Acknowledgement

This Mini Project report would not have been come into reality without the able guidance, support and wishes of all those who stand by us in the development. We wish to give our special thanks to our guide **Dr. Mansi Subhedar** for her timely advice and guidance.

We humbly thank our project Coordinator **Prof. Pooja Kulkarni** and Head of the Department, **Dr. Mansi Subhedar** for their valuable guidance and unending support. We would like to thank our Principal, **Dr. Jagdish W. Bakal** for his constant encouragement throughout the course. The cheerful spirit they radiated all the time fueled our desire to excel in the work that we had undertaken.

We acknowledge all the faculty members of the Department of Electronics and Computer Science for their help and suggestions during various phases of this project work.

Vignesh Vane
Diksha Modi
Karunesh Ghadage
Nayan Patil

Abstract

Emotion recognition plays a crucial role in human-computer interaction, enhancing applications in healthcare, security, and user experience. This project focuses on developing a multimodal emotion recognition system that predicts human emotions by analyzing physiological and visual data. The system integrates data from a temperature sensor, camera, SpO_2 sensor, and heart rate monitor to improve the accuracy of emotion detection.

The methodology involves data collection from these diverse sources, followed by preprocessing and feature extraction. Machine learning algorithms are then applied to classify emotions into predefined categories such as happiness, sadness, anger, fear, and neutral states. By combining physiological signals with facial expressions, the system achieves a more comprehensive and accurate analysis of emotional states.

The proposed multimodal emotion recognition system demonstrates how integrating multiple data sources enhances the reliability and robustness of emotion prediction. This work has potential applications in mental health monitoring, affective computing, and advanced human-computer interfaces.

List of Figures

Contents

Chapter 1

Introduction

1.1 Introduction

Emotions are fundamental to human experience and influence decision-making,communication, and behavior. With the rise of intelligent systems, accurately recognizing and interpreting human emotions has become an essential goal in fields like healthcare, human-computer interaction, and security. Traditional emotion recognition methods often rely on a single modality, such as facial expressions or speech, which can be limited by environmental factors or individual differences..

Multimodal Emotion Recognition addresses these challenges by combining data from multiple sources to improve the accuracy and robustness of emotion detection. This project integrates facial expressions, voice analysis, heart rate, SpO_2 levels, and body temperature to capture a comprehensive representation of emotional states. By leveraging Machine Learning techniques, this system enhances the ability to interpret complex emotional cues from diverse physiological and behavioral data.[1]

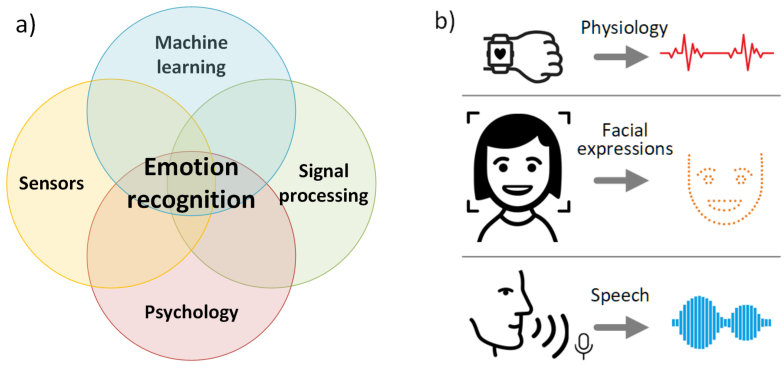

Figure 1.1: Emotion Recognition

1.1.1 Need and scope of the Project

With the increasing integration of artificial intelligence in daily life, there is a growing need for systems that can understand and respond to human emotions effectively. Emotion recognition is critical for applications requiring personalized interactions and enhanced decision-making capabilities.[1]

Need:

- Enhanced Human-Computer Interaction: Emotion-aware systems improve user experiences by adapting responses to emotional states.
- Healthcare Monitoring: Early detection of emotional distress can aid in mental health assessment and intervention.
- Security and Surveillance: Identifying stress or anxiety in high-risk environments enhances safety and threat detection.

Scope:

- Multimodal Integration: Combines facial, vocal, and physiological data for improved accuracy.
- Machine Learning Models: Utilizes advanced models like CNN for facial analysis and explores other techniques for different modalities.

- Diverse Applications: Useful in mental health monitoring,smart environments,adaptive user interfaces, and beyond.
- Future Improvements: Provides a foundation for expanding to real-time systems.

Figure 1.2: Emotion Relationship

1.1.2 Motivation

- Mental health disorders are rising, but early intervention is lacking.
- Single-modality systems (self-reporting, facial recognition, voice analysis) struggle with accuracy in real-world settings.
- Multimodal approaches improve reliability by combining physiological, facial and voice-based emotion analysis.
- IoT and ML advancements enable continuous monitoring for stress detection, mental wellness, and smart AI applications.

