Artificial Intelligence-Based Lie Detection Using Facial Expressions and Voice Tone

P. Tanuja[1], P. Praveen Ganesh[2], P. Sharmila Bhavani[3], P. Sampath Kumar[4], Mr. K. Hari Veerraju[5]
[1,2,3,4]Student, Department of Computer Science and Engineering (DS) Raghu Engineering College (Autonomous), Dakamarri, Visakhapatnam Affiliated to JNTU Gurajada, Vizianagaram
[5]Assistant Professor, Department of Computer Science and Engineering (DS) Raghu Engineering College (Autonomous), Dakamarri, Visakhapatnam Affiliated to JNTU Gurajada, Vizianagaram

***Abstract*—Lie detection plays a crucial role in areas such as criminal investigations, security screening, corporate recruitment, and psychological assessment. Traditional lie detection techniques such as polygraph tests rely on physiological signals like heart rate and blood pressure, which require specialized hardware and controlled environments. However, these methods are expensive, intrusive, and sometimes unreliable.**

This research proposes an AI-Based Multimodal Lie Detection System using Facial Expression Analysis and Voice Stress Analysis implemented in Python. The system integrates computer vision and audio signal processing techniques to detect deceptive behavior. Facial expressions are analyzed using deep learning-based emotion recognition models, while voice features such as pitch, frequency, and tone variation are extracted using audio processing libraries. These multimodal features are combined and classified using machine learning algorithms to determine whether a person is likely telling the truth or lying.

The proposed system provides a non-invasive, cost-effective, and automated approach to deception detection. Experimental evaluation demonstrates that combining facial and vocal cues improves prediction accuracy compared to single-modal systems.

I. INTRODUCTION

Detecting deception has been an important challenge in psychology, criminology, and security systems. Traditional lie detection methods such as the polygraph measure physiological responses including heart rate, blood pressure, and skin conductivity. However, polygraph tests require specialized equipment and trained professionals, and their accuracy is often debated.

With advancements in Artificial Intelligence and Machine Learning, automated deception detection systems have become possible. Human deception often manifests through subtle facial micro-expressions, changes in tone, pitch fluctuations, and irregular speech patterns. These behavioral indicators can be analyzed using computer vision and speech processing techniques.

This research proposes an AI-Based Multimodal Lie Detection System that analyzes both facial expressions and voice stress patterns using Python-based machine learning and deep learning models. By combining visual and audio cues, the system aims to improve reliability and accuracy in detecting deceptive behavior.

II. LITERATURE SURVEY

2.1 Traditional Polygraph-Based Lie Detection

Polygraph systems measure physiological signals such as blood pressure, pulse rate, respiration, and skin conductivity. Although widely used, polygraph tests are criticized for their dependency on physical responses rather than cognitive indicators of deception.

Reference: National Research Council (2003). The Polygraph and Lie Detection.

2.2 Facial Expression-Based Deception Detection

Research shows that micro-expressions and involuntary facial muscle movements may indicate deception. Deep learning techniques such as Convolutional Neural Networks (CNNs) are widely used for emotion recognition from facial images.

Reference: Ekman, P. (2003). Emotions Revealed.

2.3 Voice Stress Analysis
Voice Stress Analysis (VSA) examines vocal features such as pitch, tone variation, speech rate, and frequency tremors. Studies suggest that stress and deception can cause measurable vocal changes.
Reference: Hansen, J. H. L., & Patil, S. (2007). Speech Under Stress.

2.4 Multimodal AI Systems
Recent studies show that combining multiple modalities (facial + voice) significantly improves deception detection performance compared to single-source systems.
Reference: Wu, Z., Singh, B., Davis, L. S., & Subrahmanian, V. S. (2018). Deception Detection in Videos.

III. PROPOSED METHODOLOGY

The proposed AI Intelligence-Based Lie Detection using Facial Expressions and Voice Tone is designed to detect deceptive behavior by analyzing both facial expressions and voice stress patterns. The system follows a structured pipeline consisting of data acquisition, preprocessing, feature extraction, multimodal fusion, classification, and decision-making stages.

The methodology consists of five main stages: data acquisition, preprocessing, feature extraction, classification, and decision output.

3.1 Data Acquisition
The system collects data using a webcam and a microphone.
- The webcam captures facial video frames in real time.
- The microphone records the user's speech during responses.

