ABSTRACT

This study explores EEG-based emotion recognition using a hybrid deep learning model of 1D-CNN and LSTM networks. The study aims to improve the classification accuracy and feature extraction efficiency of two-dimensional four-classified emotions, focusing on raw EEG data analysis without prior pre-processing or feature extraction. Experiments using the DEAP dataset demonstrate that the model is effective in automatically learning and accurately predicting emotion labels from raw data. A 5-fold cross-validation experiment is also conducted to validate the model's subject-independent generalization ability and robustness. Remarkably, the proposed model achieves an exceptional validation accuracy of 96.37%, showcasing the efficacious fusion of CNN's prowess in feature learning and LSTM's aptitude in processing sequential data. These findings are instrumental in substantiating the hypothesis that the integration of CNN and LSTM into a cohesive hybrid model can significantly propel the field of EEG-based emotion recognition forward. Thus, the study not only highlights the potential of integrating CNN and LSTM into hybrid models for EEG-based emotion recognition, but also points out new directions for the development of this area.

KEYWORD: 1D-CNN, LSTM, DEAP, Deep Learning, Emotion Recognition.

TABLE OF CONTENTS

DECLARATION	**ii**
APPROVAL FOR SUBMISSION	**iii**
ACKNOWLEDGEMENTS	**v**
ABSTRACT	**vi**
TABLE OF CONTENTS	**vii**
LIST OF TABLES	**ix**
LIST OF FIGURES	**x**
LIST OF SYMBOLS / ABBREVIATIONS	**xi**

CHAPTER

1	**INTRODUCTION**	**1**
	1.1 Background	**1**
	1.2 Aims and Objectives	**3**
	1.3 Thesis Content	**4**
2	**LITERATURE REVIEW**	**5**
	2.1 Overview of EEG-Based Emotion Recognition	**5**
	2.1.1 Datasets in EEG Emotion Recognition Research	**5**
	2.1.2 Classification of Emotions	**7**
	2.1.3 EEG Signal Feature Extraction	**8**
	2.1.4 Trends in Classification Techniques	**9**
	2.2 Comparative Studies of Deep Learning	**11**
	2.2.1 CNN	**11**
	2.2.2 DBN	**12**
	2.2.3 LSTM, RNN	**13**
	2.2.4 Hybrid Model	**14**
	2.2.5 Subject-Independent Emotion Recognition Systems	**15**

3	**RESEARCH METHODLOGY**	**17**
	3.1 Research Question	**17**
	3.2.1 Electroencephalogram	**18**
	3.2.2 Data Description	**20**
	3.2.3 Label Description	**21**
	3.3.1 Data Epoching and Extraction	**22**
	3.3.2 Feature Extraction	**23**
	3.3.3 Data Preprocess	**24**
	3.3.3.1 Z-Score	**24**
	3.3.3.2 Min-Max Scaling	**25**
	3.4 Proposed Model	**26**
	3.4.1 1D-CNN	**26**
	3.4.1.1 One-Dimensional Convolutional Layer (Conv1d)	**27**
	3.4.1.2 ReLU Activation Function	**27**
	3.4.1.3 MaxPool1d Pooling Layer	**27**
	3.4.1.4 Dropout Layer	**27**
	3.4.2 LSTM	**28**
	3.4.3 Fully Connected (FC) Layer	**29**
	3.5 Cross-validation	**30**
	3.6 Early Stopping	**31**
4	**RESULT AND DISCUSSION**	**32**
	4.1 Experiment Setup	**32**
	4.2 Experiment Results	**33**
	4.3 Comparison with Published Studies	**35**
	4.4 Discussion	**36**
5	**CONCLUSION**	**38**
	5.1 Conclusion	**38**
	5.2 Limitation and Future Work	**39**
	5.3 Data Availability Statement	**40**
REFERENCES		**41**
APPENDICES		**47**

LIST OF TABLES

Table 2.1:	Datasets used in EEG Emotion Recognition Research	6
Table 2.2:	Overview of EEG Signal Feature Extraction Methods	9
Table 3.1:	Characteristics of EEG Wave Frequency Bands	19
Table 3.2:	DEAP Dataset Structure	21
Table 4.1:	Performance Metrics of Purposed Model	34
Table 4.2:	Comparison of Published Studies on DEAP using Deep Learning	36

LIST OF FIGURES

Figure 2.1:	2-Dimensional Emotion Model	8
Figure 2.2:	3-Dimensional Emotion Model	8
Figure 3.1:	EEG Wave Frequency Bands	19
Figure 3.2:	Different Channel Settings for EEG Electrodes	20
Figure 3.3:	4-Class Emotion Classification Based on Valence and Arousal	21
Figure 3.4:	Sample of Raw Data	22
Figure 3.5:	Sample of FFT Feature Extraction Data	24
Figure 3.6:	Architecture of Proposed Hybrid Model (CNN)	28
Figure 3.7:	Architecture of Proposed Hybrid Model (LSTM)	29
Figure 4.1:	Training & Validation Accuracy	33
Figure 4.2:	Training & Validation Loss	33

LIST OF SYMBOLS / ABBREVIATIONS

α	Alpha
β	Beta
δ	Delta
γ	Gamma
θ	Theta

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DL	Deep Learning
DT	Decision Tree
ECG	Electrocardiogram
EEG	Electroencephalogram
FFT	Fast Fourier Transform
HCI	Human-Computer Interaction
HMI	Human-Machine Interaction
KNN	K-Nearest Neighbor
LSTM	Long Short-Term Memory
ML	Machine Learning
NB	Naive Bayes
RNN	Recurrent Neural Network
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 Background

The significance of understanding human emotions spans multiple disciplines, including Psychology, Neuroscience, Artificial Intelligence (AI), and Human-Computer Interaction (HCI). Emotions are complex in nature and have a profound influence on aspects of human cognition, decision making, and overall behavior (Koole & Rothermund, 2022). These emotional states are not only central to daily human interactions and social communication, but also play a subtle yet critical role in influencing responses and actions. Accurately recognizing emotions is essential in the healthcare sector for effective patient care and treating mental health conditions (Awais et al., 2021). The field of education also benefits from an understanding of emotional states, as this knowledge can significantly improve teaching methods and promote better learning outcomes (Sarmiento-Calisaya, Calcina-Ccori, & Parari, 2022). Additionally, in the entertainment industry, a deep understanding of audience emotions is key to creating content that is more engaging and provides an immersive experience (Wickramasinghe et al., 2021). The importance of emotion recognition is highlighted across various sectors, emphasizing the need for ongoing research and development in this area.

In the fields of artificial intelligence and robotics, accurately interpreting emotions can significantly enhance the effectiveness of human-machine interfaces, fostering more intuitive and empathetic interactions (Pal, Mukhopadhyay, & Suryadevara, 2021). A notable application can be observed in customer service, where AI-driven chatbots and virtual assistants embedded with emotion recognition capabilities effectively analyze customer sentiment through voice modulation and text analysis. Andrade and Tumelero (2022) noted that these advances could lead to more personalized and empathetic interactions, which in turn significantly increase customer engagement and satisfaction. Meanwhile, in the field of mental health care, AI is revolutionizing the approach to emotion recognition. Innovative AI applications are being developed to detect emotional distress or mental health issues by analyzing speech patterns and facial expressions (Vasaikar et al., 2022). AI systems are able to detect subtle changes in vocal tonality or facial micro-expressions that may indicate depression or anxiety, facilitating early detection and intervention. This capability of accurately recognizing and analyzing human emotions transcends various sectors, driving the development of advanced emotion recognition methods in the field of AI, reflecting the profound and far-reaching implications of this technology.

In the evolving landscape of Emotion Recognition using AI, the analysis of emotions through Electroencephalogram (EEG) signals has become a pivotal area of research. This surge in interest largely stems from advancements in machine learning (ML) and deep learning (DL), which have notably enhanced the capacity to decode complex EEG data. As Cai et al. (2021) elucidate in their review, machine learning in EEG emotion recognition predominantly employs traditional algorithms necessitating manual feature extraction from EEG signals. These conventional methods typically involve the incorporation of statistical or frequency-based attributes, subsequently analyzed through classifiers like Support Vector Machines (SVMs) or Decision Trees (DTs). While these techniques exhibit a degree of effectiveness, their reliance on expert knowledge for feature selection and their vulnerability to the inherent variability in EEG signals present notable limitations. Conversely, deep learning offers in a transformative approach to EEG emotion recognition. DL models, exemplified by Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are equipped to autonomously extract features directly from raw EEG data (Shen et al., 2020). This represents a substantial progression, effectively circumventing the labor-intensive and subjective nature of manual feature extraction that characterizes traditional ML methods. Furthermore, deep learning models are adept at capturing more intricate and abstract data representations, crucial for discerning the nuanced patterns associated with emotional states in EEG signals, thereby heralding a new era in the field of emotion recognition using AI.

In conclusion, the integration of deep learning with EEG for emotion recognition is a dynamic and evolving field. Despite challenges that need to be carefully addressed, it has great potential for breakthroughs in understanding human emotions through brain activity.

1.2 Aims and Objectives

This research aims to advance the field of emotion recognition using electroencephalogram (EEG) signals by developing a robust hybrid architecture model. The primary objective is to investigate the effectiveness of a combined one-dimensional convolutional neural network (1D-CNN) and long short-term memory (LSTM) network in accurately recognizing emotional states. Uniquely, this study adopts for a two-dimensional, four-category emotion model to address the gaps in more comprehensive emotion classification, a departure from the prevalent focus on binary valence and arousal categorization. A variety of dataset processing methods are employed, encompassing data preprocessing techniques, feature extraction methods, and different EEG channel selections, all aimed at demonstrating the superiority of the hybrid model in feature extraction and overall classification accuracy. Furthermore, to investigate the model's generalizability and robustness across different subjects, subject-independent experiments are conducted. Overall, the objectives of this study can be summarized as follows:

To develop and optimize a hybrid deep learning model, integrating 1D-CNN and LSTM networks for EEG-based emotion recognition. To evaluate model performance on various EEG data processing methodologies, including different preprocessing techniques, feature extraction methods, and EEG channel selections. To validate the model's performance in a subject-independent setting. This involves using cross-validation techniques to ensure the model's generalizability and robustness across different individuals.

1.3 Thesis Content

This thesis is organized into five chapters, beginning with an in-depth exploration of the background and challenges in the field of EEG-based emotion recognition through a comprehensive literature review. Then, the experimental materials and methods are described, and the design and application of the proposed model are elaborated. Subsequently, the paper focuses on summarizing and analyzing the experimental results, providing a critical discussion. The thesis ends with a conclusion of the study, summarizing the main findings and contributions, and suggesting avenues for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of EEG-Based Emotion Recognition: Developments and Trends

This Literature Review is dedicated to a thorough examination of the advancements and methodologies in the realm of EEG-based emotion recognition, emphasizing the integration and comparison of deep learning techniques. It aims to provide a detailed exploration of the current trends, developments, and challenges in extracting and interpreting EEG signals for robust emotion recognition, particularly focusing on the application of various deep learning models. The review will critically analyze how different approaches, from traditional machine learning techniques to innovative deep learning strategies, have contributed to the field, offering insights into their efficacy and limitations. Special attention is given to the comparison between single and hybrid deep learning models, as well as the emerging role of transformer models in enhancing the accuracy and generalizability of emotion recognition systems. The review is organized to systematically cover the historical development of EEG-based emotion recognition, the evolution of emotion classification standards, and the comparative effectiveness of diverse deep learning models, thereby providing a comprehensive overview of the current state and potential future directions of this rapidly evolving field.

