Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

EEG Emotion Classification based on Valence and Arousal using DEAP Dataset

Dibya Thapa[1*], Rebika Rai[2]
[1] *Department of Computer Engineering, Sikkim Institute of Science and Technology, India, dibya.smit@gmail.com.*
[2] *Department of Computer Applications, Sikkim University, India, rrai@cus.ac.in.*

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Emotion Recognition using Electroencephalography (EEG) signals has emerged as an important area of research due to its wide applications related to mental health monitoring, affective computing and Human Computer Interaction (HCI). However, majority of high performing emotion recognition systems are computationally intensive, lack interpretability making them ineffective for real time application. In this paper, we present a simple yet effective model for EEG emotion classification using an ensemble of basic machine learning classifiers. The model is enhanced with frequency augmentation technique which efficiently classifies the emotional states based on Valence and Arousal while maintaining interpretability and reducing complexity. A comprehensive analysis comparing the effectiveness of various feature sets, such as Statistical methods, Hjorth parameters, Entropy-based features, Frequency band power, Fractal Dimension has been performed on Database for Emotion Analysis using Physiological Signals (DEAP) dataset. The model was trained using K-Nearest Neighbors (kNN), Random Forest (RF), and CatBoost (CB) classifiers along with their ensemble model utilizing the strength of each algorithm. Results demonstrate high accuracy in classifying Valence (91%) and Arousal (89%) dimensions. We demonstrate through our work that even with basic machine learning classifiers, high accuracy can be achieved for Valence and Arousal labels using the EEG signals.

Keywords: Emotion Recognition, EEG, Human Computer Interaction, DEAP, Ensemble Model, Frequency Augmentation Technique, Machine Learning.

1. INTRODUCTION

Emotions can be defined as physical or mental state due to neurophysiological changes and greatly influence decision making process. These when positive, lead to improved health but when negative may lead to various health issues. Hence, any artificial system in the field of medicine, education, entertainment etc. interacting with humans must be able to automatically and efficiently recognize such emotions. Recognition of emotions may be done through non-physiological signals such as facial expression, actions, voice and through various physiological signals such as EEG, Electrocardiogram (ECG), and Electrooculogram (EOG) etc. Physiological signals commonly used for emotion recognition such as EEG can yield relatively objective and accurate results [1].

A number of well-established models have been established to classify or recognize emotions. Discrete models prefer to classify emotions into disparate categories such as happy, sad, surprise, disgust etc. Dimension model uses multiple dimensions or scales based on Arousal (calm or excited emotion) and Valence (positive or negative emotion). One of the popular Dimensional Model is the Russel's 2D model 1983 [2] used in our study. The model as displayed in Figure 1 is described by four quadrants and two axes. The horizontal axes denoted by Valence describe whether the emotion experienced has a positive or negative effect. On the other hand, the vertical axes denoted by Arousal describes the state of calmness or excitement experienced by the humans.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

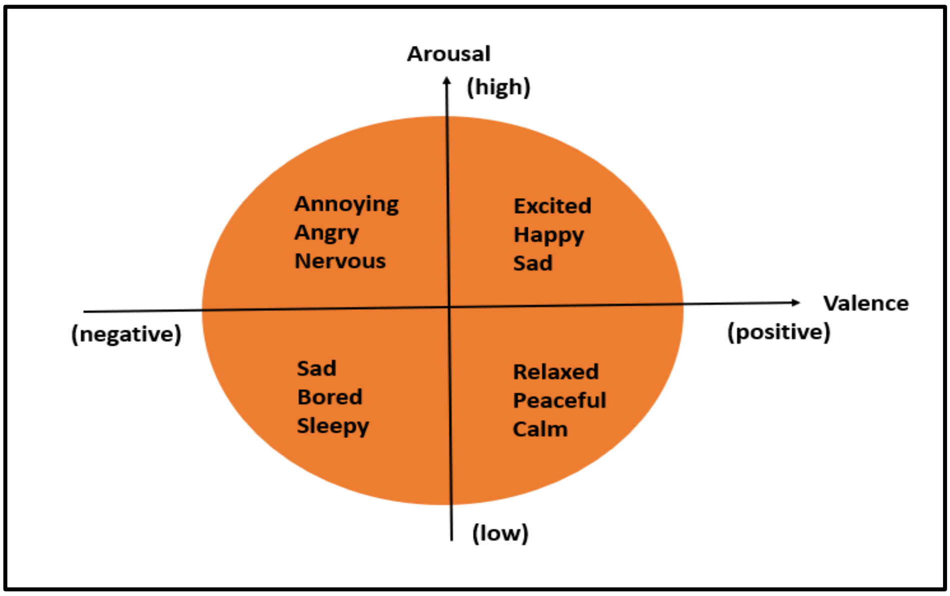

Figure 1. Russel's 2D Dimension Model for Emotion Recognition

The general approach to Emotion classification involves a series of well-defined tasks as depicted in Figure 2. EEG signals collected are pre-processed which involves removing unwanted artifacts and noise such as eye movements, down sampling and re-referencing. The next step consists of feature extraction and selection, where information most relevant to the study are selected. Classification of the pre-processed signals is done using Machine Learning (ML) or Deep Learning (DL) architectures. A number of architectures have been proposed for performing emotion recognitions. These architectures although powerful are based on complex Deep Learning architectures requiring large datasets, are computationally intensive, lack interpretability and unsuitable for cost effective deployment. Few of the main challenges faced in the literature review are dealing with highly imbalanced dataset, selecting robust and simple classification algorithms, selecting the optimal features and understanding methodological limitations.