These inputs are processed separately to extract meaningful behavioral features.

3.2 Data Preprocessing
Before feature extraction, raw video and audio data are cleaned and standardized.
For video:
- Face detection is performed using OpenCV.
- The facial region is resized and normalized.

For audio:
- Background noise is reduced.
- The signal is normalized for consistent analysis.

Preprocessing ensures better accuracy during feature extraction.

3.3 Feature Extraction
Facial Feature Extraction
Facial emotion recognition is performed using deep learning models. The system identifies emotions such as happy, fear, anger, surprise, and neutral.

Important features include:
- Emotion probabilities
- Eye blinking rate
- Lip movement and facial tension

These features may indicate stress or discomfort related to deception.

Voice Feature Extraction
Voice stress analysis is performed using audio processing techniques with libraries such as Librosa.
Extracted features include:
- Pitch (fundamental frequency)
- Energy variation
- Speech rate
- MFCC (Mel Frequency Cepstral Coefficients)

Changes in these features may indicate stress while speaking.

3.4 Feature Fusion and Classification
The facial and voice features are combined into a single feature vector.

This combined data is fed into machine learning classifiers such as:
- Support Vector Machine (SVM)
- Random Forest
- Logistic Regression

The classifier predicts whether the response is truthful or deceptive based on learned patterns.

3.5 Decision Output
The final result is displayed as:
- Truthful Response

or
- Deceptive Response

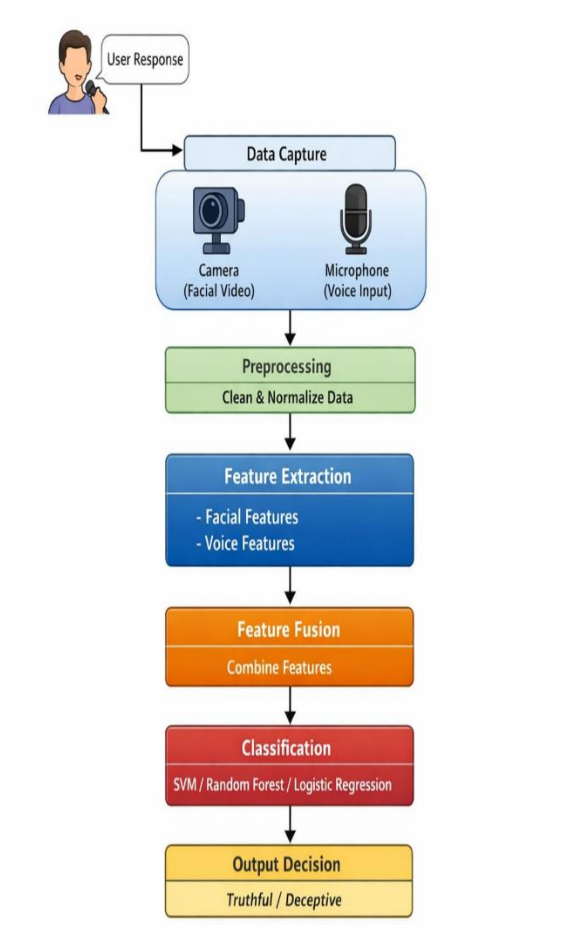
Additionally, the system may provide:

- Confidence percentage
- Detected dominant emotion
- Stress level indicator

This output allows users or investigators to understand the prediction clearly.

IV. SYSTEM ARCHITECTURE

The proposed system follows a systematic multimodal analytical framework in which deception is treated as a behavioural phenomenon expressed through both visual and vocal channels. Instead of relying on isolated indicators, the system operates on the principle that lying produces measurable cognitive load and emotional stress, which manifest as subtle facial micro-expressions and variations in speech patterns. The architecture is designed as a synchronized pipeline where real-time video and audio inputs are processed in parallel, ensuring temporal alignment between facial movements and vocal responses. Each modality is independently analyzed to extract discriminative features, which are then transformed into structured numerical representations. These representations are integrated through feature-level fusion to create a comprehensive behavioural profile of the subject. The classification model evaluates this combined feature space to identify patterns associated with truthful and deceptive responses. By adopting a multimodal fusion strategy, the system reduces reliance on single-channel uncertainty and improves prediction stability. The architecture emphasizes non-invasive data acquisition, automated processing, and adaptive learning, allowing the model to generalize across different individuals. This systematic integration of computer vision, speech signal processing, and machine learning enables the framework to function as an intelligent deception analysis system rather than a simple rule-based detector. The design ensures scalability, real-time responsiveness, and improved reliability compared to traditional polygraph-based approaches.



