DEEP LEARNING BASED MULTIMODAL EMOTIONAL RECOGNITION

Major Project Report

Submitted by

Kota Abhishek (B201030EC)

Padi Pranav Reddy (B201013EC)

Poluparthi Venkata Chandra Manikumar (B200985EC)

Pradeepkumar Puchala (B200986EC)

In partial fulfillment for the award of the Degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY, CALICUT

NIT CAMPUS P.O., CALICUT

KERALA, INDIA 673601.

NATIONAL INSTITUTE OF TECHNOLOGY, CALICUT

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

*This is to certify that the major project report entitled "**DEEP LEARNING BASED MULTIMODAL EMOTIONAL RECOGNITION**" is a bonafide record of the Project done by **Kota Abhishek** (B201030EC), **Padi Pranav Reddy** (B201013EC), **Poluparthi Venkata Chandra Manikumar** (B200985EC), **Pradeepkumar Puchala** (B200986EC), under our supervision, in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Electronics and Communication Engineering** from **National Institute of Technology Calicut**, and this work has not been submitted elsewhere for the award of a degree.*

Dr.Gangireddy Narendrakumar Reddy
Assistant Professor, ECE,
National Institute of Technology Calicut

Dr. Jaikumar M.G
HOD ECED,
NIT Calicut

Place: Calicut
Date: 07 - 05 - 2024

ACKNOWLEGEMENT

We would like to take this opportunity to express our deepest gratitude to everyone who helped us in completing the project. We have great pleasure in expressing my gratitude and obligations to Dr.Gangireddy Narendrakumar Reddy, Assistant Professor , Department of ECE, for his valuable guidance, and suggestions to make this work a success. We would also like to thank our project coordinator, Dr. P. S. Sathidevi, and the evaluation committee for giving their suggestions during the evaluations to build our project better. We express our gratitude to Dr. Jaikumar M. G. Head of the Department, Department of ECE, for his wholehearted cooperation and encouragement. We also acknowledge our gratitude to other faculty members in the Department of Electronics and Communication Engineering, family, and friends for their wholehearted cooperation and encouragement. Lastly, we thank the almighty, my parents, and my friends for their constant encouragement, without which this project would not be possible.

Kota Abhishek
Padi Pranav Reddy
Poluparthi Venkata Chandra Manikumar
Pradeepkumar Puchala

May 2024
Calicut

ABSTRACT

Emotion recognition has attracted significant interest due to its wide range of applications across various domains, including human-computer interaction, healthcare, and affective computing. Recent advancements have highlighted the potential of integrating multiple physiological signals to enhance the accuracy and robustness of emotion recognition systems.

In this project, we aim to develop a Deep Learning-based multimodal emotion recognition system utilizing a combination of Photoplethysmography (PPG), Electroencephalography (EEG), and Galvanic Skin Response (GSR). Our objective is to detect emotions such as happiness, calmness, excitement, fatigue, depression, frustration, and anger.

Each physiological signal provides unique insights: PPG offers information about cardiovascular changes, EEG indicates brainwave patterns, and GSR reflects sympathetic nervous system activity.

Initially, we applied the k-Nearest Neighbors (KNN) model individually for EEG, GSR, and PPG signals, achieving varying accuracies for valence and arousal prediction. The accuracies ranged from approximately 49% to 58% for valence and 53.12% for arousal. Combining these signals improved accuracy significantly, with the combined model reaching an accuracy of 58.83% for valence and 50.89% for arousal, indicating the importance of multimodal analysis.

Additionally, we employed a specialized 4DCRNN model for EEG signals, obtaining accuracies of 94.33% for arousal and 94.23% for valence prediction. Further, we explored Long Short-Term Memory (LSTM) models, using raw EEG, GSR, and PPG signals separately and in combination. By leveraging feature extraction techniques, we observed enhanced accuracy across all cases, with accuracies ranging from approximately 67% to 78.3%, emphasizing the efficacy of deep learning approaches.

This project demonstrates the effectiveness of combining physiological signals and advanced machine learning models for emotion recognition. The integration of diverse signals and sophisticated models lays a foundation for developing practical emotion recognition systems with broad applications in various domains.

To achieve our goals, we will leverage standard publicly available databases for model training and validation. Additionally, our framework will be optimized and deployed on a resource-constrained Raspberry Pi board platform to enable real-time applications. After implementing on Raspberry Pi, we got valence accuracy of 70.0% and arousal accuracy of 76.6%.

Through this comprehensive approach, we aim to develop a robust and efficient emotion recognition system that can be deployed across various real-world scenarios, enhancing human-machine interactions and facilitating personalized healthcare interventions.

CONTENTS

List of Figures

List of Tables

List of Abbreviations

CNN	Convolutional Neural Network
DEAP	Dataset for Emotion Analysis using Physiological Signals
ECG	Electrocardiogram
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
GSR	Galvanic Skin Response
HCI	Human Computer Interaction
KNN	K-Nearest Neighbours
LSTM	Long Short-Term Memory
PPG	photoplethysmogram
RNN	Recurrent Neural Network
ReLu	Rectified Linear Units
REM	Rapid Eye Movement
SVM	Support Vector Machine
4D-CRNN	4 Dimensional Convolutional Recurrent Neural Networks

CHAPTER 1

Introduction

The Deep Learning-Based Multimodal Emotional Recognition project marks a significant step forward in understanding human emotions. It uses deep learning algorithms and multiple types of data inputs to achieve this. This project aims to improve emotional recognition by combining different signals such as EEG, PPG, and GSR.

Recognizing emotions is important in various fields like healthcare, customer service, and human-computer interaction. However, current systems often struggle to accurately understand complex emotional cues and details across different types of data. By using deep learning techniques, this project hopes to overcome these challenges, leading to a more accurate understanding of human emotions.

The project recognizes that advances in deep learning can greatly enhance the reliability and accuracy of emotional recognition systems. By analyzing data from multiple sources with advanced neural networks, the goal is to automate emotional recognition, make the system more adaptable to different situations, and reduce errors made by humans.

At its core, the project focuses on the latest technology, specifically exploring how deep learning can recognize complex emotional states from different types of data. It assumes that the input data is clear and well-organized, without any extra noise or irrelevant information. Additionally, the effectiveness of the deep learning models depends on training them with large datasets covering a wide range of emotional expressions and situations.

By connecting advanced technology with emotional understanding, this project

not only pushes the boundaries of what's possible but also has the potential to change how we understand and interact with emotions in the digital age. Through ongoing improvements and innovation, the goal is to create systems that are more empathetic and responsive, truly connecting with human experiences and emotions.