2.1.1 Datasets in EEG Emotion Recognition Research

In the specialized domain of EEG emotion recognition, the role of various datasets has been pivotal in advancing research forward and establishing benchmarks for system performance. These datasets, which are critical to the field, offer researchers access to labeled EEG data that is correlated with diverse emotional states. They not only serve as a basis for developing and testing emotion recognition algorithms, but also provide a standard for comparative analysis across studies. Table 2.1 shown below highlights some of the most well-konwn and frequently utilized datasets in this field, each with its unique characteristics and contributions to the advancement of EEG-based emotion recognition research. These datasets, including DEAP, SEED, and DREAMER, have been instrumental in enabling researchers to explore and refine EEG-based methods for emotion detection, providing invaluable resources for the ongoing evolution of this field.

DEAP (Koelstra et al., 2011)	The DEAP dataset is known for its comprehensive approach to EEG-based emotion recognition, integrating EEG, peripheral physiological, and audiovisual data. This multimodal dataset provides a unique advantage over others by offering a broader perspective on emotional states. The inclusion of four-dimensional emotion labels (Arousal, Valence, Dominance, and Liking) enhances its utility for in-depth emotion analysis. DEAP's multiple data types make it particularly useful for studies that require a holistic view of emotional responses.
SEED (Duan et al., 2013)	The SEED (SJTU Emotion EEG Dataset) dataset focuses on EEG signals recorded from subjects while watching movie clips, a method that targets specific emotional responses. This dataset is unique in its approach because it uses a controlled environment to induce and capture emotions, making it ideal for research exploring the impact of visual stimuli on emotional states. SEED's methodology provides a streamlined dataset for researchers interested in the direct effects of controlled stimuli on EEG-based emotion recognition.
DREAMER (Katsigiannis & Ramzan, 2017)	The DREAMER dataset is notable for its inclusion of both EEG and Electrocardiogram (ECG) signals, providing a dual perspective on emotional state analysis. This dataset is particularly valuable for its use of low-cost, off-the-shelf

	devices for data collection, demonstrating an emphasis on accessibility and practicality in emotion recognition research. DREAMER's unique approach of combining EEG and ECG data provides a comprehensive view of the physiological aspects of emotional responses, suitable for multifaceted emotion recognition studies.

Table 2.1: Datasets used in EEG Emotion Recognition Research

2.1.2 Classification of Emotions: Discrete and Multi-Dimensional Models

Emotion recognition is commonly categorized into two principal models: Discrete and Dimensional. The Discrete model posits that humans experience a set of fundamental emotions. A classic example of this model is Ekman's (1992) framework, which identifies six basic emotions: Anger, Disgust, Fear, Happiness, Sadness, and Surprise. In contrast, the Dimensional approach, exemplified by 2-Dimensional (2D) model purposed by Russell et al. (2005), has gained prominence in EEG Emotion Recognition research. This model categorizes emotions along two primary axes: Valence, denoting the positive or negative nature of an emotion, and Arousal, indicating the level of intensity. The 2D emotion model provides a nuanced understanding of emotions, surpassing the discrete model's capabilities. It serves as a foundational framework for emotion studies, the development of emotion recognition systems, and the design of emotion regulation interventions. However, the 2D emotion model has its limitations. While it effectively distinguishes between positive and negative emotions, it struggles to differentiate between emotions sharing similar valence and arousal, such as joy and surprise. To address the need for more refined emotion categorization, Mehrabian (1996) introduced a 3D emotion model, which delineates emotions along three fundamental axes: Valence, Arousal, and Dominance. By encompassing these three dimensions, this framework captures the complexity of emotional experiences more comprehensively, offering a robust tool for analyzing emotional reactions across diverse contexts.

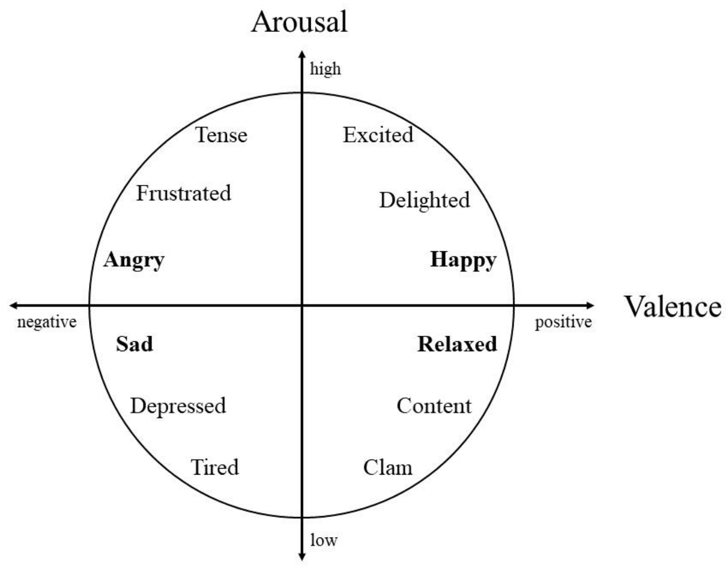

Figure 2.1: 2-Dimensional Emotion Model

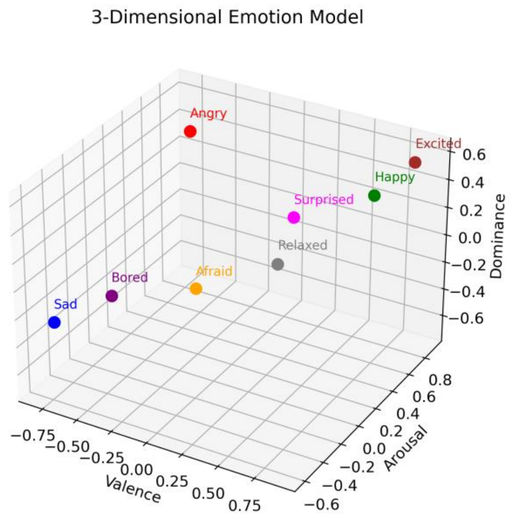

Figure 2.2: 3-Dimensional Emotion Model

2.1.3 EEG Signal Feature Extraction

Feature extraction plays a pivotal role in AI-based algorithms, with distinct approaches in Machine Learning (ML) and Deep Learning (DL). In ML algorithms, feature or pattern extraction is manually conducted before applying to the algorithm, whereas DL algorithms inherently possess the capability to extract features and perform classification, as noted by Kamble and Sengupta (2023). Recent literature, such as the studies by Li et al. (2022), indicate a variety of employed features: Time domain features as originally identified by Hjorth (1970), Frequency domain features detailed by Valdés et al. (1992), and Time-Frequency domain features expanded upon by Roach and Mathalon (2008). These varying approaches to feature extraction and their respective utilizations are concisely summarized in Table 2.2.

Feature Type	Principles and characteristics	Common methods of feature extraction
Time domain	Directly analyze time series data, focusing on changes in signal amplitude over time.	Higher-Order Crossing (HOC) Fractal Dimension (FD)
Frequency domain	Converts a signal into frequency components and analyzes the strength of each frequency component.	Power Spectral Density (PSD) Fast Fourier Transform (FFT)
Time-Frequency domain	Combines the time and frequency domains to provide information about the signal in both time and frequency.	Discrete Wavelet Transform (DWT) Short-Time Fourier Transform (STFT)

Table 2.2: Overview of EEG Signal Feature Extraction Methods

2.1.4 Trends in Classification Techniques for Emotion Recognition

Feature extraction plays a pivotal role in converting raw EEG signals into informative representations that capture relevant patterns and characteristics associated with emotions. This step is fundamental in determining the accuracy of emotion recognition models. However, as Dadebayev et al. (2022) highlight, there is a notable absence of a universally superior feature extraction methodology or classification technique in this domain. This observation underscores the importance of considering deep learning approaches in evaluating the reliability of emotion recognition models. Deep learning methods exhibit unique strengths in handling feature extraction tasks, particularly in the context of EEG signal processing. The complexity of emotions, which encompass a range of intricate cognitive and physiological processes, adds to the challenges in their computational recognition and analysis. This complexity makes the generation and interpretation of emotions inherently subjective, posing significant challenges for traditional machine learning approaches. In light of this, it becomes increasingly evident that relying solely on classical machine learning techniques for developing an emotion recognition method might be insufficient. To address these challenges, Houssein (2022) advocates for the adoption of advanced machine learning methodologies. This includes deep learning, which offers sophisticated pattern recognition capabilities, and transferable machine learning techniques, which can leverage pre-trained models to enhance performance and generalizability in emotion recognition tasks. Such advanced approaches are essential for developing more accurate and robust emotion recognition systems, capable of effectively interpreting the subtle and complex nature of human emotions through EEG signals. This holistic approach, integrating both innovative deep learning strategies and transferable techniques, represents a promising direction in the ongoing pursuit of more effective emotion recognition methodologies.

Machine learning has gained increasing attention in the field of EEG-based emotion recognition. However, this field faces numerous challenges. Traditional machine learning approaches may have limitations, particularly in the aspect of EEG feature extraction. Traditional methods struggle to ensure the extraction of all relevant features, and the processing of a large number of feature channels makes real-time emotion recognition applications challenging (Nath et al., 2020). As a result, researchers are exploring alternative models or solutions, such as deep learning, that deviate from traditional machine learning. Suhaimi et al. (2020) have noted that between 2016 and 2019, traditional machine learning algorithms like Support Vector Machines (SVMs), K-Nearest Neighbors (KNNs), and Naive Bayes (NBs) were primarily used for emotion recognition. Deep learning algorithms were applied comparatively little during this period. However, the research landscape experienced a notable shift as reported by Torres et al. (2020). Starting from 2020, there has been an accelerated integration of deep learning algorithms in emotion recognition studies. This shift is underscored by the more recent analysis of Vempati et al. (2023), indicating a growing preference for deep learning techniques, particularly Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs), and Long Short-Term Memory (LSTMs) networks, reinforcing their status as the preferred methodology in contemporary emotion recognition research.

