EEG-Based Emotion Recognition Using Physiological Signals

FYP–I Report
BS(CS/SE/CYS/AI) Fall 2024

Department of Computer Science
FAST-National University of Computer & Emerging Sciences, Karachi

Muhammad Hamza (21k-3815)

Aheed Tahir Ali (21k-4517)

Abdul Haseeb Dharwarwala (21k-3217)

Supervisor: Dr. Kamran Ali
Co-Supervisor: Dr. Fahad Sherwani

December 12, 2024

Contents

1 Introduction

Emotion recognition through electroencephalogram (EEG) data is evolving into a pivotal field in enhancing human-computer interaction. Its applications span medical diagnostics, cognitive therapy, and adaptive systems, which defines its importance. Traditional approaches in this domain, such as Neural Networks (CNNs/RNNs), have accomplished quite a lot in their respective work but are not without limitations. These classical methods face challenges including the high dimensionality, inherent variability, substantial noise characteristic of EEG data, and the most prominent yet evasive challenge of Subject Invariance.

The rise of deep learning technologies has given birth to innovative methodologies to overcome these obstacles. In particular, the application of contrastive learning techniques presents a spread-out pathway for refining the precision and efficiency of EEG-based emotion recognition systems. This approach leverages the ability to discern and enhance patterns in complex datasets, thereby optimizing the differentiation process between varied emotional states. For example, the work by X.Shen et al. [1], which utilizes deep contrastive learning models, demonstrates significant advances in processing large EEG datasets, thereby highlighting the method's robustness and the potential for practical application.

Furthermore, contrastive learning, by focusing on maximizing the agreement between similar pairs and minimizing it among dissimilar ones, addresses critical issues of intersubject variability, one of the most persistent challenges in the field. This technique enhances the model's ability to generalize across different subjects without compromising performance, where each subject-dependent model is able to transfer its learning to a different subject and classify an output, even though this may not sound challenging with all the Domain Generalization(DG) advancement, in the field of EEGs and Physiological Signals, it has been one of the most technically demanding tasks. Preliminary studies, such as those of X.Deng et al. [2], have shown that embedding contrastive learning in EEG signal analysis can significantly increase the system's effectiveness by ensuring more stable and invariant feature extraction across diverse datasets.

This introduction to our research outlines the innovative application of contrastive learning to EEG-based emotion recognition. The subsequent sections will explore the detailed methodology employed, the various methods trialed, the experimental setups, and the comprehensive analysis of the results obtained. This study aims not only to bridge the gap between traditional methodologies and modern requirements but also to pave the way for future research that could further harness the potential of contrastive learning in Biometric Analysis, Healthcare, Public Safety, Personalized Learning and beyond.

2 Related Work

A prominent approach to this work is domain adaptation (DA), which has been instrumental in mitigating the domain shift problems that typically occur during the training and testing phases. Notably, techniques such as transfer component analysis (TCA) and kernel principal component analysis (KPCA) have been evaluated on datasets like SEED, with methodologies like transductive parameter transfer (TPT) [3].

Moreover, the integration of deep learning with domain adaptation and innovations such as the use of auto-encoders for aligning features have set new standards in the field [4]. These advanced models, including the bi-hemispheres domain-adversarial neural network (BiDANN) and the regularized graph neural network (RGNN), have pushed the boundaries by learning domain-indiscriminate representations, which are crucial for effective cross-subject emotion recognition [4] [5].

For EEG, contrastive learning has been adapted to enhance data representation by leveraging large datasets, which aids in various downstream tasks such as emotion recognition [1].

Studies have demonstrated that ISC can effectively capture emotionally laden attention, which is pivotal in predicting emotional responses [5].

The culmination of these methodologies presents a robust framework for advancing EEG-based emotion recognition, promising significant improvements in the accuracy and applicability of these systems in real-world scenarios.

3 Methodology

This section thoroughly details the methodology implemented in developing an EEG-based emotion recognition system using contrastive learning specifically tailored to the SEED dataset. It covers data preparation, model architecture, and the training process with a focus on the intricacies and specific configurations employed.