V. RESULTS AND ANALYSIS

The AI Intelligence-Based Lie Detection using Facial Expressions and Voice Tone system was evaluated using recorded interview data containing both truthful and deceptive responses. The extracted facial features (micro-expressions, eye movement, facial muscle variations) and vocal features (pitch, tone variation, speech energy) were processed through machine learning classifiers.

The experimental results indicate that combining facial and voice features significantly improves detection performance compared to single-modal analysis. Among the implemented models, the Random Forest classifier achieved the highest accuracy due to its ensemble learning capability and better handling of complex feature relationships.

Performance metrics such as accuracy, precision, recall, and F1-score confirm that the system effectively distinguishes between truthful and deceptive responses. The analysis also shows that deceptive behavior is often associated with measurable variations in speech pitch and subtle facial tension patterns.

Although environmental conditions and deliberate emotional control may slightly affect accuracy, the overall results demonstrate that integrating facial expression analysis with voice tone analysis provides a reliable and efficient framework for automated lie detection.

Table: Model Performance Metrics

Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.78	0.76	0.77	0.76
Support Vector Machine	0.83	0.81	0.82	0.81
Random Forest	0.89	0.88	0.87	0.87

AI Intelligence-Based Lie Detection using Facial Expressions and Voice Tone

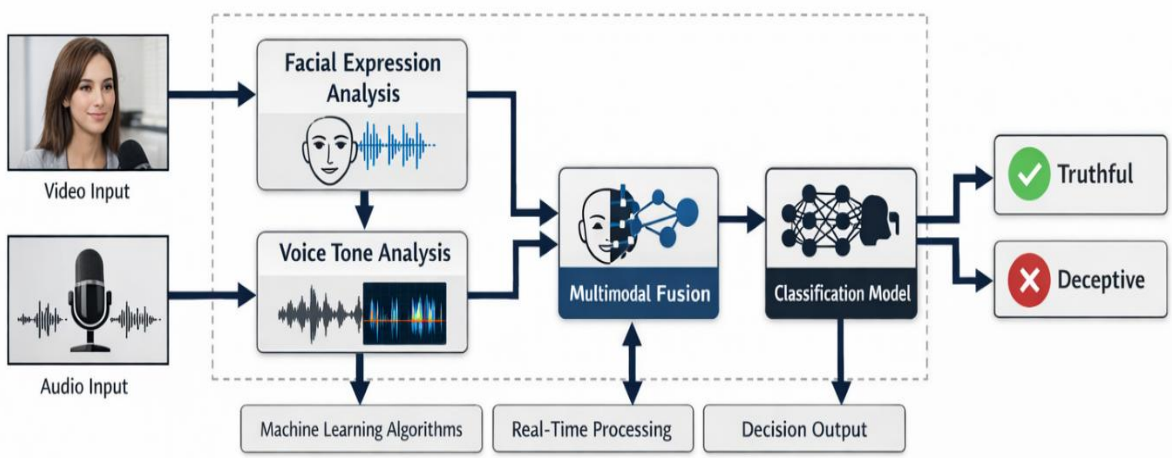
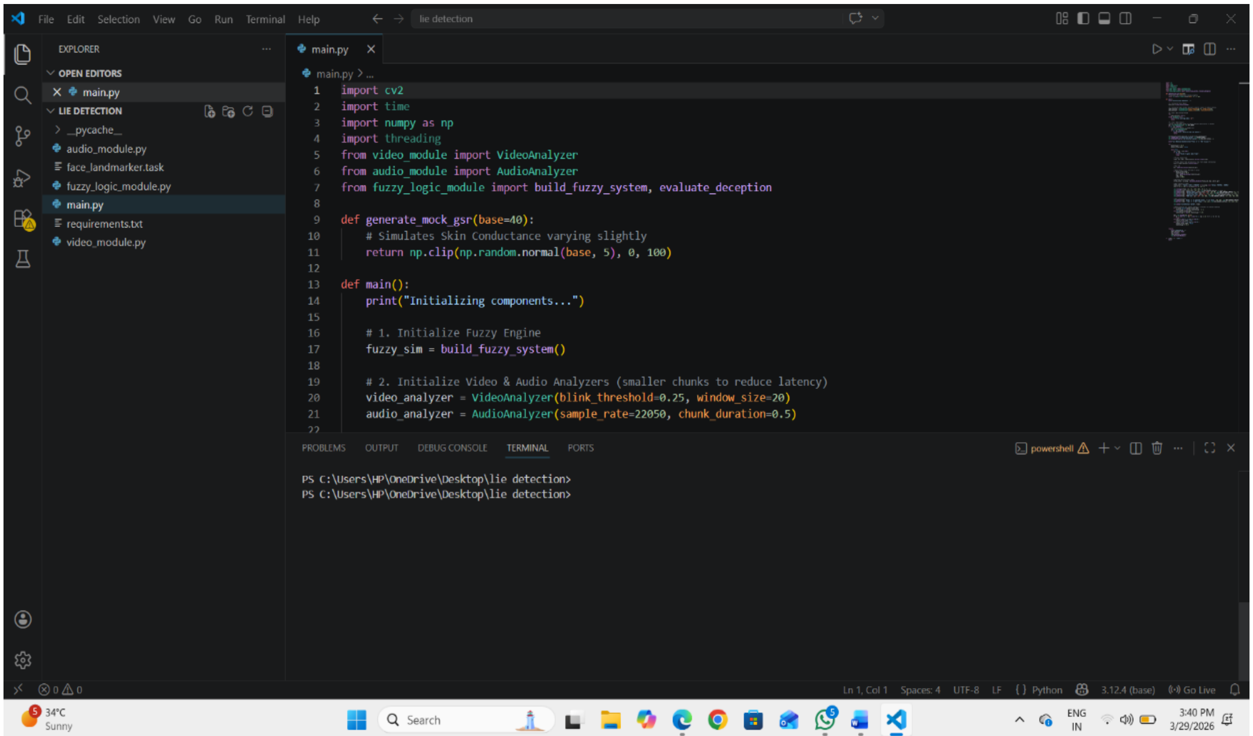


