**EMOTION DETECTION FROM AUDIO AND VIDEO DATA**

**ABSTRACT**

Emotion detection from video and audio data is an important field of research in artificial intelligence, combining computer vision, speech processing, and machine learning techniques. The ability to accurately recognize emotions has broad applications in human-computer interaction, mental health assessment, education, and customer service. This project focuses on developing a multimodal emotion detection system that integrates both visual and audio data to achieve higher accuracy and reliability.

For the front end, a user-friendly interface is designed using HTML, CSS, and JavaScript. The interface allows users to upload video or audio files for analysis or enable live camera and microphone input for real-time emotion detection. The visual output displays detected emotions with corresponding confidence levels and graphical representations, ensuring a seamless user experience.

For the back end, the system utilizes Python with frameworks such as Flask or Django to handle data processing and communication between the front end and machine learning models. Video data is processed using convolutional neural networks (CNNs) to analyze facial expressions and classify emotions into categories like happiness, sadness, anger, fear, surprise, and neutrality. Audio data is analyzed using Mel Frequency Cepstral Coefficients (MFCCs) for feature extraction, and deep learning models such as long short-term memory (LSTM) networks are employed to detect emotional tones in speech.

The integration of video and audio models is achieved through a fusion technique that combines outputs from both modalities, ensuring improved accuracy and robustness. The system is trained and evaluated on publicly available datasets, including RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) and FER2013 (Facial Expression Recognition).

The results demonstrate that this multimodal approach significantly outperforms unimodal systems relying solely on video or audio inputs. The combination of a responsive front end and a robust back end makes the system suitable for real-time emotion detection applications in healthcare, education, customer engagement, and virtual assistants. Future work will focus on improving real-time performance and integrating contextual analysis to detect subtle and mixed emotions effectively.