Chapter 2

Literature Review

2.1 Literature Review

Emotion recognition has become vital for intelligent systems in healthcare, smart environments, and mental wellness. Traditional unimodal approaches—relying on a single signal like facial expression or speech—often fall short in real-world scenarios, motivating the shift toward multimodal methods.

- Unimodal Emotion Detection

  Early systems focused on single modalities such as facial expressions using datasets like FER2013 and techniques like LBP [2], but struggled with lighting and occlusion. Audio-based models using MFCCs and pitch trained on datasets like TESS , SAVEE , and CREMA-D faced issues with noise [3]. Physiological signals like ECG and SpO offered more stable data, but lacked contextual nuance and were sensitive to individual variation [4].

- Multimodal Fusion Techniques

  To improve reliability, multimodal fusion has been explored. Tensor Fusion Networks (TFN) combined visual, audio, and physiological inputs for richer emotion recognition [8], while datasets like CosMIC [9] introduced multi-signal integration. Transformer based models further improved performance across modalities [10], but practical deployment remains limited due to data and computational constraints. Systems like those by Ayubi et al. [11] focused on wearable data but lacked vocal and facial context.

- Deep Learning Advances

  CNNs and LSTMs have enabled automatic feature extraction, improving performance. For example, neural networks applied to ECG data reached 92% accuracy in controlled conditions [4]. Visual models trained via transfer learning on FER2013 achieved 70–75% [6], and speech-based models reached 65–70% depending on noise[7] . Transformer based models [10] show promise but require significant data and tuning.

  Most systems remain unimodal, limiting adaptability to real-time use.

These studies highlight the effectiveness of integrating facial, vocal, and physiological data for comprehensive emotion recognition.

2.2 Problem Definition

Accurate emotion recognition is challenging due to the limitations of using a single data source. This project aims to develop a Multimodal Emotion Recognition system that integrates facial expressions, voice analysis, heart rate, SpO_2 levels, and body temperature to improve the accuracy and reliability of emotion detection using Machine Learning techniques. The goal is to enhance emotional understanding across diverse applications by combining multiple physiological and behavioral signals.

2.3 Objectives

- To develop a multimodal system that integrates facial expressions, voice analysis, heart rate, SpO_2, and temperature for accurate emotion recognition, and to implement a Convolutional Neural Network (CNN) for facial emotion recognition.
- To explore and apply suitable machine learning models for analyzing voice and physiological data , and to fuse multiple data sources for improved accuracy and robustness in emotion prediction.
- To evaluate the system's performance using appropriate metrics and validate its effectiveness across different emotional states, and to create a foundation for future improvements in real-time emotion recognition systems and broader emotion categories.
- To design a web dashboard for real-time visualization of classified emotions, and to integrate Firebase for secure storage and efficient access to datasets.

Chapter 3

Design Methodology

3.1 Hardware Components

- **ESP32**
  - Type: Microcontroller with integrated Wi-Fi and Bluetooth.
  - Features: Dual-core processor, low power consumption, multiple GPIO pins,uses I^2C interface for easy connection between sensors

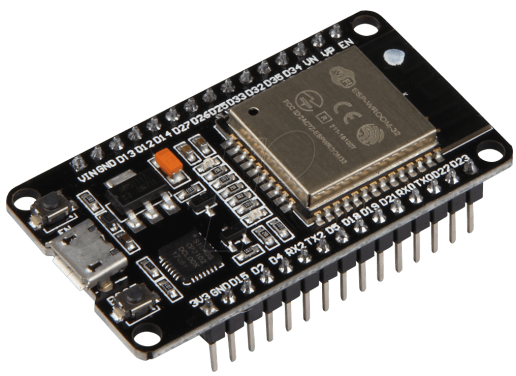

Figure 3.1: ESP32

.

- **MAX30102**
  - Type: Pulse oximeter and heart rate sensor.
  - Features: Measures SpO_2 (blood oxygen levels),heart rate using infrared and red LEDs and uses the I^2C (Inter-Integrated Circuit) interface for easy connection to microcontrollers

Figure 3.2: MAX30102

- **DS18B20**
  - Type: Digital temperature sensor.
  - Features: Accurate temperature readings (-55°C to +125°C), operates on 1-wire protocol.

Figure 3.3: DS18B20

- **Camera**
  - Type: Digital camera module for image capture.
  - Features: Captures still images and video.
- **Microphone**
  - Type: Audio input device (analog or digital).
  - Features: Captures sound waves and converts them to electrical signals.

3.2 Software Components

- **Arduino IDE:**An open-source platform to write and upload code to Arduino boards, used for collecting sensor data like temperature, SpO_2, and heart rate. .