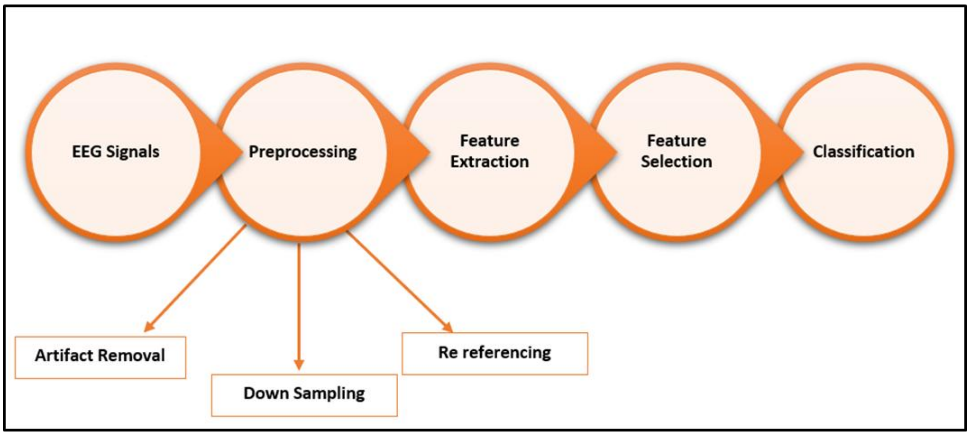

Figure 2. An Overview of EEG Classification

The primary contributions of this research are as follows:

1. Development of Lightweight Emotion Recognition framework that leverages simple, traditional machine learning classifiers instead of computationally intensive Deep Learning models.
2. Addressing of highly imbalanced datasets such as DEAP through frequency augmentation technique using Fast Fourier Transform (FFT) therefore building a more generalized and robust model.
3. Conduction of extensive experiments to compare the effectiveness of multiple feature sets including Statistical metrics, Hjorth parameters, Band Power, Entropy, and Fractal Dimensions—on Valence and Arousal classification.

4. Superior performance on the DEAP dataset [3] through soft voting ensemble of KNN [4], CatBoost [5] and Random Forest classifiers [6], with a significant classification accuracy without added complexity.

The remainder of the paper as displayed in Figure 3 is organized as follows. Section 2 discusses the literature review. The methodology including dataset description, preprocessing, augmentation technique, features extracted and classifiers used are described in Section 3. Section 4 presents the results of the various experiments. The final Section 5 presents the study's conclusion and indicates future work.

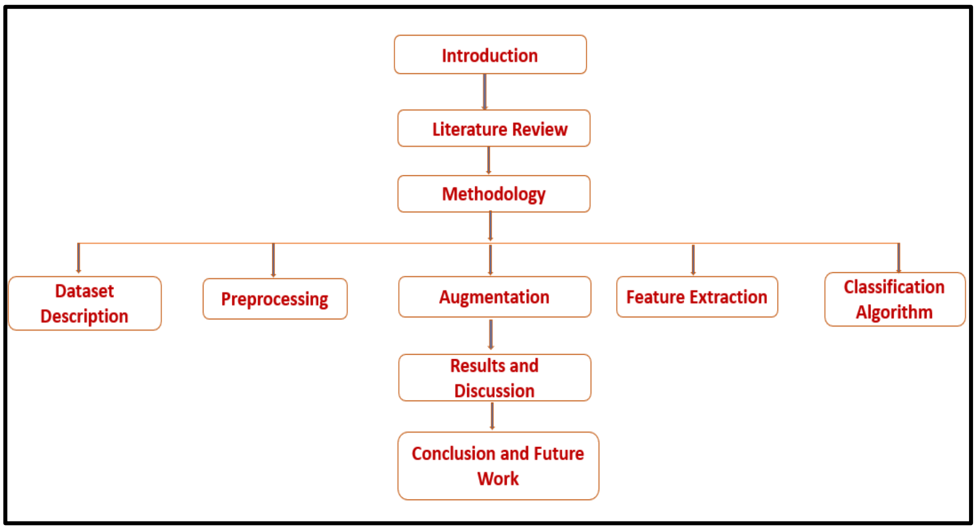

Figure 3. Organization of the Paper

2. RELATED WORK

C. Godin *et al.* [7] investigated the relevance of particular features for emotion classification purpose. They suggested that a lot of features proposed in the earlier studies were highly correlated. The study revealed that most relevant feature for Valence was Zygomatic EMG and for Arousal it was found to be Eye blinking rate for the DEAP dataset. Anuchin Chatchinarat *et al.* [8] used Support Vector Machine (SVM) as a closed box and a second classifier, Classification and Regression Trees (CART) to extract rules from the classifier. The authors suggest only using classifiers like Support Vector Machine is not suitable for medical application where a better understanding of how a classifier makes a decision is required. The model has been used for performing binary classification of EEG signals (positive, negative). SVM-CART model displayed accuracies of 53.1%, 55.81%, and 66.4% for Arousal, Valence and Dominance respectively. Yan Song *et al.* [9] employed various temporal, frequency based and spatial features on DEAP dataset to achieve an accuracy of 95.70% for Arousal and 95.33% for Valence. They proposed a novel model named Efficient Channel Attention-Convolutional Recurrent Neural Network (ECA-CRNN) integrating the efficient channel attention (ECA-Net), a modified combination of Convolutional Neural Network (CNN) and Gated Circulation Unit (GRU). Abhishek Iyer *et al.* [10] performed testing on SJTU Emotion EEG Dataset (SEED) and DEAP dataset while extracting the frequency-based features and differential entropy from EEG signals while using CNN, Long Short-Term Memory (LSTM) and CNN-LSTM hybrid models as classifiers. The highest accuracy reported was for DEAP dataset of 97.16%. Xingyi Wang *et al.* [11] utilized the self-supervised multitask CNN on SEED dataset to achieve an accuracy of 98.65% whereas on DEAP dataset the model yields an accuracy of >90%. Differential entropy was extracted as a feature from SEED dataset whereas it was Differential Entropy and five frequency bands for DEAP dataset. Md. Rabiul Islam et al. [12] proposed the usage of deep machine learning-based model with Convolutional Neural Network (CNN). EEG data were converted to Pearson's Correlation Coefficient (PCC) featured images of channel correlation of EEG sub-bands. The model was deployed on DEAP dataset and maximum accuracy of 78.22% on valence and 74.92% on arousal were obtained using the internationally authorized DEAP dataset. Aasim Raheel et al. [13] in their work used physiological signals such as Photoplethysmography (PPG), EEG and Galvanic Skin Response (GSR) observed in response to a tactile enhanced multimedia with an aim to give the users a real-world sensation while engaging with the multimedia content. Frequency domain and Time domain features were extracted to achieve an accuracy of 79.76% for four emotions with the K nearest neighbor classifier. A total of 34 features were

