Decoding Emotions: An In-Depth Exploration of Sentiment Detection through EEG Signal Analysis

Elaheh Toulabinejad, Kiana Aghakasiri
Department of Computing Science
University of Alberta
Edmonton, Alberta
toulabin@ualberta.ca , kaghakas@ualberta.ca

Abstract

Emotions are a fundamental aspect of human life that influence our daily experiences. Extensive research has been conducted to recognize human emotions. Within this field, AI tools have been developed to enhance our ability to recognize emotions. Previous research has employed AI techniques to analyze EEG signals, which represent brain activity, for emotion recognition. In this study, our primary objective is to utilize these AI methods, including machine learning and deep learning, to identify emotions from EEG signals. Additionally, we aim to investigate how the selection of specific EEG signal features affects the performance of our models. Furthermore, we seek to explore neural activity within specific brain regions among individuals as they experience emotions, providing valuable insights into the underlying cognitive processes.

1 Introduction and Related Works

Emotions play an important role in influencing human actions, decisions, and communications [1]. There is a need for emotion recognition in fields such as medicine, psychology, and human-computer interaction. As technology has advanced, various tools and methods have been developed to detect emotions. Previous works have attempted to detect emotions using behavior, speech, and facial expressions. As some studies suggest, when processing different emotional stimuli, patterns of activation are observed in certain areas of the brain [2]. Therefore, methods to detect these psychological signals have been at the center of attention [3].

The Electroencephalogram (EEG) is a non-invasive imaging technique that scans the electrical activity of the brain. When EEG signals are recorded from the scalp, they are referred to as electroencephalograms. If recorded from the cortical surface of the brain, they are called electrocardiograms [4]. These EEG signals reflect the neural activity in the brain, particularly during different emotional states, providing an unbiased means of data collection that is accessible at all times[5]. They offer valuable insights into how the brain functions, making them a crucial tool in fields such as psychology, medicine, and human-computer interaction, neuroscience.

Machine learning (ML) and deep learning (DL) models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Graph CNNs, are common tools used in emotion extraction from EEG signals [5, 6, 7, 8]. In these methods, after pre-processing the data and extracting and selecting the features, the recognition of emotions is done with the help of classification models. [5].

In [9], the authors conducted an analysis of the preprocessed SEED-IV and DEAP data using a range of features in both the frequency and time domains. They used a Support Vector Machine (SVM) as a supervised machine learning algorithm to categorize emotional states into four distinct classes. Deep learning frameworks have also been used to classify EEG signals to recognize self-induced emotions

[10, 11]. For instance, in [10], the authors demonstrated that by choosing a channel selection method based on the statistical information of the raw EEG signal, computational complexity can be reduced without compromising classification accuracy. In [11], the authors employed a low-complexity DL method to recognize emotions through EEG analysis of the cerebral cortex's electrical responses. What distinguishes this work from previous studies is its feature improvement method aimed at enhancing the learning process during classification tasks. Through the utilization of multiclass genetic programming with multidimensional populations (M3GP), this paper endeavors to construct the most suitable features by selecting, combining, and eliminating the initial features. In another study [12], the authors recorded EEG signals in real-time while subjects watched videos and extracted four types of features from this data. They utilized PCA to select important features from the dataset and subsequently employed three deep learning classifiers for emotion recognition: the Gated Recurrent Unit (GRU), CNN, and Deep Emotion Recognizer (DER). In [13], the authors employed Fast Fourier Transformation to extract frequency domain features, convolutional layers for deep features, and complementary features from one of the most popular EEG benchmark datasets, DEAP. They also introduced a CNN model called M1, which is a heavily parameterized model for binary classification and outperforms all previous state-of-the-art models. Subsequently, they proposed a lightly parameterized CNN model, M2, which achieved accuracy close to that of M1. Some studies have directed their attention toward analyzing group EEGs rather than individual EEGs to identify human group emotional states [14].

Furthermore, two novel graph models have been introduced showing promising results in recognizing emotions across different subjects [7, 15]. In [7], the authors introduced an advanced model called the self-organized graph neural network (SOGNN) for cross-subject EEG emotion recognition. They evaluated the model on SEED and SEED-IV datasets and achieved cross-subject accuracy of 86.81% on the SEED dataset and 75.27% on the SEED-IV dataset. In [15], the authors proposed a regularized graph neural network (RGNN) for EEG-based on emotion recognition which has the basis of two models of the simple graph convolution network (SGC) [16] and spectral graph convolution. They presented two regularizers, namely NodeDAT and EmotionDL, to improve the robustness of the model against cross-subject EEG variations and noisy labels.

The primary focus of this study is to employ machine learning techniques for decoding tasks in emotion recognition through EEG recordings. Additionally, we aim to investigate the role of features and frequency bands in enhancing this process.

2 Data

SEED dataset was gathered with ESI NeuroScan System. Fifteen Chinese subjects (7 males and 8 females; MEAN: 23.27, STD: 2.37) with self-reported normal or corrected-to-normal vision and normal hearing took part in the experiments. 15 emotional video clips from Chinese films were chosen to elicit each of three emotions: positive, neutral, and negative, and each of them lasted for about 4 min [17]. These videos are selected with three features:

- Being understood without explanation.
- Not being too long to cause fatigue.
- Eliciting only one target emotion.

There are 15 trials in each experiment. In each experiment, there were 5 seconds for a hint before showing the video, 45 seconds for self-assessment, and 15 seconds for rest after videos in each session [17]. The self-assessment component follows Philippot's guidelines [18], involving three key inquiries: 1) their actual emotional response to the film clip; 2) whether they had previously viewed the clip; 3) Whether they have understood the clip. The SEED dataset is organized into two main folders: "Preprocessed_Data" and "Extracted_Features."

2.1 Preprocessed_Data

The raw data was downsampled to a 200 Hz sampling rate. Subsequently, a bandpass filter in the frequency range of 0.3 to 50 Hz was applied to filter out noise and remove artifacts. Following these filtering steps, EEG segments corresponding to the duration of each movie were extracted and were given in Preprocessed data [17].

