MULTI-CLASS EMOTION RECOGNITION FROM EEG USING 3D-CNN AND RNN

THESIS REPORT

Submitted in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Electronics and Communication Engineering with specialization in Communication Systems by the A P J Abdul Kalam Technological University

by

SURUMI SHAJAHAN

Reg.No TKM21ECCS13

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

TKM COLLEGE OF ENGINEERING

KOLLAM 691005

MAY 2023

MULTI-CLASS EMOTION RECOGNITION FROM EEG USING 3D-CNN AND RNN

THESIS REPORT

Submitted in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Electronics and Communication Engineering with specialization in Communication Systems by the A P J Abdul Kalam Technological University

by

SURUMI SHAJAHAN

Reg.No TKM21ECCS13

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

TKM COLLEGE OF ENGINEERING

KOLLAM 691005

MAY 2023

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

TKM COLLEGE OF ENGINEERING

KOLLAM 691 005

CERTIFICATE

Certified that this Project report titled **"MULTI-CLASS EMOTION RECOGNITION FROM EEG USING 3D-CNN AND RNN"** is a bonafide record of the work done by **SURUMI SHAJAHAN** (Reg.No.TKM21ECCS13) under my supervision, in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Electronics and Communication Engineering with specialization in Communication Systems by the A P J Abdul Kalam Technological University.

Guide	**Coordinator**
Dr. ANZAR S M	**Dr. NISHANTH N**
Assistant Professor	Professor
Dept. of ECE, TKMCE	Dept. of ECE, TKMCE

HoD

Prof. SHABEER S

Head, Dept. of ECE

TKMCE

Acknowledgements

At the outset, I consider it my duty to thank Almighty God for giving me the necessary wisdom to successfully complete this project presentation.

I thank **Prof. SHABEER S**, HOD, Department of Electronics and Communication, for his encouragement and support.

I express my sincere thanks to our PG coordinator, **Dr. NISHANTH N**, Professor, Department of Electronics and Communication Engineering, for the support and encouragement during the course of this presentation.

I take this opportunity to express my sincere gratitude and profound thanks to my guide, **Dr. ANZAR S M**, Assistant Professor, Department of Electronics and Communication, for his advice, supervision and patience during the course of project preparation and presentation and for providing me guidance and critical inputs in the preparation and presentation of my project.

I would also like to express my sincere gratitude to all my teachers, friends and my parents for their much needed support during the preparation and presentation of the project.

SURUMI SHAJAHAN

TKM21ECCS13

ABSTRACT

Emotion recognition from electroencephalography (EEG) signals has attracted considerable attention in recent years, as it offers a non-invasive and objective means of assessing emotional processing. In this paper, we propose a novel hybrid model that combines the power of three-dimensional Convolutional Neural Networks (3D-CNN) and Recurrent Neural Networks (RNN) for emotion recognition from EEG signals. Our model extracts spatiotemporal features using a 3D-CNN and captures temporal dependencies using an RNN, achieving state-of-the-art performance in recognizing emotions such as happiness, sadness, anger, and fear. Emotion recognition is a critical aspect of human communication and interaction, and its accurate identification has significant implications in various fields, including psychology, neuroscience, and affective computing.

We present a novel method for emotional classification, which is evaluated on the DEAP dataset. Our method achieves high accuracy in binary classification of valence and arousal, with scores of 95.45%, 96.63%, and 97.43%, respectively. In the four-class classification task, our models perform with similar accuracy. However, we note that only four emotion spaces are found with binary classification, whereas 8-class classification is more precise. Therefore, we extend our method to 8-class classification and achieve a promising accuracy of 94.83%.To extract features from the EEG, physiological, and video signals in the DEAP dataset, we use Fast Fourier Transformation (FFT). We employ the same 3D-CNN+RNN architecture for all four classification models, which contributes to the consistency of our results.Overall, our experiments demonstrate the effectiveness of our proposed method for emotional classification, and provide evidence that it can perform well in both binary and multi-class classification scenarios.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Emotion is a fundamental aspect of human experience that is essential to our well-being and ability to function in the world. It is a complex psychological and physiological state that involves a wide range of subjective feelings, such as happiness, contentment, envy, jealousy, anxiety, guilt, shame, and many others.It can be triggered by a variety of internal and external factors, such as thoughts, memories, events, or stimuli in the environment, and they play an important role in shaping our thoughts, behavior, and social interactions.

Facial expressions, speech, eye blinking, and physiological signals can be used to detect human emotions. However, the first three approaches are susceptible to subjective influences of the participants and can deliberately disguise their emotions. On the other hand, physiological signals such as electroencephalograms (EEG), electrooculography (EOG), and blood volume pressure (BVP) are produced spontaneously by

Figure 1.1: EED Data Acquisition

the human body. Physiological signals are more objective and reliable.EEG signal, among all the physiological signals, originates directly from the human brain which directly reflect changes in human emotional states. Therefore, researchers are interested in studying human emotions through the EEG signal. EEG possesses intrinsic characteristics such as nonlinearity, nonstationarity, and multichannel properties. The nonlinear dynamics and nonstationary processes of EEG are represented by the temporal information. The spatial information, on the other hand, captures the interdependence between electrodes or different regions of the brain.

Machine learning and deep learning have also been applied in the field of emotion recognition from EEG signals. Electroencephalography (EEG) is a non-invasive technique for measuring and recording electrical activity in the brain. EEG signals are obtained by placing electrodes on the scalp and measuring the electrical potentials generated by the neurons in the brain as shown in Figure 1.1. This involves training models to automatically detect and classify emotional states based on the complex and dynamic patterns of EEG signals.Several methods have been proposed for EEG-based emotion recognition, including traditional machine learning methods such as support vector machines (SVM), k-nearest neighbors (KNN), and random forests, as well as deep learning-based methods such as convolutional neural networks (CNN), long short-term memory (LSTM), and attention-based models. The choice of method depends on various factors, including the nature of the dataset, the number of features, and the computational resources available.One of the challenges in EEG-based emotion recognition is the nonlinearity and nonstationarity of EEG signals, which require specialized preprocessing techniques and feature extraction methods.Numerous EEG datasets are readily accessible to the public, while some researchers choose to utilize their own dataset. Prominent public datasets, including DEAP, SEED, and MAHNOB, have been utilized in various studies. For this particular research, the DEAP dataset was employed.

Emotion recognition from EEG signals has become a topic of significant interest in the field of affective computing, as it offers a non-invasive and objective means of assessing emotional states. However, the high dimensionality and temporal complexity of EEG signals present significant challenges in accurately classifying emotions.To address this challenge, we propose a 3DCNN+RNN hybrid model that combines the power of 3D-CNN and RNN for emotion recognition from EEG signals. The model leverages the power of three-dimensional Convolutional Neural Networks (3D-CNN) and Recurrent Neural Networks (RNN) to extract spatiotemporal features and capture temporal dependencies, respectively. Experimental results demonstrate that our model achieves state-of-the-art performance in recognizing emotions such as happiness, sadness, anger, and fear from EEG signals, outperforming other state-of-the-art models. In addition to achieving high recognition accuracy, our proposed method is capable of accurately recognizing emotions further into eight divisions.

Chapter 2

Literature Review

Over the past few years, there has been a significant amount of research focused on EEG-based recognition of human emotions. These studies have explored a variety of approaches to feature extraction, channel selection, and classification methods in order to accurately recognize emotions from EEG signals. By comparing and contrasting these different approaches, researchers can better understand the strengths and limitations of each method, ultimately leading to more effective and efficient emotion recognition techniques.