CHAPTER 2

Project Overview

2.1 Problem Statement

The objective is to develop a Deep Learning-based Multimodal Emotional Recognition system, incorporating input from physiological signals like electroencephalogram (EEG), galvanic Skin Response (GSR), photoplethysmography (PPG) , with a focus on leveraging Raspberry Pi for physical implementation. This system aims to accurately identify and interpret human emotions across multiple modalities, utilizing advanced algorithms and neural networks. By integrating Raspberry Pi, the project ensures practicality and accessibility, enabling realworld deployment of emotion recognition capabilities in various environments.

2.2 Background

Emotional recognition has emerged as a significant area of research with implications across various domains, including healthcare, human-computer interaction(HCI), and affective computing. Understanding and interpreting human emotions play a crucial role in improving communication, enhancing user experience, and developing empathetic technologies.

Traditionally, emotion recognition relied heavily on manual observation and subjective interpretation. However, advancements in technology, particularly in the fields of signal processing and machine learning, have paved the way for more automated and objective methods of emotional analysis.

Previous studies have explored different modalities for emotion recognition, including facial expressions, vocal intonations, physiological signals, and textual data. Each modality offers unique insights into emotional states, with potential applications ranging from mental health monitoring to personalized user interfaces.

Despite progress in the field, challenges remain in achieving accurate and robust emotion recognition across diverse contexts and populations. Variability in individual expression, cultural differences, and environmental factors can all impact the effectiveness of emotion recognition systems.

2.3 Objective

The objective of the Deep Learning-Based Multimodal Emotional Recognition project is to develop an innovative system that accurately recognizes and interprets human emotions by leveraging deep learning algorithms and multimodal data inputs. Our primary aim is to enhance the accuracy of emotional recognition by creating deep learning models capable of effectively analyzing physiological signals such as EEG, PPG, and GSR to identify a wide range of emotional states.

Additionally, we seek to automate the emotional recognition process by designing a system that can autonomously extract nuanced emotional cues from multimodal data inputs, thereby streamlining the process and reducing the need for manual intervention. Furthermore, we aim to improve the adaptability of the system by training the deep learning models on comprehensive datasets encompassing diverse emotional expressions and contexts, ensuring its effectiveness across different cultural, linguistic, and situational factors.

Ultimately, our goal is to revolutionize emotional recognition by integrating advanced deep learning techniques with multimodal data inputs, providing a more nuanced and accurate understanding of human emotions in domains such as healthcare, customer service, and human-computer interaction. Through this project, we aim to contribute to the advancement of research and development in the field of emotional recognition, offering innovative methodologies and techniques that can be applied to future endeavors.

2.4 Motivation

Our project is fueled by the desire to improve how technology understands human emotions. Current methods struggle to accurately capture the complexities of human feelings. We aim to use advanced techniques like deep learning to make this process better.

The main drive behind our project is to automate emotional recognition, reducing the need for manual intervention. By training our system with diverse examples, we hope to make it effective across various situations and for different people.

We're motivated to enhance emotional recognition because we believe it can have a positive impact in areas like healthcare and customer service. Our goal is to create a system that can truly understand human emotions and be helpful in real-world scenarios.

In essence, our motivation lies in making technology more empathetic and useful in understanding human emotions.

2.5 Proposed Plan

1. **Data Collection:**

   Gather the DEAP dataset, comprising EEG, physiological, and video signals, along with participant ratings and other metadata.

2. **Initial Signal Analysis:**

   Begin with a detailed analysis of each signal (EEG, PPG, GSR) individually to understand their characteristics and potential for emotion classification.

3. **Machine Learning with Single Signals:**

   Implement machine learning algorithms such as K-Nearest Neighbors (KNN) on each signal independently to establish baseline performance. Evaluate the performance of KNN and other baseline algorithms for each signal.

4. **Combining Single Signals:**

Explore methods to combine information from individual signals effectively. Experiment with feature fusion techniques or simple concatenation to combine features extracted from each signal.

5. **Spatial Feature Extraction with EEG:**

   Focus on EEG signals and investigate spatial features by applying techniques like 4D Convolutional Recurrent Neural Networks (4DCRNN). Train and evaluate 4DCRNN models to capture spatial patterns in EEG data for emotion classification.

6. **Temporal Modeling with LSTM:**

   Recognize the temporal dynamics present in physiological signals and videos. Implement Long Short-Term Memory (LSTM) models to capture temporal dependencies in the data, particularly suitable for sequences like EEG and PPG.

7. **Individual Signal Modeling with LSTM:**

   Develop LSTM models for each signal individually to exploit temporal information.

   Train and evaluate LSTM models for EEG, PPG, and GSR signals separately to understand their predictive capabilities. Combining Temporal Features: Combine the temporal features extracted from individual LSTM models. Explore techniques such as feature concatenation or attention mechanisms to integrate features from different signals effectively.

8. **Overall Model Fusion:**

   Integrate spatial features from EEG (4DCRNN) and temporal features from LSTM models for EEG, PPG, and GSR. Design a comprehensive fusion strategy to combine information from all signals effectively.

9. **Evaluation and Comparison:**

   Evaluate the performance of the combined models using appropriate metrics such as accuracy, precision, recall, and F1-score. Compare the performance of different models (KNN, LSTM, combined models) to identify the most effective approach for emotion classification.

10. **Discussion and Conclusion:**

    Discuss the results obtained from the various models and fusion techniques. Highlight the strengths and limitations of each approach and provide insights for future research. Conclude with recommendations for real-world applications and further improvements in emotion classification using multimodal signals.

CHAPTER 3

Literature Survey

The Deep Learning-Based Multimodal Emotional Recognition project merges deep learning algorithms with EEG, PPG, and GSR signals to revolutionize emotion recognition. Traditional systems struggle to interpret emotional cues accurately across different modalities, prompting the need for advanced techniques. By leveraging deep learning, this project aims to automate emotion recognition, enhance adaptability, and minimize errors. It focuses on analyzing multimodal data inputs with neural networks, assuming clear and structured data. The models are trained on diverse datasets to capture a wide range of emotional expressions. Through this literature survey, we explore existing research aligning with our project's goals, paving the way for groundbreaking advancements in emotional recognition technology.

shen [1] EEG-based emotion recognition explore methods that integrate frequency, spatial, and temporal information to boost accuracy. Deep learning, notably CNNs for spatial and frequency features and RNNs like LSTMs for temporal dependencies, dominates this research. Benchmarking on datasets like DEAP validates these models' effectiveness. Approaches that fuse multiple modalities consistently outperform single-focus methods, highlighting the importance of comprehensive feature representation. Continued advancements focus on novel architectures and evaluation techniques to deepen our understanding of neural correlates of emotion.

