

Multimodal Emotion Detection System 1. Project Title **Real-Time Multimodal Emotion Detection Using Facial Expressions, Voice Signals, and Textual Sentiment Analysis**. 2. Overview This project focuses on building an intelligent system capable of detecting human emotions by combining three sources of information: facial expressions, vocal characteristics, and textual sentiment. By integrating these modalities, the system achieves a higher level of accuracy compared to traditional single-modality approaches. 3. Why This Topic Is Important and Innovative Most systems only analyze a single source (face OR voice OR text). Multimodal fusion is still emerging. Emotion understanding plays a crucial role in mental health, user experience, education, and customer support. Combining multiple signals reduces misclassification and increases reliability. Real-time processing opens doors for interactive and intelligent applications. 4. Technical Description The project uses advanced machine learning and signal processing techniques. 4.1 Facial Emotion Recognition Detects facial landmarks and expressions. Uses models such as MediaPipe FaceMesh, ML Kit Face Detection, or CNN-based classifiers. Common emotions detected: happiness, sadness, anger, fear, surprise, disgust, neutral. 4.2 Voice Emotion Recognition Extracts features such as pitch, tone, energy, silence duration, and spectral properties. Uses audio processing tools like librosa. SVM, CNN, or RNN models classify vocal emotional states. 4.3 Text Emotion/Sentiment Analysis Processes user sentences using Natural Language Processing. Uses models such as BERT, DistilBERT, or RoBERTa. Detects tone, sentiment polarity, and emotional intent. 4.4 Multimodal Fusion The system combines outputs from all three channels into a single emotion prediction. Feature-level fusion: concatenation of feature vectors. Decision-level fusion: combining classifier decisions. Advanced fusion: attention-based neural networks. 5. Target Applications **Mental Health:** Emotion monitoring during therapy sessions. **Smart Assistants:** Adaptive responses based on user emotion. **Customer Support:** Stress or anger detection for improved service. **Education:** Detecting confusion or disengagement in students. **Security:** Alert systems for detecting distress. 6. Technologies and Tools **Machine Learning:** TensorFlow, PyTorch, Scikit-learn **Vision:** MediaPipe, ML Kit, OpenCV **Audio:** librosa, PyDub **NLP:** BERT, HuggingFace Transformers **Backend:** Python, Flask, FastAPI **Frontend:** Flutter or React **Deployment:** Firebase, Docker, or local server 7. Expected Deliverables A multimodal dataset or dataset integration. Emotion classification models for vision, audio, and text. A full fusion model. A live dashboard or mobile app for real-time emotion tracking. Performance evaluation and comparison between modalities. A final report and presentation. 8. Conclusion This project represents a cutting-edge approach to human-computer interaction and emotional AI. By merging visual, audio, and textual information, the system provides a deeper and more accurate understanding of human emotions. Its applications span healthcare, education, industry, and smart technologies, making it a strong and modern topic for academic or professional projects.