Figure 3.4: Arduino

- **Python:**A versatile programming language used for processing data, implementing machine learning models, and integrating with Firebase. .

.

Figure 3.5: Python

- **Firebase:**A cloud-based platform used to store and retrieve real-time data, including sensor inputs and emotion predictions.

Figure 3.6: Firebase

- **Web Dashboard:**An interactive interface that displays real-time emotion predictions using data from Firebase for easy monitoring.

3.3 Design Methodology

The figure 3.7 illustrates the end-to-end architecture of a real-time multimodal emotion recognition system. Data is collected using sensors and a camera connected to an ESP32, with physiological, facial, and vocal inputs processed separately. The ESP32 transmits data to Firebase, which is then fetched by a Python backend for preprocessing and emotion classification using machine learning models. Results from facial, vocal, and physiological analysis are fused to generate a final emotion prediction, which is visualized on a React.js dashboard for real-time monitoring.

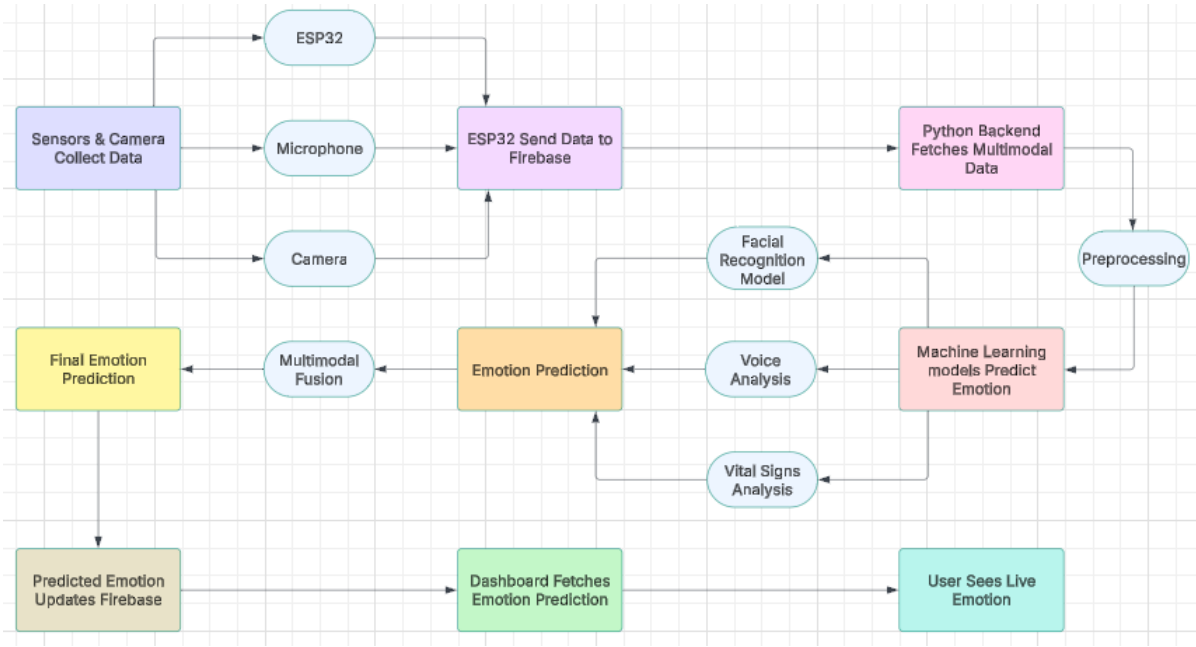

Figure 3.7: System Workflow for Multimodal Emotion Recognition

3.4 Circuit Diagram

Figure 3.8 consists the circuit connects the ESP32 (U1) microcontroller with two sensors:

- MAX30102 (U3) for heart rate and SpO_2
- DS18B20 (U2) for body temperature

Connections:

- MAX30102:
  - SCL → GPIO22 (ESP32)
  - SDA → GPIO21 (ESP32)
  - GND → GND
  - VDD → 3.3V (via VDD pin)