M. Khateeb [2] Emotion recognition via EEG signals presents challenges with existing models, often limited by controlled stimuli and a restricted range of emotion classes. Identifying crucial EEG features is essential for accurate classification. The DEAP dataset, capturing physiological responses to real-world stimuli, provides

valuable data for research. Recent studies have delved into multi-domain features like time, wavelet, and frequency to enhance classification accuracy. The proposed model endeavors to classify nine emotion classes using SVM, achieving an average accuracy of 65.92% through cross-validation. This accuracy signifies significant progress in affective computing, promising richer human-computer interaction experiences. Further advancements in EEG-based emotion recognition hold potential for diverse applications, from healthcare to entertainment, by enabling systems to better understand and respond to human emotions in real time.

S. Koelstra et al [3] outlines the creation and analysis of a multimodal dataset for studying human affective states. EEG and peripheral physiological signals were recorded from 32 participants as they watched 40 one-minute excerpts of music videos. Participants also rated each video on various affective dimensions including arousal, valence, like/dislike, dominance, and familiarity. Additionally, frontal face video was recorded for 22 participants. A novel method for stimuli selection was proposed, incorporating affective tags from last.fm, video highlight detection, and an online assessment tool. The abstract describes an extensive analysis of participant ratings and investigates correlations between EEG signal frequencies and affective ratings. Methods for single-trial classification of arousal, valence, and like/dislike ratings using EEG, physiological signals, and multimedia content analysis are presented, along with decision fusion techniques for combining modalities. The dataset is made publicly available for further research, encouraging the testing of affective state estimation methods by other researchers.

Ayata D [4] delve into the intricate relationship between human emotions and computational algorithms, with a specific focus on emotion recognition through Galvanic Skin Response (GSR) signals. Emotions, as pivotal aspects of human life, drive behavior and decision-making, making their accurate recognition a significant area of research. Leveraging biomedical signal processing techniques, the study delves into feature extraction methods, including time domain analysis, wavelet transformation, and Empirical Mode Decomposition (EMD), to extract relevant features from GSR signals. These features serve as inputs for machine learning algorithms such as k-Nearest Neighbors, Decision Trees, Random Forests, and Support Vector Machines, renowned for their efficacy in pattern recognition tasks. By employing these algo-

rithms, the study aims to classify emotional states, particularly valence and arousal, with high accuracy. Remarkably, the research achieves impressive accuracy rates of 81.81% for arousal and 89.29% for valence, underscoring the effectiveness of the developed algorithms and feature extraction methods. Through interdisciplinary keywords like biomedical signal processing, emotion recognition, and machine learning, the study underscores its interdisciplinary nature, bridging domains such as physiology, computer science, and pattern recognition to deepen our understanding of human emotions.

N. Y. Oktavia [5] EEG stands out for its reliability in understanding emotions processing. Studies frequently employ custom EEG datasets, often comprising data from approximately 12 participants, to train and test emotions recognition algorithms. Time domain features extracted from EEG signals, including mean, standard deviation, and peak count, play a crucial role in distinguishing emotional states, with achieved accuracies reaching up to 87.5%. Notably, alpha and beta frequency bands are commonly explored, with the combination of both demonstrating better accuracy compared to individual bands, resulting in an observed increase of approximately 20%. Machine learning classifiers, particularly Naïve Bayes, are commonly utilized, achieving accuracies around 87.5%. Evaluation of models typically involves split testing options, commonly utilizing a 66% split to ensure robustness. Understanding the neural correlates of emotions through EEG contributes significantly to enhancing user experience and interaction with computational systems in HCI. Ongoing research efforts aim to refine algorithms and explore additional physiological signals to further improve emotions recognition accuracy beyond current levels.

Research in EEG-based emotion recognition has witnessed significant advancements, integrating various modalities to enhance accuracy. Deep learning, particularly CNNs for spatial and frequency features and RNNs like LSTMs for temporal dependencies, has been prominent, validated by benchmarking on datasets like DEAP. Multi-modal approaches consistently outperform single-focus methods, emphasizing comprehensive feature representation. Challenges persist with existing models due to limited stimuli and emotion classes, necessitating the identification of crucial EEG features. Recent studies have explored multi-domain features to improve classification accuracy, with SVM achieving 65.92% average accuracy in classifying nine

emotion classes. Additionally, a multimodal dataset incorporating EEG, peripheral physiological signals, participant ratings, and face video was created, enabling extensive analysis and correlation investigation between EEG signal frequencies and affective ratings. Custom EEG datasets, often with approximately 12 participants, have been utilized, with time domain features and machine learning classifiers such as Naïve Bayes achieving accuracies up to 87.5%. Ongoing efforts aim to refine algorithms and explore additional physiological signals to further enhance emotion recognition accuracy, facilitating richer human-computer interaction experiences and applications across various domains from healthcare to entertainment.

CHAPTER 4

Technical Theory

4.1 EEG (Electroencephalography):

EEG measures electrical activity in the brain using electrodes placed on the scalp. It records the summed electrical potentials of thousands to millions of neurons firing synchronously increased alpha wave activity in the frontal cortex is linked to relaxation, while beta-wave activity may signify arousal or cognitive engagement. Feature extraction involves power spectral density analysis ,time-frequency analysis to capture relevant patterns.

1. **Alpha Waves (8-12 Hz):**

   Alpha waves are dominant in the EEG when a person is awake but relaxed with their eyes closed. They are typically observed in the occipital region of the brain (at the back of the head) .Alpha waves are associated with a state of relaxation, calmness, and non-arousal. They often increase during meditation or deep relaxation.

2. **Beta Waves (12-30 Hz):**

   Beta waves are prominent in the EEG when a person is awake and engaged in active mental tasks, such as problem-solving, decision-making, or focused attention.They are often higher in frequency and amplitude during periods of mental activity and concentration.

3. **Gamma Waves (30-100 Hz):**

Gamma waves are the highest frequency brainwaves observed in the EEG. Gamma waves are thought to be involved in binding together different sensory inputs and integrating information across brain regions.

4. **Delta Waves (0.5-4 Hz):**

   Delta waves are the slowest frequency brainwaves observed in the EEG. They are most prominent during deep sleep stages, particularly during non-(rapid eye movement) sleep.Delta waves are associated with deep relaxation, unconsciousness, and restorative sleep. They play a crucial role in the maintenance of sleep and the restoration of bodily functions.

