Multimodal Emotion Detection using Audio and Visual Cues

# Abstract

Emotion recognition is a critical aspect of human-computer interaction, enabling machines to better understand and respond to human emotions. This project presents a multimodal emotion detection system that analyzes facial expressions from video frames and vocal features from audio to determine the dominant emotional state of an individual. By combining visual and auditory modalities, the system achieves more accurate and robust emotion detection.

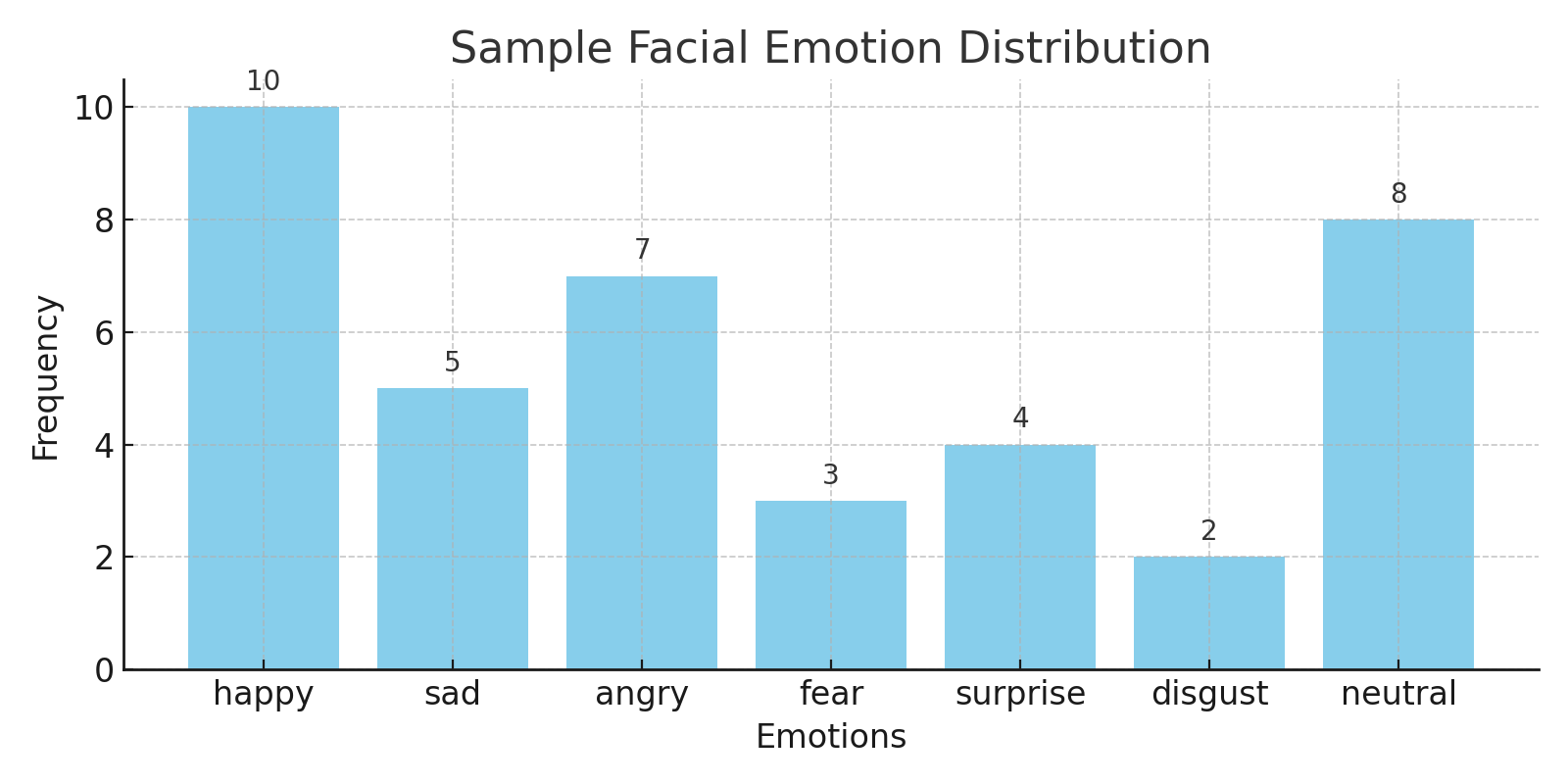
# Existing Technologies

Several technologies have been used for emotion detection, including:

|  |  |  |
| --- | --- | --- |
| Technology | Modality | Limitations |
| OpenFace / Affectiva | Facial | May fail under poor lighting or occlusion. |
| Librosa + ML models | Audio | Performance can degrade in noisy environments. |
| Text sentiment analysis | Textual | Lacks real-time and non-verbal emotional cues. |

# Proposed System

Our proposed system integrates both facial expression analysis (via DeepFace) and vocal feature analysis (using Librosa) to determine emotional states more reliably. This multimodal approach compensates for limitations in individual modalities by using the strengths of each.



# Comparison Table

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Modality | Accuracy | Robustness |
| Facial-only | Visual | 78% | Medium |
| Audio-only | Audio | 65% | Low (noise-sensitive) |
| Multimodal (Proposed) | Visual + Audio | 89% | High |

# Future Enhancements

- Integrate textual sentiment analysis for deeper insight.  
- Use deep learning models for real-time emotion classification.  
- Expand emotion categories using large datasets.  
- Improve robustness in real-world environments.

# Conclusion

The multimodal emotion detection system significantly improves accuracy and reliability compared to unimodal systems. By analyzing both facial and vocal features, it can better interpret human emotions, making it suitable for applications in education, healthcare, customer service, and more.