- DS18B20:
  - DQ → GPIO32 (ESP32)
  - VDD → 3.3V
  - GND → GND

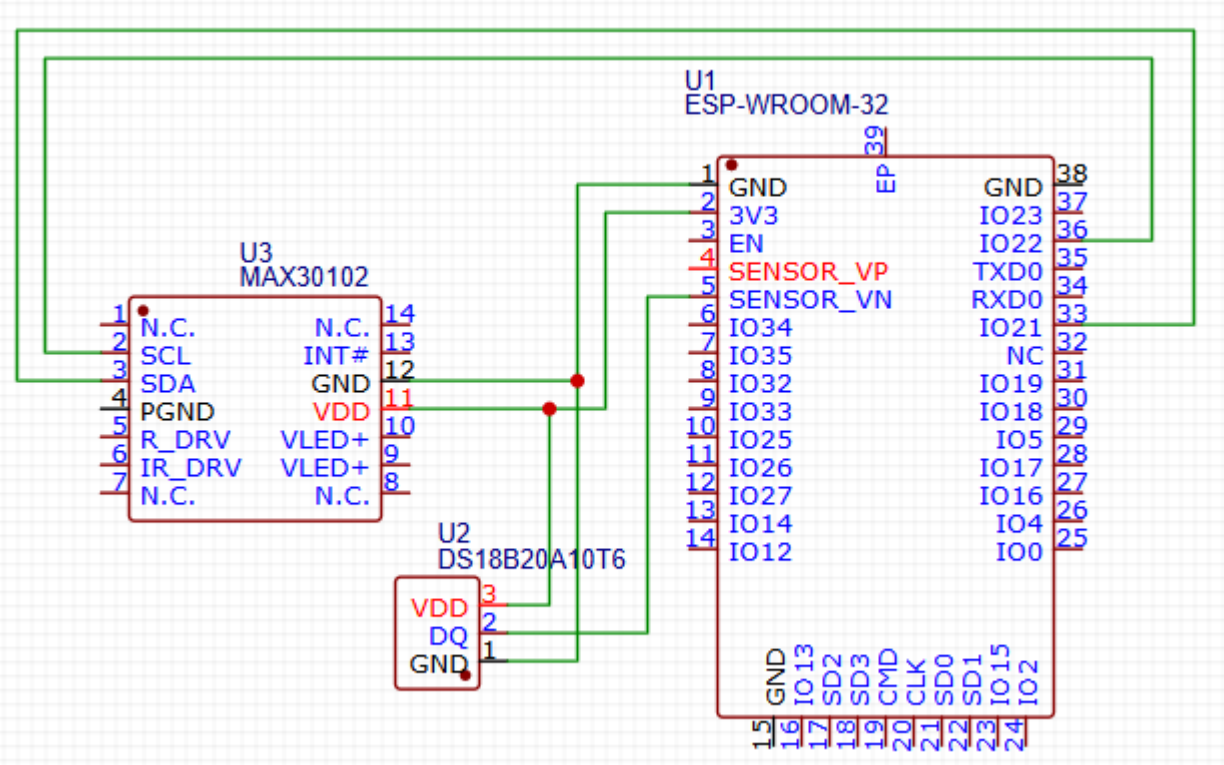

Figure 3.8: ESP32-Based Sensor Interface for Emotion Recognition

3.5 Implementation

1. **Data Collection**

   - Sensor data (body temperature, SpO_2, heart rate) is gathered using an ESP32 microcontroller.[4].
   - Facial expressions are captured through a webcam.
   - Voice data is taken from public datasets like CREMA-D[14] , SAVEE [15], and TESS[16] for training.

2. **Data Processing**

   - Sensor Data: Sent from Arduino to Firebase at 1 Hz using the FirebaseESP32 library.[5].
   - Voice Features: Extracted using Librosa (MFCCs, pitch, RMS, spectral centroid).
   - Facial Images: Detected via Haar Cascade, resized to 48×48, normalized, and flattened using OpenCV.[6].

3. **Machine Learning Model**

   - Model Types:
     - CNN for facial emotion recognition.[6]
     - LSTM for vocal emotion recognition.LSTM[6]
     - Random Forest for physiological emotion detection.[4]
   - Fusion Strategy: Weighted combination: 60% Physiological, 20% Facial, 20% Vocal.
   - Training Frameworks: Used TensorFlow , Scikit-learn , and KerasTensorFlow[6], Scikit-learn[13], and Keras[9]

     .

4. **Real-Time Data Handling**

   - Data Synchronization: Python script aggregates and timestamps data from three Firebase sources every 5 seconds.[5]
   - Unified Output: Compiled into a single JSON object and pushed to a central Firebase database.
   - Live Updates: Ensures real-time responsiveness of predictions and input streams.

5. **Web Dashboard**

   - Developed using React.js.[7].
   - Features:
     - Real-time emotion classification display.
     - Graphical visualization of physiological data using React-chartjs-2.
     - Live webcam feed.
   - Connects to Firebase via Firebase JS SDK for dynamic and real-time updates.[5]

3.6 Use Case

Emotion-Aware Virtual Assistant for Elderly Care

Actors:

- Elderly User – Senior citizen using a smart assistant for daily support.
- Caretaker – Family member or healthcare provider monitoring the user remotely.
- Emotion Recognition System – Backend system collecting and interpreting emotional signals.