2.2 Comparative Studies of Deep Learning in EEG Emotion Recognition

2.2.1 CNN

Convolutional Neural Networks (CNNs) have increasingly become the focus of EEG-based emotion recognition due to their exceptional ability to analyze and interpret EEG signals. Originating in the domain of computer vision (Krizhevsky et al., 2017), the application of CNNs has been effectively extended to EEG analysis. This achievement is attributed to their proficiency in capturing spatial dependencies within EEG data, which is a robust capability for understanding the featuress. The work of Cimtay and Ekmekcioglu (2020) was instrumental in showcasing the versatility of CNNs, particularly in cross-subject and cross-dataset training within EEG-based emotion recognition. Utilizing datasets such as SEED, DEAP, and LUMED, their research emphasized CNNs' remarkable performance consistency across varying subjects and datasets, thereby reinforcing the adaptability of CNNs in diverse emotion recognition contexts. Building upon these insights, Wang et al. (2020) delved into the realm of transfer learning techniques applied to CNNs, specifically targeting the challenges posed by individual differences in cross-dataset emotion recognition studies. Their approach, centered on the refinement of deep neural networks through fine-tuning, further elucidated the inherent strength of CNNs in identifying pertinent EEG channels and features—a cornerstone for precise emotion recognition. Furthering the understanding of CNNs' capabilities, Chen et al.

(2022) conducted a study that exemplified the utility of CNNs in extracting both frequency and spatial information from EEG signals, achieving significant accuracy in emotion recognition. This research once again underscored the integral role of CNNs in feature extraction within the EEG domain. Similarly, Mahmoud et al. (2023) also demonstrated the effectiveness of CNNs in feature extraction and emotion recognition, further affirming the robustness of CNNs in this field. In summary, Convolutional Neural Networks stand as a formidable approach in the field of EEG-based brainwave recognition. Their inherent capacity to process spatial dependencies and develop hierarchical representations categorizes them as an indispensable tool in the realm of EEG analysis. The accumulated research in this area not only highlights the effectiveness of CNNs, but also points to their growing importance in advancing the understanding and capabilities of researchers in the field of emotion recognition from EEG.

2.2.2 DBN

Deep Belief Networks (DBNs) are a class of deep learning models that are composed of multiple layers of hidden units. DBNs are specifically designed to learn hierarchical representations of data and have been successfully applied in various domains, including image and speech recognition (Hinton, Osindero, & Teh, 2006). In the context of EEG emotion recognition, DBNs have demonstrated potential in capturing the intricate patterns and dynamics present in brainwave signals. Through their deep architecture, DBNs have the capability to automatically extract meaningful features from raw EEG data, thereby facilitating the development of more precise and reliable emotion recognition models. Wang and Shang (2013) demonstrated the potential of DBNs in predicting emotions by learning features from raw physiological data. While DBN-based methods showed limitations in predicting valence, they performed well in predicting arousal and preference. To enhance performance, a combination of DBN-learned features and hand-designed features was proposed. In another study, Zheng and Lu (2015) investigated the application of

DBNs in EEG emotion recognition. Their findings indicated the presence of neural features associated with positive, neutral, and negative emotions, which were common among individuals. DBN models exhibited higher accuracy and lower standard deviation compared to shallow models like KNN and SVM. In summary, DBNs have shown promise in emotion recognition, extracting informative features from physiological and EEG data. Combining DBN-learned features with hand-designed features improves performance and DBN models outperform shallow models like KNN and SVM. These findings contribute to the advancement of emotion recognition research and highlight the potential of DBNs in this field.

2.2.3 LSTM, RNN

Long Short-Term Memory (LSTM) networks and Recurrent Neural Networks (RNNs), as advanced neural network architectures, are adept at capturing the sequential and temporal nuances in data. RNN, introduced by Schuster and Paliwal (1997), have been instrumental across various fields, including natural language processing and time series analysis. LSTM, introduced by Hochreiter and Schmidhuber (1997), represent an evolution in this area, specifically designed to overcome the vanishing gradient challenge inherent in traditional RNNs. This innovation enables them to effectively model long-term dependencies in data sequences. In the field of EEG emotion recognition, the application of LSTM-RNNs has shown promising results. For instance, a study by Chai et al. (2022) demonstrates the capability of LSTM-RNNs in processing EEG signals and accurately classifying emotions. Their research particularly highlighted the efficiency of these networks in determining emotions through directed functional connectivity, achieving notable accuracy in both valence and arousal detection. This study not only provides insight into the brain's coordination during various emotional states but also underscores the ability of LSTM-RNNs in managing complex temporal data, thereby enhancing emotion classification accuracy. Further solidifying their utility, recent advancements by Wang et al. (2023) illustrate the successful application of RNN and LSTM in capturing the dynamic temporal aspects of brainwave signals. These models excel in identifying patterns and correlations within sequential EEG data, which is pivotal for accurate emotion state recognition. By exploiting the inherently sequential nature of EEG signals, RNNs and LSTMs effectively model temporal dependencies, facilitating the extraction of salient features critical for accurate emotion classification.

2.2.4 Hybrid Model

Li et al. (2016) developed a novel hybrid deep learning model: C-RNN, blending the strengths of CNN and RNN architectures for emotion recognition using multichannel EEG signals. The model combines the strengths of CNN and RNN architectures, utilizing CNN's ability to capture inter-channel correlations and RNN's proficiency in learning long-term dependencies from sequential data. Demonstrating its efficacy in pilot-level sentiment recognition tasks, the C-RNN offers real-time emotion monitoring capabilities by providing predictions for each time step as well as for the entire trial. Complementing this, Jeevan et al. (2019) conducted a study that underscored the superiority of similar models over traditional and deep learning models in emotional task recognition, emphasizing the enhanced emotion classification accuracy of their proposed model. Building upon these advancements, Zhang et al. (2020) contributed significantly to the domain of EEG-based emotion recognition. Their research employed a 1D-CNN combined with an LSTM network to perform four-category emotion classification using the DEAP dataset. Their novel approach effectively utilized the 1D-CNN for capturing spatial features from EEG signals, while the LSTM was proved to be crucial for recognizing temporal patterns in the data. Moreover, Samavat et al. (2022) further contributed to this evolving field with their hybrid multi-input model, which synergizes a CNN with a bidirectional LSTM (Bi-LSTM) network. This model, optimized through adaptive regularization techniques, excels in parsing EEG data, showcasing significant advancements in efficient emotion recognition systems. Building on these foundations, Iyer et al.

(2023) proposed an integrated model that combines CNNs, LSTM networks, and hybrid models for EEG-based emotion recognition. Their extensive experimentation revealed that this integrated model surpasses the performance of standalone CNN and LSTM models, highlighting the combined approach's enhanced accuracy and robustness in emotion recognition tasks. Collectively, these studies mark significant strides in multimodal deep learning for EEG-based emotion recognition, showcasing the compelling benefits of integrating CNN and LSTM networks within hybrid models.

2.2.5 Subject-Independent Emotion Recognition Systems

A major challenge in EEG emotion recognition is the inter-subject variability. This variability includes physiological and neurological differences between individuals, as well as individual differences in the way they express their emotions. These differences lead to difficulties when performing emotion classification and recognition, as the model needs to adapt to the characteristics and patterns of different subjects. For example, there may be differences in brain structure and function between individuals, leading to different expressions in the EEG signal. In addition, the way emotions are experienced and expressed varies depending on individual differences, which further adds to the challenge of emotion recognition. Consequently, it is necessary to consider the distinctions between subject-dependent (SD) and subject-independent (SI) training schemes (Pandey & Seeja, 2022). However, Ravi et al. (2020) found that many current studies favor the SD method based solely on its "high accuracy" without conducting controlled experiments to compare SD and SI approaches or consider the limitations of the SD method. While the SD approach may achieve superior accuracy, it requires individual calibration, which can lead to issues such as poor user compliance. The lack of consideration for individual differences in SD approaches emphasizes the importance of subject-independent (SI) training schemes in EEG emotion recognition. Although the SI approach may slightly trail SD in terms of inter-individual emotion recognition accuracy, its value lies in its generalization ability and robustness. Moreover, in specific medical scenarios, only the SI training scheme can meet the requirements. Thus, the insufficient research on SI schemes hampers progress in the field of EEG emotion recognition. Cimtay and Ekmekcioglu (2020) emphasize the challenging nature of subject-independent EEG emotion recognition. Addressing the varying distributions of spectral band power and its derivatives across subjects and datasets necessitates the utilization of deeper networks, i.e. deep learning. This further underscores the pivotal role of deep learning in advancing the field of EEG-based emotion recognition.

Subject-independent emotion recognition, as noted by Pandey and Seeja (2021), remains an area with relatively few studies and exhibits limited classification accuracy. Earlier work by Jirayucharoensak et al. (2014) using the DEAP dataset reported a subject-independent deep learning accuracy of 53.0% for Valence and 52.0% for Arousal. Subsequently, Rayatdoost and Soleymani (2018) advanced these results slightly, achieving 59.22% for Valence and 55.70% for Arousal. In their study, Pandey and Seeja explored the effectiveness of CNN models trained on DEAP, SEED, and both datasets combined. Notably, they discovered that models trained on one dataset and tested on another, a cross-database criterion, showed improved generalization. This approach led to higher accuracies than previous studies, with Valence reaching 61.50% and Arousal 58.50%. Further, in their subsequent work in 2022, Pandey and Seeja utilized a Deep Neural Network (DNN) that employed the Variational Mode Decomposition method (Dragomiretskiy & Zosso, 2013) for feature extraction, achieving even higher accuracies of 62.50% for Valence and 61.25% for Arousal. These recent advancements indicate a gradual increase in the accuracy of subject-independent emotion recognition. However, the average accuracy in this domain still significantly lags behind that of subject-dependent studies. This gap underscores the need for further exploration and research in subject-independent emotion recognition to bridge the existing research gaps in this field.

CHAPTER 3

RESEARCH METHODLOGY

3.1 Research Question

The formulation of key research questions is an important aspect of this study, aiming to reveal the efficacy and potential of hybrid deep learning models for EEG-based emotion recognition. These questions aim to address key aspects of emotion recognition using deep learning, evaluating the performance of the models relative to traditional methods, the ability to automate feature extraction, and the generalizability and robustness across different topics. The key questions of this research can be summarized as follows:

1. **Comparison with Traditional Methods**: The first research question investigates whether the hybrid deep learning model improves the accuracy of emotion recognition compared to traditional Machine Learning (ML) approaches. This question is fundamental in determining the practical value of the proposed model in EEG-based emotion recognition. It aims to determine whether the integration of CNN and LSTM networks, as seen in the hybrid model, leads to significant improvements in correctly identifying emotional states from EEG data.

2. **Effectiveness in Automatic Feature Extraction**: Another key question explores the effectiveness of the hybrid model in automatically extracting relevant features from EEG signals. This aspect is essential as it addresses the model's ability to autonomously identify patterns and correlations within the EEG data, which are indicative of various emotional states. The model's success in this area would demonstrate its ability to minimize the requirement for manual feature engineering, thus simplifying the process of emotion recognition.

3. **Subject-Independent Performance**: The final research question delves into the subject-independent performance of the proposed model. It aims to assess the universality and robustness of the model across different individuals. This question is vital for evaluating the model's applicability in real-world scenarios, where it must

accurately recognize emotions across a diverse population with varying EEG signal characteristics.

Collectively, these research questions guide the study towards a comprehensive understanding of the proposed hybrid deep learning model's potential to revolutionize EEG-based emotion recognition. They help in critically analyzing the model's performance, its advancements over existing methods, and its applicability in diverse, real-world situations.