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

extracted from frequency, time and nonlinear domain to classify emotion based on Valence Arousal model using Artificial Neural Networks (ANN) [14]. Study was performed using different scalp regions frontal, parietal, temporal and occipital. The highest accuracy of 93.25% was reported while using the frontal region electrodes. Two different algorithms namely Normalized Window Short Time Fourier transform (NW-STFT) algorithm and the Cognitive Mapping-based Hebbian Learning (CM-HL) algorithm were proposed by P. Santhiya and S. Chitrakala [15]. While the training of the model was performed using the Database for Emotional Analysis in Music and Electroencephalogram Recordings (DREAMER) dataset the testing was performed on DEAP and A Dataset for Affect, Personality and Mood Research on Individuals and Groups (AMIGOS). Different accuracies reported were 85% forAngry,89% for Calm, 88% for Sad,90% for Fear, 91% for Happy. Classification accuracy using different feature sets were investigated in terms of valence and arousal by Rajamanickam Yuvara et al. [16]. Feature sets selected were statistical feature, fractal dimension, Hjorth parameters, Higher Order Spectra (HOS) and the ones derived using wavelet analysis. Classifiers employed during the study were SVM and CART across MANHOB-HCI, DEAP, SEED, AMIGOS and DREAMER as datasets. Accuracy of 85.06% and 84.55% was reported for valence and arousal respectively.

3. METHODOLOGY

The flowchart of the proposed model is illustrated in Figure 4. As shown in the figure, emotion recognition begins with acquisition of EEG signals from the DEAP dataset. This is followed by preprocessing, augmentation to address the limited size of dataset, feature extraction and finally, classification. Each intermediate step is explained in detail in the following sections.

3.1 DEAP Dataset Description

In order to facilitate the process of development of efficient affective recognition algorithms, benchmark datasets have been created and maintained by different research teams [3,17,18]. Among them, one of the most widely used, the DEAP dataset has been used for our study. As shown in Figure 4 a, the dataset consists of peripheral, biological and EEG signals from 32 participants, while each watched 40 1 min long music videos. The signals were recorded at a sampling rate of 512 Hz. The participants rated the videos in terms of Valence, Arousal, Dominance and Liking on a scale of 1-9. Among these Valence which reflects how positive or negative the emotion is and Arousal which defines the level of intensity (calmness to excited) has been used in our study. DEAP original recording contains 32 .bdf files (BioSemi's data format generated by the Actiview recording software), each with 48 recorded channels at 512Hz. (32 EEG channels, 12 peripheral channels, 3 unused channels and 1 status channels) among which only the 32 EEG channels have been selected to carry out our study.

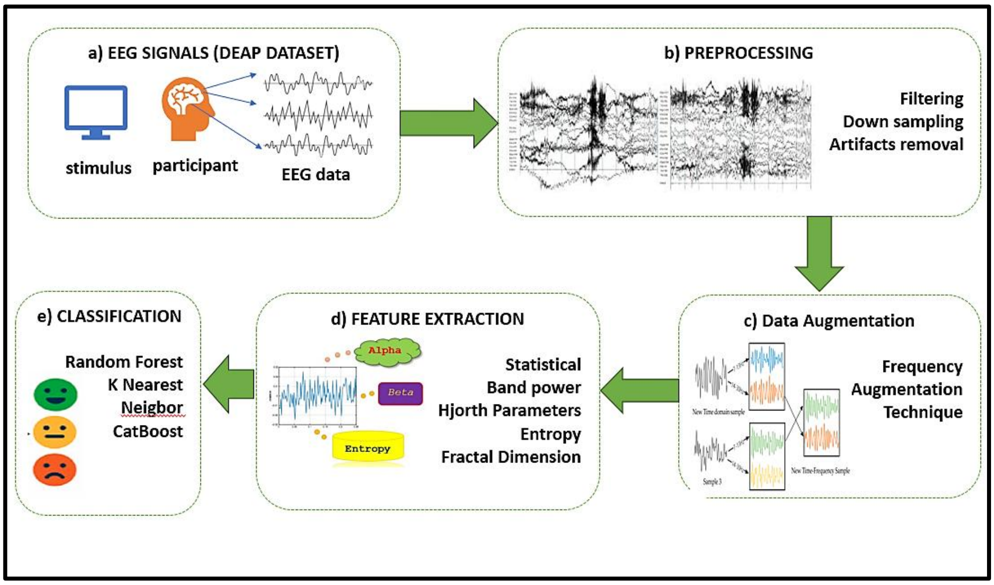

Figure 4. Flowchart of the proposed approach

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

3.2 Preprocessing

EEG signals are accompanied by a lot of other artifacts such as eye blink, muscle movement, noise, hence EEG analytics requires removal of such noise to avoid misleading conclusions. Researchers in their study have demonstrated how lack of attention to the early stages of preprocessing leads to reduction of signal to noise ratio and introduces unwanted artifacts in the data. Early-stage preprocessing in our study involved line noise removal, down sampling of dataset from 512 Hz to 128 Hz.