Scenario:

An elderly user interacts with a smart assistant embedded with sensors (SpO_2, heart rate, temperature)[4], a microphone, and a camera. Throughout the day:

1. The system passively collects real-time physiological and behavioral data via ESP32 and sends it to Firebase.
2. The Python backend processes the data, and machine learning models analyze facial expressions, voice tone, and vital signs[6] to assess the user's emotional state.
3. The web dashboard displays current emotions like stress, happiness, or sadness.
4. If anxiety, loneliness, or distress is detected, the system alerts the caretaker.

Outcome:

The caretaker is notified of emotional changes, allowing them to check in with the user or take action if needed. This enables continuous emotional support and improves quality of life and safety for the elderly.

Chapter 4

Results

The Fig. 4.1 contains the Real-Time image of circuit connection between the sensors (MAX30102, DS18B20) and the ESP32 microcontroller.

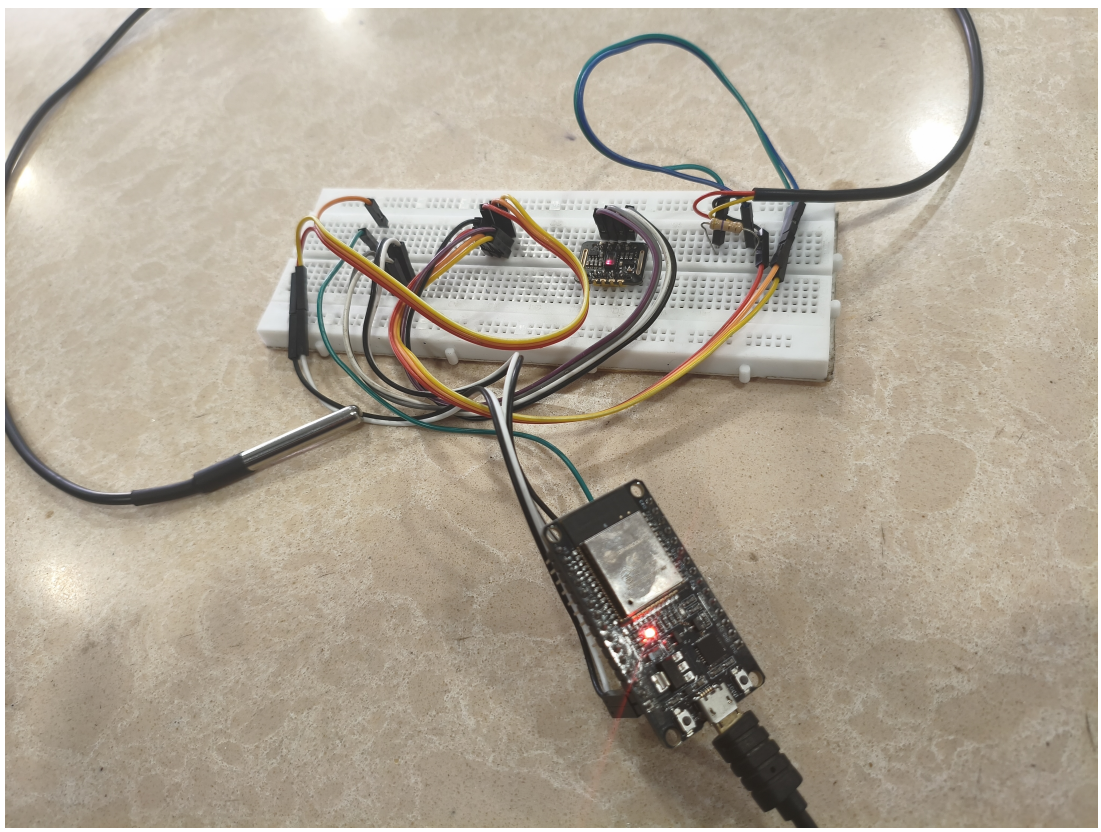

Figure 4.1: Circuit Implementation/Circuit Connection

Figure 4.2 illustrates the finalized user interface, providing a real-time visualization of the predicted emotional state.