The dataset comprises 45 individual MATLAB (.mat) files, each corresponding to a unique experimental session. Participants engaged in the experiment three times, with an interval of approximately one week between sessions, ensuring diverse and longitudinally captured responses. Within each subject's file, a structured set of 16 arrays is present, encapsulating segmented and preprocessed EEG data for the 15 trials in a given experiment and a key array for labels categorizing emotional states: -1 for negative, 0 for neutral, and $+1$ for positive emotional states.

2.2 Extracted_Features

To facilitate further analysis, each channel of the EEG data was then divided into non-overlapping epochs, each lasting 1 second. This segmentation allowed for a more detailed examination of temporal patterns. Within the extracted features form, we had precomputed features, including the widely-used differential entropy (DE) features firstly introduced to EEG-based emotion recognition, proposed in [19] which extends the concept of Shannon entropy to measure the complexity of continuous random variables. DE exhibits a balance in discriminating EEG patterns between low and high-frequency energy, making it particularly suitable for EEG-based emotion recognition. The formula for DE calculation is defined as:

$$h(X) = -\int f(x)\log_2(f(x))\,dx$$

Additionally, considering the effectiveness of asymmetrical brain activity in emotion processing, differential asymmetry (DASM) and rational asymmetry (RASM) features were computed as the differences and ratios between DE features of 27 pairs of hemispheric asymmetry electrodes. The electrode pairs are given in the Table 1. RASM and DASM are defined as follows[17]

$$DASM = DE(X_{left}) - DE(X_{right})$$

$$RASM = DE(X_{\text{left}})/DE(X_{\text{right}})$$

Pair No.	1	2	3	4	5	6	7	8	9
Left	FP1	F7	F3	FT7	FC3	T7	P7	C3	TP7
Right	FP2	F8	F4	FT8	FC4	T8	P8	C4	TP8
Pair No.	10	11	12	13	14	15	16	17	18
Left	CP3	P3	O1	AF3	F5	F7	FC5	FC1	C5
Right	CP4	P4	O2	AF4	F6	F8	FC6	FC2	C6
Pair No.	19	20	21	22	23	24	25	26	27
Left	C1	CP5	CP1	P5	P1	PO7	PO5	PO3	CB1
Right	C2	CP6	CP2	P6	P2	PO8	PO6	PO4	CB2

Table 1: 27 Pairs of hemispheric asymmetry electrodes [19]

Furthermore, Differential caudality (DCAU) features as the differences between DE features of 23 pairs of frontal-posterior electrodes were extracted. The 23 electrode pairs are given in the Table 2. DCAU is defined as

$$DCAU = DE(X_{\text{frontal}})/DE(X_{\text{posterior}})$$

Moreover, power spectral density (PSD), defined as the distribution of signal power over frequency, was extracted as baseline for comparison. PSD is a common signal processing technique that distributes the signal power over frequency and depicts the quantification of EEG signals.

3 Methods

Our main goal is to identify patterns and connections related to emotions in the brain signals. In order to do so, we use preprocessed data and extracted features to categorize EEG signals according to the associated emotions.

Pair No.	1	2	3	4	5	6	7	8
Frontal	FT7	FC5	FC3	FC1	FCZ	FC2	FC4	FC6
Posterior	TP7	CP5	CP3	CP1	CPZ	CP2	CP4	CP6
Pair No.	9	10	11	12	13	14	15	16
Frontal	FT8	F7	F5	F3	F1	FZ	F2	F4
Posterior	TP8	P7	P5	P3	P1	PZ	P2	P4
Pair No.	17	18	19	20	21	22	23	
Frontal	F6	F8	FP1	FP2	FPZ	AF3	AF4	
Posterior	P6	P8	O1	O2	OZ	CB1	CB2	

Table 2: 23 Pairs of hemispheric asymmetry electrodes [19]

Classification Algorithms In this study, we employ three classifiers: Support Vector Machine (SVM), Logistic Regression (LR), and k-Nearest Neighbors (KNN). The choice of these classifiers is due to their widespread use in similar studies, where they have demonstrated effectiveness in classifying EEG signals [20, 15, 21, 22]. Moreover, we selected these classifiers for their distinct strengths: SVM excels with complex relationships, LR maintains simplicity, and KNN is adept at identifying patterns in diverse EEG data.

Evaluation To assess the performance of our classification models, we use accuracy. Model accuracy refers to the model's capability of correctly classifying instances in the test dataset. Accuracy is calculated as follows:

$$Accuracy = \frac{NumberofCorrectPredictions}{TotalNumberofPredictions}$$

where *Number of Correct Predictions* represents the instances in the test set that the model accurately classified, while the *Total Number of Predictions* denotes the overall number of instances in the test set.

To evaluate our classifiers, we implement a 5-fold cross-validation setup. This method establishes a reliable and robust validation mechanism, ensuring the generalizability of our results. During this process, the data is divided into five folds, and each classifier is tested against unique combinations of training and validation sets. While these methods generally apply to all our experiments, minor variations in how we select test and train data may occur. We will provide specific explanations for such changes in each experiment if they arise. We used the default settings for all our classifiers and did not perform any hyperparameter tuning.

4 Results

Our experiment was organized around proposing questions and subsequently finding corresponding responses. This section guides you through this structured process.[1]

4.1 Can emotional information be extracted from raw EEG data?

Method In the first experiment, we investigated the feasibility of extracting emotional information from preprocessed data. The preprocessed dataset is substantial, consisting of 62 channels with approximately 47,000 data points for each channel, posing a considerable challenge due to its large volume. Given our limited modeling resources, we opted to limit the data used. To achieve this, we concentrated solely on the final 20 seconds of each trial for the initial experiment of the first subject during the classification process. We shuffled this dataset and partitioned 80% of the data for training, reserving the remaining 20% for testing. Furthermore, to enhance the reliability of our results, we employed a 5-fold cross-validation setup.