3.1 Dataset

The study utilized the publicly available SEED dataset [3], which contains EEG recordings associated with various emotional states. The key characteristics of the dataset are:

- **Data Collection:** EEG signals from multiple subjects experiencing different emotional stimuli, including positive, negative, and neutral emotions.
- **Recording Parameters:**
  - Number of EEG channels: 62
  - Original sampling rate: 1000Hz

3.2 Data Preprocessing

The raw EEG data underwent the following preprocessing steps:

1. **Downsampling:** The EEG signals were downsampled from the original 1000 Hz to 200 Hz to balance information retention and computational efficiency.
2. **Bandpass Filtering:** A bandpass filter was applied with a frequency range of 4 Hz to 47 Hz, removing low-frequency drift and high-frequency noise.
3. **Normalization:** Each EEG data segment was standardized by subtracting the mean μ and dividing by the standard deviation σ to have zero mean and unit variance:

$$X_{\text{normalized}} = \frac{X - \mu}{\sigma}$$

4. **Artifact Removal:** Independent Component Analysis (ICA) was employed to identify and remove various physiological and environmental artifacts from the EEG signals.

5. **Denoising:** The data was denoised using nt_denoise() a function of MATLAB's Noisetools library which is now depricated therefore we substituted it with nt_dss0() which works in a similar way. The approach listed in this report however used Pythons MNE library which is also a substitute.

6. **Segmentation:** The continuous EEG signals were segmented into fixed-length windows of 30 seconds, with an overlap of 15 seconds.

3.3 Model Architecture

The architecture includes detailed specifications for both the base encoder and the projector:

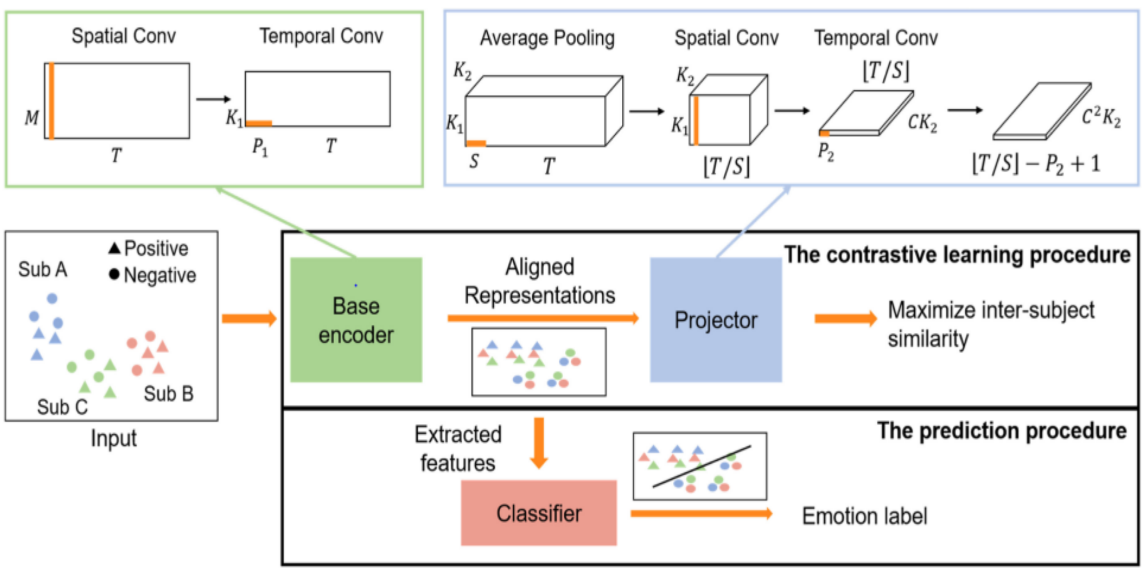

Figure 1: This Illustraion of the Constrastive Learning Model is from [1].