3.3 Data Augmentation

One of the major problems in the field of EEG emotion recognition is the limited availability of dataset and the class imbalance that resides within them, limiting the generalizability of models trained on them. Since collection of such medical data is resource intensive, requires expertise knowledge, specialised equipment and controlled environments, researchers have come up with alternate solutions to address the challenges. Advanced Deep Learning techniques such as Generative Adversarial Networks (GANs) have been used extensively to generate synthetic EEG data. Such networks consist of the generator and the discriminator model which compete against each other to produce realistic synthetic samples. However, such models require large computational resources and long training time making them expensive. Such networks also lack interpretability acting as “black boxes”, failing to explain how the synthetic data resembles the original neurophysiological patterns. To address these challenges, we utilized Frequency Domain Augmentation technique which could efficiently handle the data scarcity and imbalance issues. The process begins with the application of FFT converting the original time domain EEG signals into their frequency components, performing a detailed analysis of the signal’s frequency characteristics. Next, the frequency domain components are then modified using techniques such as scaling, shifting or adding noise to specific frequency bands by a random factor which in turn leads to augmented EEG signals but retains the original features.

3.4 Feature Extraction

The accuracy of the classifier has a direct correlation with the features extracted and hence becomes a crucial aspect of the emotion recognition process. Short Time Fourier Transform (STFT), Wavelet Transform (WT), Power Spectral Density (PSD), Sample Entropy (SE) are some of the powerful techniques for feature extraction. EEG feature extraction can be performed in the time, frequency and time-frequency domain. Frequency domain analysis provides information contained in the frequency domain of the EEG signals using Statistical and Fourier Transformation methods, Power Spectral Analysis being the most common one. In order to carry out our research work the following features have been extracted:

3.4.1 Statistical features

They are one of the major divisions of the feature set of the EEG features. The physiological signals like EEG are complex and non-stationaryhence some statistical features like Power Spectral Density (PSD) and Spectral Entropy (SE) are widely known features in emotion recognition. The statistical feature selected for our work are:

Mean*:* It denotes the average of the values of the various data points of the EEG signals and is defined as:

$$\mu = \frac{1}{T} \sum_{t=1}^{T} x(t) \tag{1}$$

T here represents the total number of samples whereas x(t) is the EEG signal with respect to time.

Standard Deviation*:* It measures the amount of variation or dispersion from the mean value. A low value of standard deviation indicates closeness of the data points around the mean whereas a high value indicates a large dispersion. It is denoted by:

$$\sigma_x = \left[\frac{1}{N-1}\right] \sum_{n=1}^{N} . (X_n - \mu_x) \tag{2}$$

X here denotes random variable with mean value μ.

Skewness: Skewness measures the symmetry of the data. Symmetric data has a skewness value near zero and skewness for normal distribution is zero. Data that are skewed left will have negative value of skewness whereas positive values of the skewness mean data are skewed right.

3.4.2 Hjorth Parameters

They describe the EEG signal in time domain in terms of activity, mobility and complexity.

Activity parameter describes the variance in the signal and is described by

$$\text{Activity} = \frac{1}{T}\sum_{t=1}^{T}(x(t)-m)^2 \tag{3}$$

Mobility, on the other hand describes the mean frequency and is estimated as

$$\text{Mobility} = \sqrt{\frac{\text{var}\left(\frac{dx(t)}{dt}\right)}{\text{var}(x(t))}} \tag{4}$$

Complexity is one of the parameters that provides an idea of the bandwidth of the EEG signal. It is expressed as

$$\text{Complexity} = \frac{Mobility(\frac{dx(t)}{dt})}{Mobility(x(t))} \tag{5}$$

x(t) and dx(t) represent the signal and mean of the first derivative.

3.4.3 Fractal Dimension

Fractal Dimension denotes the structural complexity of the EEG signal. They are the repetition of self-similar structures in a signal and can be considered as smaller yardsticks that are used repetitively to measure the entire region.

3.4.4 Band Power Features

The frequency range of a normal healthy human EEG lies between 1-30 Hz with an amplitude varying from 20-100 µV, which could be further subdivided into Alpha (8 – 13 Hz), Beta (13-30 Hz), Delta (0.1- 4 Hz), Theta (4 - 8 Hz) and Gamma (30-100 Hz). The various frequency bands are closely associated with different emotion states, with alpha and beta representing relaxation and focus, while theta and gamma associated with deeper relaxation and higher cognitive processing respectively, therefore, their deeper understanding becomes crucial for understanding brain functions and various neurological disorders. Hence, in order to carry out our experiments, the DEAP dataset signals have been subdivided into theta, alpha, beta and gamma signal bands. Further, for each band, their associated band powers have been calculated. The band power features of EEG signals have been found to be very significant in recognizing emotional variation after stimulation. Power can be calculated for each of the bands filtered signals using equation (6).

$$\text{Power} = \frac{1}{\text{N}}\sum_{\text{t}=0}^{\text{N}}(\text{x}(\text{t}))^2 \tag{6}$$

3.4.5 Entropy

It is a measure of the randomness of a given signal. Entropy of EEG signal defines the variability within the EEG signal and is one of the significant features used for emotion classification. Entropy can be represented as

$$e = -\sum_{1}^{n} x^2 - log\,(x^2) \tag{7}$$

Permutation Entropy, Spectral Entropy and Singular Value Decomposition Entropy have been calculated to carry out our study.