Results In Figure 1, we showcase the classification performance of three classifiers on a specific segment of the preprocessed dataset. Notably, SVM and KNN exhibit accuracies of 28.78% and

[1]The source code is available at: `https://github.com/ellietoulabi/eeg_emotion_detection`

34.81%, respectively, which are close to random assignment given our dataset's three labels. In contrast, LR outperforms with an accuracy of 54.07%.

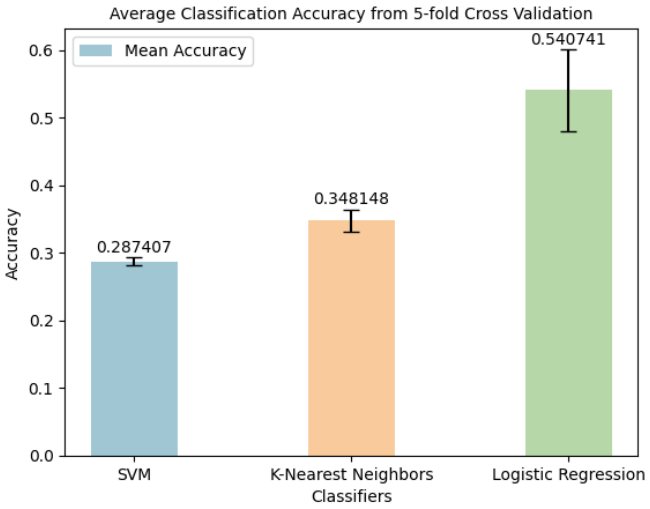


Figure 1: Classification accuracy of preprocessed EEG data.

4.2 Can emotional information be derived from features extracted from EEG data?

Method Considering the size of the original dataset, we chose the features dataset over the preprocessed data for further experiments. Our experimental design to classify features included two distinct approaches: the first involved utilizing all features for classification, while the second focused on classification based on a single feature at a time. We shuffled the data and dedicated 80% for training our model and 20% for testing. To strengthen the robustness of our results, we tested our model in a 5-fold cross-validation setup.

Results The average classification accuracy using all features is 29%, 46%, and 37% for SVM, KNN, and LR, respectively. Notably, the accuracy for SVM and LR is comparable to a scenario where labels are chosen by chance, while KNN outperforms them (Figure 2 (a)). Figure 2 (b) illustrates that each feature can achieve a different accuracy based on the emotion-related information it encompasses. However, the overall trend suggests that using a single feature is likely to improve accuracy compared to employing all features (See Table 3).

	asm_LDS	asm_movingAve	dasm_LDS	dasm_movingAve	dcau_LDS	dcau_movingAve
SVM	0.48	0.48	0.35	0.36	0.66	0.66
KNN	0.38	0.39	0.4	0.36	0.39	0.4
LR	0.71	0.75	0.59	0.59	0.82	0.86
	de_LDS	de_movingAve	psd_LDS	psd_movingAve	rasm_LDS	rasm_movingAve
SVM	0.74	0.74	0.28	0.29	0.73	0.73
KNN	0.77	0.79	0.47	0.45	0.67	0.68
LR	0.85	0.88	0.38	0.36	0.68	0.74

Table 3: Average classification accuracy over each feature with shuffled data in train/test sets.

4.3 How does the presence of overlapping videos in both the training and testing data impact the classification of emotions?

Method We hypothesize that including overlapping videos in both the training and test datasets might introduce bias to the model. This bias could lead the model to capture patterns related to the unique characteristics of the videos and use them in the classification rather than recognizing emotion-related features. Such a scenario might result in an overly optimistic evaluation of the

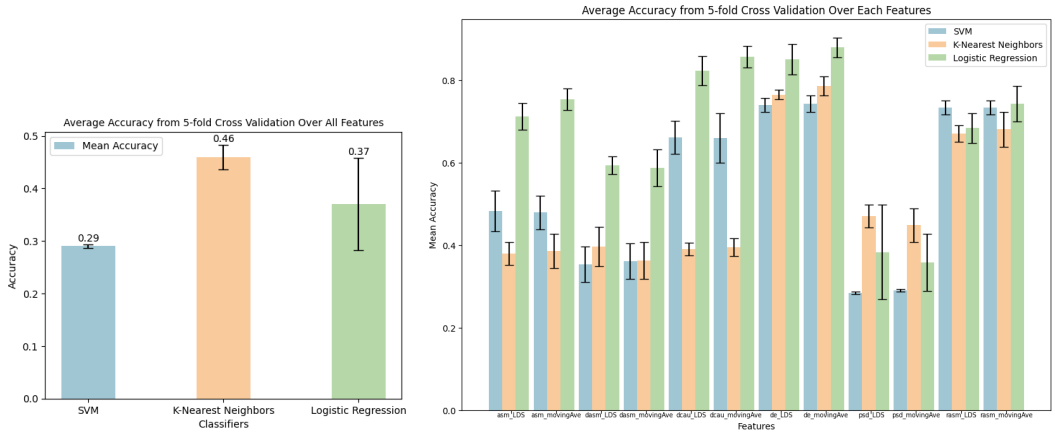


Figure 2: Classification accuracy of extracted features dataset. (a) Classification accuracy using all features for data classification. (b) Classification accuracy using only one feature for data classification. Each set of three bars adjacent to each other corresponds to using one feature across all three classifiers.

model's performance. To assess this hypothesis, we trained a model while ensuring independence between the training and test sets. For this purpose, 3 videos from all experiments of all subjects were assigned to the test set, while the other 12 videos were assigned to the training set. Given that consecutive videos in the dataset do not share the same label, this selection ensures that these videos cover all available labels. Moreover, this distribution maintains an 80/20 ratio between the training and test sets. In each fold of the 5-fold cross-validation setup, the classifier is tested on three videos and trained on the remaining 12 in the first fold, followed by testing on the next three and training on the other 12 videos in the subsequent folds. This experiment also employed two distinct approaches: the first involved including all features for classification, while the second focused on classifying based on a single feature at a time.