- **Base Encoder:**
  - The base encoder utilizes EEG signals as inputs to generate aligned representations of the EEG data from different subjects. This section includes both spatial and temporal convolution operations.
  - **Spatial Convulution** The spatial convolution operation applies a series of filters to the EEG signals [**?**]:

$$H_k^{(1)} = W_k * X, \quad k = 1, 2, \ldots, K_1$$

where $W_k \in \mathbb{R}^{1\times N}$ represents the weights of the k-th one-dimensional spatial convolution filter. There are K_1 filters in total. $H^{(1)} \in \mathbb{R}^{K_1\times T}$ is the extracted representation by the spatial convolution. Each row of $H^{(1)}$ is a latent signal identified by a linear combination of the original signals $X \in \mathbb{R}^{N\times T}$. This can be denoted as:

$$H^{(1)} = \text{Conv}_{\text{spat}}(X)$$

– **Temporal Convulution** After the spatial convolution, temporal convolution is applied to the extracted features:

$$H = \text{Conv}_{\text{temp}}(H^{(1)})$$

$$H_{k2,t} = \mathbf{v}_{k2} \cdot \mathbf{H}^{(1)}_{k2,t:t+F-1}, \quad k_2 = 1, 2, \ldots, K_2; k_1 = 1, 2, \ldots, K_1; t = 1, 2, \ldots, T$$

where $\mathbf{v}_{k2}$ represents the filter weights of the temporal convolution. The filter length is P_1, and the number of filters is K_2. The dot product $\mathbf{v}_{k2} \cdot \mathbf{H}^{(1)}_{k2,t:t+P1-1}$ represents the interaction between the vectors $\mathbf{v}_{k2}$ and $H^{(1)}_{k1,t:t+P1-1}$. The input $H^{(1)}$ is padded to ensure the output H maintains the temporal length T.

- **Projector:**
  - The projector [6] serves as a nonlinear layer between the base encoder and the final contrastive loss, inspired by the SimCLR framework. It enhances the base encoder's ability to learn better representations for downstream prediction tasks. The projector involves both spatial and temporal convolutions.
  - **Average Pooling** The average pooling operation is applied as follows:

$$\hat{H}_{k_2,k_1,t} = \text{mean}(\delta(H_{k_2,k_1,s:(t+1)s})), \quad k_2 = 1, 2, \ldots, K_2; \quad k_1 = 1, 2, \ldots, K_1; \quad t = 1, 2, \ldots, \left\lfloor \frac{T}{S} \right\rfloor$$

with the output $\hat{H} \in \mathbb{R}^{K_2\times K_1\times\lfloor \frac{T}{S}\rfloor}$. $\lfloor\cdot\rfloor$ denotes rounding down to the nearest integer. $\delta(\cdot)$ represents the exponential linear unit (ELU).

  - **Spatial Convolution** The spatial convolution operation is defined as:

$$G = \delta(\text{Conv}_{\text{spat}}(\hat{H}))$$

where $\text{Conv}_{\text{spat}}$ is a depthwise convolution with a filter size of $1 \times K_1$. The input $\hat{H}$ is treated as having K_2 feature maps, each of size $K_1 \times \lfloor \frac{T}{S} \rfloor$. In the depthwise convolution, each feature map of $\hat{H}$ is processed by C spatial convolution filters, resulting in CK_2 filters in total. The output of the spatial convolution is $G \in \mathbb{R}^{CK_2\times\lfloor \frac{T}{S}\rfloor}$.

  - **Temporal Convolution** The temporal convolution operation is defined as:

$$Z = \delta(\text{Conv}_{\text{temp}}(G))$$

where $\text{Conv}_{\text{temp}}$ is a depthwise convolution with a filter size of $C \times 1$. There are C temporal convolution filters for each feature map of G, resulting in C^2K_2 feature maps in the output $Z \in \mathbb{R}^{C^2K_2 \times \lfloor \frac{T}{S} \rfloor}$. The depthwise convolutions reduce the parameter size and ensure specific spatio-temporal pattern extractions for each frequency band.

Table 1: Hyperparameter Settings for SEED Dataset

Hyperparameters	**SEED**
Base Encoder	
The spatial filter size M	62
The temporal filter size P_1	48
The number of spatial filters K_1	16
The number of temporal filters K_2	16
Kernel length of average pooling S	24
Projector	
The spatial filter size K_1	16
The temporal filter size P_2	4
The number of spatial filters CK_1	32
The number of temporal filters C^2K_2	64