3.2.1 Electroencephalogram

Brain signal collection can be categorized into two main approaches: invasive and non-invasive. Invasive methods involve direct extraction of signals from the cortical surface of the brain, while non-invasive techniques utilize external sensors to capture the brain's electrical activity. Among these non-invasive methods, the Electroencephalogram (EEG) is particularly prominent (Kamble & Sengupta, 2023). EEG is the most common non-invasive brain imaging technique. It evaluates brain electrical activity using electrodes placed on the scalp (Rivera-Tello et al., 2022). It captures the brain's five primary electrical patterns or brainwaves: Alpha (α), Beta (β), Delta (δ), Gamma (γ) and Theta (θ) waves. Each brainwave type, with its distinct frequency band, correlates to specific mental states and cognitive functions, as depicted in Figure 3.1 and Table 3.1. EEG's ability to identify and analyze these distinct brainwaves makes it an invaluable tool in various applications, including the study of brain function, the diagnosis of neurological disorders, and the development of brain-computer interfaces. By providing a detailed analysis of these brainwaves, EEG offers profound insights into an array of brain states, ranging from deep sleep to heightened concentration, thereby enhancing our understanding of the intricate dynamics of the human brain.

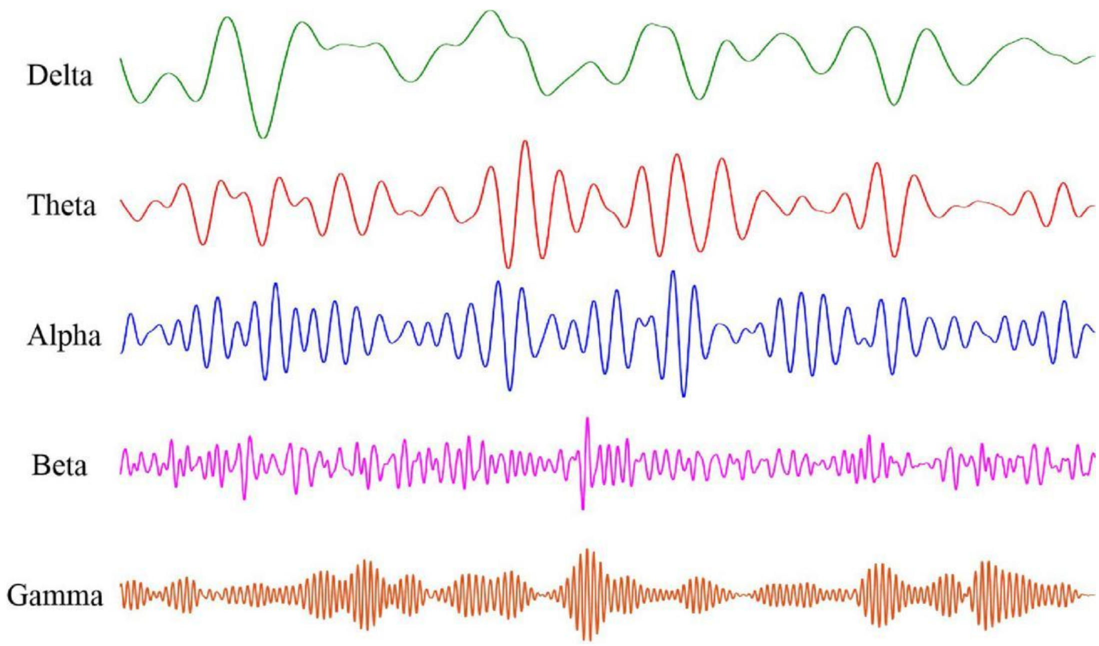

Figure 3.1: EEG Wave Frequency Bands

Brain Wave	Frequency (Hz)	Characteristic
Delta (δ)	0.5 - 4	Predominantly appear during deep sleep. Most prominent in the first two stages of Non-Rapid Eye Movement (NREM) sleep.
Theta (θ)	4 - 8	Typically linked to light sleep, relaxation, and enhanced creativity. Related to memory, learning, and emotional regulation.
Alpha (α)	8 - 12	Present in a relaxed, wakeful state, especially when eyes are closed. Often linked to the brain's resting state and energy restoration.
Beta (β)	13 - 35	Appear during alert and focused states of consciousness. Linked to cognitive tasks, thinking, decision-making, and active communication.
Gamma (γ)	>35	Associated with higher-level information processing, learning, and cognitive functioning. Related to perception, consciousness, and flexible attention regulation.

Table 3.1: Characteristics of EEG Wave Frequency Bands

3.2.2 Data Description

This study utilized the DEAP dataset from Queen Mary University of London as the primary resource. DEAP is a multimodal dataset meticulously curated for the analysis of human affective states using EEG signals (Koelstra et al., 2011). It combines EEG data with peripheral physiological and audiovisual recordings from participants who viewed a sequence of music video clips. The dataset includes EEG recordings from 32 participants, with each participant watching 40 music videos, each lasting one minute. This approach produced a comprehensive data set, rich in both physiological signals and subjective ratings. Participants rated each video on a scale of 1 to 9 based on excitement, capturing dimensions such as arousal, valence, liking, and dominance. The EEG data was meticulously recorded using a 32-channel biosignal acquisition system, ensuring the detailed capture of brain activity.

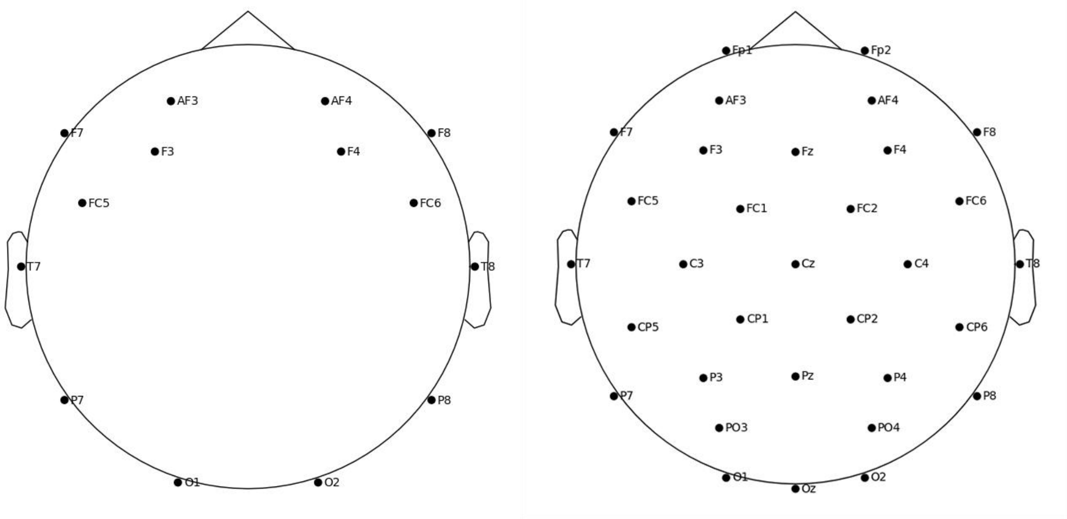

Figure 3.2: Different Channel Settings for EEG Electrodes

Structurally, the DEAP dataset is organized to facilitate ease of use. It includes preprocessed and raw data files. The preprocessed data, already filtered, downsampled, and segmented into epochs, is primed for immediate analysis. For this study, preprocessed data files were utilized, encompassing a total of 32 sets of EEG data corresponding to the experiments conducted on 32 subjects (s01 to s32). Each data set contains two arrays: data and labels, one for the EEG data and the other for the associated labels.

Array	Array Shape	Array Contents
Data	40 * 40 * 8064	video/trail * channel * data
Label	40 * 4	video/trail * label (valence, arousal, dominance, liking)

Table 3.2: DEAP Dataset Structure

3.2.3 Label Description

In the DEAP dataset, each video/trial was rated on a scale ranging from 1 to 9. Therefore, in this study, the discrimination thresholds for valence level and arousal level were set to 5, and then the four emotions were categorized and transformed into discrete labeling format using one-hot coding form as: High-Valence High-Arousal (HVHA), Low-Valence High-Arousal (LVHA), High-Valence Low-Arousal (HVLA), and Low-Valence Low-Arousal (LVLA), denoted as [1,0,0,0], [0,1,0,0], [0,0,1,0] and [0,0,0,1], as shown in Figure 3.2.

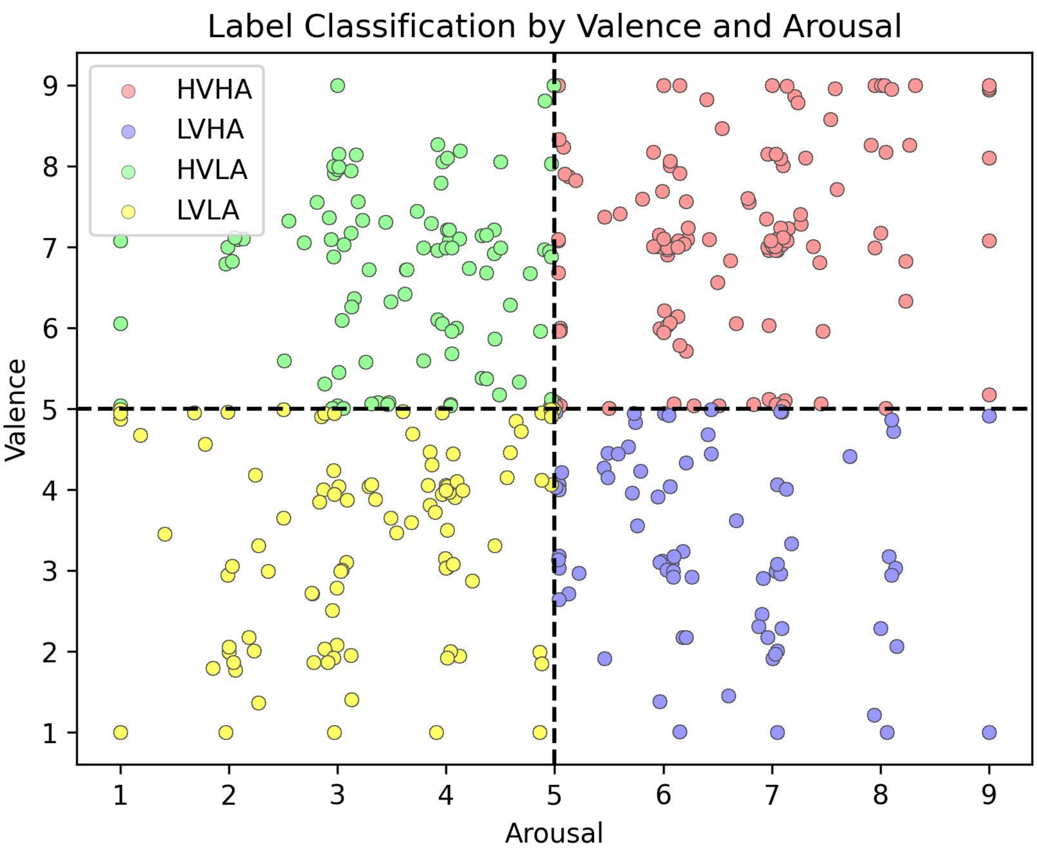

Figure 3.3: 4-Class Emotion Classification Based on Valence and Arousal

3.3.1 Data Epoching and Extraction

In this study, the focal procedure involves epoching and extraction of EEG signals from the DEAP dataset. The procedure commences with segmenting the continuous EEG signals into specific epochs. For analytical precision, a window size of 256 data points is employed, corresponding to a 2-second epoch, based on the sampling rate of 128 Hz. This window size chosen to balance the need for comprehensive data capture against the practicality of dataset manageability. To capture the dynamic character of EEG signals, the data is updated at greater frequency, utilizing a step size of 16 data points. This results in an update every 0.125 seconds, allowing for a detailed and dynamic representation of EEG signal variation. Subsequent to the temporal segmentation, the EEG data is processed for specific channels and predefined frequency bands. The selected bands, denoted by [4, 8, 12, 16, 25, 45], cover five distinct EEG frequency ranges. This selection is pivotal in discerning various cerebral activities and emotional states, each band providing insights into different aspects of brain function. This methodical approach in segmenting and processing EEG data is instrumental in unveiling the intricate correlations between EEG patterns and emotional responses, ensuring a robust and thorough analysis.