Results As shown in Figure 3, the average accuracy of utilizing all features in this scenario is 6% and 4% lower than when having the same videos in both the test and train sets for KNN and LR, respectively. However, this accuracy increased by 4% in the case of SVM. Additionally, in 28 out of 36 cases, the accuracy of classification using a single feature demonstrates a decrease (Table 4).

	asm_LDS	asm_movingAve	dasm_LDS	dasm_movingAve	dcau_LDS	dcau_movingAve
SVM	0.27	0.25	0.33	0.33	0.44	0.43
KNN	**0.38**	0.36	**0.47**	**0.41**	**0.45**	**0.42**
LR	0.32	0.32	0.38	0.36	0.45	0.47
	de_LDS	de_movingAve	psd_LDS	psd_movingAve	rasm_LDS	rasm_movingAve
SVM	0.31	0.32	**0.33**	**0.33**	0.31	0.31
KNN	0.29	0.28	0.4	0.38	0.27	0.26
LR	0.31	0.29	0.33	**0.33**	0.25	0.25

Table 4: Average classification accuracy over each feature with independent train/test videos. Numbers highlighted in bold indicate increased accuracy compared to having the same videos in both test and train sets.

4.4 How does the presence of overlapping subjects in both the training and testing data impact the classification of emotions?

Method We hypothesized that the inclusion of subjects in both the training and test data could introduce bias to the model, causing it to capture patterns related to the unique characteristics of the subject and include them in the classification rather than recognizing emotion-related features. This bias could result in an overly optimistic evaluation of the model's performance. To test this

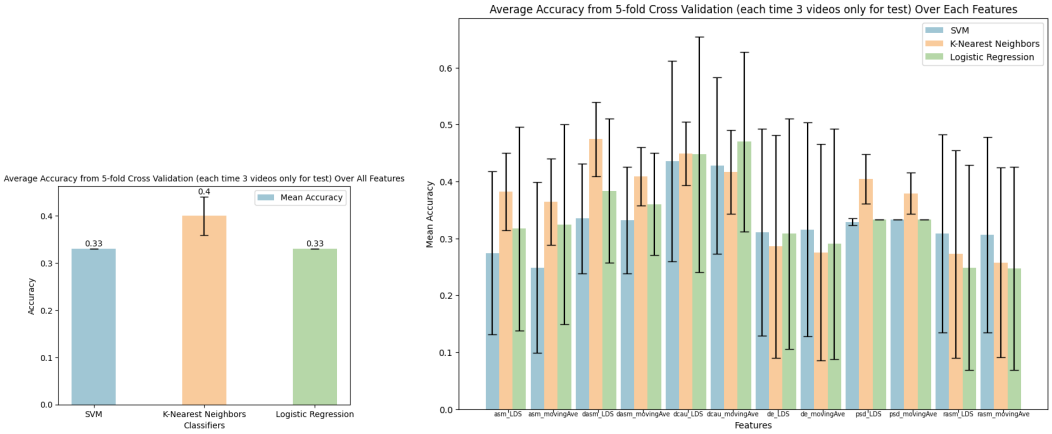


Figure 3: Classification accuracy of extracted features dataset using independent videos for training and test. (a) Classification accuracy using all features for data classification. (b) Classification accuracy using only one feature for data classification. Each set of three bars adjacent to each other corresponds to using one feature across all three classifiers.

hypothesis, we trained a model by ensuring independence between the training and test sets. In this experiment, we chose 12 subjects for training and 3 subjects for testing, maintaining an 80/20 ratio between the training and test sets. Two distinct approaches were employed for this experiment: the first included all features for classification, while the second focused on classifying based on a single feature at a time. The evaluation was performed using a 5-fold cross-validation setup. In initial fold of this cross-validation setup, the classifier was tested on three subjects and trained on the remaining 12 during the first fold. Subsequently, in the following folds, the process involved testing on the next three subjects and training on the remaining 12.

Results As shown in Figure 4, the average accuracy of utilizing all features in this scenario is 4% and 8% higher than when having the same subjects in both the test and train sets for SVM and LR, respectively. However, this accuracy decreases by 1% in the case of KNN. Additionally, in 22 out of 36 cases, the accuracy of classification using a single feature demonstrates an increase (Table 5).

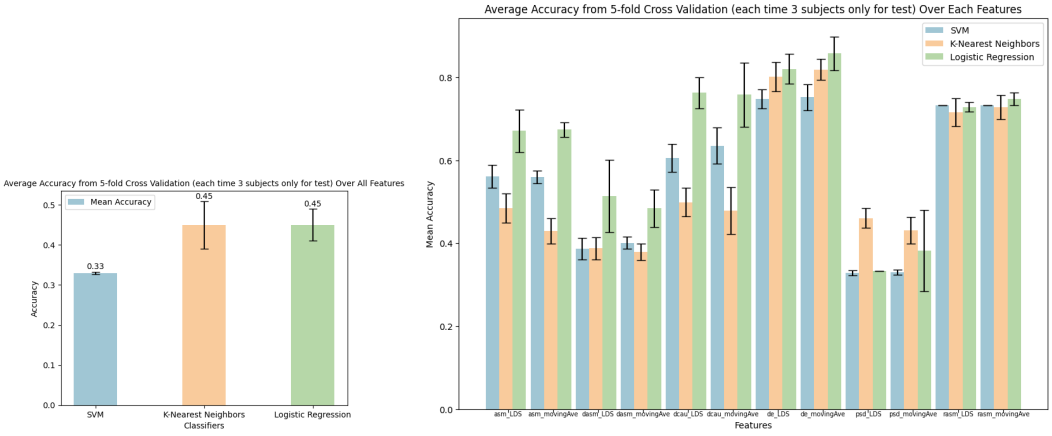


Figure 4: Classification accuracy of extracted features dataset using independent subjects for training and test. (a) Classification accuracy using all features for data classification. (b) Classification accuracy using only one feature for data classification. Each set of three bars adjacent to each other corresponds to using one feature across all three classifiers.