3.5 Classification

Classification is an important step in EEG based emotion recognition, where extracted features are used to train a model accurately to distinguish different emotional states. Various Machine Learning (ML) algorithms have been widely used to classify emotions. Among them, RF, CB, and kNN and finally, an ensemble of the three have been used in our study. These classifiers are not only easier for training and implementation purpose, but also offer greater transparency, allowing for better understanding of feature importance and decision-making processes.

3.5.1 K Nearest Neighbor

Simple, straightforward yet a very powerful algorithm that classifies data points based on the majority class of their 'k' nearest neighbors in the feature space. The given algorithm is particularly very effective in EEG classification because of its ability to handle high dimensional data without making any assumptions about the underlying data distribution. Its capability of handling noisy data, often present in EEG signals, makes it a popular choice for real time applications in brain computer interfaces. The parameters considered in our study for kNN are as follows:

- n_neighbor=5: It is the number of nearest neighbors to consider for classification.
- weights=uniform: It signifies that the neighbors contribute equally to the classification decision.
- metric=euclidean: It denotes that distance metric used to calculate the distance between points is Euclidean distance.

3.5.2 CatBoost

It is a gradient boosting algorithm often used for categorical features that reduces the requirement for extensive processing, possessing the ability to work with small datasets. With regards to EEG classification CB can effectively model complex relationships in the EEG data and makes it suitable where interpretability and performance are vital. The parameters set in our study are:

- iterations=1000: It denotes the number of boosting iterations or trees to be built. A higher number of iterations can improve model performance but may also increase the risk of overfitting if not managed properly.
- learning_rate=0.1: The learning rate controls the contribution of each tree to the final model. A smaller learning rate requires more iterations to converge but can lead to better generalization. The value of 0.1 is a common starting point.
- depth=6: This parameter specifies the depth of the trees. A depth of 6 is often used as a balance between model complexity and the risk of overfitting. Deeper trees can capture more intricate patterns but may also lead to overfitting.
- l2_leaf_reg=3: This is the L2 regularization coefficient for leaf weights. It helps prevent overfitting by penalizing large weights in the model. A value of 3 is a typical choice to maintain a balance between fitting the training data and generalizing to unseen data.
- random_seed=42: This parameter sets the random seed for reproducibility. By fixing the seed, you ensure that the results can be replicated across different runs of the model.
- verbose=100: This parameter controls the verbosity of the output during training. Setting it to 100 means that the model will output training progress every 100 iterations, allowing you to monitor the training process.

3.5.3 Random Forest

The Random Forest Classifier is an ensemble technique. The classifier generates a number of decision tress during training and aggregates their predictions to improve the efficiency and accuracy of the model. The classifiers' ability to assess feature importance, significantly assists in identifying the most relevant features for EEG emotion classification tasks. The parameters used for our study are:

- n_estimators=100: The number of trees in the forest is set to 100.
- max_depth=None: The maximum depth of the trees is not specified, allowing them to grow until all leaves are pure or contain fewer than min_samples_split samples.
- random_state=42: This ensures reproducibility of results by initializing the random number generator with a fixed seed.

4. RESULTS AND DISCUSSION

The results described in the following section have been derived using the methods and dataset described in Section 3. The key performance measures in our study to evaluate the performance of various models are accuracy measures such as precision, recall, f1-score, support, average accuracy, confusion matrix and Receiver Operating Characteristic (ROC) Curve. Some of these have been described in detail in Table 1.

Table 1: Key Performance Measures

Sl No	Metric	Formulation	Description
1	Accuracy	TP+TN+FP+FN /TP+TN	Accuracy is the ratio of correctly predicted instances to the total number of predictions.
2	Precision	TP/TP+FP	Measures the accuracy of positive predictions. Tells how many of the predicted positive instances were actually correct.
3	Recall/Sensitivity	TP/TP+FN	It measures the ability to find all relevant instances in the dataset. It tells us how many of the actual positive instances were correctly identified.
4	F1-score	2*{Precision*Recall/Precision + Recall}	Defines the harmonic mean of Precision and Recall
5	Support	___	It is the actual occurrences of each class in the dataset.
6	Confusion Matrix	TP FN FP TN	Compares the actual target values with those predicted by the Machine Learning model
7	ROC Curve	$TPR = \frac{TP}{TP + FN}$ $FPR = \frac{FP}{FP + TN}$	ROC curves provide an understanding of the model performance across range of decision boundaries. The area under ROC curve known as AUC indicates the discriminatory ability of the model.

TP=True Positive, TN= True Negative, FP=False Positive, FN= False Negative

4.1 Emotion Classification Accuracy -Valence Dimension

Table 2 displays the classification accuracies achieved for the Valence dimension across various feature sets using various Machine Learning models- kNN, CB, RF and their ensemble.

- Statistical Features: These features provided the highest accuracy for Valence classification using the ensemble (91%) and RF (90%) suggesting that basic statistical features such as mean, median, mode etc significantly contribute in pattern capturing for Valence related EEG variations.
- Hjorth Parameters: The Hjorth parameters on the other hand, provided relatively low accuracy across all models (69%) using RF indicating their drawback on for Valence classification.
- Fractal Dimension: Fractal Dimension shows strong performance for EEG based Valence classification, especially with CB (86%) and ensemble (89%). This also means that non linear features perform well in EEG based emotion recognition.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

- Band Power: Consistent performance with accuracies of 89% for CB, RF and ensemble was achieved using the band power, indicating the effectiveness of frequency domain features, especially while frequency- based augmentation is utilized.
- Entropy: Performance is consistently poor with the entropy-based features across all models with a highest score of only 60% using Ensemble method.