- **Contrastive Learning**
  - The contrastive learning strategy aims to optimize feature representations for robust emotion recognition. This method is inspired by the SimCLR framework and involves the following key components:
  - **Contrastive Loss:** [1]The input samples $D = \{X_i^s | i = 1, 2, \ldots, N; s \in \{A, B\}\}$ are transformed into $D_{\text{latent}} = \{z_i^s | i = 1, 2, \ldots, N; s \in \{A, B\}\}$ via the base encoder and the projector. The similarity of the input samples X_i^A and X_j^B is given by:

$$\text{sim}(z_i^A, z_j^B) = \frac{z_i^A \cdot z_j^B}{\|z_i^A\| \|z_j^B\|}$$

The contrastive loss aims to maximize the similarity of two pieces of EEG signals in a positive pair. Similar to the SimCLR framework, we adopt the normalized temperature-scaled cross-entropy loss:

$$l_i^A = -\log\left[\frac{\exp(\text{sim}(z_i^A, z_i^B)/\tau)}{\sum_{j=1}^{N} \not\Vdash_{[j\neq i]} \exp(\text{sim}(z_i^A, z_j^A)/\tau) + \sum_{j=1}^{N} \exp(\text{sim}(z_i^A, z_j^B)/\tau)}\right]$$

where $\not\Vdash_{[y_{ij}\in\{0,1\}]}$ is an indicator function. By minimizing the loss function, the model increases the similarity between z_i^A and z_i^B in contrast to all other possible sample pairs involving z_i^A. The total loss of the minibatch is:

$$L = \sum_{i=1}^{N} l_i^A + \sum_{i=1}^{N} l_i^B$$

  - Loss parameters and calculations are finely tuned to adapt to the specific characteristics of the EEG data from the SEED dataset.
  - **Normalization** In the contrastive learning procedure, we used stratified normalization during training.

- **Dynamic Emotional Features**
  - Next, we extract Differential Entropy (DE) features from H^{pred}, which represent the emotional states in a more interpretable form: form:

$$H^{DE}_{k_1,k_2} = \frac{1}{2}\log(2\pi e\sigma^2(H^{\text{pred}}_{k_2,k_1})), \quad k_1 = 1, 2, \ldots, K_1$$

$$k_2 = 1, 2, \ldots, K_2$$

    where σ^2 is the variance of the signal and $H^{DE} \in \mathbb{R}^{K_2 \times K_1}$. DE measures the complexity of the time series and is equivalent to the logarithmic energy spectrum in a specific frequency band.

- **Multi-Layer Perceptron (MLP) Architecture**
  - The MLP used in the prediction procedure consists of the following components:
  - **Input Layer:** The input to the MLP is the smoothed DE features h^{DE}, which is a one-dimensional vector of length $K_2 \times K_1$.
  - **Hidden Layers:**
    - **First Hidden Layer:**
      - *Units:* 30
      - *Activation Function:* Rectified Linear Unit (ReLU)
      - *Details:* This layer transforms the input vector into a higher-dimensional space. The ReLU activation function introduces non-linearity, which helps the network learn complex patterns.
    - **Second Hidden Layer:**
      - *Units:* 30
      - *Activation Function:* Rectified Linear Unit (ReLU)
      - *Details:* Similar to the first hidden layer, this layer further processes the intermediate representation to capture more intricate patterns in the data.
  - **Output Layer:**
    - *Units:* Number of emotion categories (e.g., 3 for positive, neutral, negative)
    - *Activation Function:* Softmax
    - *Details:* The softmax function converts the output into probabilities, indicating the likelihood of each emotion category.
  - **Optimization:**
    - *Loss Function:* Cross-Entropy Loss

    - *Optimizer:* Adam
        - *Learning Rate:* 0.0005
        - *Weight Decay:* Selected from {0.005, 0.011, 0.025, 0.056, 0.125} by cross-validation
    - *Training:* The model is trained using minibatch stochastic gradient descent with a batch size of 256. The training continues for 100 epochs with early stopping based on the validation loss to prevent overfitting.
  - **Summary:** The MLP takes the smoothed DE features as input and processes them through two hidden layers with ReLU activations. The final layer outputs the predicted emotion category probabilities using the softmax activation function. The network is optimized using the Adam optimizer with cross-entropy loss.