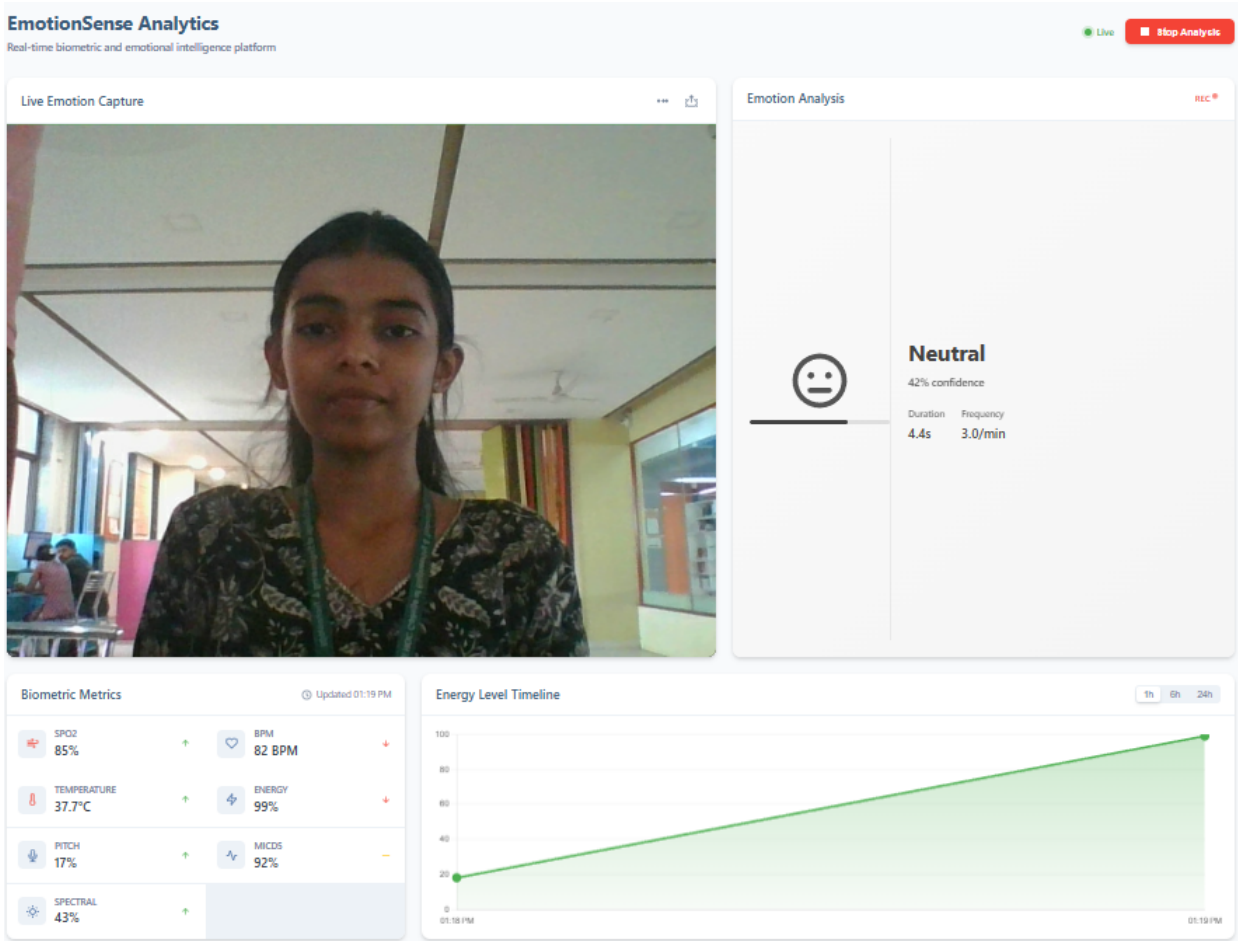

Figure 4.2: Emotion Prediction Result Interface

Vocal Analysed Parameters

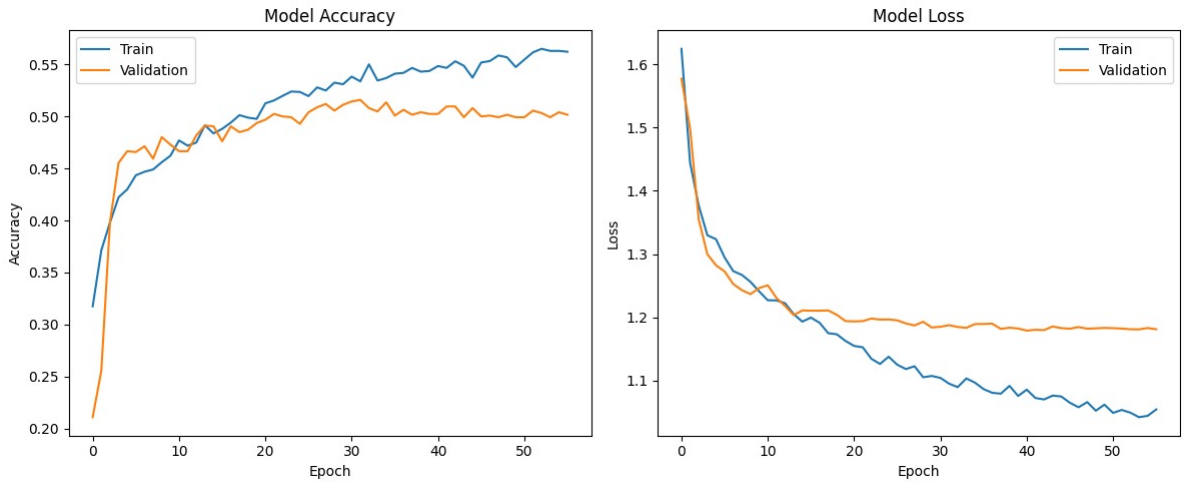

Figure 4.3: Vocal Feature Extraction Graph

Physiology Analysed Parameters

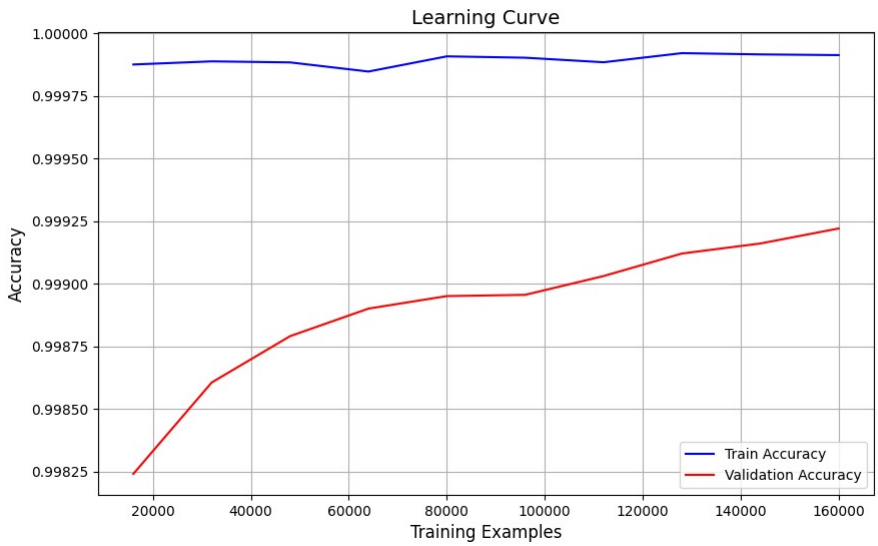

Figure 4.4: Physiology Emotion Model Learning CurveS

Facial Analysed Parameters

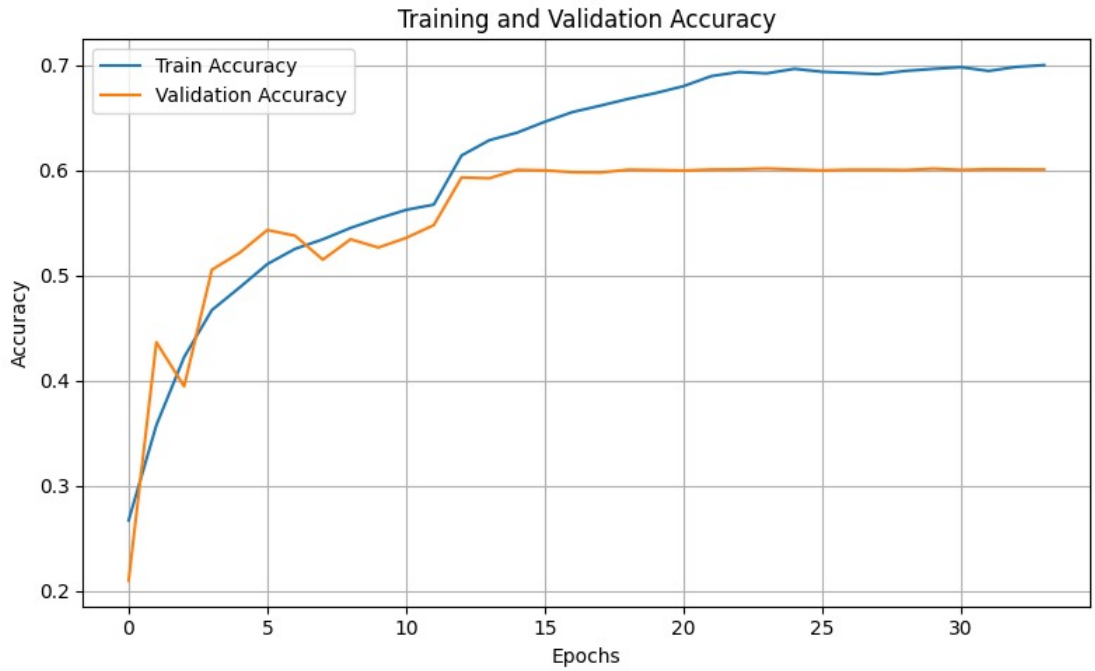

Figure 4.5: Facial Features Extraction Graph

Performance Metrics

Table 4.1: Modality-wise Performance Matrix

Modality	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Facial (CNN)	60.18%	61.02%	60.18%	60.36%
Vocal (LSTM)	62%	60%	60%	59%
Physio (RF)	99.93%	99.93%	99.93%	99.93%
Fusion	83.95%	84.8%	84.9%	84.7%

Chapter 5

Conclusion

The Multimodal Emotion Recognition System successfully integrates physiological and behavioral data to enhance the accuracy of emotion detection. By leveraging multiple data sources—including temperature,SpO_2,heart rate, facial expressions and voice sentiment—the system overcomes the limitations of unimodal approaches. The implementation of machine learning models and data fusion techniques ensures a more comprehensive analysis of emotional states.