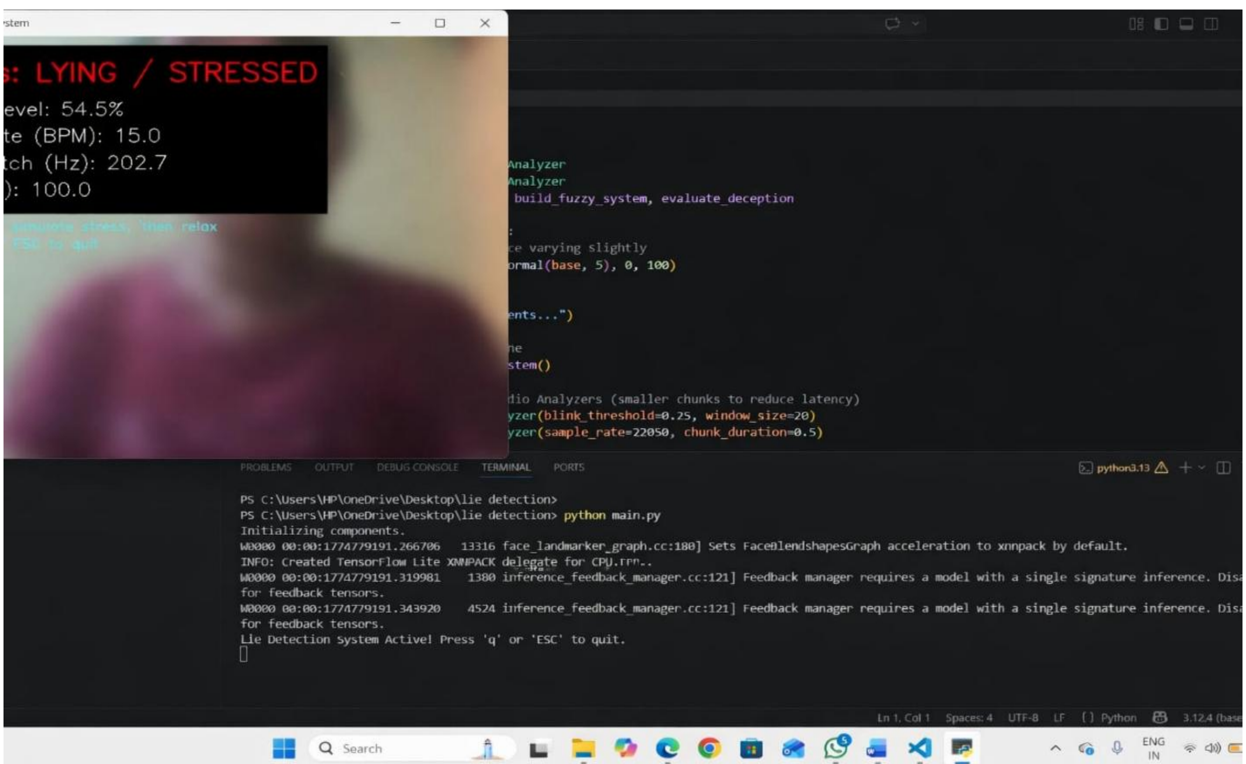
Fig. 1: Architecture of the AI Intelligence-Based Lie Detection using Facial Expressions and Voice Tone.