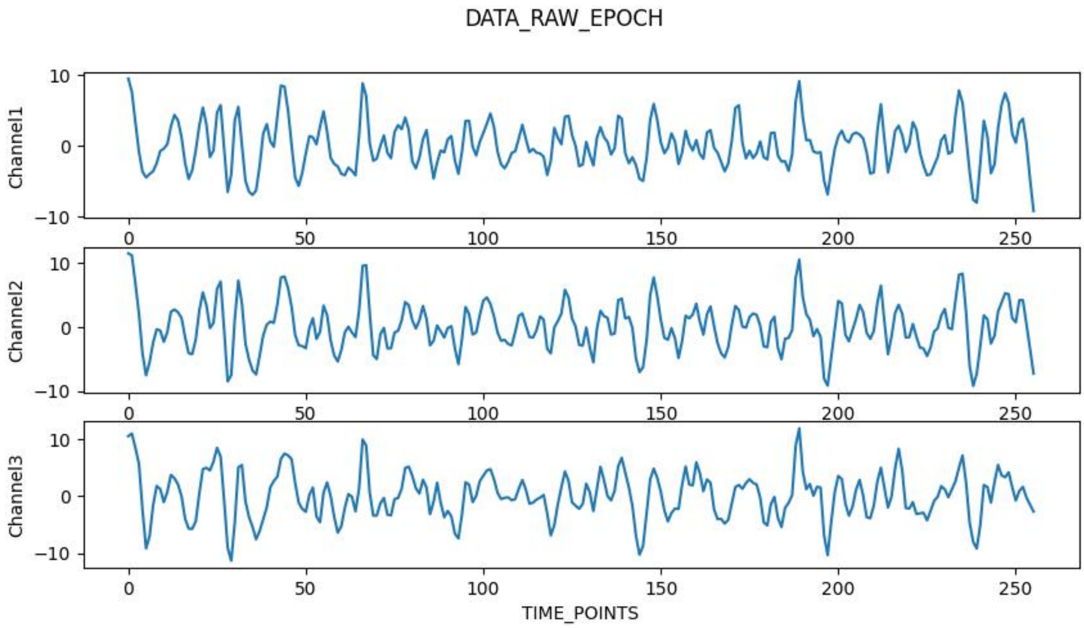

Figure 3.4: Sample of Raw Data

3.3.2 Feature Extraction

In this study, the Fast Fourier Transform (FFT) is utilized to process the raw EEG data obtained through epoching, transforming it to emphasize frequency domain features. FFT is a widely used algorithm for computing the Discrete Fourier Transform (DFT) and its inverse, efficiently converts a signal from its original time domain into a representation in the frequency domain. The primary function of FFT in EEG analysis is to decompose the brain's electrical signals into constituent frequencies. This is particularly important in emotion recognition, as different emotional states are often characterized by distinct patterns in specific EEG frequency bands (Saxena et al., 2020). By applying FFT to EEG data, one can isolate the frequency components such as alpha, beta, theta, delta, and gamma waves, each of which is associated with different mental states and cognitive processes. FFT facilitates the extraction of frequency features from raw EEG signals. These features, which are derived from the frequency domain, provide valuable insights into the brain's response to various emotional stimuli. For instance, certain frequency bands might be more active or show specific patterns during different emotional experiences. The ability of FFT to effectively and efficiently analyze these frequency-based changes makes it a powerful tool in the identification and classification of emotional states based on EEG data.

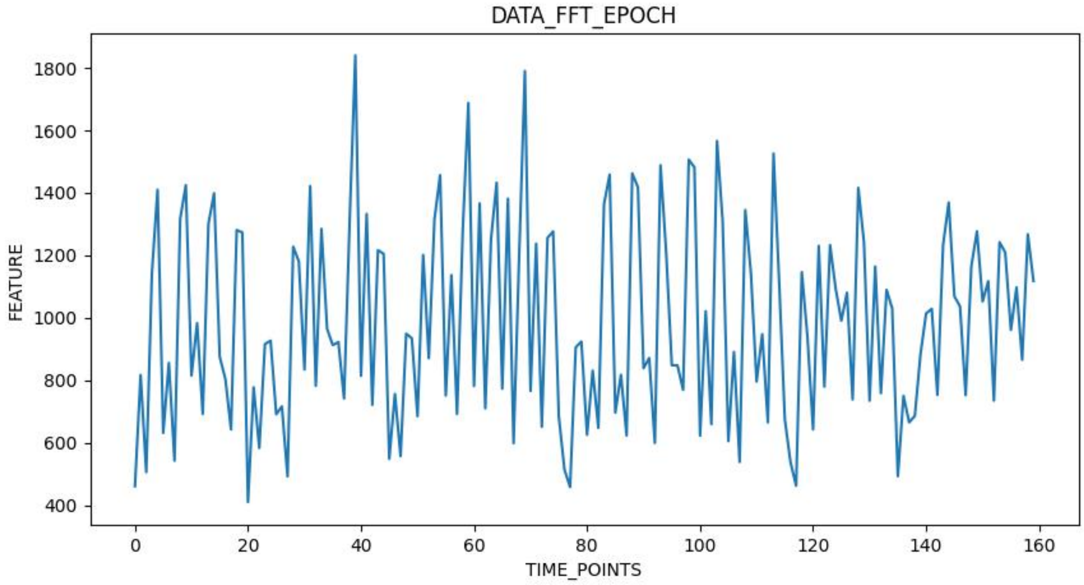

Figure 3.5: Sample of FFT Feature Extraction Data

3.3.3 Data Preprocess

In this study, a two-step process was used in order to obtain further preprocessed data. Initially, z-score standardization was applied to the dataset, followed by min-max scaling, as called normalization. This two-step approach was designed to try to compare the performance of the proposed model with respect to the preprocessed data and the raw data.

3.3.3.1 Z-Score

In this study, Z-Score standardization was used as the first step in the preprocessing step to deal with the distributional patterns of the data. Z-Score standardization is a statistical method that adjusts the features of a dataset to have a mean (μ) of 0 and a standard deviation (σ) of 1. This is achieved through the formula shown below, where X represents the original value and Z denotes the standardized value. Standardization plays a substantial role in EEG signal processing because signal amplitude and frequency are subject to inherent variability across subjects and across sessions with the same subject. By standardizing these signals, Z-Score ensures a consistent scale for input into the model, thereby enhancing the model's learning efficiency and reducing bias. This approach also plays a crucial role in preventing model overfitting and improving its generalization capabilities.

$$Z = \frac{(X - \mu)}{\sigma}$$

3.3.3.2 Min-Max Scaling

Min-Max Scaling, also known as normalization, is the second preprocessing step after Z-Score standardization of the EEG signal. This method involves rescaling the range of features to scale the data within a specified range, typically between 0 and 1. The formula below describes the Min-Max process, where X is an original value, X_{Norm} is the normalized value, and X_{Max} and X_{Min} are the maximum and minimum values in the feature, respectively. The application of Min-Max Scaling is also pivotal in EEG signal processing, after standardization through the Z-Score method, which ensures a uniform mean and standard deviation, Min-Max Scaling adjusts these standardized values to a defined range. This scaling is beneficial for EEG data because it preserves the original distribution of the signals while aligning them within a common scale. Additionally, the use of Min-Max Scaling can improve the stability and performance of learning algorithms by reducing their sensitivity to the scale of features. In EEG-based emotion recognition, where data from various channels might exhibit differing ranges, normalization ensures that these variations do not skew the analysis, thereby enabling the model to more accurately interpret the underlying emotional states from the EEG signals. This uniform scaling approach contributes to the accuracy and robustness of the emotion recognition model, ensuring that the subtleties of EEG data are effectively captured and analyzed.

$$X_{Norm} = \frac{(X - X_{Min})}{(X_{Max} - X_{Min})}$$

3.4 Proposed Model

In this research, a novel hybrid model combining the strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks is proposed for EEG-based emotion recognition, The model structure with parameter settings can be found in Appendix A. The model comprises 3 layers of 1D-CNNs followed by 2 layers of LSTM, creating a sophisticated architecture for analyzing EEG signals. This hybrid model, combining the feature extraction capabilities of CNNs and the sequential data analysis strengths of LSTMs, presents a powerful approach for EEG-based emotion recognition. The integration of these two neural network architectures in the proposed model offers a promising approach to improve the accuracy and effectiveness of emotion recognition from EEG data. The detailed description of each of the two models will be presented in the following sections.

3.4.1 1D-CNN

Convolutional Neural Networks (CNNs) are a class of deep neural networks known primarily for their ability to process data with a grid-like topology, such as images. Since CNNs are good at automatically detecting important features, it is extremely effective to use a convolutional layer to filter the input to obtain useful information. The one-dimensional CNNs (1D-CNNs) are particularly useful for time-series data, such as EEG signals, by adeptly capturing temporal features along with spatial relationships within the data. This functionality is particularly advantageous in EEG signal analysis, where distinguishing the temporal dynamics and spatial correlations of brainwave patterns is key to accurate analysis. In the proposed hybrid model of this study, a series of three structurally analogous 1D-CNN layers are used in the initial phase. These layers are intricately designed for autonomous feature extraction from EEG signals. Each layer in this sequence is comprised of a Conv1d layer for feature extraction, followed by a ReLU activation function, a MaxPool1d pooling layer for dimensionality reduction, and a Dropout layer for model regularization. Specifically, the structure of each layer is described below:

3.4.1.1 One-Dimensional Convolutional Layer (Conv1d)

The Conv1d layer is configured to extract temporal features from the EEG signal. The first layer has 1024 output channels, employs a 75-size kernel and a stride of 1, with padding set to 'same', ensuring comprehensive feature capture. Subsequent layers (1024 to 512 and then 512 to 256 output channels) with reduced kernel sizes (50 and 25 respectively) allow for more refined feature extraction.