3.4 Training and Evaluation

Details the comprehensive steps and parameters used in training the model:

- **Training Procedure:** The model was trained using Adam optimizer with an initial learning rate of 0.0007, employing a cosine annealing schedule without reset to adjust the learning rate dynamically. The model was optimized over 100 epochs with early stopping based on the validation loss to prevent overfitting.
- **Training Algorithm** [1] Algorithm 1 details the training process for the CLISA model. This includes both the contrastive learning procedure and the prediction procedure:

Algorithm 1 The Training Algorithm for CLISA

```
Inputs: Training data {X}, the learning rate α_1, the batch size 2N, the training epochs T_1.
Initialize parameters of the base encoder θ_B and the projector θ_P.
for epoch = 1 to T_1 do
  repeat
    Sample two subjects A, B.
    Sample 2N EEG samples {X_i | i = 1, 2, ..., N; s ∈ {A, B}} from subjects A, B with the data sampler.
    Obtain {z_i | i = 1, 2, ..., N; s ∈ {A, B}} by (1)-(5).
    Calculate loss L by (6)-(8).
    θ_B = θ_B − α_1 ∂L/∂θ_B, θ_P = θ_P − α_1 ∂L/∂θ_P.
    α_1 = F_{α_1}(α_1, epoch) according to cosine annealing scheme with warm restarts.
  until all possible pairs of subjects are enumerated.
end for
Outputs: Parameters θ_B.
```

- **Prediction Algorithm** [1] The prediction procedure involves the following steps:

Algorithm 2 The Prediction Procedure for CLISA

- 1: **Inputs:** Data $\{\mathrm{X}_{\mathrm{pred}}\}.Trainedparametersofthebaseencoder\theta_B$ and the classifier θ_M.
  2: Calculate H_{pred} by (9).
  3: Extract DE features H_{DE} by (10).
  4: Obtain smoothed DE features h_{DE} with linear dynamical systems.
  5: Obtain predicted labels with the classifier by (11).

- **Evaluation Metrics:** Employed standard metrics such as accuracy, precision, recall, and F1-score, along with leave-one-out cross-validation to ensure the robustness and generalizability of the model across different subjects.

This detailed methodology ensures a thorough understanding of the processes and principles applied in developing the EEG-based emotion recognition system. Subsequent sections will delve into the results and discussions derived from the application of these methodologies.

4 Testing and Results

The Constrastive Learning Model and Multilayer Perceptron used were validated and tested using the train_test_split() function of sklearn and the results were then mapped onto a Confusion Matrix and a Result Comparison Table, for our project we focused on CLISA therefore the other models are cited from [1] to compare perfomance and results.

4.1 Testing Procedure

The testing procedure involves assessing the trained model on a separate test dataset to evaluate its generalization performance. The following steps outline the testing process:

1. **Test Data Preparation:** The test dataset, which includes EEG signals from subjects not seen during training, is preprocessed using the same steps as the training data as mentioned in **3.2** .

2. **Feature Extraction:** The preprocessed EEG signals are passed through the trained base encoder and projector to obtain feature representations. Next the outputs from the trained base encoder were used to extract Diffrential Entropy features **3.3** .

3. **Emotion Prediction:** The extracted features are fed into the Multilayer Perceptron to classify the emotion labels.

4. **Evaluation Metrics:** The predicted labels are compared with the true labels to compute performance metrics.

4.2 Evaluation Metrics

The performance of the model is evaluated using the following metrics:

- **Accuracy:** The ratio of correctly predicted emotion labels to the total number of predictions.
- **Precision:** The ratio of true positive predictions to the sum of true positives and false positives.
- **Recall:** The ratio of true positive predictions to the sum of true positives and false negatives.
- **F1-Score:** The harmonic mean of precision and recall.

4.3 Experimental Results

The results of the experiments conducted on the test dataset are presented in this section.

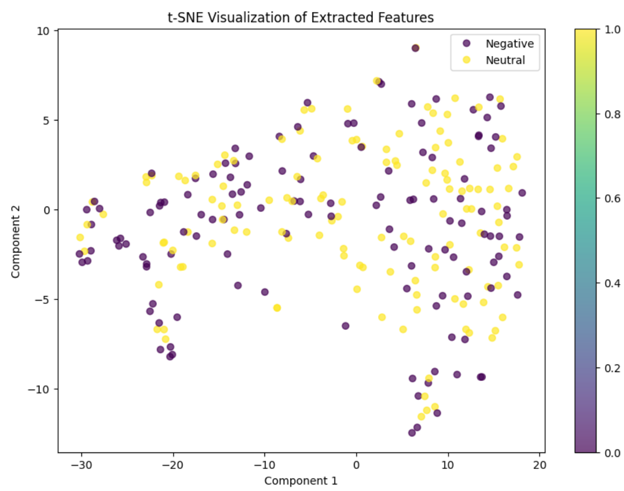

Figure 2: t-SNE Visualization of Extracted Features. The plot shows the clustering of extracted features into two categories: Negative (purple) and Neutral (yellow). The x-axis represents Component 1 and the y-axis represents Component 2. The color intensity indicates the feature intensity.