Real-time classification and visualization on a web dashboard, with Firebase as a cloud backend, make the system highly responsive and accessible. The project demonstrates significant potential for applications in healthcare, human-computer interaction, and mental health monitoring.

Moving forward, improvements in model accuracy,additional physiological signals,and expanded real-world testing can further enhance the system's reliability and usability. This project lays the groundwork for future advancements in emotion-aware AI systems,contributing to smarter, more intuitive human-computer interactions.

Chapter 6

Applications and Future Scope

■ Applications

- Healthcare and Mental Health Monitoring
  - Assists in detecting stress, anxiety, and emotional disorders.
  - Provides real-time emotional feedback for mental health assessment.
  - Helps doctors and therapists track emotional well-being remotely.
- Education and E-Learning
  - Monitors student engagement and emotional states during online learning.
  - Helps teachers adapt content delivery based on emotional feedback.
- Automotive and Safety Applications
  - Detects driver fatigue or stress to prevent accidents.
  - Can be used in emotion-aware in-car assistants.

■ Future Scope

- Real-World Deployment and Testing
  - Implement in real-time environments such as hospitals, smart homes, and workplaces.
  - Perform large-scale user studies to validate system accuracy.
- Mobile and IoT Integration
  - Develop mobile applications for continuous emotion tracking.
  - Extend support for IoT-based wearable devices for seamless monitoring.
- Cross-Platform and AI Integration
  - Integrate with AI assistants (Alexa, Google Assistant, Siri) for better human-machine interaction.
  - Expand compatibility with smart home and healthcare systems.

References

1. M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," Proceedings of the IEEE, vol. 91, no. 9, pp. 1370–1390, 2003

2. G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915–928, 2007.

3. Z. Zhang and B. Schuller, "Speech emotion recognition: Features and classification models," Speech Communication, vol. 114, pp. 3–20, 2019.

4. S. Kumar, P. Bharti, and V. Nagar, "ECG-based emotion recognition using wavelet transform and neural networks," International Journal of Computer Sciences and Engineering, vol. 8, no. 4, pp. 23–27, 2020.

5. A. Dhall et al., "Emotion recognition in the wild challenge: EmotiW 2013," in Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI), 2013.

6. L. Cohn, M. Schuller, and B. Schuller, "Multimodal emotion recognition: State-of-the-art and open challenges," IEEE Transactions on Affective Computing, vol. 11, no. 1, pp. 74-90, 2020.

7. R. Mathew and K. Sharma, "Building real-time dashboards with React.js and Firebase," International Journal of Computer Applications, vol. 11, no. 4, pp. 55-64, 2019.

8. L. Wang et al., "Real-time emotion recognition system using deep learning on multimodal data," Journal of Real-Time Image Processing, vol. 16, no. 2, pp. 329-340, 2020.

9. R. Tripathi et al., "Transformer-based multimodal emotion recognition," arXiv preprint arXiv:2111.10202, 2021.

10. A. Zadeh et al., "Multimodal Sentiment Intensity Analysis in Videos: Facial, Verbal, and Physiological Fusion," arXiv preprint arXiv:1810.04635, 2018.

11. D. Ghosal et al., "CosMIC: A Multimodal Emotion Recognition Dataset," arXiv preprint arXiv:2006.08129, 2020.

12. J. Ayobi et al., "Mental health monitoring using wearable and mobile sensors," Multimodal Technologies and Interaction, vol. 6, no. 3, pp. 28, 2022.

13. M. Sambare, "FER2013 Facial Expression Recognition Dataset," Kaggle, Available: https://www.kaggle.com/datasets/msambare/fer2013

14. , "Toronto Emotional Speech Set (TESS)," Kaggle, Available: https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess

15. E. Lok, "Surrey Audio-Visual Expressed Emotion (SAVEE)," Kaggle, Available: https://www.kaggle.com/datasets/ejlok1/surrey-audiovisual-expressed-emotion-savee

16. E. Lok, "CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset," Kaggle, Available: https://www.kaggle.com/datasets/ejlok1/cremad

17. N. Ayub, "Human Vital Sign Dataset," Kaggle, Available: https://www.kaggle.com/datasets/nasirayub2/human-vital-sign-dataset