Xiang Li *et.al.*, [1] proposed an effective ensemble method to solve the problem of imbalance between several classes. An effective recursive technique for imbalanced data regression and multiclass classification is proposed. The LSTM-RNN model employed by Alhagry *et.al.*, [2] highlighted the significance of emotion recognition systems based on Human-Computer Interaction (HCI). Unlike other studies that focus on only two levels (arousal and valence), here they identified three levels valence arousal and liking. They achieved impressive accuracy rates of 85.65%, 85.45%, and 87.99% for the valence, arousal, and liking categories, respectively. Notably, they used an end-to-end approach that eliminated the need for feature extraction methods, as deep learning algorithms can perform both feature extraction and classification in a single step

George *et.al.*, [3] utilized the SVM method in their study, which yielded a high overall accuracy of 92%. To determine the features, they employed the DCT method, in the DEAP dataset. The researchers concluded that the Fast Fourier Transform (FFT) statistical features for detecting emotions resulted in 92% higher accuracy.

Rahul Sharmaa *et.al.*, [4] achieved an accuracy of 82.01% using Long short-term memory (LSTM) and a ten-fold cross-validation technique, EEG signals were decomposed using Discrete Wavelet Transform (DWT). The study focused on the recognition of emotions within two dimensions - arousal and valence - and four corresponding quadrants, including LaLv (low arousal, low valence), HaLv (high arousal, low valence), LaHv (low arousal, high valence), and HaHv (high arousal, high valence).

Wang *et.al.*, [5] introduced a model called the Channel-fused dense convolutional network (CDCN), which includes a 1D convolution layer and 1D dense layer. To extract features, they utilized differential entropy (DE) and focused on four emotions. The performance of their model was evaluated on two datasets, the SEED dataset and DEAP dataset, achieving accuracies of 90.6% and 92.58%, respectively.

Lin *et.al.*, [6] proposed a method for classifying emotional states using deep convolutional neural networks (CNNs) trained end-to-end on the DEAP dataset. The researchers transformed the dataset into six grayscale images that contained both time and frequency domain information. These images were then used to extract features before being trained using the AlexNet model. The study achieved an accuracy of 87.30% for arousal and 85.50% for valence.

The graph convolutional neural networks (GCNN) model employed by Zheng *et.al.*, [7] create an EEG-based emotion recognition model, specifically applied to the DEAP database. They segmented the data and extracted differential entropy features before using a merged approach of GCNN and LSTM known as ECLGCNN. This method proved effective, achieving high accuracies of 90.45% and 90.60% for valence and arousal labels in subject-dependent trials, and 85.04% in independent trials. However, there is a need to reduce the computational complexity of this method by exploring additional feature extraction techniques.

Y Cimtay *et.al.*, [8] utilized a CNN Deep Learning method in their research work without any manual pre-extraction methods. They used two datasets, DEAP and SEED. In the SEED dataset, they achieved 86.56% accuracy for two classes of emotions and 78.34% accuracy for three classes of emotions. For the DEAP dataset, their method achieved 72.81% accuracy for the two emotion states of valence and arousal.

Zhang *et.al.*, [9] explored the use of multiple deep learning models, including DNN, CNN, LSTM, and a CNN-LSTM hybrid model, for EEG-based emotion recognition.

They used the DEAP dataset and extracted several features, such as mean, standard deviation, max value, min value, skewness, and kurtosis. Results showed that the CNN and CNN-LSTM models performed the best, achieving 90.12% and 94.17% accuracy, respectively.

Wei *et.al.*, [10] developed an emotion recognition model that can identify three emotions (positive, neutral, and negative) using a SEED dataset. The model utilizes Simple Recurrent Unit (SRU) models, which are capable of processing sequence data. Additionally, the model employs four features across five frequency bands and extracts time, frequency, and nonlinear features using the Dual-Tree Compound Wave Transfer (DT-CWT) method. The model achieves an accuracy of 80.02%, but its development relies on a trial-and-error approach.

In a recent study, Acharya *et.al.*, [11] compared the effectiveness of LSTM and CNN models for carrying out emotion classification tasks using the DEAP dataset. The Fast Fourier Transform method was employed for feature extraction. Results showed that the LSTM model outperformed the CNN model, achieving an accuracy of 88.6% for the liking emotion classification task, while the CNN model achieved an accuracy of 87.2%. A taxonomy of related studies is shown in 2.1.

In the publication Zhuang *et.al.*, [12] the researchers identified a gap in the application of neural patterns in subjective emotion recognition systems. The researchers found that electrodes placed on the temporal, frontal, and occipital lobes resulted in better recognition of high-frequency EEG signal features. Using these features, the researchers were able to classify six primary emotions - fear, joy, sadness, disgust, neutrality, and anger. The features were extracted using the Short-Time Fourier Transform (STFT) algorithm and the Support Vector Machine (SVM) method was used for classification. These findings highlight the potential for utilizing neural patterns in subjective emotion recognition systems and provide insights into the optimal placement of electrodes and feature extraction methods for this purpose.

Table 2.1: Taxonomy of related studies

Paper	Contribution	Algorithms used	Results	Limitation
Xiang Li [1]	Pre-processing technique involves utilizing wavelet and scalogram transforms to convert the multi-channel EEG signals into grid-like frames.	Hybrid deep learning model that combines the CNN and RNN	The proposed method has been shown effective in the trial-level emotion recognition task	Insufficient training data.
Alhagry et.al. [2]	Presents an end-to-end deep learning neural networks method to recognize emotion from raw EEG signals.	LSTM-RNN is used to learn features from EEG signals, the dense layer is used for classification	Average accuracy of 86% is obtained for valence, arousal and liking.	3 - class of classification
George et.al [3]	Box-and-whisker plot is used to select the optimal features	SVM classifier for training and testing the DEAP dataset	The experimental results show that the proposed method exhibits **92.36%** accuracy using skewness, kurtosis, and wave entropy features	Only two class of classification.
Rahul Sharmaa et.al [4]	The third-order cumulants (ToC) are used to explore the nonlinear dynamics of each sub-band signal in higher dimensional space	The LSTM is used to retrieve the emotion variation from the optimized data corresponding to the labeled EEG signals	82.01% average classification accuracy with 10-fold cross-validation technique corresponding to 4-labeled emotions classes	Deep algorithms parameters constraints make the process slightly slow.
Wang et.al [5]	Proposes a cross-datasets emotion recognition method of deep model transfer learning	Residual block based deep convolutional neural network (CNN)	Average accuracy of 82.84% and concludes that the high frequency bands are more favorable for emotion recognition.	Only studied the transfer learning method of fine-tuning deep neural networks.
Lin et.al [6]	Transform different frequency band into six gray images, contains time and frequency domain information.	end-to-end learning of deep convolutional neural network (CNN)	Average accuracy comes to around 86%	The difficulty here is to develop and fit a generalized classifier that will work well for all individuals.
Zheng et.al [7]	Parallel GCNNs are constructed to extract graph domain features from each feature cube.	Proposed a fusion model of LSTM and GCNN for emotion classification (ECLGCNN)	The average accuracy of 84.81% and 85.27% in subject-independent experiments.	Only explored the effectiveness of ECLGCNN on the binary classification of emotions.
Y Cimtay et.al [8]	Extra pooling and dense layers are added to the pretrained model in order to increase its depth, so that the classification capability is enhanced.	pretrained state-of-the-art Convolutional Neural Network (CNN) architectures	It produces a mean prediction accuracy of 81.8% with a standard deviation of 10.9, post media filtering increases the mean accuracy by approximately 4.5%	One of the most important issues is the data themselves and the quality and reliability of the labels of the data.
Zhang et.al [9]	Comparison experiments with different models, including epoch, learning rate, and dropout probability, were also conducted in the paper	Comparison of DNN, CNN, LSTM), and a hybrid model of CNN and LSTM (CNN-LSTM).	Experimental results show that the CNN and CNN-LSTM models had high classification performance, and their accurate extraction rate of RAW data reached 90.12 and 94.17%, respectively.	Obtain more data on EEG signals and implement other EEG-based emotional recognition models with more variables considered.
Wei et.al [10]	Simple Recurrent Unit models were established considering time changing in emotion.	EEG signals were decomposed and reconstructed using DT-CWT transform.	This method led to an average accuracy of 85.65%, 85.45% and 87.99% with low/high arousal, valence and liking classes, respectively.	The machine learning model constructed in this paper was subject-specific. The well-learned parameters are appropriate for one subject, but may not suitable for others.