	asm_LDS	asm_movingAve	dasm_LDS	dasm_movingAve	dcau_LDS	dcau_movingAve
SVM	**0.56**	**0.56**	**0.39**	**0.4**	0.61	0.64
KNN	**0.48**	**0.43**	0.39	**0.38**	**0.5**	**0.48**
LR	0.67	0.67	0.51	0.48	0.76	0.76
	de_LDS	de_movingAve	psd_LDS	psd_movingAve	rasm_LDS	rasm_movingAve
SVM	**0.75**	**0.75**	**0.33**	**0.33**	**0.73**	**0.73**
KNN	**0.8**	**0.82**	0.46	0.43	**0.72**	**0.73**
LR	0.82	0.86	0.33	**0.38**	**0.73**	**0.75**

Table 5: Average classification accuracy over each feature with independent train/test subjects. Numbers highlighted in bold indicate increased accuracy compared to having the same subjects in both test and train sets.

4.5 Why does the overall classification accuracy improve when subject bias is removed, contrary to the expected decrease?

Method Despite our expectation that removing subject bias would lead to a decrease in accuracy, our model showed an unexpected improvement. This paradox arises from the fact that our model was tested on entirely new data in terms of subject characteristics, making classification more challenging. The experimental setup is similar to Section 4.4. To ensure the proper functioning of our model, we implemented cross-validation by independently splitting the data into train and test subjects in each fold. Then, we conducted an experiment where, at each step, one test subject was replaced with a random subject from the training data used to train the model. In the first step, one test subject was replaced, followed by two in the second step, and three in the third step. If the model works properly, the accuracy should increase as the number of test data instances that the model is familiar with grows.

Results As shown in Figure 5, the inclusion of more subjects in the test set, which the model has encountered previously, leads to an increase in classification accuracy for KNN and LR. However, in SVM, there is minimal change observed.

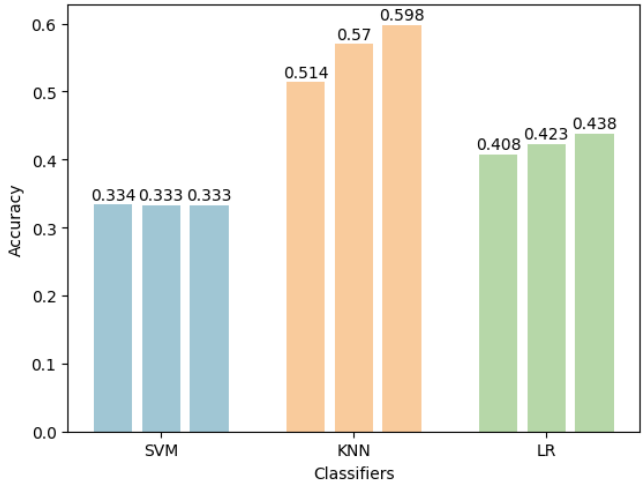


Figure 5: Classification accuracy at consecutive steps for all classifiers.

4.6 Which features carry the most informative content for classification purposes, and does classifying based on these features lead to an improvement in classification accuracy?

Method Based on insights from previous experiments, we aimed to identify features that contain significant emotional information. The expectation was that by selectively choosing features, we could improve the accuracy of our model. Based on the information from Table 3, 4, and 5, four features, *de_LDS*, *de_movingAve*, *rasm_LDS*, and *rasm_movingAve*, were selected for their potential to provide

important information in our classification task. In this section, we used a 5-fold cross-validation setup similar to the one used in Section 4.4.

Results As anticipated, the utilization of selected features results in an increase in classification accuracy by 40%, 37%, and 38% for SVM, KNN, and LR, respectively compared to using all features (Figure 6).

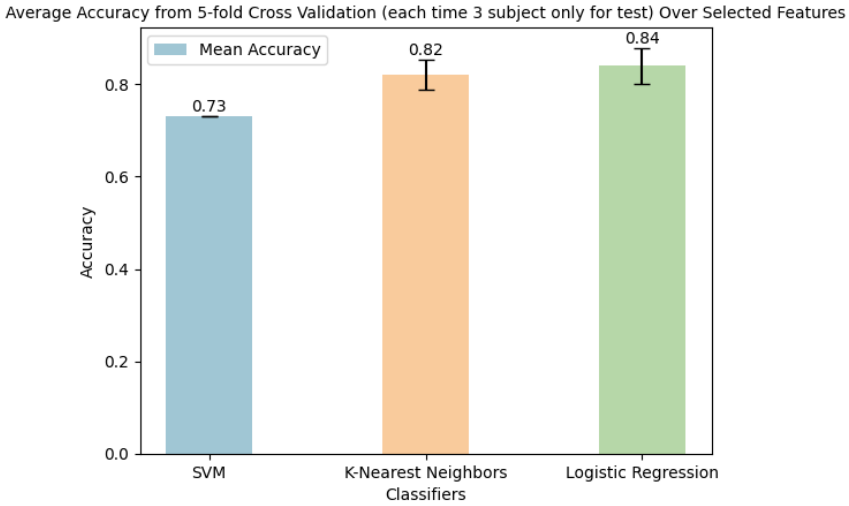


Figure 6: Classification accuracy of selected features.

4.7 Which frequency bands carry the most informative content for classification purposes?

Method Previous studies have indicated that emotion-related information is more recognizable in higher frequency bands, particularly Beta and Gamma [17, 19]. We graphed the *de* feature from the first experiment of a random subject in all frequency bands and channels. We noticed patterns in higher frequency bands, possibly linked to emotional responses (Figure **??**).

To investigate this hypothesis in the context of our dataset, we structured an experiment where, at each iteration, a single feature was employed for data classification within each frequency band, utilizing all three classifiers. The train-test split and cross-validation setup followed the methodology outlined in Section 4.4. This design allowed us to systematically evaluate the impact of individual features across various frequency bands on the classification process, providing insights into the patterns of emotional information within our dataset.