3.4.1.2 ReLU Activation Function

ReLU Activation Function: The ReLU function is incorporated after Conv1d in each layer, introducing nonlinearity and enabling the model to learn complex feature representations. Its effectiveness in mitigating the gradient vanishing problem significantly improves model training efficiency. The implementation of the ReLu function can be found in the following equation:

$$\mathrm{ReLU}(X) = \max(0, X)$$

3.4.1.3 MaxPool1d Pooling Layer

MaxPool1d Pooling Layer: This layer is configured after the ReLu function to reduce the dimensions of the features. The MaxPool1d layer configured in this study is applied with a kernel size of 2 and a step size of 2. By adopting the strategy of selecting the maximum value in each localized region, it not only reduces the computational burden but also helps to reduce overfitting.

3.4.1.4 Dropout Layer

Dropout Layer: Set at a rate of 0.2, the Dropout layer randomly deactivates 20% of the neurons during each training iteration. This approach significantly increases the model's ability to generalize and protects against overfitting.

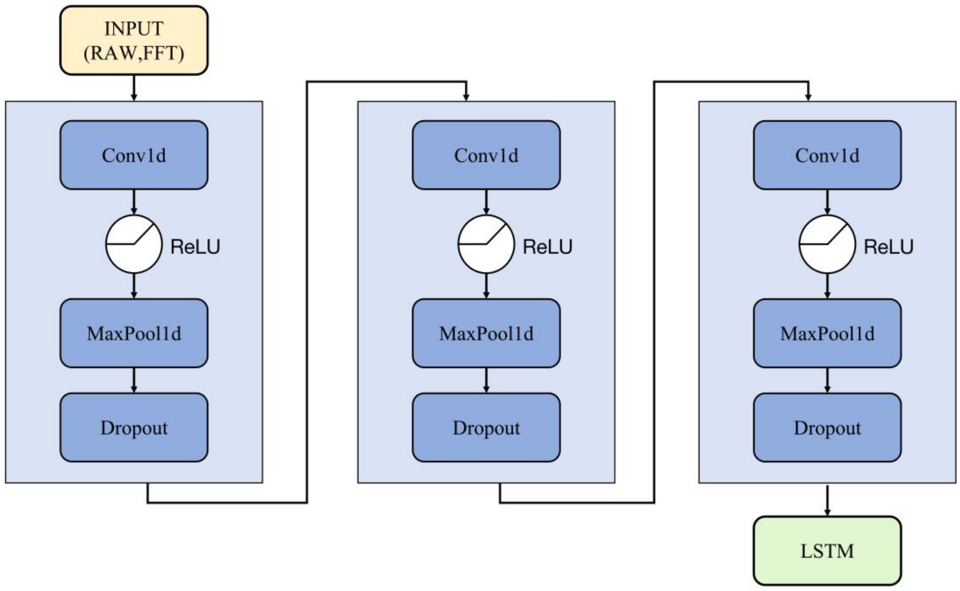

Figure 3.6: Architecture of Proposed Hybrid Model (CNN)

3.4.2 LSTM

Following the CNN layers in this study, Long Short-Term Memory (LSTM) networks are employed to enhance the model's capacity to analyze EEG signals. While CNNs are adept at extracting local spatial features from data, they often fall short in capturing the long-term dependencies present in sequential data. This gap is effectively bridged by the integration of LSTM layers, which are specifically designed to address this limitation. LSTM networks, a type of Recurrent Neural Network (RNN), are renowned for their ability to remember information over extended periods. This characteristic is particularly beneficial in the context of EEG signal analysis, where understanding the temporal progression and dependencies of brainwave patterns is crucial. The LSTM layers in this model are structured to sequentially process the features extracted by the CNN layers, thereby capturing the temporal dynamics within the EEG data.

The model incorporates two LSTM layers. The first LSTM layer has an input and output size of 256 units, designed to maintain a high level of feature detail after processing by the CNN layers. This layer is set with 'batch_first=True' to ensure compatibility with the input data structure. The second LSTM layer reduces the output size to 128 units, further refining the data processing and facilitating a more nuanced interpretation of the temporal features. This reduction also aids in mitigating computational complexity while preserving the essential temporal information.

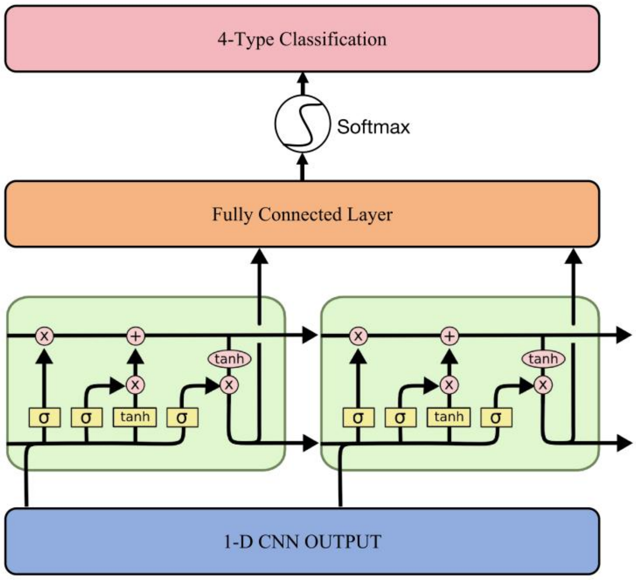

Figure 3.7: Architecture of Proposed Hybrid Model (LSTM)

3.4.3 Fully Connected (FC) Layer

In the architecture of the proposed model, the last component consists of Fully Connected (FC) layers, which are used to synthesize the feature aspects extracted and processed by the previous CNN and LSTM layers. Fully connected layers, as the name suggests, connect every neuron in one layer to each neuron in the subsequent layer. This comprehensive connectivity ensures that the model can consider all the features together to make informed predictions. The primary function of these layers in deep learning models, particularly in EEG signal analysis, is to integrate the learned representations and map them to the final output space, which in this case is the classification of emotional states.

In this study, the Fully Connected layers are meticulously structured to optimize the classification process. The sequence begins with an input size calculated based on the output of the LSTM layers, then followed by a series of linear transformations and non-linear activations, designed to progressively refine the feature representation. The first linear layer reduces the dimensionality from the LSTM output size to 256, followed by a ReLU activation function that introduces non-linearity and aids in mitigating the vanishing gradient problem. Subsequent linear layers further reduce the dimensions from 256 to 128, and then to 64, each followed by ReLU activations. This gradual reduction in size helps in distilling the most salient features for the final classification task. The last linear layer transforms these features into the final output size, equal to the number of classes, which is then passed through a Softmax activation. The Softmax function converts the output into probability distributions, making the model's predictions interpretable for emotion classification. This structured approach in the fully connected layers ensures that the complex and rich representations learned from the EEG data are effectively translated into precise and meaningful emotional state predictions.

3.5 Cross-validation

Cross-validation is a statistical technique used to assess the generalizability of a predictive model by partitioning the original dataset into a training set to train the model and a testing set to evaluate it. In the context of EEG emotion recognition, where data is often complex and highly variable, cross-validation becomes essential. It helps in ensuring that the model is not overfitting to a specific subset of data and is capable of performing consistently across different sets of EEG readings. This is particularly important as EEG datasets can exhibit significant variability across individuals and sessions, making it imperative to test models across various subsets of data to verify their reliability and accuracy in real-world scenarios (Cimtay & Ekmekcioglu, 2020; He, Zhong, & Pan, 2022). In this study, a 5-fold cross-validation technique has been implemented to validate the proposed hybrid deep learning model's performance, particularly focusing on its subject-independent capabilities. In 5-fold cross-validation, the dataset is randomly divided into five equal parts. In each iteration, four parts are used for training the model, and the remaining part is used for testing. This process is repeated five times, with each part being used exactly once as the test set. This technique is integral to the study as it allows for a comprehensive evaluation of the model across different subsets of data, simulating a subject-independent scenario. It ensures that the model's performance is not biased towards any specific part of the dataset and can generalize well across various EEG signal characteristics.

3.6 Early Stopping

Early stopping is a widely used regularization technique in deep learning and other AI domains to prevent overfitting during model training. It involves monitoring the model's performance on a validation set during the training process and terminating the training early if the performance stops improving. The rationale behind early stopping is based on the observation that, as training progresses, models tend to first learn general patterns (which are beneficial for performance on unseen data) and then start overfitting to the noise and idiosyncrasies of the training data. By halting the training process at the point where performance on the validation set begins to degrade, early stopping effectively prevents the model from learning these irrelevant patterns, thereby ensuring better generalization to new, unseen data (Gupta & Sharma, 2022). In this study, early stopping with a patience parameter set to 10 was incorporated into the training process of the proposed hybrid deep learning model for EEG-based emotion recognition. Patience is a hyperparameter that defines the number of epochs to wait for an improvement in the model's performance on the validation set before stopping the training. In this case, if the model's performance on the validation set does not improve for 10 consecutive epochs, the training is halted. This implementation ensures that the model does not overfit to the peculiarities of the training data, thus maintaining its ability to generalize effectively to new data. This is particularly important in the context of EEG data, which is highly subject-specific and can vary greatly from one individual to another. By preventing overfitting, early stopping contributes significantly to the robustness and reliability of our model in accurately recognizing emotions across different subjects, thereby enhancing the model's practical applicability in real-world scenarios.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Experiment Setup

In this experimental setup, the study validates the efficacy of the proposed 1DCNN-LSTM model in four-class EEG-based emotion recognition and explores its subject-independent predictive performance on the DEAP dataset. The experiments were conducted using Google Colab, equipped with Tesla V100-SXM2-16GB GPU and 50GB of system RAM. The software environment consisted of Python 3.10.12, with PyTorch 2.1.0 utilized for constructing the neural network models. The experiment's parameters included a dropout rate of 0.2 and the use of the Adam optimizer with a learning rate of 1e-5 to optimize training efficiency and model performance. For the empirical component of this study, data from 10 subjects were carefully selected from the DEAP dataset to conduct a series of experiments. These experiments were designed to assess the model's performance across various configurations, thereby evaluating its adaptability and effectiveness in different contexts. The comparative analyses were as follows:

1. **Data Processing Impact**: The model's performance was evaluated by comparing results obtained from preprocessed data, which underwent standardization and normalization, against those derived from unprocessed raw data. This comparison aimed to gauge the impact of preprocessing steps on model efficacy.
2. **Channel Number Variation**: Another aspect of the study involved comparing the model's performance using 32-channel raw data against 14-channel configuration. This experiment was intended to understand the influence of channel count on the model's accuracy and efficiency.
3. **Feature Extraction Techniques**: Lastly, the study compared the model's performance on raw data with its performance on data processed through FFT feature extraction. This comparison was used in determining the effectiveness of Feature Extraction in model's emotion recognition capabilities.

To evaluate the model's classification performance and its ability to predict across different subjects, a 5-fold cross-validation technique was employed. This involved dividing the samples into five subsets, with one subset used as the test set and the remaining four as the training set. This process was repeated five times, ensuring each subset was used as the test set once, to assess the model's predictive ability and robustness in various experimental environments.