4.2 PPG (Photoplethysmography):

A PPG signal is a non-invasive optical measurement technique used to detect blood volume changes in the microvascular bed of tissue. It is commonly measured by placing a light source, typically an LED, and a photodetector on the skin.The PPG waveform, represents the pulsatile component of the arterial blood volume and can be used to estimate various physiological parameters such as heart rate, blood pressure, and even oxygen saturation. Motional arousal can lead to changes in peripheral blood flow, affecting PPG signals.For instance, increased sympathetic nervous system activity during stress or excitement can cause vasoconstriction, resulting in PPG amplitude changes.Time-domain features are extracted to characterize physiological responses associated with emotional states.

4.3 GSR (Galvanic Skin Response):

GSR stands for Galvanic Skin Response,it is a measure of the electrical conductance of the skin, which varies with the moisture level of the skin. GSR is typically measured by attaching electrodes to the skin, often on the fingers or palms.When a person experiences physiological or emotional arousal, such as stress, excitement, or anxiety, the activity of the sympathetic nervous system increases, leading to an increase in sweat gland activity and consequently an increase in skin conductance.GSR is used in various fields for research and applications, including psychology, neuroscience and

human-computer interaction. In psychology and neuroscience, GSR can be used to study emotional responses, arousal levels, and stress reactions. In human-computer interaction, GSR can be used as an input modality to infer the user's emotional state and adapt system responses accordingly.Higher arousal levels, such as stress or excitement, result in increased GSR.

4.4 DEAP Dataset:

The DEAP (Database for Emotion Analysis using Physiological Signals) dataset is a widely used benchmark dataset in the field of affective computing and emotion recognition. It was created to facilitate research on the relationship between physiological signals and human emotions.

0	0	AF3	FP1	FPZ	FP2	AF4	0	0
F7	F5	F3	F1	FZ	F2	F4	F6	F8
FT7	FC5	FC3	FC1	FCZ	FC2	FC4	FC6	FT8
T7	C5	C3	C1	CZ	C2	C4	C6	T8
TP7	CP5	CP3	CP1	CPZ	CP2	CP4	CP6	TP8
P7	P5	P3	P1	PZ	P2	P4	P6	P8
0	PO7	PO5	PO3	POZ	PO4	PO6	PO8	0
0	0	CB1	O1	OZ	O2	CB2	0	0

Figure 4.1: The compact 2D map of 62 channels

EEG: 32-channel EEG signals recorded at a sampling rate of 512 Hz.

PPG: Photoplethysmography signals recorded at a sampling rate of 64 Hz.

ECG: Electrocardiography signals recorded at a sampling rate of 64 Hz.

GSR: Galvanic Skin Response signals recorded at a sampling rate of 64 Hz.

In addition to the physiological signals, the DEAP dataset includes self-reported ratings of emotional valence (ranging from 1 to 9) and arousal (ranging from 1 to 9) for each video stimulus. These ratings provide ground truth labels for training and evaluating emotion recognition algorithms.

4.5 Features:

1. **Differential Entropy (DE):**

   Differential Entropy (DE) is a measure of uncertainty for continuous random variables, applied here to EEG signals reflecting brain activity.In emotional analysis, DE helps reveal patterns linked to various emotional states, providing insights into brain activity variability.We calculated DE from EEG signals to capture distinctive features related to valence and arousal.

   $$DE = \log(2 * \pi * e * \text{variance})/2 \tag{4.1}$$

   In each participant's video, there are 63 seconds of EEG signal data. The initial 3 seconds represent baseline data, and the following 60 seconds are trial data.For each segment of EEG data, we apply a Butterworth filter to divide it into four frequency bands: theta(4-8 Hz), alpha(8-12Hz), beta(12-30Hz), and gamma(>30Hz).

   We extract Differential Entropy (DE) features from each frequency band using a 0.5- second window and store them in vectors. We later normalize these vectors and organize the DE features into 2D maps and stack them, creating a 4D structure.

2. **Power Spectral Density (PSD):** Power Spectral Density (PSD) is a tool for frequency- domain analysis, depicting power distribution across signal frequencies.PSD unveils the intensity of neural activity in diverse frequency bands.

$$PSD(f_k) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f_k n/T} \right|^2 \tag{4.2}$$

$Here : x[n]$ is the signal in the discrete time domain.

f_k is the frequency corresponding to the k-th DFT bin.

N is the number of samples in the signal.

T is the total duration of the signal.

3. **Mean PPG (Photoplethysmography):**
   Represents the average value of the PPG signal over time.Provides insight into the average blood volume changes in the peripheral blood vessels.

4. **Standard Deviation of PPG:**
   Measures the variability or dispersion of the PPG signal values around the mean.Indicates the degree of fluctuation in the PPG signal, which may reflect changes in cardiac activity or vascular tone.

5. **Minimum PPG:**
   Represents the lowest value observed in the PPG signal.Indicates the minimum blood volume or intensity recorded during the measurement period

6. **Maximum PPG:**
   Represents the highest value observed in the PPG signal.Indicates the maximum blood volume or intensity recorded during the measurement period.

7. **Range of PPG:**
   Calculated as the difference between the maximum and minimum values of the PPG signal. Provides information about the amplitude or magnitude of blood volume changes.

8. **Skewness of PPG:**

   Measures the asymmetry of the PPG signal distribution around its mean. Positive skewness indicates that the tail of the distribution is longer on the right side, while negative skewness indicates a longer tail on the left side.

9. **Kurtosis of PPG:**

   Measures the peakedness or flatness of the PPG signal distribution. High kurtosis indicates a sharp peak and heavy tails, while low kurtosis indicates a flatter distribution.

4.6 Raspberry Pi

In selecting a hardware platform for deployment, the Raspberry Pi stands out due to its versatility, computational power, and strong community support. Unlike Arduino or ESP platforms, Raspberry Pi offers a complete computing environment, making it suitable for a wide array of projects, including those requiring machine learning inference and real-time data processing.

Advantages of Raspberry Pi:

Processing Power: Raspberry Pi boards are equipped with powerful ARM-based processors, enabling them to handle complex algorithms and deep learning models effectively.

Flexibility: With support for various operating systems and programming languages like Python and C/C++, Raspberry Pi offers developers the flexibility to choose the environment best suited for their applications.

Connectivity: Raspberry Pi boards provide extensive connectivity options such as Wi-Fi, Bluetooth, USB, and Ethernet, facilitating seamless communication with external devices and networks.

Community and Resources: Raspberry Pi boasts a vibrant community of developers and enthusiasts, offering access to extensive documentation, tutorials, forums, and open-source projects, which accelerates development and troubleshooting processes.

