo-labels is critical for effective learning.

5.4. Convex Combination Generation

The **MixUp augmentation** technique creates convex combinations of labeled and unlabeled data [13]. By linearly interpolating inputs and their labels, MixUp:

1. Enforces consistency regularization, ensuring that predictions remain robust under augmented conditions.
2. Prevents overfitting by creating diverse synthetic examples.
3. Aligns embeddings for labeled and unlabeled samples, reducing domain discrepancies.

5.5. Pairwise Embedding Alignment

Pairwise embedding alignment involves aligning the feature representations of labeled and unlabeled data:

- **Gradient Reverse Layer (GRL)**: Reverses gradients during training to confuse the domain discriminator, making labeled and unlabeled embeddings indistinguishable.
- **Hinge Loss**: Minimizes distribution divergence between labeled and unlabeled embeddings, encouraging alignment.

5.6. Training Strategy

The proposed framework employs simultaneous training of:

1. An emotion classifier for recognizing emotions.
2. A domain discriminator for aligning labeled and unlabeled data distributions.

A warm-up function gradually increases the unsupervised loss contribution, allowing the model to first focus on supervised training and then incorporate pseudo-labels as confidence improves.

6. Model Architecture

The architecture comprises:

1. **Encoder**: Two 1D convolutional layers with Batch Normalization and Leaky ReLU activation. Each convolutional layer reduces the input dimension and captures critical features.
2. **Embedding Layer**: Flattens the encoded features for further processing.
3. **Classifier**: A fully connected layer followed by a dropout layer and a final linear layer producing k output classes.
4. **Discriminator**: Two fully connected layers with ReLU and dropout, culminating in a binary classification layer for domain discrimination.

The model balances complexity with efficiency, leveraging convolutional layers for feature extraction and adversarial domain adaptation.

7. Comparative Analysis

The proposed framework was evaluated against existing SSL methods, including MixMatch, FixMatch, and AdaMatch:

- **MixMatch**: Combines pseudo-labeling, sharpening, and MixUp, but suffers from sensitivity to pseudo-label quality.
- **FixMatch**: Employs confidence thresholds and strong augmentations but is less adaptable to domain shifts.
- **AdaMatch**: Incorporates adaptive thresholds and distribution alignment but is computationally intensive.

The proposed method outperforms these approaches by integrating task-specific enhancements tailored for EEG data, offering robustness to noise and domain shifts.

8. Results

The results demonstrate the effectiveness of the proposed method compared to existing semi-supervised learning frameworks (MixMatch, FixMatch, and AdaMatch) across varying numbers of labeled samples per class. With fewer labeled samples, the proposed method achieves competitive accuracy, surpassing other methods by maintaining robustness against noise and leveraging domain alignment. As the number of labeled samples increases, the performance gap between methods narrows; however, the proposed method consistently achieves the highest accuracy across all scenarios. Notably, the use of advanced augmentations, pseudo-labeling, and pairwise alignment allows the proposed method to handle domain shifts effectively and generalize better, particularly when the labeled data is scarce.

The experimental results demonstrate the effectiveness of the proposed method compared to existing semi-supervised learning (SSL) approaches. Key observations include:

- **Performance with Limited Labels:** The proposed method consistently achieves higher accuracy across all label settings, particularly in low-label scenarios (e.g., 1-label and 3-label cases). This highlights its ability to leverage unlabeled data effectively using robust pseudo-labeling and domain alignment strategies.
- **Comparison with MixMatch:** While MixMatch performs well as the number of labels increases, it struggles with lower label counts, reflecting its sensitivity to pseudo-label quality. In contrast, the proposed method demonstrates stability and adaptability in such scenarios.
- **Comparison with FixMatch:** FixMatch, though simple and efficient, lags behind due to its reliance on fixed confidence thresholds for pseudo-labeling. This approach is less effective for complex EEG data with domain shifts.
- **Comparison with AdaMatch:** AdaMatch shows robust performance, but its computational complexity limits scalability. The proposed method outperforms AdaMatch in most label settings while maintaining computational efficiency.
- **Scalability with Increasing Labels:** As the number of labeled samples increases, all methods show improved accuracy, with the performance gap narrowing. However, the proposed method retains a slight edge, demonstrating its robustness even with ample labeled data.
- **Generalization and Robustness:** The combination of strong and weak augmentations, MixUp, and adversarial domain alignment enables the proposed method to generalize well across labeled and unlabeled data, handling noise and domain shifts effectively.

These observations validate the proposed framework's suitability for EEG-based emotion recognition, particularly in scenarios with limited labeled data. Its task-specific enhancements and efficient training strategy provide a balanced approach for real-world applications.