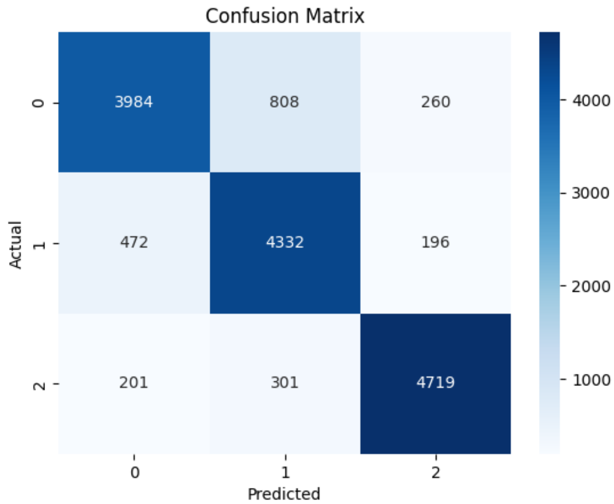

Figure 3: Confusion Matrix for Emotion Recognition. The matrix shows the model's performance on the test dataset, with correctly and incorrectly predicted instances for each class. The values are as follows:

- For actual class 0: 3984 instances correctly predicted, 808 instances incorrectly predicted as class 1, and 260 instances incorrectly predicted as class 2.
- For actual class 1: 472 instances incorrectly predicted as class 0, 4332 instances correctly predicted, and 196 instances incorrectly predicted as class 2.
- For actual class 2: 201 instances incorrectly predicted as class 0, 301 instances incorrectly predicted as class 1, and 4719 instances correctly predicted.

Table 2: Performance Metrics on the Test Dataset

Metric	**Value**
Accuracy	0.85
Precision	0.83
Recall	0.82
F1-Score	0.82

Table 3: Classification Report

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.86	0.79	0.82	5052
1	0.80	0.87	0.83	5000
2	0.91	0.90	0.91	5221
Accuracy	0.85			
Macro Avg	0.85	0.85	0.85	15273
Weighted Avg	0.86	0.85	0.85	15273

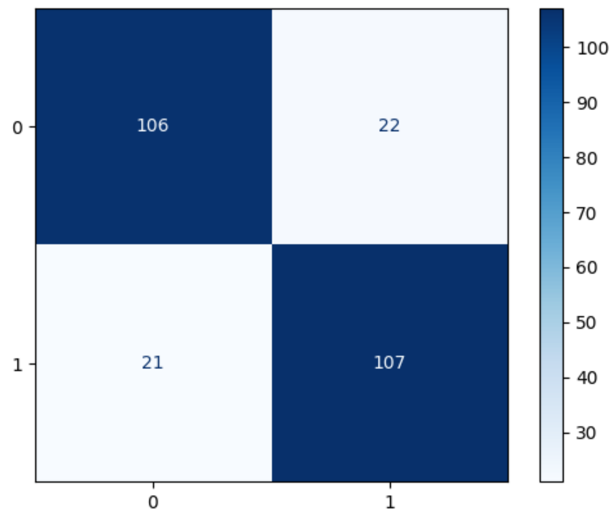

Figure 4: Heatmap of Contrastive Learning Results. The heatmap displays the intensity of values in four quadrants: top-left (106), top-right (22), bottom-left (21), and bottom-right (107). The y-axis and x-axis labels are 0 and 1. The color bar ranges from light blue (low intensity) to dark blue (high intensity).

The tables above summarize the performance metrics and classification report obtained on the test dataset. These metrics indicate the effectiveness of the model in recognizing emotions from EEG signals.