The confusion matrix and related ROC curves are depicted in Figure 5 (a, b, c, d, e) and Figure 6 (a, b, c, d, e). From the given figures it is evident that the statistical features along with ensemble model delivers the most accurate and balanced results (Figure 5a). Fractal (Figure 5c) follow closely, promising equivalent performance and reliability. Band power features (Figure 5d), performed exceptionally well, supporting the usefulness of frequency-based features. Hjorth parameters (Figure 5b) provide suboptimal results while Entropy based features (Figure 5e) show the poorest results.

The AUC for statistical features and ensemble method combination, reaches 0.96 and ROC curve almost touches the top-left corner (Figure 6a), indicating high performance and excellent sensitivity and specificity. For the Hjorth parameters using RF classifier the curve is closer to the diagonal (Figure 6b), indicating moderate discriminating power. As suggested by the confusion matrix the ROC curve for fractal dimension is close to top-left corner (Figure 6c), suggesting very strong classification performance. AUC value for Band power reaches 0.93 (Figure 6d), with its' shape almost identical to Figure 6c. Entropy measures ROC curves (Figure 6e) indicates poor separation between the classes.

Table 2: Classification accuracies achieved for the Valence dimension using various feature sets and classifiers.

Features	Model accuracy (Valence)			
	KNN	CB	RF	Ensemble
Statistical	61	89	90	**91**
Hjorth parameter	53	66	**69**	65
Fractal dimension	79	86	85	**89**
Band power	64	**89**	**89**	**89**
Entropy	59	59	58	**60**

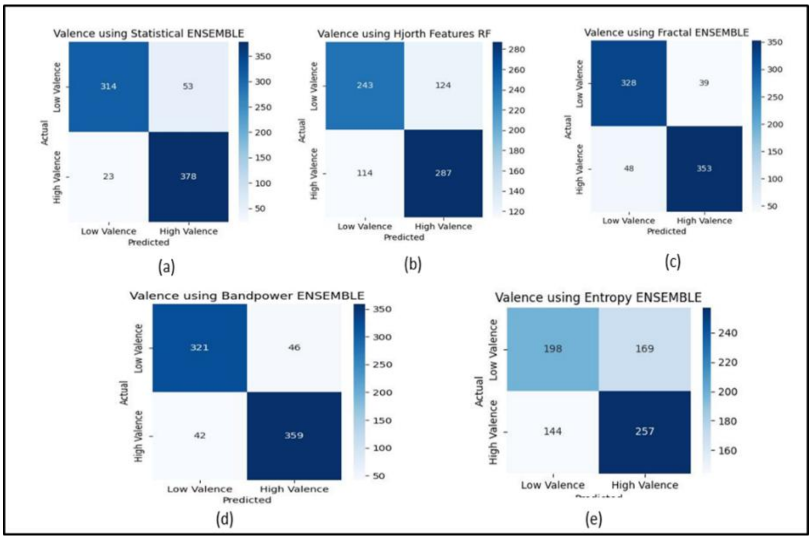

Figure 5. Confusion matrix for the highest accuracies received using various features and classifiers for Valence Label.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

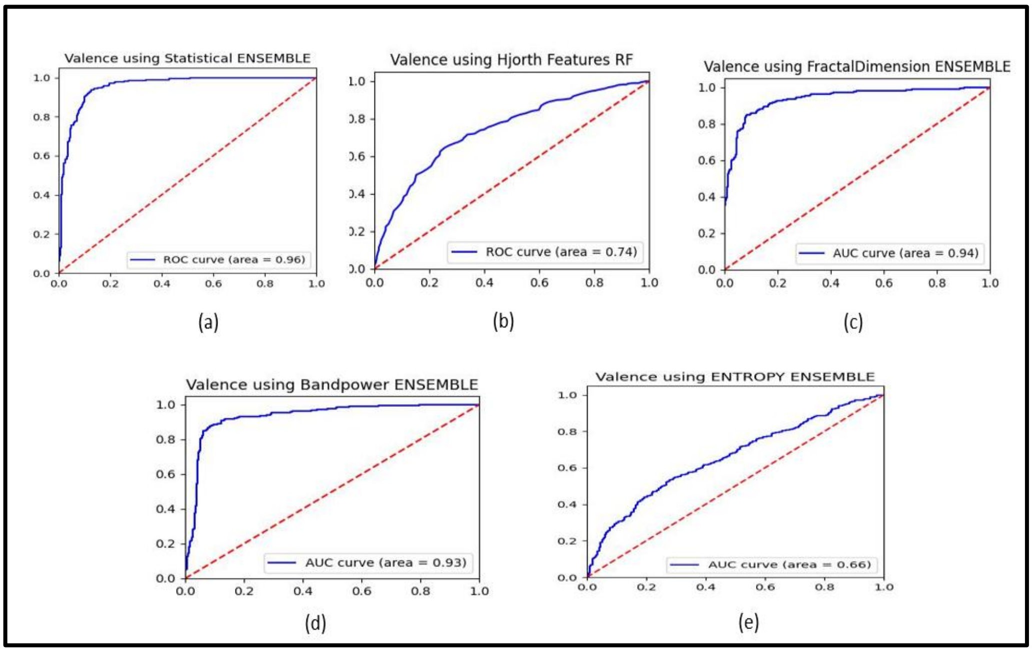

Figure 6: ROC Curves for the highest accuracies received using various features and classifiers for Valence.

4.2 Emotion Classification Accuracy -Arousal Dimension

Table 3 displays the classification accuracies achieved for the Arousal dimension across various feature sets using various Machine Learning models- kNN, CB, RF and their ensemble.