4.2 Experiment Results

The outcomes of these experiments, detailing training and validation accuracy as well as loss metrics, are systematically documented. These results are visually represented in Figure 4.1 and Figure 4.2 which illustrates the variation of accuracy and loss across epochs during the model training process.

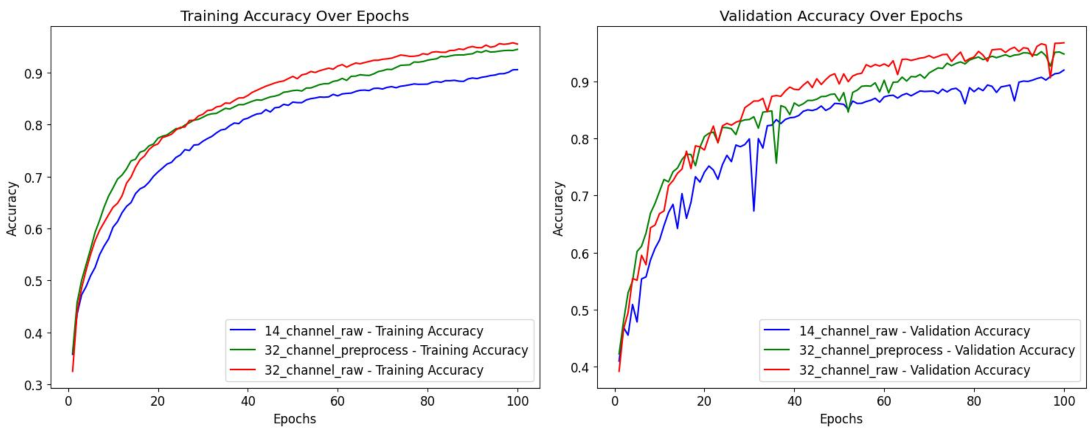

Figure 4.1: Training & Validation Accuracy

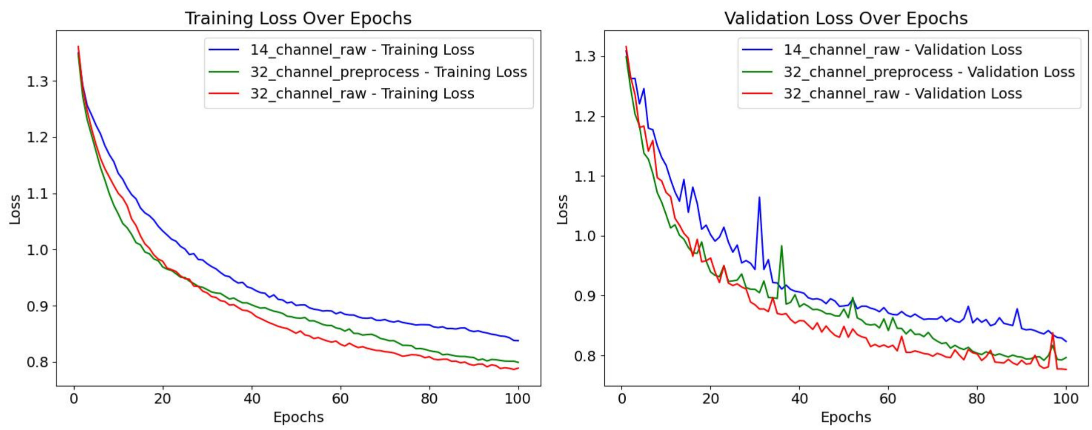

Figure 4.2: Training & Validation Loss

Additionally, Table 4.1 provides a comprehensive comparison of these metrics across the various experimental setups, offering a clear and structured overview of the model's performance under different conditions.

Experiment	Training Accuracy (%)	Training Loss	Validation Accuracy (%)	Validation Loss
32-channel Raw Data	95.50	0.7890	96.73	0.7765
32-channel Preprocessed Data	94.84	0.7955	95.68	0.7868
14-channel Raw Data	90.56	0.8379	91.96	0.8233
32-channel FFT Feature Data	88.86	0.8590	89.73	0.8455

Table 4.1: Performance Metrics of Purposed Model

In this study's result analysis, the performance of the 1DCNN+LSTM hybrid model with the 32-channel Raw Data configuration notably surpassed other experimental setups. This configuration achieved the highest training and validation accuracies of 95.50% and 96.73%, respectively, along with the lowest losses of 0.7890 and 0.7765, showcasing the model's adeptness in effectively learning and interpreting intrinsic features within the raw EEG data. This superior performance, achieved without preprocessing or external feature extraction, underscores the model's inherent capability to discern critical emotional cues from unprocessed EEG signals. In contrast, the 32-channel preprocessed Data, subjected to standardization and normalization, exhibited a decline in performance, with training and validation accuracies dropping to 94.84% and 95.68%, respectively. This suggests that preprocessing might have altered or eliminated essential EEG signal characteristics vital for accurate emotion recognition. Notably, when analyzing the 14-channel raw data, the performance decline was marginal (approximately 5%), despite a substantial reduction in the number of channels. This observation suggests that crucial emotional information in the EEG data is concentrated within specific channels, and the model effectively learns and retains predictive power even with reduced data. Contrary to expectations, the model's performance on the 32-channel

FFT Feature Data, with the lowest accuracies of 88.86% in training and 89.73% in validation, revealed that FFT feature extraction might have oversimplified the EEG data, leading to a loss of vital information necessary for precise emotion classification. These findings demonstrate the effectiveness of the 1DCNN+LSTM model in processing raw EEG data for emotion recognition tasks. Furthermore, they challenge the initial assumption that preprocessing and feature extraction are essential for improved model performance and underline the robustness and adaptability of the model in EEG-based emotion recognition tasks.

Additionally, the study incorporated a 5-fold cross-validation technique to rigorously evaluate the model's ability for subject-independent emotion prediction. This method was essential for a comprehensive analysis of the model's consistency and reliability across a diverse range of subjects. By partitioning the data into five subsets and iteratively using each as a test set while training on the remaining, the model's performance was thoroughly vetted under varying conditions. The averaged outcomes of these cycles, exhibiting a training accuracy of 90.66% and a validation accuracy of 91.68%, these metrics not only attest to the model's precision but also speak volumes about its capacity to generalize effectively across different datasets. Such robust performance in subject-independent scenarios is crucial, considering the variability inherent in EEG data from different individuals. These findings, coupled with the earlier observations, collectively underscore the hybrid model's potential in offering reliable, accurate, and generalizable solutions for EEG-based emotion recognition, marking a valid and significant experimental outcome of the study.

4.3 Comparison with Published Studies

To contextualize and validate the efficacy of the proposed model in this study, a comparative analysis was conducted with several existing studies that utilized the DEAP dataset for 4-class emotion recognition using deep learning. This comparison aimed to position the current study's findings within the broader research landscape, providing a comprehensive perspective on its performance relative to established benchmarks. Table 4.2 presents a detailed overview of these published studies, including their models and results, alongside the outcomes of this study.

Reference	Model	Accuracy(%)
Zhang et al. (2020)	1D-CNN + LSTM	94.17
Marjit et al. (2021)	Multi Layer Perceptron (MLP)	83.52
Ahmed & Sinha (2021)	Long Short-Term Memory (LSTM)	90
Hou et al. (2023)	Space-to-Depth (S2D),Residual Feature Pyramid Network (RFPN)	93.56
Islam & Lee (2023)	3D-CNN	94.78
This study	1D-CNN + LSTM	96.73

Table 4.2: Comparison of Several Published Studies on DEAP using Deep Learning

4.4 Discussion

In this section, we delve into a comprehensive analysis of the experimental results, underscoring the strengths and distinct characteristics of the proposed 1D-CNN + LSTM model. This analysis is based on its performance across a variety of experimental setups and its demonstrated robustness through 5-fold cross-validation. Primarily, the model exhibited exceptional performance in the 32-channel raw data, achieving the highest training and validation accuracies of 95.50% and 96.73%, respectively. This result highlights the model's remarkable ability to effectively interpret raw EEG signals for emotion recognition. Notably, even in the 14-channel raw data experiments, the model showed only a minimal decrease in performance, underscoring its efficient learning capabilities even with a reduced number of channels. Furthermore, when comparing the model's performance on raw data with that on preprocessed and feature-extracted data, it becomes evident that the model is more adept at autonomously learning data features. This capability potentially reduces the need for extensive data preprocessing, thereby simplifying the model's training and application in practical scenarios. Finally, the impressive cross-validation results, with a training accuracy of 90.66% and a validation accuracy of 91.68%, reinforce the model's reliability and adaptability across various datasets and conditions.

The effectiveness of the proposed 1D-CNN + LSTM model is highlighted when compared to previous research that used the DEAP dataset for emotion recognition. The model's validation accuracy in this study has achieved 96.73%, surpasses those of prior investigations that have utilized diverse deep learning frameworks, including Multi Layer Perceptron (MLP), Long Short-Term Memory (LSTM), Space-to-Depth (S2D) combined with Residual Feature Pyramid Networks (RFPN), and 3D-CNN, where accuracies ranged between 83.52% and 94.78%. This comparative evaluation not only underscores the model's preeminent performance in emotion classification but also distinctly showcases its pivotal contribution to the discipline. It elucidates the model's pioneering approach in processing EEG data for emotion recognition, which is a testament to its innovation. This represents a notable stride forward in harnessing deep learning methodologies for navigating the complexities of this specialized field.

CHAPTER 5

CONCLUSION

5.1 Conclusion

In conclusion, this study initially provided a detailed introduction to the direction and background of EEG-based emotion recognition, defining the research aims and objectives. Then, a comprehensive literature review enhanced the understanding of the challenges and advancements in this field, covering essential concepts of EEG signals, datasets mainly used as EEG sources, emotional modeling approaches, and the evolution from machine learning to deep learning methods. The review also highlighted the emergence of deep learning due to its efficiency and effectiveness, while also acknowledging existing challenges and areas needing resolution in this domain. Methodologically, the study delineated the materials and methods used in experiments, which can help to better understand the experimental process and the significance of each step. The experiments, designed based on the DEAP dataset and four-class emotion model with two dimensions, thoroughly explored the potential ability of EEG based emotion recognition of the proposed 1D-CNN + LSTM hybrid model. The experimental analysis provided insights into the model's performance under various conditions and highlighted its ability to learn efficiently from raw EEG data. The integration of 5-fold cross-validation confirmed the reliability and adaptability of the model across subjects. Importantly, these results emphasized the hybrid model's ability to autonomously extract features and analyze time-series data, supporting the initial hypothesis about the benefits of CNN and LSTM. Finally, a comparative evaluation with related studies further validated the proposed model's outstanding performance, achieving a leading accuracy rate of 96.73%. This not only demonstrates the model's superior performance but also its significant contribution to the advancement of EEG-based emotion recognition research.