Chapter 3

Methodology

In the previously presented work, researchers proposed various methods to extract different features from raw EEG signals. Different types of classifiers were applied to extracted features to recognize emotion. In order to improve the accuracy of emotion recognition system, In this work, a new method is proposed to recognize emotion from raw EEG signals as shown in Figure 3.1 directly by using an end-to-end deep learning approach. The proposed method improves the accuracy as discussed in Section V. The proposed model illustrated in Figure 3.2 represents the overall flow of our work. [2].

Initially, the EEG signals were accumulated, and then underwent thorough pre-prosessing to ensure data quality. Next, a feature extraction process was performed on the pre-prosessed data, to identify the relevant features for our analysis. From the extracted features, we carefully selected the most appropriate ones to be fed into the classifier. Finally, we used a 3DCNN+RNN classifier to effectively classify the selected features and accurately recognize emotions.

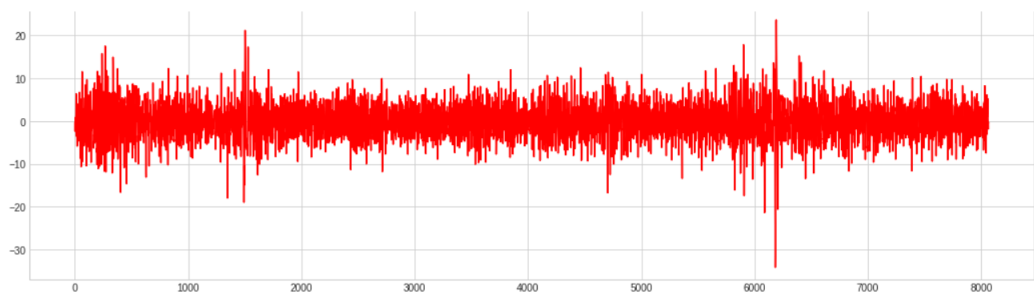

Figure 3.1: Raw EEG signal

Figure 3.2: Proposed Model

3.1 Data Characterisation

For our study, we utilized the DEAP dataset as a source for evaluating emotions. This dataset contains EEG and peripheral physiological signals that were recorded from 32 participants while they watched 40 music videos. Specifically, the dataset includes 32-channel EEG signals and eight-channel peripheral physiological signals, with the EEG signals used for emotion recognition while the peripheral physiological signals were disregarded. The emotional music videos consist of 40 one-minute clips, and participants were asked to rate their levels of arousal, valence, liking, and dominance for each video on a scale of 1 to 9 [13].

Table 3.1: Data included in each subject file

Array Name	**Array Shape**	**Description**
Data	$40 \times 40 \times 8064$	video/ trial $\times$ channel $\times$ data
Label	40×4	video/ trial $\times$ label

In many studies conducted on emotion classification based on an EEG signal, a 2D arousal-valence emotion description model [14] is used to represent different emotions in a 2D plane, as depicted in Figure 4.6. The VA dimensional model utilizes a two-dimensional function to represent all emotional states, comprising of two continuous dimensions. The first dimension, valence, reflects the level of pleasure or displeasure of an emotion. The second dimension, arousal, assigns a numerical value to the level of mental activity, which ranges from low engagement to ecstasy. By using this model, emotions can be classified into four distinct categories based on their combinations of valence and arousal: high arousal / low valence (HALV), high arousal / high valence (HAHV), low arousal / low valence (LALV), and low arousal / high valence (LAHV). Nonetheless, this method of representing emotions may not be entirely precise for certain emotions.Russell and Mehrabian [15] observed that fear and anger share a common characteristic of high arousal and low valence, leading to their classification

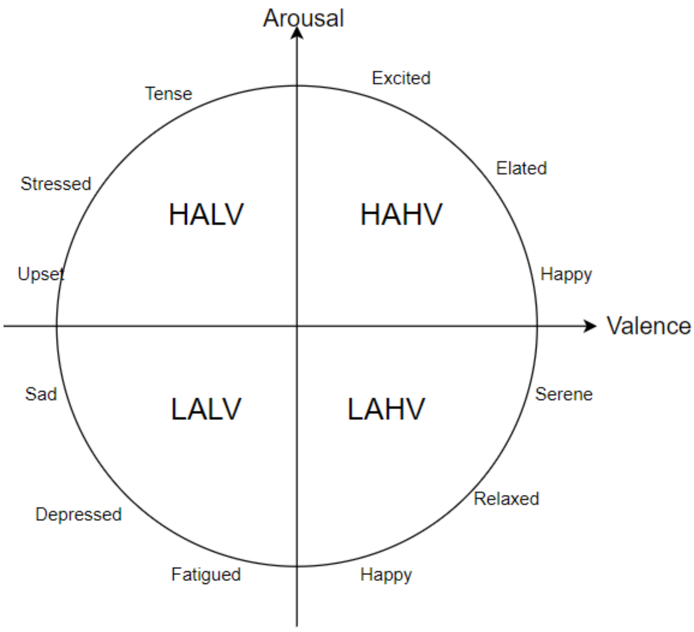

Figure 3.3: Valence Arousal 2-D model

in the same category. However, to achieve more precise distinctions between various emotional states, an extension of the VA model has been proposed. The Valence-Arousal-Dominance (VAD) model employs an extra continuous dimension, namely dominance, to represent emotional states. Dominance reflects the degree of control in a situation, ranging from completely dominant to submissive. The research of Russell and Mehrabian on the Three-Factor Theory demonstrated that dominance is necessary and sufficient to define all emotional states. In our study, we utilize the additional information of dominance to represent emotions using the VAD model.

3.2 Data Pre-Processing

During the experiment, the EEG signals were sampled at 512 Hz and then down sampled to 128 Hz. Then, averaged to the common reference, after that, eye artifacts were removed, and a high-pass filter was applied. Additionally, electrooculography (EOG) artifacts were removed using the blind source separation technique. The pre-processed EEG data for each trial contains 60 seconds of trial data and 3 seconds of baseline data.Each subject file contained two arrays, and the data format of these files is detailed in Table 2.1 . In order to process the continuous EEG signal, the technique of epoching is employed, which involves extracting specific time windows from the signal. For this purpose, we first extract 2-second wavelength segments from each signal, with a time step of 2 seconds. This process yields a total of 31 signals for each channel, with 256 data points representing each epoch, which corresponds to 2 seconds of the original signal.

Apart from processing the input data, we also processed the output data in this study. The original valence, arousal, and dominance ratings of the DEAP dataset were initially scored on a scale ranging from 1 to 9. However, to classify the data, we divided these ratings into three distinct classes using a threshold value of 5. As such, if the valence, dominance and arousal ratings were below the threshold value, we encoded them as 0 (representing low valence/arousal/dominance). Conversely, if the ratings were equal to or greater than the threshold value, we encoded them as 1 (representing high valence/arousal/dominance). In the following, we form the final label by combining these categories according to the representation of the VAD model.This threshold-based encoding approach was used to facilitate the classification process.