- Statistical Features: For the Arousal label too, the statistical features provided the most promising results with accuracies reaching up to 89% with both CB, RF and the ensemble indicating the strong relationship between statistical features and Arousal label.
- Hjorth Features: Hjorth parameters performed slightly better for the Arousal label than in the Valence label, but still comparatively weaker than other features.
- Fractal Dimension: In the Arousal dimension, fractal dimension provided quite satisfactory performance using CB (88%), RF (87%) and the Ensemble (87%) as with the Valence label. This reinforces the robustness of fractal dimension in modelling the complex EEG patterns.
- Band Power: Along with statistical features, Band Power provided excellent performance of 89% with CB RF and with the ensemble technique. Band power seems to be especially effective when used in combination with the frequency augmentation technique since it enhances the frequency domain representations.
- Entropy: Similar to the Valence dimension, entropy provides relatively weak yet improved performance with a highest value of 62% using the ensemble model indicating, entropy is not a strong predictor of Arousal or Valence.

The confusion matrix and related ROC curves are depicted in Figure 7 (a, b, c, d, e) and Figure 8 (a, b, c, d, e). Statistical features along with ensemble model yield accurate and balanced results (Figure 7a). Fractal features with CatBoost (Figure 7d) and Band power features (Figure 7d), also showed promising results, supporting the usefulness of frequency-based features. Hjorth parameters (Figure 7b) provide suboptimal results while Entropy based features (Figure 7e) performs the weakest.

The AUC for statistical ensemble method combination, nearly touches the top left corner, with AUC value of 0.95 (Figure 8a), indicating high performance and excellent sensitivity and specificity. Hjorth parameters using CatBoost provides AUC value of 0.75(Figure 8b), indicating moderate discriminating power. With AUC of 0.93 (Figure 8c), Fractal CatBoost model suggest very strong classification performance. Band power AUC reaches an impressive value of 0.92 (Figure 8d) using the ensemble model. Entropy measures ROC curves are relatively flat with AUC value of 0.63 (Figure 8e) indicating the models poor performance.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 3: Classification accuracies achieved for the Arousal dimension using various feature sets and classifiers.

Features	Model accuracy (Arousal)			
	KNN	CB	RF	Ensemble
Statistical	63	**89**	**89**	**89**
Hjorth parameter	53	**70**	**70**	68
Fractal dimension	63	**88**	87	87
Band power	68	**89**	**89**	**89**
Entropy	61	61	58	**62**

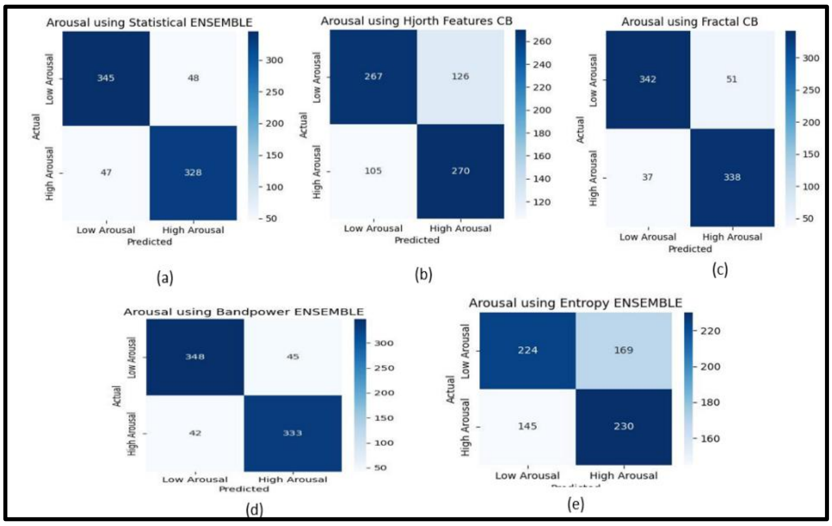

Figure 7: Confusion matrix for the highest accuracies received using various features and classifiers for Arousal Label.

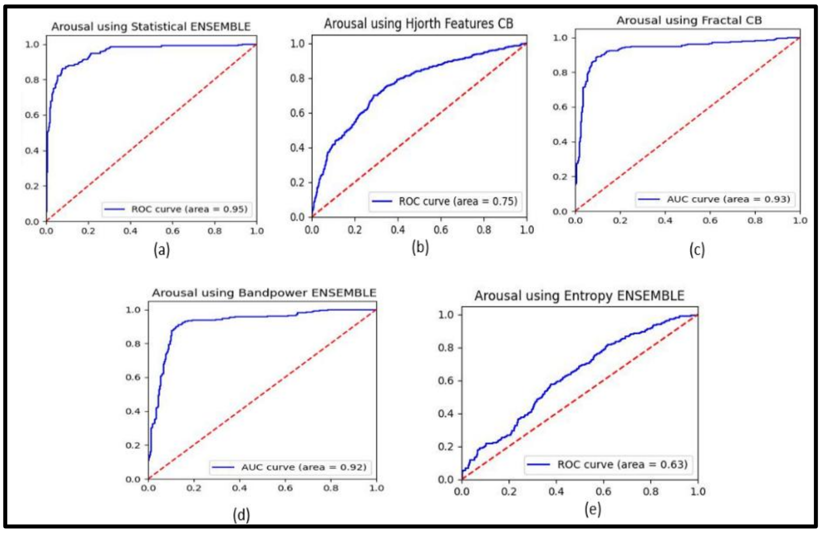

Figure 8: ROC Curves for the highest accuracies received using various features and classifiers for Arousal.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The analysis performed clearly indicates, statistical features and band power consistently provides the highest classification accuracy across the Valence and Arousal dimension reaching 91% for some cases. Next to statistical and band power, fractal dimension provided satisfactory performance with the highest value recorded of 88%. On the other hand, features like entropy and Hjorth parameters were less reliable as standalone features for EEG emotion classification. The results also demonstrate that the combination of band power features along with ensemble learning techniques have remarkable potential for EEG based emotion recognition.