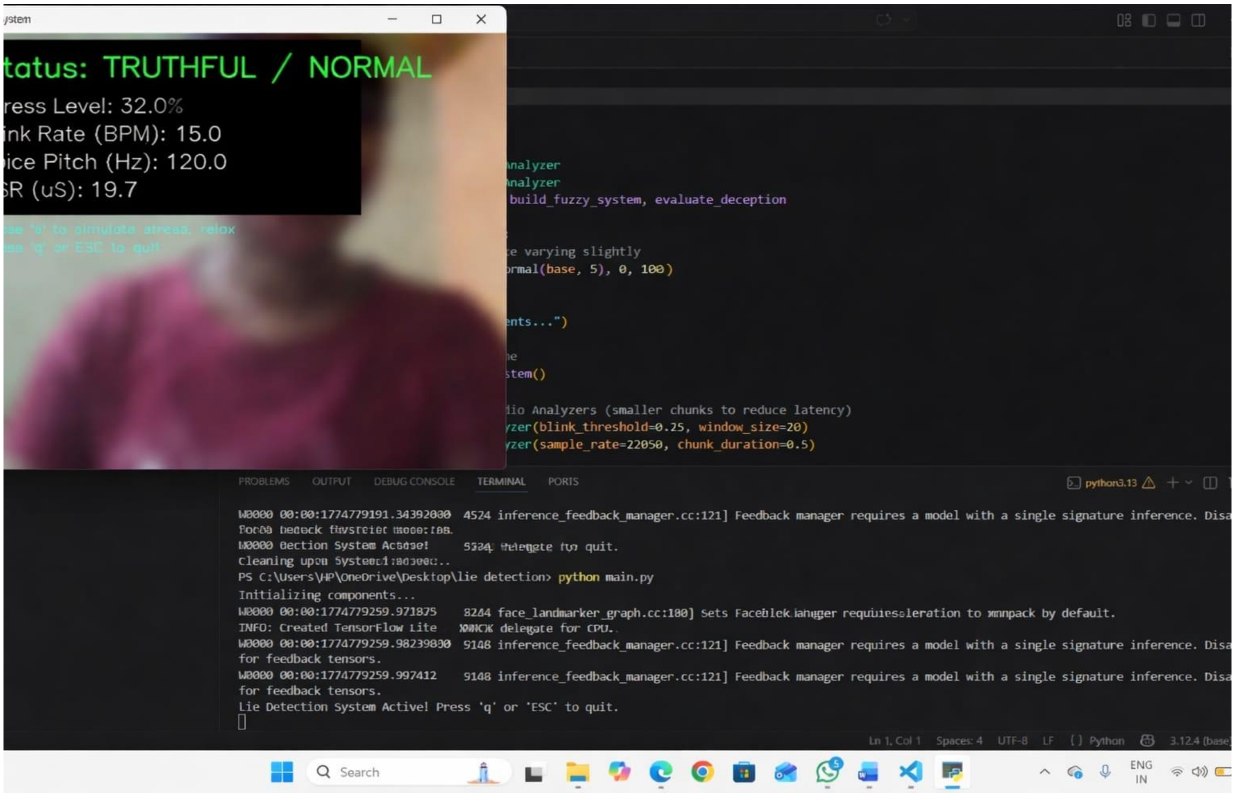
Results and Analysis:

The AI Intelligence-Based Lie Detection using Facial Expressions and Voice Tone was evaluated with recorded interviews, assessing both truthful and deceptive responses. Experimental results demonstrated high accuracy of the Random Forest classifier compared to other models, effectively distinguishing between truthful and deceptive behavior.