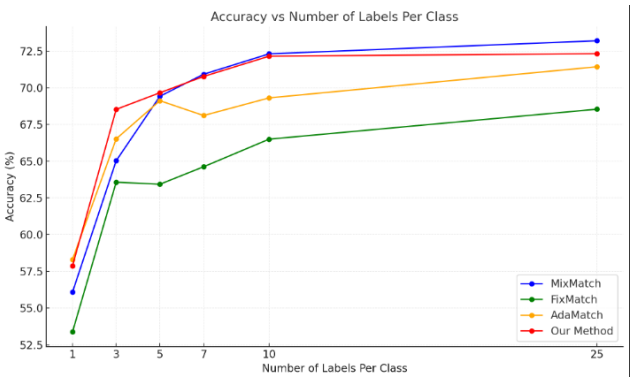

Figure 2. Accuracies of different SSL architectures with different n-labeled per classses

n-labeled per class

	1-labels	3-labels	5-labels	7-labels	10-labels	25-labels
MixMatch	56.08(15.92)	65.03(15.79)	69.42(16.31)	**70.92(16.02)**	**72.31(16.27)**	73.20(15.19)
FixMatch	53.37(17.33)	63.57(15.57)	63.43(16.26)	64.62(15.57)	66.50(15.91)	68.54(15.58)
AdaMatch	**58.30(15.95)**	66.52(16.58)	69.12(16.45)	68.11(15.80)	69.31(16.87)	71.43(16.04)
Our Method	57.87(18.09)	**68.53(16.34)**	**69.66(15.96)**	70.77(16.47)	72.15(16.08)	**72.32(16.15)**

Figure 3. Accuracies of different SSL architectures

9. Conclusion

This study presents a comprehensive framework for enhancing EEG-based emotion recognition using semi-supervised learning (SSL) techniques. By addressing the challenges of limited labeled data, high dimensionality, and domain shifts, the proposed method demonstrates robust performance across varying levels of labeled samples. The exploration and analysis of EEG data provided valuable insights into its processing and interpretation for practical applications.

Advanced SSL concepts such as pseudo-label generation, domain alignment, and data augmentation were effectively implemented and evaluated. The framework's use of techniques like MixUp, strong and weak augmentations, and adversarial training allowed for a balanced integration of labeled and unlabeled data, improving model generalization and robustness.

The experiments conducted on the SEED-IV dataset validated the efficacy of the proposed approach, showcasing its superiority over existing methods such as MixMatch, FixMatch, and AdaMatch. The results highlight the framework's adaptability and scalability, making it a promising

solution for real-world scenarios with limited labeled EEG data.

Additionally, this research facilitated the development of professional skills such as collaborative problem-solving, effective communication, and structured research presentation. The findings underscore the importance of leveraging unlabeled data and task-specific enhancements for advancing emotion recognition technologies.

References

[1] D. Rodenburg P. Hungler A. Bhatti, B. Behinaein and A. Etemad. Attentive cross-modal connections for deep multimodal wearable-based emotion recognition. *arXiv preprint arXiv:2108.02241*, 2021. 1

[2] F. Pereira A. Sepas-Moghaddam, A. Etemad and P. L. Correia. Facial emotion recognition using light field images with deep attention-based bidirectional lstm. In *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pages 3367–3371, 2020. 1

[3] I. Goodfellow N. Papernot-A. Oliver D. Berthelot, N. Carlini and C. A. Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *Proc. Adv. Neural Inf. Process. Syst.*, pages 5049–5059, 2019. 4

[4] K. Sohn N. Carlini D. Berthelot, R. Roelofs and A. Kurakin. Adamatch: A unified approach to semi-supervised learning and domain adaptation. *arXiv preprint arXiv:2106.04732*, 2021. 4

[5] Y. Bengio D. Erhan, A. Courville and P. Vincent. Why does unsupervised pre-training help deep learning? In *Proc. 13th Int. Conf. Artif. Intell. Statist.*, pages 201–208, 2010. 2

[6] J. E. Van Engelen and H. H. Hoos. A survey on semi-supervised learning. *Mach. Learn.*, 109(2):373–440, 2020. 2

[7] A. Sepas-Moghaddam Y. Zhang G. Zhang, V. Davoodnia and A. Etemad. Classification of hand movements from eeg using a deep attention-based lstm network. *IEEE Sensors J.*, 20(6): 3113–3122, 2019. 2

[8] A. Hajavi and A. Etemad. Knowing what to listen to: Early attention for deep speech representation learning. *arXiv preprint arXiv:2009.01822*, 2020. 1

[9] M. Z. Uddin S. Huda A. Almogren M. M. Hassan, M. G. R. Alam and G. Fortino. Human emotion recognition using deep belief network architecture. *Inf. Fusion*, 51:10–18, 2019. 1

[10] R. W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, USA, 2000. 1

[11] P. Sarkar and A. Etemad. Self-supervised ecg representation learning for emotion recognition. *IEEE Trans. Affective Comput.*, 13(3):1541–1554, 2020. 1

[12] Y. Lu B.-L. Lu W.-L. Zheng, W. Liu and A. Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.*, 49(3):1110–1122, 2019. 2

[13] Wei Wei, Jiahuan Zhou, and Ying Wu. Beyond empirical risk minimization: Local structure preserving regularization for improving adversarial robustness. *arXiv preprint arXiv:2303.16861*, 2023. 4

[14] G. Zhang and A. Etemad. Rfnet: Riemannian fusion network for eeg-based brain-computer interfaces. *arXiv preprint arXiv:2008.08633*, 2020. 2

[15] G. Zhang and A. Etemad. Distilling eeg representations via capsules for affective computing. *arXiv preprint arXiv:2105.00104*, 2021. 1, 2

[16] Guangyi Zhang, Vandad Davoodnia, and Ali Etemad. Parse: Pairwise alignment of representations in semi-supervised eeg learning for emotion recognition. *IEEE Transactions on Affective Computing*, 13(4):2185–2200, 2022. 2

[17] W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, pages 1–13, 2018. 3

[18] W.-L. Zheng and B.-L. Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Mental Develop.*, 7(3):162–175, 2015. 2