Results In the case of the *de* feature, it is observed that Beta and Gamma frequency bands consistently exhibit the highest accuracy. For the *asm* and *dasm* features, the Theta and Alpha frequency bands tend to demonstrate higher accuracy. However, no consistent patterns are evident across all cases (See Figure 8.

5 Discussion

Emotion detection is crucial in understanding human behavior and facilitating effective communication, as emotions significantly influence decision-making and social interactions. Leveraging tools such as brain recording, particularly Electroencephalography (EEG), has become crucial in recent years for its non-invasive nature and ability to provide real-time insights into neural patterns. Brain recordings offer valuable insights into the complex dynamics of emotional responses.

In this study, we employed diverse methodologies to extract emotion-related insights from EEG signals. Our approach included utilizing preprocessed data and extracted features from the SEED dataset. To fully understand the topic, we designed multiple experiments framed as decoding tasks. In

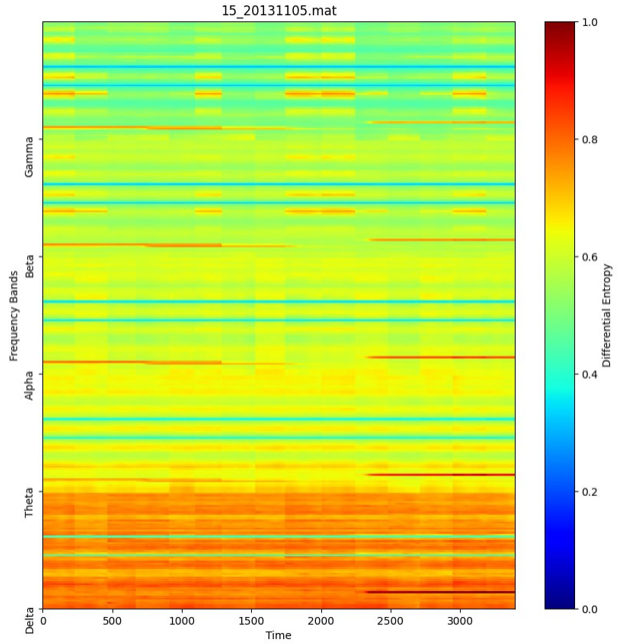


Figure 7: DE Feature Plot. The vertical axis represents all channels across various frequency bands, while the horizontal axis represents time steps.

the classification process, we employed SVM, LR, and KNN as classifiers to categorize EEG signals based on their corresponding emotions. This approach allowed us to explore and grasp the complex connection between EEG patterns and emotional states, giving us valuable insights into how machine learning can help classify emotions.

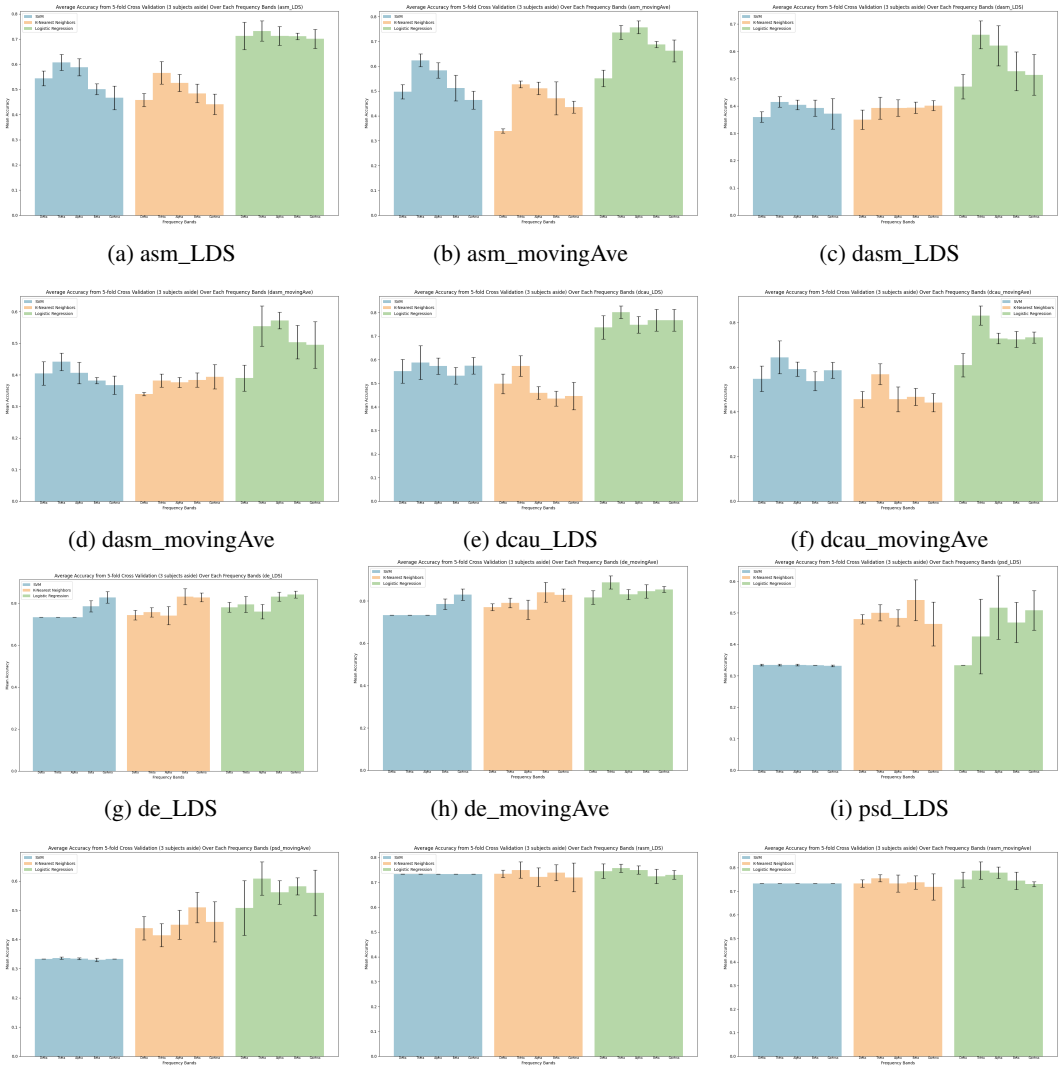
Our results indicate that our models were able to effectively classify EEG recordings into different emotions. Among the methods used, LR demonstrated superior performance in this task. We propose that the low accuracy can be attributed to the limited number of records, given our choice of a small subset of preprocessed data, despite the extensive length of EEG recordings for each record.