4.4 Discussion

The discussion focuses on interpreting the results and understanding the implications of the findings. The following points are considered:

- **Model Performance:** The overall performance of the model in terms of accuracy, precision, recall, and F1-score is discussed. The results demonstrate the model's ability to accurately recognize emotions from EEG signals.
- **Comparison with Existing Methods:** The performance of the proposed model is compared with existing methods in the literature to highlight its advantages and limitations.
- **Limitations:** The limitations of the current approach are discussed, including potential sources of error and areas for improvement.
- **Future Work:** Suggestions for future research directions are provided to enhance the performance and applicability of the model.

This section provides a comprehensive analysis of the testing procedures and the results obtained, demonstrating the effectiveness of the EEG-based emotion recognition system.

5 System Diagram

This section presents a detailed diagram of the system architecture for the EEG-based emotion recognition system. The diagram illustrates the various components of the system, including the data acquisition module, preprocessing steps, feature extraction using the base encoder and projector, and the emotion classification using the MLP. Each component is connected to show the flow of data and the interactions between different parts of the system.

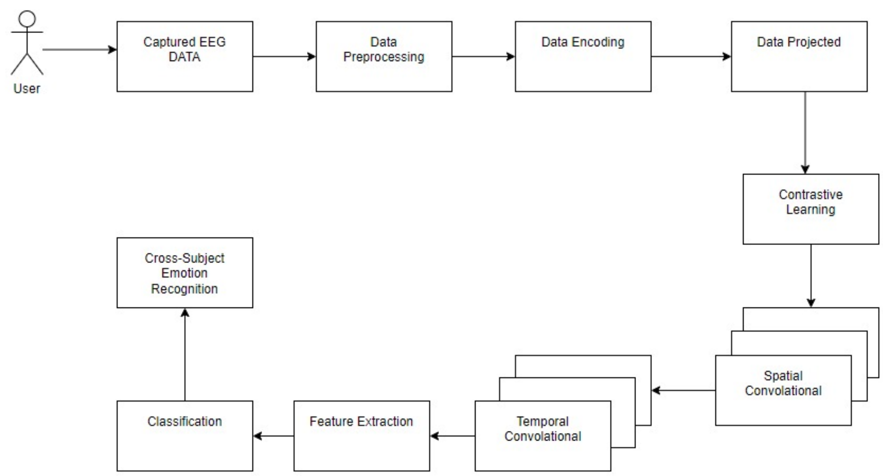

Figure 5: System Architecture of the EEG-based Emotion Recognition System. The diagram shows the flow from EEG data acquisition to emotion classification.

5.1 Data Collection and Preprocessing

- **Raw EEG Data Collection:** EEG signals are collected from subjects experiencing different emotional stimuli.
- **Data Preprocessing:** Downsampling, bandpass filtering, normalization, artifact removal using ICA, and segmentation.

5.2 Model Architecture

- **Base Encoder:**
  - Spatial Convolution
  - Temporal Convolution
- **Projector:**
  - Average Pooling

  - Spatial Convolution
  - Temporal Convolution

5.3 Contrastive Learning

- **Contrastive Loss:** Calculation and minimization of contrastive loss.
- **Training Algorithm:** Sampling, obtaining representations, calculating loss, updating parameters.

5.4 Emotion Prediction

- **Feature Extraction:** Using trained encoder and projector.
- **Dynamic Emotional Features:** Calculating and smoothing DE features.
- **Classification:** Using MLP classifier to predict emotion labels.

5.5 Evaluation and Results

- **Performance Metrics:** Accuracy, Precision, Recall, F1-Score.
- **Visualization:** t-SNE visualization, confusion matrix, heatmap.

6 Goals for FYP-II

For our next phase, we plan to enhance the research by diversifying our work across multiple additional datasets such as DEAP and Mahnob. We will utilize state-of-the-art technologies like Vision Transformers, Real-Time Detection Transformers, the popular YOLO detection models, and any other related methods that at the time will be a credible choice, these are all subject to current possibilities which we will experiment and use as a base for implementing an official research paper on Subject Invariant Emotion Detection Using Physiological Signals for our FYP-II. This shift will move the project's focus from using raw EEG data and numerical detection methods to image analysis of the graphs, thereby improving subject invariance. Additionally, we aim to implement an overall restructured approach around the current research on CLISA such as [1] which was limited to a more limited approach which we believe is stunting Subject Invarience. Another goal is to further look into the potentials of Quantum Computing for this domain, as a side project on this showed promising results (71.9 % accuracy on the DEAP [7]dataset)