3.3 Feature Extraction

In EEG-based emotion recognition research, feature extraction is a crucial step in reducing the dimensionality of the feature space to obtain a set of highly effective features for classification, thereby increasing the accuracy of classification. EEG signals can be analyzed using three main types of features: time-domain, frequency-domain, and time-frequency domain features [16]. To extract these features, various methods are available such as fast Fourier transform (FFT), wavelet transform (WT), and time-frequency distribution (TFD). However, in this project, a 3D CNN and RNN are used, along with FFT as a feature extraction method. FFT is the most widely used due to its speed, efficiency, and accuracy in computing Fourier transform sequences. By extracting relevant frequency components from the signal, FFT can provide informative features that can be used to distinguish between different emotional states. The equation for FFT is shown in Figure 3.4

Frequency-domain analysis is used to divide the EEG signal into different frequency bands such as delta, theta, alpha, beta, and gamma for feature extraction.

$$F(x) = \sum_{n=0}^{N-1} f(n) e^{-j2\pi(x\frac{n}{N})}$$

Figure 3.4: Equation for the Fourier transformation

The FFT is a faster version of the Fourier transformation that involves taking the Fourier transform over short windows of the signal, reducing the time needed for the transformation [17]. This approach enables us to obtain highly informative features for classification, thereby improving the accuracy of EEG-based emotion recognition systems.

3.4 Emotion Classification Model

Our proposed hybrid model consists of a 3DCNN that extracts spatiotemporal features from raw EEG signals and an RNN that captures temporal dependencies.The 3DCNN+RNN hybrid model for Emotion Recognition from EEG is a cutting-edge classification model that combines the power of three-dimensional Convolution Neural Networks (3D-CNN) and Recurrent Neural Networks (RNN) for the recognition of emotions from electroencephalography (EEG) data. 3D-CNN is a type of deep learning neural network used for processing 3D data such as videos and medical images. It is an extension of the traditional 2D convolutional neural networks that are widely used in computer vision tasks like image classification and object detection. It uses 3D convolutional kernels to extract spatial and temporal features from the input data. The output of the 3D convolutional layers is typically fed into fully connected layers for classification or regression tasks.

RNN which is a type of deep learning neural network that is designed to process sequential data. Unlike traditional feed-forward neural networks, which process data inputs independently, RNNs have a recurrent connection that allows them to process sequences of inputs in a dynamic and temporal manner.RNNs are particularly useful for processing sequential data such as natural language, speech, time series, and music. They work by maintaining a hidden state that captures the context of the previous inputs in the sequence. The hidden state is updated at each time step by taking a weighted sum of the current input and the previous hidden state.

There are several types of RNN architectures, including the basic RNN, the LSTM (Long Short-Term Memory) network, and the GRU (Gated Recurrent Unit) network. The LSTM and GRU networks are more advanced architectures that are designed to address the vanishing gradient problem that can occur in basic RNNs. Researchers are utilizing EEG signals to address the issue of time correlation among emotions. Since emotions usually endure for varying lengths of time and not just for a brief moment, understanding the connection between different segments of emotions over time can significantly enhance the accuracy of emotion recognition.

The model utilizes a 3DCNN for extracting high-level features from raw EEG signals, and a RNN for capturing temporal dependencies between successive EEG samples. The 3D-CNN first learns spatiotemporal features by convolving over the EEG data in three dimensions (time, frequency, and channel), which are then fed into the RNN for further processing. The RNN then leverages its ability to model sequence data to analyze the temporal dynamics of the EEG signals, and classify the emotion states based on the learned features. The model is trained using a cross-entropy loss function and optimized using stochastic gradient descent.

System Architecture: The proposed model is a 3D-CNN based network architecture that can be optimized for a given problem to achieve more accurate results. Three-dimensional CNNs consist of 3D convolution and 3D pooling operations. To select an optimal chain of different layers, one needs to carefully consider the specific problem and experiment with various combinations. The current model consists of three sets of convolutional layers, max-pooling layers, and dropout layers followed by an RNN layer and a dense layer using the sigmoid activation function.

The input size of the model is defined as $c \times t \times h \times w$, representing the channels, length, height, and width of the 3D EEG stream, respectively. In this study, the default size of the 3D EEG stream used is $6 \times 64 \times 128$, where 6 is the number of consecutive frames processed at once, 32 is the number of channels, and 128 is the number of samples in a frame.

The proposed network architecture can be further optimized by carefully designing the convolution blocks. In each 3D convolution block, a 3D convolution layer is used with a $3 \times 3 \times 3$ kernel and a stride of 1. This is followed by a 3D batch normalization

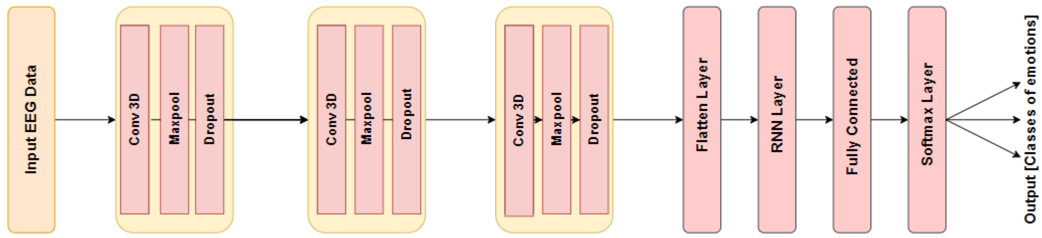

Figure 3.5: Overview Of The Proposed Model

layer, a rectified linear unit (ReLU) activation function, and a 3D max-pooling layer. ReLU is used as the activation function in both convolution layers due to its simplicity and efficiency.To prevent overfitting, a 3D max-pooling layer is used, followed by a dropout layer with a rate of 0.2, which randomly drops 20% of the input neurons. Finally, the model includes two fully connected blocks and a fully connected layer to process the extracted features. A detailed illustration of the proposed network architecture is shown in Figure 3.5.

In summary, optimizing the network architecture by experimenting with different combinations of layers, kernel sizes, and activation functions can improve the accuracy of the model for a specific problem. The proposed 3D-CNN based network architecture can be further optimized by carefully designing the convolution blocks and adding layers to prevent overfitting.

3.5 Classification and Evaluation

Deep learning models are evaluated based on several performance metrics, which may vary depending on the specific task and application. The performance of this model was measured by a set of criteria including accuracy, precision, recall and f-score. Detection errors were measured through the confusion matrix. The performance results of the proposed method were compared with the results of other existing models.

- **Confusion Matrix**:

  A confusion matrix is a table that summarizes the performance of a classification model by comparing its predicted results against the actual results. It helps us visualise the classification performance of each predictive model. The matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 3.6: Confusion Matrix

contains four possible outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

There are 4 different outcomes that can occur in a given column, as shown in 3.6. A true positive result is one where both the actual and predicted classification are positive (1,1), which means that the classifier has correctly identified the positive sample. The positive sample is incorrectly classified as negative by the classifier when the actual classification is positive and the predicted classification is negative (1.0), which is called a false negative result. The four outcomes can be formulated in a 2×2 confusion matrix.