Terminal and VNC Viewer: For deploying our project on the Raspberry Pi, we primarily utilized the terminal interface for executing commands, managing files, and installing dependencies. Additionally, we leveraged VNC Viewer for graphical desktop access to the Raspberry Pi, allowing for remote desktop control and visualization of graphical user interfaces (GUIs) and applications running on the Raspberry Pi's desktop environment.

0000xtbfProject Implementation on Raspberry Pi: With the hardware setup and software configuration complete, we deployed our deep learning model deep learning based multimodal emotional recognition. Using the terminal interface and VNC Viewer, we executed commands to load the trained model and got the required accuracy for emotion recognition.

CHAPTER 5

Process Involved

In this section, we delve into the design principles and technical aspects underpinning our emotion classification system utilizing multimodal physiological signals, namely Electroencephalography (EEG), Photoplethysmography (PPG), and Galvanic Skin Response (GSR). Our methodology comprises signal preprocessing, feature extraction, and the application of deep learning techniques.

5.1 Data Preprocessing:

Our initial phase revolves around preparing the DEAP dataset for analysis. This process begins with downsampling the data to 128Hz, ensuring uniformity and reducing computational load while retaining essential information. Subsequently, we segment the dataset into 60-second trials, enabling focused analysis of discrete time periods. To ensure data integrity, we meticulously scrutinize and eliminate artifacts that may skew results, thereby enhancing the reliability of subsequent analyses. Additionally, we apply standard preprocessing techniques to refine the data further. This includes implementing bandpass filtering to isolate frequencies of interest, effectively removing noise and extraneous signals. Furthermore, baseline correction is employed to normalize the data, mitigating potential biases introduced during data acquisition. Through these rigorous preprocessing steps, we aim to optimize the quality and consistency of the dataset, laying a robust foundation for subsequent emotion classification.

5.2 Feature Extraction:

For each signal type (EEG, PPG, GSR), we extract a plethora of statistical features aimed at encapsulating pertinent information for emotion classification. These statistical features encompass fundamental metrics such as mean, standard deviation, minimum, maximum, peak-to-peak amplitude, skewness, and kurtosis.

5.3 Normalization:

The process of ensuring uniformity across feature distributions in machine learning is crucial for equitable contribution during model training. To achieve this, normalization techniques such as Z-score normalization are employed. Z-score normalization transforms each feature such that it has a mean of 0 and a standard deviation of 1, aligning their scales and distributions. By standardizing the features in this way, the dominance of certain features due to their scale is mitigated, leading to a more balanced training process. This normalization technique ensures that all features contribute equally to the learning process, preventing biases that may arise from differences in feature scales. Additionally, Z-score normalization aids in improving the convergence of optimization algorithms during training, as features with consistent scales facilitate smoother optimization. Furthermore, normalized features are easier to interpret, as they are all on the same scale, enhancing the understanding of their relative importance in the model. Overall, Z-score normalization enhances model performance, convergence, and interpretability, making it a fundamental step in the preprocessing pipeline of machine learning tasks.

5.4 Model Selection:

1. **K-nearest neighbors (KNN):**

   K-nearest neighbors (KNN) is a simple and popular machine learning algorithm used for classification and regression tasks. K-Nearest Neighbors (KNN) algorithm plays a pivotal role in classifying physiological signals. KNN is a non-parametric and instance-based learning method that relies on proximity

to determine classifications. By considering the 'k' nearest data points in the feature space, KNN effectively captured patterns in our physiological data, resulting in commendable accuracy, recall, F1 score, and precision. This approach demonstrated its adaptability and effectiveness in handling the complexity of physiological signal classification.

2. **4D convolutional recurrent neural network(CRNN) Modeling:**

   The 4D CRNN (Four-Dimensional Convolutional Recurrent Neural Network) is a deep learning architecture designed for processing spatiotemporal data, such as videos or volumetric data.The 4D-CRNN model integrates frequency, spatial, and temporal information in a 4D feature structure, employing a deep fusion of CNN and LSTM for effective learning.The 4D CRNN operates on four dimensions: width, height, time (frames), and channels.

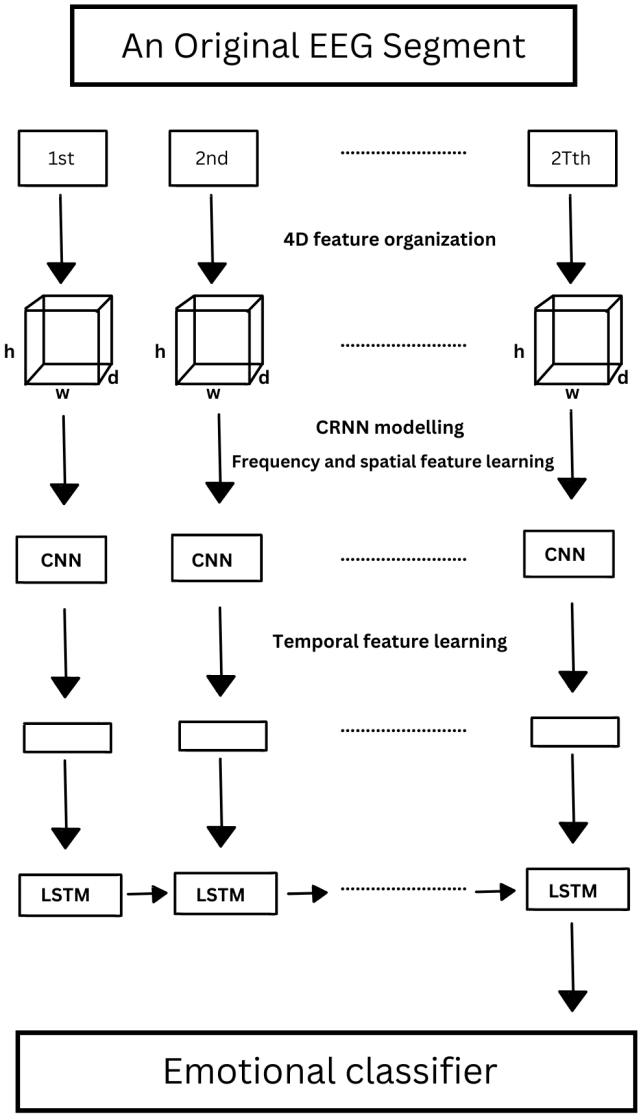

Figure 5.1: An overview of the proposed EEG-based emotion recognition framework using 4D-CRNN