Handling preprocessed data needs substantial resources, and we demonstrated an efficient alternative by successfully classifying signals based on the extracted features. In this experiment, KNN exhibited superior performance among the classifiers. Additionally, we observed that certain features played a more significant role in the classification process than others. However, a challenge emerged as some videos and subjects were included in both training and test sets, leading us to hypothesize that the alleged strength of our network might be influenced by biases introduced by these subjects and videos.

Thus, we trained models on independent training and test sets in terms of both videos and subjects. When it came to videos, we noticed a decrease in accuracy, indicating that the bias introduced by these videos did not exist anymore to assist the model in prediction. However, surprisingly, in terms of subjects, we observed higher accuracy. This suggests that the decision to independently select subjects had a positive impact on the model's performance, in contrast to the negative effect observed with videos.

Since this phenomenon does not conform to our understanding of subject bias, we sought to assess the accuracy of our model's training by testing it with an increasing amount of data seen during training. We expected that if there were issues with the model, its performance should not improve, but rather show a decline. The results, however, revealed that the model learned effectively, demonstrating enhanced accuracy for data observed during training, in contrast to our initial assumption.

Furthermore, our findings demonstrate that utilizing just four features—*de_LDS*, *de_movingAve*, *rasm_LDS*, and *rasm_movingAve*—proves to lead to higher accuracy in classification compared to other features. Building models based on these features enables accurate information classification. The selection of these features is attributed to their ability to carry the most important information



(a) asm_LDS (b) asm_movingAve (c) dasm_LDS

(d) dasm_movingAve (e) dcau_LDS (f) dcau_movingAve

(g) de_LDS (h) de_movingAve (i) psd_LDS

(j) psd_movingAve (k) rasm_LDS (l) rasm_movingAve

Figure 8: Classification accuracy of individual features across five frequency bands using SVM, KNN, and LR classifiers. Each set of five bars corresponds to the accuracy of classification for a specific feature across all five frequency bands using a particular classifier

about signals, while simultaneously removing features that might offer irrelevant data, potentially misleading the model and increasing its complexity.

Additionally, our study on information in different frequency bands showed that there are no consistent patterns. Some frequency bands have more information about certain features while others have less. For example, Beta and Gamma frequency bands contain a lot of information about *de_LDS* and *de_movingAve*. This aligns with what previous studies have found [17, 19].

In conclusion, our research leads to improved methods for detecting emotions from EEG signals. This work contributes to understanding both machine learning methods and how the brain handles emotions. Additionally, our findings generate insights into how the brain processes emotions, revealing variations in the spread of information across different frequency bands and features.

5.1 Ethics

The ethical principles in ML and AI are complex and multifaceted. These principles should be considered during the design, development, and deployment of AI systems [23].

In our project, we utilized SEED dataset to conduct our research, where ethical considerations should be of utmost importance. Prior to participating, all subjects provided informed consent, understanding the nature of the experiments, including exposure to emotional movie clips and EEG data recording. Personal privacy was safeguarded through anonymization, and cultural sensitivity was considered. We also took into account the potential for subject and video bias, which can lead to models not representative of the broader population, leading to unfair outcomes.

Another ethical considerations in our study revolve around the similar nature of our subjects, consisting exclusively of individuals from a specific university in China. This limitation in geographic and cultural diversity raises concerns about the generalizability of our findings. For example, emotion expression and interpretation can vary significantly across cultures, potentially introducing bias into our models that may not extend well to more diverse populations.

Furthermore, it is crucial to acknowledge that our EEG data was gathered in 2013. Since then, there have been advancements in technology, protocols, and methods, potentially making our data less applicable to contemporary tests. This temporal gap introduces a challenge in ensuring the relevance and reliability of our findings.

Another potential bias relates to the assumption that all subjects experience the same emotions when exposed to the same video stimuli. This aggregation bias neglects the potential influence of various factors on individual interpretations. It is crucial to recognize the multifaceted nature of emotional responses and consider a more proper approach to analysis.

Additionally, our evaluation methodology relies solely on accuracy, potentially overlooking the importance of alternative measures in certain situations. This may lead to evaluation bias, necessitating a broader consideration of performance metrics that align with the specific context and goals of the study.

Moreover, to mitigate deployment bias, it is vital to emphasize that our models should be viewed as tools to assist humans rather than as decision-making entities. Interpretation by qualified individuals is essential, especially in critical settings and environments, to avoid relying only on automated outputs for vital decisions. This precaution is crucial for ensuring responsible and ethical deployment of our models.

6 Future Work

Given additional time and unlimited resources, there are several avenues to explore and enhance our project. The first step is conducting a more extensive exploratory data analysis (EDA) on both datasets to gain deeper insights.

Furthermore, we can use various classification methods to enhance the classification of preprocessed EEG data by involving more subjects. Experimentation with data reduction techniques is also another option. One potential approach is exploring the classification performance on datasets where trials with similar emotions are averaged for each experiment, creating a more manageable dataset. Another

option is to break down trial signals into smaller sections, treating each section as an individual data point, thereby increasing the dataset size with shorter lengths. However, the lack of available videos in our research limited our ability to identify specific time intervals when emotions are likely to peak. Access to video resources in the future could provide valuable insights by selecting and studying intervals aligning with emotional peaks.