- **Accuracy**: A measure of how closely a value matches the real number. The proportion of all things that are correctly classified is this number. The accuracy was calculated by dividing the number of true results by the total number of results. Accuracy = (TP + TN) / (TP + TN + FP + FN)

- **Precision**: Percentage of the total number of positives that have been accurately classified as such. It measures the proportion of true positives to the total number of true and false positives. The ratio of true positives to the sum of true and false positives is called precision. Precision = TP / (TP + FP)

- **Recall**: It is the number of items correctly identified as positive in relation to the total number of true positives. To measure the model's ability to identify all relevant instances, we calculated the recall. This measures the proportion of true positive predictions among all actual positive instances in the dataset. Recall = TP / (TP + FN)

- **F-Measure**: The F-measure is the weighted harmonic mean of precision and recall The F-score or F-measure is a measure of the accuracy of a test. It is calculated from the precision and recall of the test. The F-score was calculated as the harmonic mean of precision and recall. F1-score = 2 * ((Precision * Recall) / (Precision + Recall))

Chapter 4

Result and Discussions

Automatic recognition of emotions using EEG recordings and nonlinear metrics has become an active area of research. In this section, we present our proposed emotion recognition scheme and its performance compared to state-of-the-art methods. We conducted three types of experiments to evaluate our models, including two-level classification on the DEAP dataset, the results of all the Machine Learning models used were examined and compared, comparison with deep learning methods, and a novel approach focusing on different regions of the brain and also different frequency bands of EEG to classify human emotions into high/low valence and high/low arousal. Our results showed differences in classification accuracy across brain regions. The results of the study are divided into subsections in which the performance of each model is assessed in detail. A statistical investigation of the effectiveness of the model in determining the relevance of features is also conducted. The proposed model utilizes the VAD dimensional space to classify emotions into eight distinct classes. By leveraging this framework, the model is able to capture the valence, arousal, and dominance dimensions that are commonly associated with emotional experiences. This approach allows for a more nuanced and comprehensive understanding of emotional states, and enables the model to achieve greater accuracy in its predictions.

4.1 Comparison Of Binary Class Classification

Binary classification of emotions as valence and arousal using EEG involves predicting whether a given EEG signal corresponds to high or low levels of valence and arousal. Valence refers to the pleasantness or unpleasantness of an emotion, while arousal refers to the level of activation or excitement associated with the emotion. There are many machine learning algorithms that can be used for this task. The choice of algorithm depends on the specific task and the characteristics of the data. It's also important to carefully pre-process and feature engineer the data to ensure that the algorithms perform well.

At first the goal of the study was to compare different machine learning methods for emotion recognition from EEG signals. Four algorithms were selected for comparison: Random Forest(RF), Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Dtree and Random Forest(RF) are used to classify emotional. By comparing the accuracy of the results as shown in figure 4.1 of previous studies, we conclude that although the dataset is the same, there are different levels of classification accuracy. Due to the different techniques of extracting features from EEG signals, the different methods of classifying EEG data, and their different parameters. The EEG signals were preprocessed to extract relevant features such as power spectral density and wavelet entropy. These features were used to train and test the models on a dataset consisting of emotional responses from participants. Compared with different other models K-NN achieve an testing accuracy of 63.1%. Cross-validation and feature selection techniques can be used to improve the performance of the algorithms.

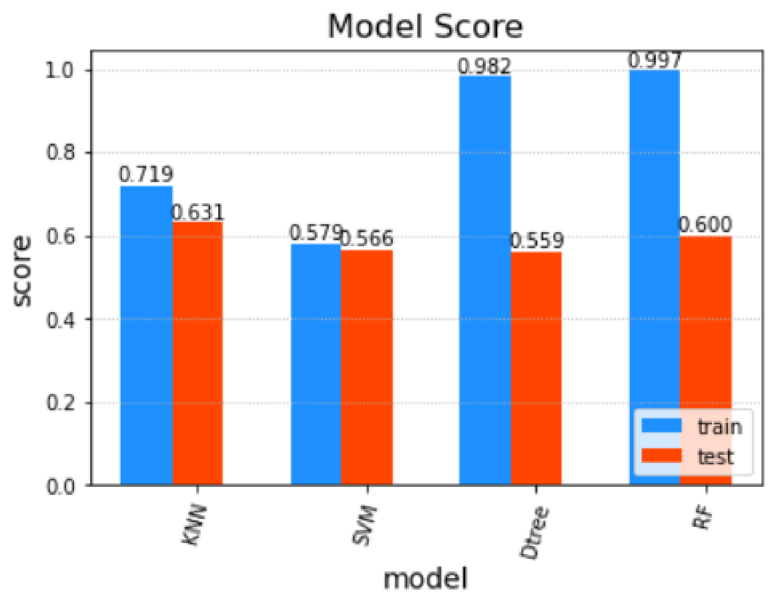

Figure 4.1: Comparison of Machine Learning Models

4.2 Binary Class Classification - Proposed Model

The combination of 3DCNN and RNN has been shown to be effective for emotion recognition from EEG signals, as it allows the model to extract both spatial and temporal features from the input data and model the dynamic changes in emotion over time. Firstly, we will elaborate on the performance of both models used in this study based on the accuracy and loss value of the model as shown in Figure 4.2 and Figure 4.3.