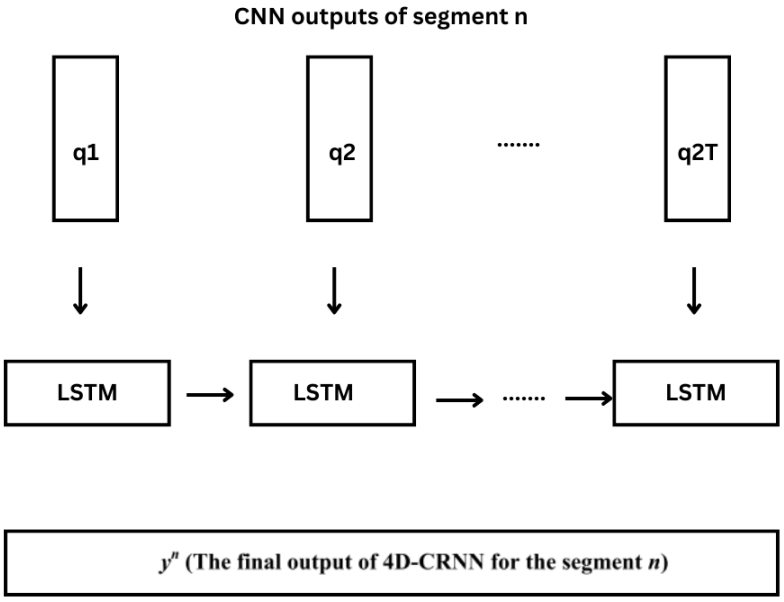

Figure 5.2: The structure of LSTM module for temporal feature learning

3. **Frequency and Spatial Feature Learning:**

   In this approach, we leverage a Convolutional Neural Network (CNN) to extract both frequency and spatial information from each temporal slice of the 4D structure, denoted as X_n. Unlike conventional CNN architectures that typically include pooling layers after each convolutional layer, we opt for a single pooling layer positioned after the last convolutional layer. This design choice aims to balance parameter reduction and information preservation, crucial for our small-sized 2D maps in X_n.

   This model consists of four convolutional layers (Conv1 to Conv4), a max-pooling layer (Pool), and a fully-connected layer (FC). Conv1 employs 64 feature maps with a 5x5 filter. Successive layers (Conv2 and Conv3) utilize 128 and 256 feature maps, respectively, with 4x4 filters. Conv4 employs 64 feature maps with a 1x1 filter, strategically employed to fuse feature maps from the prior convolutional layer. Zero-padding and Rectified Linear Units (ReLU) activation functions are applied across all convolutional layers.

   Following convolutional operations, a max-pooling layer (Pool) with a 2x2 size and a stride of 2 is introduced to mitigate overfitting and enhance network robustness. Subsequently, the outputs of the pooling layer are flattened and

directed to a fully-connected layer (FC) comprising 512 units. The resulting output, denoted as Q_n, is a representation encapsulating both frequency and spatial features of the original EEG segments.

4. **Temporal Feature Learning:**

   As EEG signal inherently possess dynamic characteristics, the nuances between temporal slices within the 4D structure may conceal valuable information crucial for enhanced emotion classification. To unveil this temporal information, we incorporate a Recurrent Neural Network (RNN) equipped with Long Short-Term Memory (LSTM) cells.

   Considering a CNN output sequence $Q_n = \{q_1, q_2, ..., q_{2T}\}$, where $q_t \in R^{512}$ and $t = 1, 2, ..., 2T$, we employ LSTM layers comprising 128 memory cells to uncover temporal dependencies within each segment, as illustrated in Fig. 8. The LSTM layer's output is calculated as follows

$$i_t = \sigma(W_q * q_t + W_h * h_{t-1} + W_c * C_{t-1} + b_i) \tag{5.1}$$

$$f_t = \sigma(W_q * q_t + W_h * h_{t-1} + W_c f * C_{t-1} + b_f) \tag{5.2}$$

$$c_t = f_t * C_{t-1} + i_t * \tanh(W_q c * q_t + W_h c * h_{t-1} + b_c) \tag{5.3}$$

$$o_t = \sigma(W_q o * q_t + W_h o * h_{t-1} + W_c o * C_t + b_o) \tag{5.4}$$

$$h_t = o_t * \tanh(c_t) \tag{5.5}$$

$$y_t = W_{ho} * h_t + b_o \tag{5.6}$$

   where σ is the logistic sigmoid function, and i,f,o and c are the input gate, forget gate, output gate and cell activation vectors. The W terms are weight matrices (e.g. W_{hi} is the hidden-input weight matrix), the b terms are bias vectors (e.g. b_i is the input bias vector) respectively.

5. **Long Short-Term Memory (LSTM):**

   Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) architecture that is particularly well-suited for handling sequential data. This includes time series data, text data, and more. In the context of

deep learning-based emotional classification, LSTM networks can be employed to effectively capture temporal dependencies within the input data, which is crucial for understanding emotional context. During training, backpropagation through time optimizes LSTM parameters, enabling the model to discern nuanced emotional patterns. Evaluation metrics such as accuracy and F1-score gauge the model's performance in classifying emotions.

6. **Classifier**

   Based on the final feature representation y_n, we predict the label of the original EEG segment X_n by a linear transform approach, which can be computed as

$$\mathrm{OUT} = A * y_n + b = [\mathrm{out}_1, \mathrm{out}_2, ..., \mathrm{out}_C] \tag{5.7}$$

   where A is the transform matrix, b is the bias and C is the number of emotion category. Then, the output is fed into a softmax classifier for emotion recognition, which can be formulated as

$$P(c|X_n) = \frac{\exp(\mathrm{out}j)}{\sum i = 1^C \exp(\mathrm{out}_i)}, \quad j = 1, ..., C \tag{5.8}$$

   where $P(c|X_n)$ represents the probability of the EEG segment X_n belonging to the class c.

5.5 Model Architecture:

Our model architecture centers around Long Short-Term Memory (LSTM) neural networks, meticulously crafted to capture temporal dependencies inherent in physiological signals. Comprising multiple LSTM layers supplemented by dropout layers to forestall overfitting, our architecture adeptly assimilates sequential information from each signal to discern temporal dynamics effectively.