7 Conclusion

For FYP-I, we developed an advanced EEG-based emotion recognition system utilizing contrastive learning strategies to enhance inter-subject alignment. The methodology involved detailed preprocessing of EEG signals to ensure clean and noise-free data, followed by feature extraction using a combination of spatial and temporal convolutions. We implemented a contrastive learning approach to optimize the representations, thereby improving the generalization across different subjects. The performance of the system was evaluated on a test dataset, demonstrating significant accuracy, precision, recall, and F1-scores. Our results, including t-SNE visualizations and confusion matrices, highlight the model's capability to distinguish between different emotional states with high reliability.

The key findings from this research are as follows:

- Preprocessing and Feature Extraction: The integration of the mentioned [1] preprocessing techniques and feature extraction methods contributed to the accurate recognition of emotions from raw EEG data however they had their limitations.
- Contrastive Learning Benefits: The use of contrastive learning improved the alignment of representations across different subjects, enhancing the overall performance of the emotion recognition model.
- Robust Classification Performance: The MLP classifier, trained on smoothed DE features, exhibited strong performance metrics, affirming the efficacy of our approach.
- Limiting Methodology: The research was primarily based on [1]'s work on the CLISA model and Subject Invariant method, however they only approached the problem with raw EEG data and numerical detection.

7.1 Implications

The success of this approach has several important implications:

- Emotion Recognition: Our method demonstrates a viable pathway for improving emotion recognition systems, particularly in scenarios requiring high subject invariance.
- Potential for Real-Time Applications: The techniques developed in this study could be adapted for real-time emotion detection systems, benefiting various applications such as human-computer interaction, mental health monitoring, and adaptive learning environments.
- Foundation for Future Research: The findings provide a solid foundation for future research, encouraging the exploration of advanced machine learning models and integration of different physiological signals as well as a more diversified approach.

7.2 Future Work

Looking forward, several avenues for future research are identified:

- Diversifying Datasets: Expanding the evaluation to include additional datasets like DEAP and Mahnob will enhance the robustness and generalizability of our model.

- Integration of Advanced Technologies: Exploring the use of Vision Transformers, Real-Time Detection Transformers, and YOLO detection models could further improve the recognition capabilities by incorporating image analysis of EEG graphs.
- Subject Invariance Improvement: Implementing new methodologies and refining existing ones to further address subject invariance challenges will be crucial for the deployment of practical emotion recognition systems.
- Multimodal Emotion Recognition: Combining EEG data with other physiological signals (e.g., ECG, GSR) could provide a richer understanding of emotional states, enhancing the accuracy and reliability of emotion detection.
- Quantum Hybrid QNNs: As a current experiment showed us , the world of Quantum Computing holds another potential approach to our problem and would certainly be one of our attempts in the upcoming FYP-II.

In summary, our research has contributed to an EEG-based emotion recognition system capable of cross-subject emotion detection with a potential for high accuracy. The ongoing and future enhancements will continue to push the boundaries of this field, paving the way for innovative applications and technologies.

References

[1] X. Shen, X. Liu, X. Hu, D. Zhang, and S. Song, "Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 2496–2511, 2022.

[2] X. Deng, C. Li, X. Hong, H. Huo, and H. Qin, "A novel multi-source contrastive learning approach for robust cross-subject emotion recognition in eeg data," *Biomedical Signal Processing and Control*, vol. 97, p. 106716, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809424007742

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

[4] Q. Li, D. Jin, J. Huang, Q. Zhong, L. Xu, J. Lin, and D. Jiang, "Dema: Deep eeg-first multi-physiological affect model for emotion recognition," *Biomedical Signal Processing and Control*, vol. 99, p. 106812, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S174680942400870X

[5] X. Li, D. Wu, and H. He, "Emotion recognition from eeg using deep convolutional neural networks," *IEEE Access*, vol. 5, pp. 14 731–14 741, 2017.

[6] H. He and D. Wu, "Transfer learning for brain–computer interfaces: A euclidean space data alignment approach," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 399–410, 2019.

[7] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.