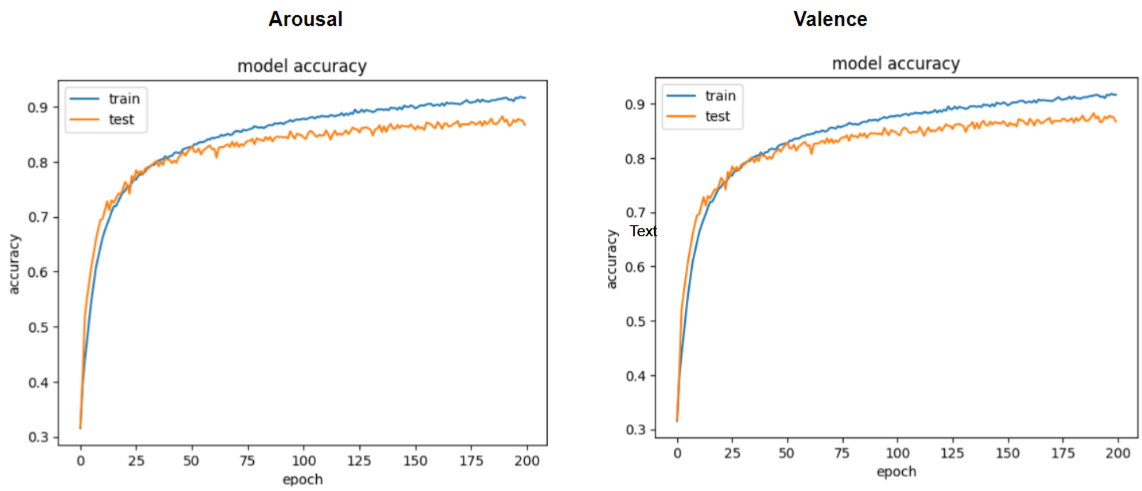

Figure 4.2: Accuracy Graph for Binary Class

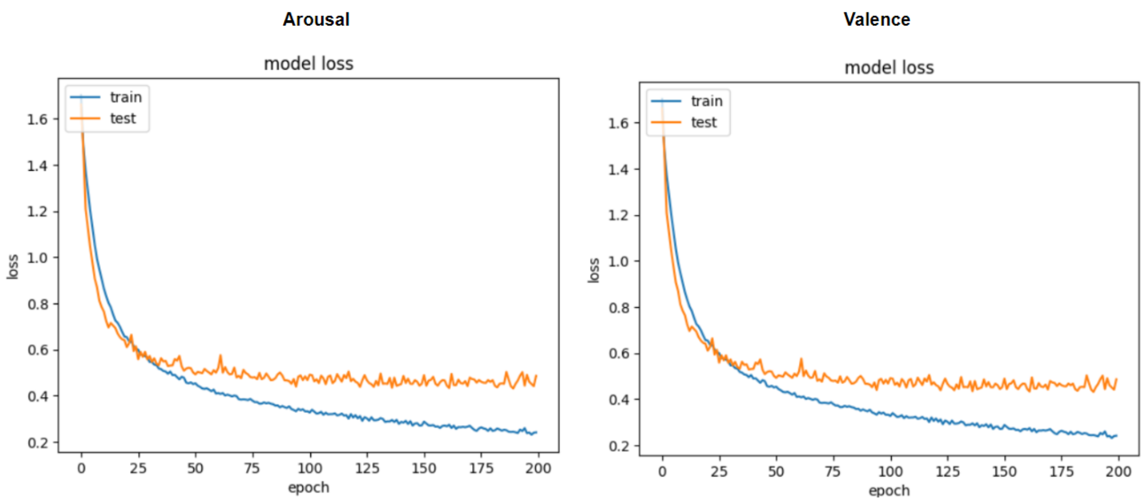

Figure 4.3: Loss Graph for Binary Class

Performance is assessed using classification accuracy, precision, recall (sensitivity), f1-score and also confusion matrix. The correlation matrix is an effective tool for finding and displaying trends in the data provided and for summarising a large data set. For our emotion prediction dataset, we plotted the correlation matrix, refer to Figure 4.4. This figure demonstrates the normalized confusion matrix of arousal and valence where arousal and valence is divided into HAHV HALV LAHV LALV. For all classes , the condition of true positive and true negative is looking better. It contains the classification report for both arousal and valence which shows the precision, recall and f1- score very clearly and also it is tabulated in the Table 4.1

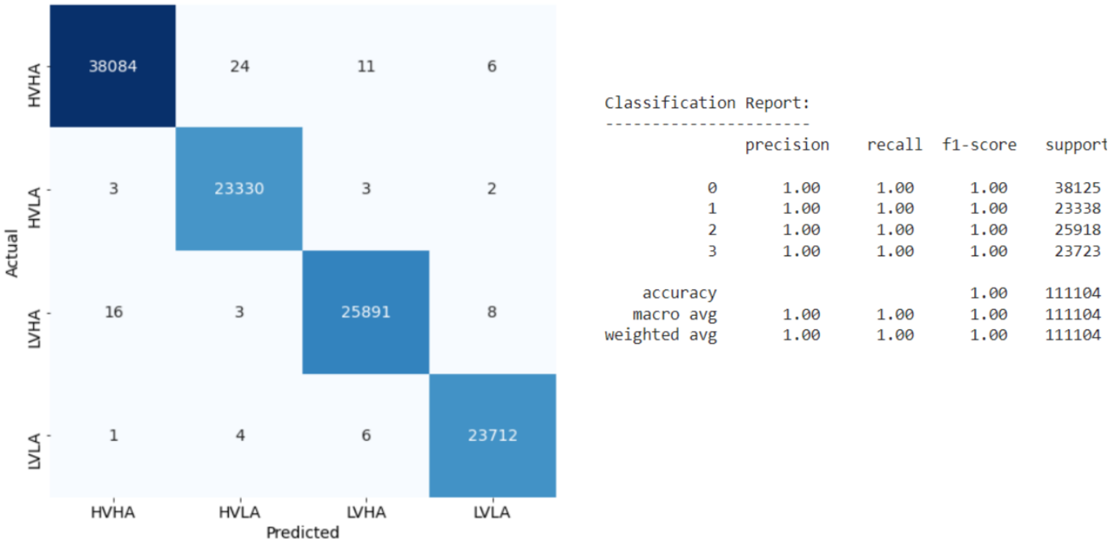

Figure 4.4: Confusion Matrix and Classification Report

Table 4.1: Classification Report Of Four Classes emotion

Group	Accuracy	Precision	Recall	f1-score
HVHA	95.45%	1.00	1.00	1.00
HVLA		1.00	1.00	1.00
LVHA	96.63 %	1.00	1.00	1.00
LVLA		1.00	1.00	1.00

4.3 Comparison Using Deep Learning Methods

Recognizing emotions is a crucial aspect of the Human-Computer Interaction process, as it enables machines to better understand and respond to human emotions. To this end, deep learning has emerged as an effective approach for categorizing feelings, leveraging its ability to process vast amounts of data with high efficiency. Compared to traditional machine learning models, deep learning architectures typically contain more layers, which allows them to learn more complex representations of data.

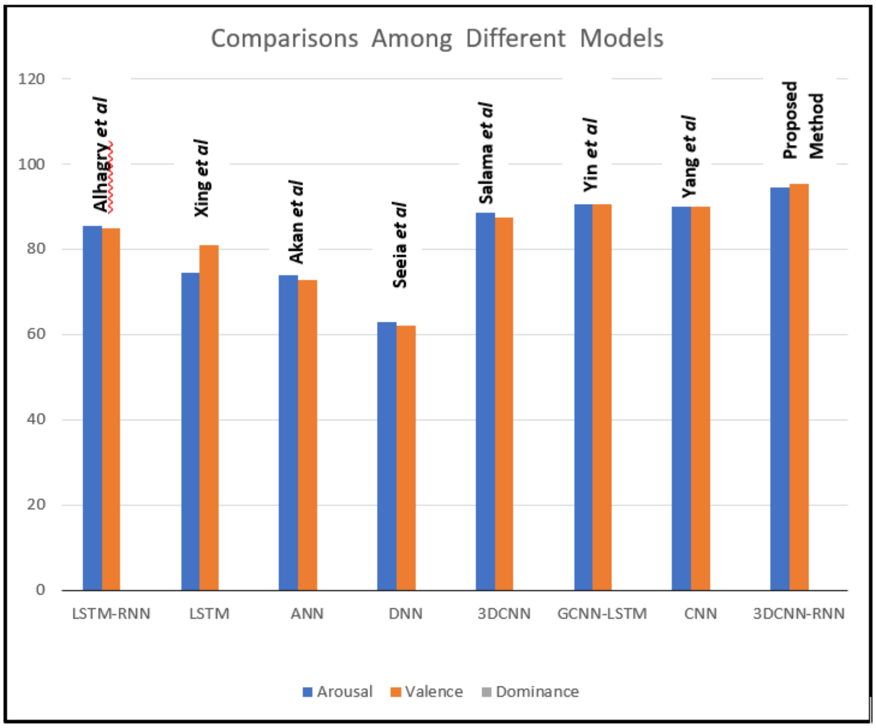

Figure 4.5: Comparison of Deep Learning Models

Figure 4.5 presents a comparison of the results obtained by different emotion classification models, demonstrating the superiority of our proposed model. We trained our model using a large dataset and a rigorous evaluation framework, which resulted in high performance and processing efficiency for emotion classification. Specifically, Table 8 summarizes the performance criteria results for our proposed model, which outperforms existing models in terms of accuracy, precision, and recall.

4.4 Eight Class Classification

Emotion recognition is a complex task that involves analyzing multiple dimensions of emotions to gain a complete understanding of the input emotional data. In fact, classifying emotions in terms of only one dimension at a time can lead to an incomplete description of the given input emotional video. According to the dimensional emotions, happiness is associated with high arousal and high valence. Therefore, if we classify this input based on valence alone or arousal alone, it would not provide a comprehensive description of the emotion being expressed. Instead, a more effective approach is to consider all the dimensions of emotions simultaneously to classify the input emotion accurately. This approach is known as multi-dimensional emotion recognition, and it involves analyzing the valence, arousal, and other dimensions such as dominance and liking/disliking. By considering all these dimensions simultaneously, we can obtain a more complete and accurate description of the input emotion.