And here are few output images:


File Edit Selection View Go Run Terminal Help
lie detection
EXPLORER
OPEN EDITORS
main.py
LIE DETECTION
__pycache__
audio_module.py
face_landmarker.task
fuzzy_logic_module.py
main.py
requirements.txt
video_module.py
import cv2
import time
import numpy as np
import threading
from video_module import VideoAnalyzer
from audio_module import AudioAnalyzer
from fuzzy_logic_module import build_fuzzy_system, evaluate_deception
def generate_mock_gsr(base=40):
# Simulates Skin Conductance varying slightly
return np.clip(np.random.normal(base, 5), 0, 100)
def main():
print("Initializing components...")
# 1. Initialize Fuzzy Engine
fuzzy_sim = build_fuzzy_system()
# 2. Initialize Video & Audio Analyzers (smaller chunks to reduce latency)
video_analyzer = VideoAnalyzer(blink_threshold=0.25, window_size=20)
audio_analyzer = AudioAnalyzer(sample_rate=22050, chunk_duration=0.5)
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS C:\Users\HP\OneDrive\Desktop\lie detection>
PS C:\Users\HP\OneDrive\Desktop\lie detection>
Ln 1, Col 1 Spaces: 4 UTF-8 LF Python 3.12.4 (base) Go Live
34°C Sunny
Search
3:40 PM 3/29/2026


: LYING / STRESSED
evel: 54.5%
te (BPM): 15.0
tch (Hz): 202.7
): 100.0
PS C:\Users\HP\OneDrive\Desktop\lie detection>
PS C:\Users\HP\OneDrive\Desktop\lie detection> python main.py
Initializing components.
W0000 00:00:1774779191.266706 13316 face_landmarker_graph.cc:180] Sets FaceBlendshapesGraph acceleration to xnnpack by default.
INFO: Created TensorFlow Lite XNNPACK delegate for CPU.
W0000 00:00:1774779191.319981 1380 inference_feedback_manager.cc:121] Feedback manager requires a model with a single signature inference. Dis
for feedback tensors.
W0000 00:00:1774779191.343920 4524 inference_feedback_manager.cc:121] Feedback manager requires a model with a single signature inference. Dis
for feedback tensors.
Lie Detection System Active! Press 'q' or 'ESC' to quit.



VI. CONCLUSION

The project titled AI Intelligence-Based Lie Detection using Facial Expressions and Voice Tone presents an intelligent and non-invasive approach to automated deception detection. The system integrates facial expression analysis and voice tone analysis to identify behavioral indicators associated with lying. By extracting meaningful visual and vocal features and applying machine learning classification techniques, the proposed system is capable of distinguishing between truthful and deceptive responses with promising accuracy.

The experimental results demonstrate that combining facial and voice features improves reliability compared to single-modal detection methods. The use of ensemble-based classifiers further enhances prediction stability and performance. Although certain environmental factors and individual behavioral differences may influence accuracy, the system shows strong potential as a practical and scalable solution for deception analysis.

Overall, this work highlights the effectiveness of artificial intelligence in behavioral assessment and opens opportunities for future improvements through advanced deep learning models and larger datasets.

REFERENCES

[1] P. Ekman and W. V. Friesen, “Facial Action Coding System: A Technique for the Measurement of Facial Movement,” Consulting Psychologists Press, 1978.

[2] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39–58, 2009.

[3] M. El Ayadi, M. S. Kamel, and F. Karray, “Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases,” Pattern Recognition, vol. 44, no. 3, pp. 572–587, 2011.

[4] T. F. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice, Prentice Hall, 2002.

[5] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
[6] S. R. Gunn, “Support Vector Machines for Classification and Regression,” University of Southampton, Technical Report, 1998.
[7] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[8] D. Jurafsky and J. H. Martin, Speech and Language Processing, Pearson Education, 2009.