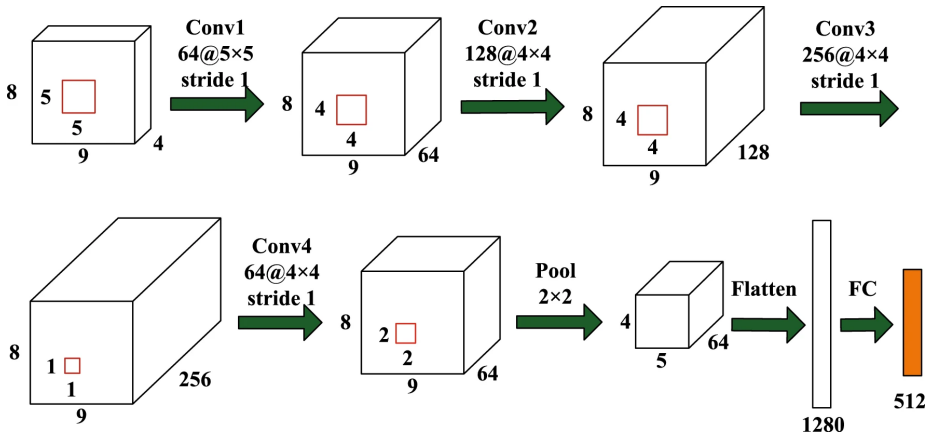

Figure 5.3: The structure of CNN module for frequency and spatial feature learning.[1]

5.6 Training Procedure:

Employing a conventional train-test split strategy, we partition the preprocessed data into training and testing sets. Leveraging the Adam optimizer and binary cross-entropy loss function, we train the LSTM models, integrating early stopping and model checkpoints to forestall overfitting and preserve optimal model iterations.

5.7 Arousal and Valence Classification:

Undertaking binary classification for both arousal and valence dimensions, we categorize arousal and valence labels through thresholding to delineate high and low emotional states, thereby facilitating binary classification tasks. Model performance is meticulously evaluated utilizing metrics such as accuracy and confusion matrices to gauge the models' efficacy in discerning emotional states.

5.8 Integration and Fusion:

In our pursuit to amalgamate information gleaned from multiple signals, we explore fusion strategies, encompassing feature concatenation and attention mechanisms. Our integrated models endeavor to harness complementary insights from EEG, PPG, and GSR signals, culminating in enhanced emotion classification capabilities.

CHAPTER 6

Results

	Accuracy	Precision	Recall	F1 Score	Confusion matrix
KNN(PPG)	0.5223	0.5593	0.5506	0.5159	[42 49] [58 75]
KNN(EEG)	0.4910	0.5363	0.5147	0.4978	[47 44] [50 83]
KNN(GSR)	0.5758	0.5515	0.6068	0.5557	[47 44] [51 82]
KNN(Combined)	0.5883	0.6096	0.5897	0.5702	[54 50] [57 95]
4d-CRNN	0.9433	-	-	-	-
LSTM Feature (EEG)	0.766	0.803	0.779	0.791	[78 26] [30 106]
LSTM Feature (GSR)	0.783	0.81	0.786	0.79	[83 21] [31 105]
LSTM Feature (PPG)	0.775	0.81	0.786	0.798	[79 25] [29 107]
LSTM Feature (All)	0.758	0.782	0.835	0.807	[60 34] [24 122]
LSTM Raw (EEG)	0.762	0.811	0.757	0.873	[80 24] [31 103]
LSTM Raw (GSR)	0.758	0.795	0.772	0.783	[77 27] [31 105]
LSTM Raw (PPG)	0.787	0.834	0.779	0.806	[83 21] [30 106]
LSTM Raw (All)	0.762	0.821	0.742	0.779	[82 22] [35 101]

Table 6.1: Performance metrics for Valence Predictions

	Accuracy	Precision	Recall	F1 Score	Confusion matrix
KNN(PPG)	0.5312	0.4261	0.64330	0.4883	[42 69] [36 77]
KNN(EEG)	0.5312	0.4662	0.6130	0.4829	[43 68] [42 71]
KNN(GSR)	0.5312	0.4532	0.6365	0.4922	[43 68] [37 76]
KNN(Combined)	0.5089	0.3920	0.5933	0.4529	[49 78] [48 81]
4d-CRNN	0.9423	-	-	-	-
LSTM Feature (EEG)	0.67	0.638	0.838	0.724	[57 59] [20 104]
LSTM Feature (GSR)	0.683	0.651	0.83	0.73	[62 54] [22 102]
LSTM Feature (PPG)	0.683	0.653	0.822	0.728	[62 54] [22 102]
LSTM Feature (All)	0.695	0.69	0.855	0.763	[49 53] [20 118]
LSTM Raw (EEG)	0.575	0.557	0.854	0.675	[32 84] [18 106]
LSTM Raw (GSR)	0.687	0.604	0.862	0.74	[58 58] [17 107]
LSTM Raw (PPG)	0.662	0.638	0.838	0.719	[55 61] [20 104]
LSTM Raw (All)	0.675	0.641	0.838	0.727	[58 58] [20 104]

Table 6.2: Performance metrics for Arousal Predictions

	Accuracy	Precision	Recall	F1 Score	Confusion matrix
LSTM Feature (EEG)	0.687	0.717	0.881	0.791	[23 56] [19 142]
LSTM Feature (GSR)	0.662	0.711	0.755	0.732	[48 45] [36 111]
LSTM Feature (PPG)	0.654	0.677	0.829	0.746	[35 58] [25 122]
LSTM Feature (All)	0.695	0.729	0.869	0.793	[27 52] [21 140]

Table 6.3: Performance metrics for Liking Predictions

	Accuracy	Precision	Recall	F1 Score	Confusion matrix
LSTM Feature (EEG)	0.804	0.838	0.901	0.869	[37 30] [17 156]
LSTM Feature (GSR)	0.775	0.842	0.857	0.827	[42 30] [24 144]
LSTM Feature (PPG)	0.791	0.827	0.886	0.856	[41 31] [19 149]
LSTM Feature (All)	0.804	0.828	0.919	0.871	[34 33] [14 159]

Table 6.4: Performance metrics for Dominance Predictions

The KNN model yielded accuracies ranging from approximately 49% to 58% for valence prediction and about 51% to 53% for arousal when applied individually to each signal. Combining all signals improved accuracy to around 59% for valence and 50.9% for arousal. Notably, the specialized 4d-CRNN model achieved exceptional accuracies of 94.33% for arousal and 94.23% for valence, demonstrating its effectiveness with EEG signals. LSTM models, trained separately for each signal type, showed accuracies ranging from approximately 66% to 78.3% for valence and around 66.2% to 68.3% for arousal when utilizing feature-extracted data. When trained on raw data, LSTM models achieved accuracies ranging from approximately 75.5% to 78.1% for valence and 57.5% to 68.7% for arousal. These results indicate the potential of combining physiological signals and employing advanced models for accurate emotion prediction across different modalities.