Classifying emotions from EEG signals is a challenging task, but recent advances in deep learning have made it possible to achieve high accuracy in this domain. One promising approach is to use the Valence-Arousal-Dominance (VAD) dimensional space, which represents emotions as points in a 3D space [18] as shown in Figure **??**, with Valence indicating how positive or negative the emotion is, Arousal indicating how intense the emotion is, and Dominance indicating how much control the person has over the emotion.

To perform 8-class classification of emotions using the VAD dimensional space, we first need to define the emotional categories we want to distinguish. Approach is, which divides emotions into four quadrants based on their Valence and Arousal levels: High Valence/High Arousal (e.g., happiness, excitement), High Valence/Low Arousal (e.g., contentment, serenity), Low Valence/Low Arousal (e.g., sadness, boredom), and Low Valence/High Arousal (e.g., fear, anger). We can further divide each of these quadrants into two subcategories based on Dominance level: High Dominance (e.g., pride, admiration) and Low Dominance (e.g., shame, submission). This results in eight emotional categories that we can classify using a deep learning model.

To train the model, we need a large dataset of EEG signals labeled with the corresponding emotional category. We can preprocess the EEG signals to extract features that capture the relevant information for emotion classification. We can then train the deep learning model, using the extracted features as input. The model should have multiple output nodes corresponding to the eight emotional categories we want to classify. We can use a softmax activation function to ensure that the output probabilities add up to 1 for each input sample. Finally, we can evaluate the performance of the model using standard metrics such as accuracy, precision, recall, and F1 score. We can also perform a cross-validation experiment to ensure that the model generalizes well to unseen data [19].

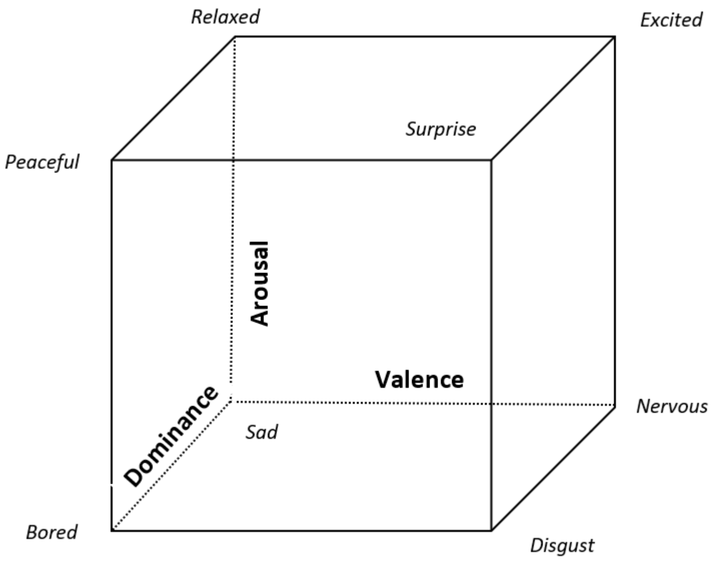

Figure 4.6: Valence Arousal Dominance Model

	precision	recall	f1-score	support
0	0.94	0.94	0.94	6283
1	0.94	0.93	0.94	7015
2	0.95	0.94	0.94	9150
3	0.94	0.94	0.94	9638
4	0.95	0.93	0.94	13542
5	0.94	0.94	0.94	14335
6	0.94	0.94	0.94	12078
7	0.94	0.93	0.93	6039
micro avg	0.94	0.94	0.94	78080
macro avg	0.94	0.94	0.94	78080
weighted avg	0.94	0.94	0.94	78080
samples avg	0.94	0.94	0.94	78080

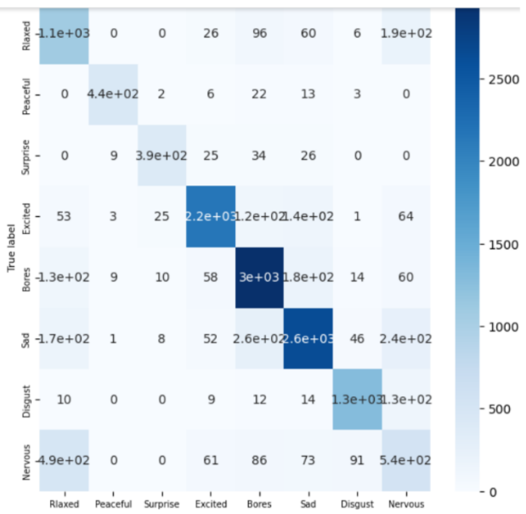

Figure 4.7: Classification Report and Confusion Matrix

Our proposed 3D emotional model represents a significant improvement over Russell's 2D emotional model by incorporating an additional dimension of dominance. Using our 3D emotional model, we have identified eight clusters of emotions, which are mapped and presented in Table 5. Each cluster corresponds to a specific emotional state, including relaxed, peaceful, bored, disgust, nervous, sad, surprise, and excited. By including dominance as an additional pattern for evaluation, our model provides a more comprehensive and nuanced understanding of emotions, allowing for a more accurate classification of emotional states.Also, the confusion matrix of 3DCNN+RNN Classifier with classification report is shown in Figure4.7. The confusion matrix

Table 4.2: 3D Emotion Assignment

Classes	Emotional State
HVLAHD	Relaxed
HVLALD	Peaceful
HVHALD	Surprise
HVHAHD	Excited
LVLALD	Bored
LVLAHD	Sad
LVHALD	Disgust
LVHAHD	Nervous

provides a comparison between actual and predicted values. This results in a 8×8 confusion matrix, which calculates an accuracy of 0.938.

Our proposed method for emotion recognition through the investigation of EEG signals achieves excellent performance in binary classification of emotions as valence and arousal. The VA model contains four classes, namely LV-LA, LV-HA, HV-LA, and HV-HA, while the VAD model includes eight classes, namely LV-LA-LD, LV-LA-HD, LV-HA-LD, LV-HA-HD, HV-LA-LD, HV-LA-HD, HV-HA-LD, and HV-HA-HD, where L denotes low, H denotes high, V denotes valence, A denotes arousal, and D denotes dominance. Table 1 provides the average recognition accuracy, precision, recall, and F1 score for each of the different classes of emotion.

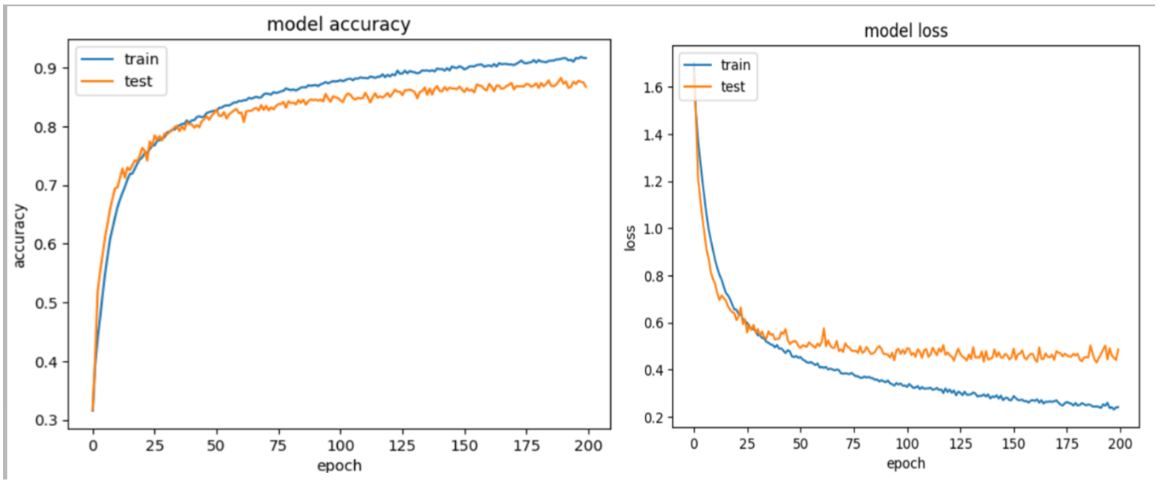

Figure 4.8: Accuracy and Loss Graph For VAD Model

Our method employs a combination of feature extraction and machine learning algorithms, such as Fast Fourier Transformation (FFT) and 3D-CNN+RNN models, to classify emotions based on EEG signals. The results of our experiments demonstrate

