

Project Title: Student's Emotion Recognition using Multimodality and Deep Learning

Abstract

With the rapid growth of online and digital learning environments, understanding students' emotional states has become essential for improving learning outcomes and providing timely instructor support. Traditional emotion detection methods rely on a single data source such as facial expressions or text, which often results in lower accuracy and limited contextual understanding. Additionally, existing systems lack real-time monitoring, adaptive multimodal fusion, and actionable insights for educators.

This project presents the development of an **AI-powered Multimodal Emotion Recognition System** designed for academic environments. The system integrates facial expressions, speech signals, and textual data to identify student emotions more accurately using deep learning techniques. **CNN-based models** are used for facial emotion classification, speech processing models analyze vocal features, and NLP techniques are applied for text emotion detection. The platform combines these modalities using adaptive fusion methods to improve performance compared to unimodal approaches.

The system supports continuous or near real-time emotion monitoring and provides a clean, user-friendly web dashboard displaying detected emotions, trend graphs, and alerts. It also generates emotion-based summaries to assist instructors in identifying student engagement and well-being. Model performance will be evaluated using standard metrics such as accuracy, precision, recall, and F1-score. By combining artificial intelligence with an intuitive interface, the platform enables data-driven decision-making in academic environments and enhances student support mechanisms.

1. User Management & Security

- **FR1:** The system shall provide a user interface that allows users to upload image, audio, or text inputs for emotion analysis.
 - **FR2:** The system shall process user inputs and display the detected emotion results on the dashboard.
-

2. Multimodal Data Acquisition & Management

- **FR3:** The system shall capture facial data through webcam or video input.
 - **FR4:** The system shall capture speech/audio input for emotion analysis.
 - **FR5:** The system shall accept textual input such as chat messages, feedback, or responses.
-

3. Facial Emotion Recognition

- **FR6:** The system shall detect faces from image or video frames automatically.
 - **FR7:** The system shall classify emotions using CNN-based deep learning models.
 - **FR8:** The system shall identify emotions such as happy, sad, angry, neutral, fear, and surprise.
-

4. Speech Emotion Analysis

- **FR9:** The system shall convert speech input into text using automatic speech recognition.
 - **FR10:** The system shall analyze transcribed speech using sentiment or emotion analysis techniques to detect emotional states.
-

5. Text Emotion Analysis

- **FR11:** The system shall analyze text data using NLP techniques.
 - **FR12:** The system shall detect emotional sentiment from textual inputs.
-

6. Multimodal Fusion

- **FR13:** The system may combine facial, speech, and text emotion outputs.
 - **FR14:** The system may use adaptive or weighted fusion techniques.
-

7. Dashboard & Visualization

- **FR15:** The system shall provide a web-based dashboard with a clean and professional interface.

- **FR16:** The dashboard shall display detected emotions.
 - **FR17:** The system shall generate charts and graphs representing emotion distribution.
 - **FR18:** The system shall generate emotion-based summaries for instructors.
-

8. Model Training & Performance Evaluation

- **FR19:** The system shall support training and testing of deep learning models.
 - **FR20:** The system shall evaluate performance using accuracy, precision, recall, and F1-score.
 - **FR21:** The system shall compare multimodal and unimodal model performance.
-

9. Documentation

- **FR22:** The system shall provide clear technical and user documentation.
- **FR23:** The system shall document model architecture, datasets, and evaluation results.