5. CONCLUSION AND FUTURE WORK

This study presents a detailed evaluation of various EEG feature extraction techniques for emotion classification, across Valence and Arousal labels. Feature extraction techniques specifically Statistical descriptors, Hjorth parameters, Fractal dimensions, Band power features, and Entropy measures were explored among which the most promising results were delivered by Statistical, Fractal, and Band power features. Each feature set was evaluated using individual simple Machine Learning classifiers viz. K -Nearest Neighbors (KNN), CatBoost (CB), Random Forest (RF) and an ensemble of all three classifiers. To mitigate the problem of limite data availability, frequency augmentation technique was applied. The ensemble models outperformed the individual models across all classes confirming the effectiveness of model integration in improving classification robustness. For instance, ensemble models using statistical and band power features provided accuracies of 91% and 89% for the Valence classification. Classification matrices and ROC curves derived for each model provided further insights into the classifier working. Ensemble models using Statistical, Band Power and Fractal Dimension features showed high true positives and true negatives revealing the efficiency of such models. Overall, this research not only highlights the most effective features but also demonstrates the importance of ensemble-based learning strategies for emotion classification. While most of the state of-the-art models are based on Deep Learning architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks with high computational complexity and lack of interpretability, our framework in contrast uses shallow, well understood Machine Learning classifiers which are computationally lightweight, require no Graphics Processing Unit (GPU) and can run efficiently on standard Central Processing Unit (CPU) which makes it appropriate for real time applications for emotion recognition.

While the results presented are quite promising, future work could expand the framework to include multi-class classification for a more detailed understanding of emotional states. Additionally, more feature sets such as wavelet transforms, graph-based features could be explored. Moreover, latest augmentation strategies can be experimented with especially for the underrepresented classes, to help address class imbalance and further improve the accuracy. Validation of such frameworks on other datasets such as SEED and DREAMER need to be performed to prove the generalizability of such models.

REFRENCES

[1] Zheng, W. L., Zhu, J. Y., Peng, Y., & Lu, B. L. (2014, July). EEG-based emotion classification using deep belief networks. In *2014 IEEE international conference on multimedia and expo (ICME)* (pp. 1-6). IEEE.

[2] Russell, J. A., & Ridgeway, D. (1983). Dimensions underlying children's emotion concepts. *Developmental Psychology*, *19*(6), 795.

[3] Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., ... & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18-31.

[4] Adhikari, S., Choudhury, N., Bhattacharya, S., Deb, N., Das, D., Ghosh, R., ... & Ghaderpour, E. (2025). Analysis of frequency domain features for the classification of evoked emotions using EEG signals. *Experimental Brain Research*, *243*(3), 65.

[5] Hossain, M. S., Biswas, S. K., Thounaojam, D. M., & Khan, A. (2024, September). Machine Learning Algorithms for Emotion Detection using EEG Signals. In *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)* (pp. 1-6). IEEE.

[6] Kumar, A., & Kumar, A. (2025). Human emotion recognition using Machine learning techniques based on the physiological signal. *Biomedical Signal Processing and Control*, *100*, 107039.

[7] Godin, C., Prost-Boucle, F., Campagne, A., Charbonnier, S., Bonnet, S., & Vidal, A. (2015, February). Selection of the most relevant physiological features for classifying emotion. In *International conference on physiological computing systems* (Vol. 2, pp. 17-25). SCITEPRESS.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[8] Chatchinarat, A., Wong, K. W., & Fung, C. C. (2017, February). Rule extraction from electroencephalogram signals using support vector machine. In *2017 9th International Conference on Knowledge and Smart Technology (KST)* (pp. 106-110). IEEE.

[9] Song, Y., Yin, Y., & Xu, P. (2023). A customized ECA-CRNN model for emotion recognition based on EEG signals. *Electronics*, *12*(13), 2900.

[10] Iyer, A., Das, S. S., Teotia, R., Maheshwari, S., & Sharma, R. R. (2023). CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings. *Multimedia Tools and Applications*, *82*(4), 4883-4896.

[11] Wang, X., Ma, Y., Cammon, J., Fang, F., Gao, Y., & Zhang, Y. (2023). Self-supervised EEG emotion recognition models based on CNN. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 1952-1962.

[12] Islam, M. R., Islam, M. M., Rahman, M. M., Mondal, C., Singha, S. K., Ahmad, M., ... & Moni, M. A. (2021). EEG channel correlation based model for emotion recognition. *Computers in Biology and Medicine*, *136*, 104757.

[13] Raheel, A., Majid, M., Alnowami, M., & Anwar, S. M. (2020). Physiological sensors based emotion recognition while experiencing tactile enhanced multimedia. *Sensors*, *20*(14), 4037.

[14] Kumar, G. S., Sampathila, N., & Martis, R. J. (2023). Classification of human emotional states based on valence-arousal scale using electroencephalogram. *Journal of Medical Signals & Sensors*, *13*(2), 173-182. Kumar, G. S., Sampathila, N., & Martis, R. J. (2023). Classification of human emotional states based on valence-arousal scale using electroencephalogram. *Journal of Medical Signals & Sensors*, *13*(2), 173-182.

[15] Santhiya, P., & Chitrakala, S. (2023). PTCERE: personality-trait mapping using cognitive-based emotion recognition from electroencephalogram signals. *The Visual Computer*, *39*(7), 2953-2967.

[16] Yuvaraj, R., Thagavel, P., Thomas, J., Fogarty, J., & Ali, F. (2023). Comprehensive analysis of feature extraction methods for emotion recognition from multichannel EEG recordings. *Sensors*, *23*(2), 915.

[17] Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). DECAF: MEG-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing*, *6*(3), 209-222.

[18] Katsigiannis, S., & Ramzan, N. (2017). DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*, *22*(1), 98-107.