Table 4.3: Performance Evaluation Of Eight Class classification

Group	**Accuracy**	**Precision**	**Recall**	**f1-score**
HVLAHD - Relaxed	94.46%	0.94	0.94	0.94
HVLALD - Peaceful		0.94	0.93	0.94
HVHALD - Surprise	95.75 %	0.95	0.94	0.94
HVHAHD - Excited	%	0.95	0.93	0.94
LVLALD - Bored	96.64 %	0.94	0.94	0.94
LVLAHD - Sad		0.95	0.94	0.94
LVHALD - Disgust	94.45%	0.94	0.93	0.94
LVHAHD - Nervous		0.94	0.94	0.94

that our method achieves high accuracy, precision, recall, and F1 score for both the VA and VAD models, as shown in Table 4.3 and also accuracy and loss plot is shown in the figure 4.8. These findings suggest that our method is effective in classifying emotions based on EEG signals and can potentially be applied in various real-world applications, such as mental health monitoring and human-computer interaction.

Chapter 5

Conclusion

We present a novel hybrid model that combines the power of 3DCNN and RNN for emotion recognition from EEG signals, which achieves high classification average accuracies of 96.11% in binary classification of valence and arousal, and four-class classification. Our experimental results show that the proposed method outperforms previous approaches, including those using handcrafted features. This suggests that the proposed model can significantly improve emotion recognition from EEG signals, producing highly accurate results compared to previous works in the same domain.

The advantage of our 3D-CNN+RNN model lies in its ability to extract spatial and temporal features in one end-to-end model. Our results demonstrate that emotions can be effectively classified into Valence-Arousal-Dominance space, and that taking more emotional features into consideration could further improve classification accuracy. By applying our approach, we introduce a method for multiple emotion recognition within the Valence-Arousal-Dominance space, with a high recognition rate of 93.73%. This indicates that our model is highly accurate and can reliably predict a person's emotional state. Moreover, through the representation of the VAD space, we introduce a method to categorize emotions into more subtle groups, making certain emotions more distinguishable. This can be a decisive advantage in practical applications. Our proposed 3DCNN+RNN models are highly effective for emotion recognition, outperforming previous research in terms of accuracy, while also being simple and light in terms of architecture.

This enhanced emotional model has the potential for numerous applications in various fields, such as psychology, neuroscience, and human-computer interaction. It can be used to develop more advanced emotion recognition systems, which can better interpret and respond to human emotions, leading to more effective and personalized user experiences. Furthermore, the model can also be used in clinical settings to better understand and treat emotional disorders, such as anxiety and depression. Overall, our proposed 3D emotional model represents a significant step forward in the field of emotion recognition and has the potential to make a meaningful impact on a wide range of applications.

Chapter 6

Future Scope

Although our proposed model has achieved significant results in emotion classification, further evaluation on other datasets and improving its robustness are necessary. Our next step is to implement a k-fold cross-validation for more accurate performance estimation. Additionally, we plan to integrate our model into a real-time application to assess emotional states of subjects. This can be applied in the human resources field to gain insights into employee emotional states.

Future studies should focus on analyzing other EEG signal attributes, such as nonlinear and spatial domain features, in addition to the FFT employed in this work. Exploration of additional EEG feature processing methods, such as multiscale principal component analysis and empirical wavelets, could be beneficial.

We aim to work with our methodology on real-time data to enable easier expression of emotions for mentally challenged and autistic individuals. In the future, we will focus on making our recognition models more efficient and portable. Our approach has broad applications in various fields, including neuroscience, psychology, and affective computing, and can contribute significantly to the development of more accurate and reliable emotion recognition systems.

References

[1] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel eeg data through convolutional recurrent neural network," in *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2016, pp. 352–359.

[2] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on eeg using lstm recurrent neural network," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.

[3] F. P. George, I. M. Shaikat, P. S. Ferdawoos, M. Z. Parvez, and J. Uddin, "Recognition of emotional states using eeg signals based on time-frequency analysis and svm classifier." *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 9, no. 2, 2019.

[4] R. Sharma, R. B. Pachori, and P. Sircar, "Automated emotion recognition based on higher order statistics and deep learning algorithm," *Biomedical Signal Processing and Control*, vol. 58, p. 101867, 2020.

[5] F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and eeg-based efdms," *Neuropsychologia*, vol. 146, p. 107506, 2020.

[6] W. Lin, C. Li, and S. Sun, "Deep convolutional neural network for emotion recognition using eeg and peripheral physiological signal," in *Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, September 13-15, 2017, Revised Selected Papers, Part II 9*. Springer, 2017, pp. 385–394.

[7] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "Eeg emotion recognition using fusion model of graph convolutional neural networks and lstm," *Applied Soft Computing*, vol. 100, p. 106954, 2021.

[8] Y. Cimtay and E. Ekmekcioglu, "Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition," *Sensors*, vol. 20, no. 7, p. 2034, 2020.

[9] Y. Zhang, J. Chen, J. H. Tan, Y. Chen, Y. Chen, D. Li, L. Yang, J. Su, X. Huang, and W. Che, "An investigation of deep learning models for eeg-based emotion recognition," *Frontiers in Neuroscience*, vol. 14, p. 622759, 2020.

[10] C. Wei, L.-l. Chen, Z.-z. Song, X.-g. Lou, and D.-d. Li, "Eeg-based emotion recognition using simple recurrent units network and ensemble learning," *Biomedical Signal Processing and Control*, vol. 58, p. 101756, 2020.

[11] A. Giri, M. V. V. Bhagavath, B. Pruthvi, and N. Dubey, "A placement prediction system using k-nearest neighbors classifier," in *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)*, 2016, pp. 1–4.

[12] N. Zhuang, Y. Zeng, K. Yang, C. Zhang, L. Tong, and B. Yan, "Investigating patterns for self-induced emotion recognition from eeg signals," *Sensors*, vol. 18, no. 3, p. 841, 2018.

[13] 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, 2011.

[14] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology*, vol. 17, no. 3, pp. 715–734, 2005.

[15] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of research in Personality*, vol. 11, no. 3, pp. 273–294, 1977.

[16] D. Nath, M. Singh, D. Sethia, D. Kalra, S. Indu *et al.*, "An efficient approach to eeg-based emotion recognition using lstm network," in *2020 16th IEEE international colloquium on signal processing & its applications (CSPA)*. IEEE, 2020, pp. 88–92.

[17] M. Akin, "Comparison of wavelet transform and fft methods in the analysis of eeg signals," *Journal of medical systems*, vol. 26, pp. 241–247, 2002.

[18] H. Dabas, C. Sethi, C. Dua, M. Dalawat, and D. Sethia, "Emotion classification using eeg signals," in *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, 2018, pp. 380–384.

[19] M. Gaertner, D. Sauter, H. Baumgartl, T. Rieg, and R. Buettner, "Multi-class emotion recognition within the valence-arousal-dominance space using eeg." in *AMCIS*, 2021.