6.0.1 Results from raspberry pi implementation

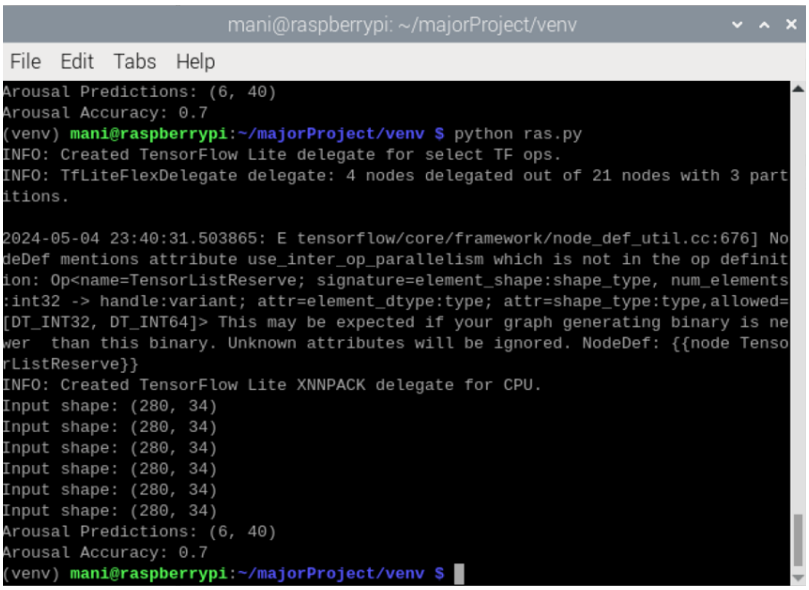

Figure 6.1: Performance metrics for Arousal Predictions On Raspberry Pi

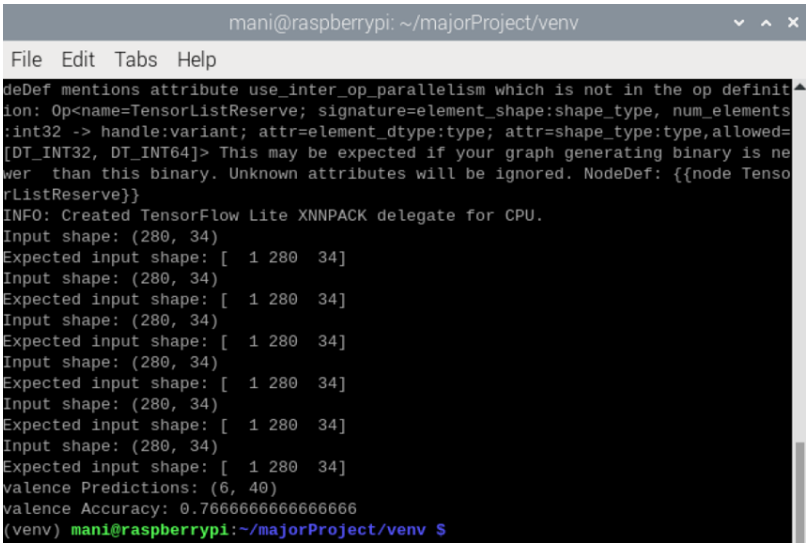

Figure 6.2: Performance metrics for valence Predictions On Raspberry Pi

The Accuracy of Valence is 76.6 % on Raspberry Pi Using LSTM Model.

The Accuracy of Arousal is 70.0 % on Raspberry Pi Using LSTM Model.

CHAPTER 7

Conclusion

In this project, we wanted to see if we could guess how people were feeling by looking at their body signals like brain waves, skin responses, and heartbeats. We started by trying a basic method called k-Nearest Neighbors on each signal EEG, GSR, and PPG alone, then combined them to see if that worked better. Later, we dug deeper by focusing on the brain waves (EEG) and using a special model called 4DCRNN.

We also trained Long Short-Term Memory (LSTM) models separately for each signal type and combined signals by using raw data and by using features like mean, standard deviation, minimum, maximum, peak-to-peak amplitude, skewness, and kurtosis . Our results showed that combining signals generally improved prediction accuracy, with deep learning models performing well.

Our findings showed that combining these signals generally improved our ability to predict emotions. Notably, the 4DCRNN model for brain waves achieved high accuracies of 94.33% for arousal and 94.23% for valence by using single EEG signal. The LSTM models, trained on different signal types, also showed promising results, with accuracies ranging from about 75.8% to 78.3%.

This project is a big step towards creating a system that can accurately understand human emotions by analyzing their body signals. By combining different signals and using advanced techniques like deep learning, we've shown that it's possible to identify emotions effectively.

In conclusion, our project represents a significant step towards developing a practical and effective Multimodal Emotional Recognition system. By combining physiological signals and deep learning techniques, we have demonstrated the potential to accurately identify human emotions across multiple modalities. Moreover, the integration of Raspberry Pi ensures the system's accessibility and suitability for real-

world deployment. After implementing on Raspberry Pi, we got valence accuracy of 70.0% and arousal accuracy of 76.6%.

Moving forward, further refinement and optimization of the system could enhance its performance and extend its applicability to a wider range of scenarios. Overall, our work lays a foundation for future research and development in the field of emotion recognition, with promising implications for various domains.

BIBLIOGRAPHY

[1] Shen, Fangyao, Guojun Dai, Guang Lin, Jianhai Zhang, Wanzeng Kong, and Hong Zeng. "EEG-based emotion recognition using 4D convolutional recurrent neural network." Cognitive Neurodynamics 14 (2020): 815-828.

[2] M. Khateeb, S. M. Anwar and M. Alnowami, "Multi-Domain Feature Fusion for Emotion Classification Using DEAP Dataset," in IEEE Access, vol. 9, pp. 12134-12142, 2021, doi: 10.1109/ACCESS.2021.3051281.

[3] S. Koelstra et al., "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," in IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18-31, Jan.-March 2012, doi: 10.1109/T-AFFC.2011.15.

[4] Ayata D, Yaslan Y, Kamaşak M. Emotion recognition via galvanic skin response: Comparison of machine learning algorithms and feature extraction methods. IU-Journal of Electrical Electronics Engineering. 2017 Mar 3;17(1):3147-56.

[5] N. Y. Oktavia, A. D. Wibawa, E. S. Pane and M. H. Purnomo, "Human Emotion Classification Based on EEG Signals Using Naïve Bayes Method," 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2019, pp. 319-324, doi:10.1109/ISEMANTIC.2019.8884224