5.2 Limitation and Future Work

This study, while achieving success in numerous areas, confronts certain limitations and suggests directions for future work. First, the reliance on a single dataset (DEAP) may limit the generalizability of the study's findings. Despite the comprehensiveness of the DEAP dataset, this study only used Valence and Arousal from this dataset as metrics for categorizing emotions, while other additional indicators such as dominance, liking, and familiarity were not used. Future research could benefit from incorporating more comprehensive datasets and employing a broader range of emotional indicators for a more nuanced understanding of human emotions. Secondly, the study focused on a two-dimensional four-category emotion model that, while more comprehensive and valid than traditional binary classification models, may not capture more subtle human emotions such as fear and panic. Future research could attempt to explore more complex emotion models to enhance the depth of emotion recognition.

For future work, expanding the study to include more diverse datasets, such as the open datasets SEED, DREAMER, which can be applied for research purposes, will help to further evaluate and improve the model's enhancement capabilities. In addition, exploring the integration of other neural network architectures or advanced deep learning techniques may improve the accuracy and efficiency of the model. For example, it has been noted that Transformer may be very effective in hierarchically learning different domain (temporal, spatial, spectral) features of data (Xu et al., 2023). Future research could also apply the model to real-world applications such as real-time emotion recognition applications and AI counseling, thus expanding the utility of the model in a variety of domains such as health care, education, and interactive media. Considering the integration of multi-modal data will also be an important direction in future research. For example, combining different emotion expressions such as facial expressions, speech features, and physiological signals may provide a more comprehensive understanding of emotion.In addition, considering the challenges of emotion recognition in cross-cultural contexts, future research could explore how to accommodate differences in emotion expression in different cultural contexts. Finally, emphasizing the importance of ethical considerations and privacy protection in the development and application of these advanced technologies will help ensure their responsible use and social acceptance.

5.3 Data Availability Statement

Publicly available dataset DEAP was used in this study. The data can be found at: https://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html

REFERENCES

Ahmed, M. Z. I., & Sinha, N. (2021). EEG-based emotion classification using LSTM under new paradigm. *Biomedical Physics & Engineering Express, 7(6),* 065018.

Awais, M., Raza, M., Singh, N., Bashir, K., Manzoor, U., Islam, S. U., & Rodrigues, J. J. (2020). LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19. *IEEE Internet of Things Journal, 8(23),* 16863-16871.

Cai, J., Xiao, R., Cui, W., Zhang, S., & Liu, G. (2021). Application of electroencephalography-based machine learning in emotion recognition: A review. *Frontiers in Systems Neuroscience, 15,* 729707.

Chai, M. T., Goh, C. M., & Aluwee, S. A. Z. S. (2022, September). Emotion recognition based on EEG directed functional connectivity using deep learning algorithm. In *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 192-197).* IEEE.

Chen, Y., Wang, D., & Wang, X. (2022). Emotion recognition model based on CLSTM and channel attention mechanism. In I*TM Web of Conferences (Vol. 47, p. 02041). EDP Sciences.*

Cimtay, Y., & Ekmekcioglu, E. (2020). Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. S*ensors, 20(7),* 2034.

Dade©bayev, D., Goh, W. W., & Tan, E. X. (2022). EEG-based emotion recognition: Review of commercial EEG devices and machine learning techniques. *Journal of King Saud University-Computer and Information Sciences, 34(7),* 4385-4401.

De Andrade, I. M., & Tumelero, C. (2022). Increasing customer service efficiency through artificial intelligence chatbot. Revista de Gestão, 29(3), 238-251.

Duan, R. N., Zhu, J. Y., & Lu, B. L. (2013, November). Differential entropy feature for EEG-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER) (pp. 81-84).* IEEE.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion, 6(3-4),* 169-200.

Gupta, G. K., & Sharma, D. K. (2022, March). A review of overfitting solutions in smart depression detection models. In *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 145-151).* IEEE.

He, Z., Zhong, Y., & Pan, J. (2022, May). Joint temporal convolutional networks and adversarial discriminative domain adaptation for EEG-based cross-subject emotion recognition. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3214-3218).* IEEE.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation, 18(7),* 1527-1554.

Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalography and clinical neurophysiology, 29(3),* 306-310.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9(8),* 1735-1780.

Hou, F., Liu, J., Bai, Z., Yang, Z., Liu, J., Gao, Q., & Song, Y. (2023). EEG-Based Emotion Recognition for Hearing Impaired and Normal Individuals With Residual Feature Pyramids Network Based on Time–Frequency–Spatial Features. *IEEE Transactions on Instrumentation and Measurement, 72,* 1-11.

Houssein, E. H., Hammad, A., & Ali, A. A. (2022). Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. *Neural Computing and Applications, 34(15),* 12527-12557.

Islam, M., & Lee, T. (2023, July). Functional Connectivity Analysis in Multi-channel EEG for Emotion Detection with Phase Locking Value and 3D CNN. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1-4).* IEEE.

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.

Jeevan, R. K., SP, V. M. R., Kumar, P. S., & Srivikas, M. (2019, April). EEG-based emotion recognition using LSTM-RNN machine learning algorithm. In *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT) (pp. 1-4).* IEEE.

Kamble, K., & Sengupta, J. (2023). A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals. *Multimedia Tools and Applications, 1-36.*

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.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60(6),* 84-90.

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.

Koole, S. L., & Rothermund, K. (2022). Coping with COVID-19: Insights from cognition and emotion research. *Cognition and Emotion, 36(1),* 1-8.

Li, X., Song, D., Zhang, P., Yu, G., Hou, Y., & Hu, B. (2016, December). Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In *2016 IEEE international conference on bioinformatics and biomedicine (BIBM) (pp. 352-359)*. IEEE.

Li, X., Zhang, Y., Tiwari, P., Song, D., Hu, B., Yang, M., ... & Marttinen, P. (2022). EEG based emotion recognition: A tutorial and review. *ACM Computing Surveys, 55(4),* 1-57.

Mahmoud, A., Amin, K., Al Rahhal, M. M., Elkilani, W. S., Mekhalfi, M. L., & Ibrahim, M. (2023). A CNN Approach for Emotion Recognition via EEG. *Symmetry, 15(10),* 1822.

Marjit, S., Talukdar, U., & Hazarika, S. M. (2021, September). EEG-Based Emotion Recognition Using Genetic Algorithm Optimized Multi-Layer Perceptron. In *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA) (pp. 304-309).* IEEE.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology, 14,* 261-292.

Nath, D., Anubhav, Singh, M., Sethia, D., Kalra, D., & Indu, S. (2020, March). A comparative study of subject-dependent and subject-independent strategies for EEG-based emotion recognition using LSTM network. In *Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis (pp. 142-147).*

Pal, S., Mukhopadhyay, S., & Suryadevara, N. (2021). Development and progress in sensors and technologies for human emotion recognition. *Sensors, 21(16),* 5554.

Pandey, P., & Seeja, K. R. (2021). Subject independent emotion recognition system for people with facial deformity: an EEG based approach. *Journal of Ambient Intelligence and Humanized Computing, 12,* 2311-2320.

Pandey, P., & Seeja, K. R. (2022). Subject independent emotion recognition from EEG using VMD and deep learning. *Journal of King Saud University-Computer and Information Sciences, 34(5),* 1730-1738.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology, 17(3),* 715-734.

Ravi, A., Beni, N. H., Manuel, J., & Jiang, N. (2020). Comparing user-dependent and user-independent training of CNN for SSVEP BCI. *Journal of neural engineering, 17(2),* 026028.

Rivera-Tello, S., Huerta-Chávez, V., & Ramos-Loyo, J. (2022). Actividad eléctrica cerebral: métodos de registro y análisis y sus implicaciones en la organización funcional del cerebro: Brain electrical activity: recording and analysis methods and its implications in the brain´ s functional organization. *e-CUCBA, (19),* 204-212.

Roach, B. J., & Mathalon, D. H. (2008). Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia bulletin, 34(5),* 907-926.

Samavat, A., Khalili, E., Ayati, B., & Ayati, M. (2022). Deep learning model with adaptive regularization for EEG-based emotion recognition using temporal and frequency features. *IEEE Access, 10,* 24520-24527.

Sarmiento-Calisaya, E., Ccori, P. C., & Parari, A. C. (2022, January). An emotion-aware persuasive architecture to support challenging classroom situations. In *2022 IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-2).* IEEE.

Saxena, A., Tripathi, K., Khanna, A., Gupta, D., & Sundaram, S. (2020). Emotion detection through EEG signals using FFT and machine learning techniques. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2019, Volume 1 (pp. 543-550).* Springer Singapore.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing, 45(11),* 2673-2681.

Shen, F., Dai, G., Lin, G., Zhang, J., Kong, W., & Zeng, H. (2020). EEG-based emotion recognition using 4D convolutional recurrent neural network. *Cognitive Neurodynamics, 14,* 815-828.

Suhaimi, N. S., Mountstephens, J., & Teo, J. (2020). EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Computational intelligence and neuroscience, 2020.*

Torres, E. P., Torres, E. A., Hernández-Álvarez, M., & Yoo, S. G. (2020). EEG-based BCI emotion recognition: A survey. *Sensors, 20(18),* 5083.

Vasaikar, S., Vartha, S., Thakare, S., & Mhatre, S. (2022, June). Classification of Anxiety in Human's using Convolution Neural Network. In *2022 7th International Conference on Communication and Electronics Systems (ICCES) (pp. 1372-1376).* IEEE.

Valdés, P., Bosch, J., Grave, R., Hernandez, J., Riera, J., Pascual, R., & Biscay, R. (1992). Frequency domain models of the EEG. *Brain topography, 4,* 309-319.

Vempati, R., & Sharma, L. D. (2023). A Systematic Review on Automated Human Emotion Recognition using Electroencephalogram Signals and Artificial Intelligence. *Results in Engineering,* 101027.

Wang, D., & Shang, Y. (2013). Modeling physiological data with deep belief networks. *International journal of information and education technology (IJIET), 3(5),* 505.

Wang, F., Wu, S., Zhang, W., Xu, Z., Zhang, Y., Wu, C., & Coleman, S. (2020). Emotion recognition with convolutional neural network and EEG-based EFDMs. *Neuropsychologia, 146,* 107506.

Wang, X., Ren, Y., Luo, Z., He, W., Hong, J., & Huang, Y. (2023). Deep learning-based EEG emotion recognition: Current trends and future perspectives. *Frontiers in Psychology, 14,* 1126994.

Wickramasinghe, W. R. M. G. K., Devduni, R. M. B., Dasanayaka, D. T. C. B., Mohomed, M. N. N., Kumari, S., & Dassanayake, T. (2021, December).

Sensor-Based Emotion Tracking System for Computer Games. In *2021 3rd International Conference on Advancements in Computing (ICAC) (pp. 389-394).* IEEE.

Xu, Y., Du, Y., Li, L., Lai, H., Zou, J., Zhou, T., ... & Ma, P. (2023). AMDET: Attention based Multiple Dimensions EEG Transformer for Emotion Recognition. *IEEE Transactions on Affective Computing.*

Zhang, Y., Chen, J., Tan, J. H., Chen, Y., Chen, Y., Li, D., ... & Che, W. (2020). An investigation of deep learning models for EEG-based emotion recognition. *Frontiers in Neuroscience, 14,* 622759.

Zheng, W. L., & Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development, 7(3),* 162-175.

APPENDIX A: 1D-CNN + LSTM Model Structure