Expanding the feature extraction process is crucial. In addition to frequency domain features, extracting time domain features and additional frequency domain features can offer a more comprehensive understanding of the data.

Additionally, fine-tuning the hyperparameters for our models can help improve our results. We have a specific focus on understanding the unexpected increase in accuracy after removing subject bias. Future work aims to delve even deeper into uncovering the reasons behind this phenomenon.

To achieve more unbiased models, we can create a model eliminating both subject and video bias. Additionally, we aim to explore the probability of predicting EEG signals of one subject based on another and identifying correlations.

Understanding that emotions can vary from person to person, even when watching the same videos, inspires us to consider combining this data with additional information. This could involve using self-assessment questionnaires, recording facial expressions, or asking individuals to describe their feelings and then extracting features from speech related to emotion. Also, including eye-tracking data could provide further insights. It's essential to acknowledge that the subjects in this study come from a limited range, and including individuals from different cultures and languages during data collection could offer a more complete understanding of emotions.

7 Contributions

The authors' individual contributions to this work are outlined below.

Elaheh Toulabinejad Literature Review and Related Works, Modeling, Model Evaluation, Analyzing Results, Paper Write-up

Kiana Aghakasiri Literature Review and Related Works, Preprocess Data, Analyzing Results, Paper Write-up

References

[1] Matteo Spezialetti, Giuseppe Placidi, and Silvia Rossi. Emotion recognition for human-robot interaction: Recent advances and future perspectives. *Frontiers in Robotics and AI*, 7, 2020.

[2] T.M. Lee, Ho-Ling Liu, C.C. Chan, S.Y. Fang, and J.H. Gao. Neural activities associated with emotion recognition observed in men and woman. *Mol Psychiatry.*, 57:1011–1019, 01 2005.

[3] Yujian Cai, Xingguang Li, and Jinsong Li. Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review. *Sensors*, 23(5), 2023.

[4] SK Pahuja, Karan Veer, et al. Recent approaches on classification and feature extraction of eeg signal: A review. *Robotica*, 40(1):77–101, 2022.

[5] Mahboobeh Jafari, Afshin Shoeibi, Marjane Khodatars, Sara Bagherzadeh, Ahmad Shalbaf, David López García, Juan M. Gorriz, and U. Rajendra Acharya. Emotion recognition in eeg signals using deep learning methods: A review. *Computers in Biology and Medicine*, 165:107450, 2023.

[6] Tong Zhang, Xuehan Wang, Xiangmin Xu, and CL Philip Chen. Gcb-net: Graph convolutional broad network and its application in emotion recognition. *IEEE Transactions on Affective Computing*, 13(1):379–388, 2019.

[7] Jingcong Li, Shuqi Li, Jiahui Pan, and Fei Wang. Cross-subject eeg emotion recognition with self-organized graph neural network. *Frontiers in Neuroscience*, 15, 2021.

[8] Xue-han Wang, Tong Zhang, Xiang-min Xu, Long Chen, Xiao-fen Xing, and CL Philip Chen. Eeg emotion recognition using dynamical graph convolutional neural networks and broad learning system. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1240–1244. IEEE, 2018.

[9] Thejaswini Shankar, Dr Kumar, and Aditya L. Analysis of eeg based emotion detection of deap and seed-iv databases using svm. *SSRN Electronic Journal*, 8, 05 2019.

[10] Yerim Ji and Suh-Yeon Dong. Deep learning-based self-induced emotion recognition using eeg. *Frontiers in Neuroscience*, 16, 2022.

[11] Adrian Rodriguez Aguiñaga, Luis Muñoz Delgado, Víctor Raul López-López, and Andrés Calvillo Téllez. Eeg-based emotion recognition using deep learning and m3gp. *Applied Sciences*, 12(5), 2022.

[12] Rajeswari Rajesh Immanuel and S. K. B. Sangeetha. Analysis of eeg signal with feature and feature extraction techniques for emotion recognition using deep learning techniques. In Nabendu Chaki, Nagaraju Devarakonda, and Agostino Cortesi, editors, *Proceedings of International Conference on Computational Intelligence and Data Engineering*, pages 141–154, Singapore, 2023. Springer Nature Singapore.

[13] Sumya Akter, Rumman Ahmed Prodhan, Tanmoy Sarkar Pias, David Eisenberg, and Jorge Fresneda Fernandez. M1m2: Deep-learning-based real-time emotion recognition from neural activity. *Sensors*, 22(21), 2022.

[14] Gaochao Cui, Xueyuan Li, and Hideaki Touyama. Emotion recognition based on group phase locking value using convolutional neural network. *Scientific Reports*, 13(1), March 2023.

[15] Peixiang Zhong, Di Wang, and Chunyan Miao. Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 13(3):1290–1301, 2020.

[16] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger. Simplifying graph convolutional networks. In *International Conference on Machine Learning*, pages 6861–6871. PMLR, 2019.

[17] Wei-Long Zheng and Bao-Liang Lu. 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, 2015.

[18] Pierre Philippot. Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and emotion*, 7(2):171–193, 1993.

[19] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for EEG-based emotion classification. In *6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84. IEEE, 2013.

[20] D. K. Kumar and J. L. Nataraj. Analysis of eeg based emotion detection of deap and seed-iv databases using svm. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(1C), May 2019. Publication Retrieval Number: A10360581C19/19©BEIESP.

[21] Hafeez Ullah Amin, Wajid Mumtaz, Ahmad Rauf Subhani, Mohamad Naufal Mohamad Saad, and Aamir Saeed Malik. Classification of eeg signals based on pattern recognition approach. *Frontiers in computational neuroscience*, 11:103, 2017.

[22] Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 2021.

[23] Harini Suresh and John Guttag. A framework for understanding sources of harm throughout the machine learning life cycle. In *Equity and access in algorithms, mechanisms, and optimization*, pages 1–9. 2021.