

International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

Mental Health Identification from Face, Text, and Voice

M. Selvi

Lecturer, Department of Computer Science and Engineering, I R T Polytechnic College, Bargur, Tamil Nadu, India

selvima1975@gmail.com

ABSTRACT: Mental health challenges such as stress, anxiety, and depression are increasing worldwide, yet many individuals hesitate to seek professional help due to stigma and accessibility issues. To address this, the proposed project introduces an AI companion for mental health identification from face, text, and voice. The system leverages multimodal data to detect emotional states and provide supportive interaction. For facial expression analysis, OpenCV is used for image processing and feature extraction, while classification is carried out using Convolutional Neural Networks (CNN) and traditional machine learning models. Text inputs, such as user messages, are analyzed using Support Vector Machine (SVM) and Random Forest algorithms to classify linguistic and sentiment patterns associated with mental well-being. For audio analysis, Mel-Frequency Cepstral Coefficients (MFCC) are extracted to capture speech features, enabling the identification of stress and emotional tone from voice signals. By combining these three modalities, the system provides a comprehensive and accurate assessment of a user's psychological state. Beyond detection, the AI companion can offer real-time coping strategies, mindfulness reminders, and motivational feedback, acting as a supplemental support tool rather than a replacement for professional therapy. With continuous monitoring, scalability, and adaptability, the system aims to promote early intervention, reduce barriers to care, and foster emotional resilience through a personalized and accessible AI-based solution.

I. INTRODUCTION

1.1 About the Project

Mental health is an increasingly critical concern in today's world, with individuals across all age groups facing stress, anxiety, depression, and other psychological challenges. Factors such as academic pressure, work-related stress, lifestyle changes, social isolation, and emotional struggles contribute significantly to declining mental well-being. While awareness about mental health has grown, barriers such as social stigma, lack of access to professional care, and limited resources prevent many from seeking timely help. In this context, Artificial Intelligence (AI) offers promising solutions for early identification and continuous monitoring of mental health conditions.

This project, titled "Mental Health Identification from Face, Text, and Voice," aims to design and implement an AI-based companion system that can analyze multimodal inputs—facial expressions, text data, and audio signals—to detect signs of stress, depression, or other mental health issues. The system not only identifies emotional states but also provides a safe, supportive platform for users to engage with, thereby promoting emotional resilience and well-being. To achieve this, the project integrates computer vision, natural language processing, and speech analysis techniques into a unified framework. Facial expressions are analyzed using OpenCV and deep learning models, which help in detecting non-verbal emotional cues such as sadness, anger, stress, or happiness. Text inputs, such as user responses or diary-style entries, are processed using Support Vector Machine (SVM) and Random Forest classifiers, enabling the identification of sentiment, thought patterns, and stress indicators. For voice analysis, Mel-Frequency Cepstral Coefficients (MFCC) are extracted from audio signals, which capture subtle variations in tone, pitch, and rhythm that

International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

often reflect psychological states. These multimodal data streams are fused to provide an accurate and holistic assessment of a user's mental health condition.

Unlike conventional systems that rely solely on questionnaires or single-mode detection, this multimodal AI companion provides 24/7 monitoring and support, reduces dependency on manual interpretation, and addresses accessibility challenges by offering a non-invasive, technology-driven approach. Furthermore, the system promotes early intervention by recognizing early warning signs before conditions escalate, thereby complementing professional mental health services.

The uniqueness of this project lies in its fusion of AI-driven identification and empathetic companionship. In addition to detection, the system can offer evidence-based coping mechanisms, including mindfulness reminders, stress management tips, and encouragement based on user mood analysis. This ensures that the solution not only identifies issues but also actively supports individuals in navigating their challenges, reducing the stigma of seeking professional therapy.

With continuous learning capabilities, the proposed system can adapt to user behavior over time, becoming more accurate and personalized. Thus, this project contributes toward the global effort of integrating AI into healthcare by offering a scalable, affordable, and accessible tool for mental health monitoring and support.

The proposed system integrates facial emotion detection using MiniXceptionNet with quiz-based analysis powered by Support Vector Machines (SVM) to provide a personalized and dynamic approach to mental health and resilience. By utilizing MiniXceptionNet, a lightweight deep learning model, the system can accurately detect and interpret facial expressions in real-time, identifying emotions such as anxiety, sadness, or happiness. This emotion recognition is then combined with a personalized quiz, where users are presented with questions based on their emotional responses and behavioral patterns. The answers are processed by an SVM classifier to classify the user's mental state, allowing the system to track emotional trends over time and provide tailored feedback on improving mental well-being.

In addition to real-time emotion detection, the proposed system adapts to users' emotional states and offers interventions such as cognitive-behavioral techniques, stress management exercises, or mindfulness practices based on the analysis. The AI companion engages users in a continuous feedback loop, suggesting strategies to build emotional resilience and providing support whenever needed. By integrating both face emotion detection and quiz-based psychological analysis, the system ensures a comprehensive, multi-dimensional approach to mental health that goes beyond static responses. This combination enables users to receive more accurate, real-time insights into their emotional well-being while empowering them with the tools to improve mental resilience and cope with stress in a personalized manner.

1.2 Purpose and Scope

The primary objective of this project is to design and develop an AI-based system that can identify mental health conditions from multimodal inputs (face, text, and voice) and provide supportive interactions to users. To achieve this goal, the project focuses on the following specific objectives:

1. To analyze facial expressions using computer vision techniques
 - Utilize OpenCV and deep learning algorithms to detect and classify emotional states (e.g., happiness, sadness, stress, anxiety) from facial images or live video.
 - Build a robust face recognition and emotion detection pipeline capable of handling variations in lighting, facial orientation, and expression intensity.
2. To process and classify textual input for mental health analysis
 - Implement Support Vector Machine (SVM) and Random Forest algorithms for sentiment analysis and text classification.

International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

- Identify linguistic features and patterns that correlate with mental health indicators such as depressive thought patterns, stress markers, or anxiety-related vocabulary.
- 3. To analyze voice input using audio processing techniques
 - Extract Mel-Frequency Cepstral Coefficients (MFCC) and other relevant acoustic features from audio samples.
 - Classify audio signals to detect emotional tone, stress levels, and vocal indicators of mental health decline.
- 4. To integrate multimodal data (face, text, and voice) for improved accuracy
 - Develop a fusion model that combines insights from all three modalities to provide a more comprehensive and reliable assessment of the user's mental health status.
 - Compare the effectiveness of single-modality vs. multimodal approaches in achieving higher detection accuracy.
- 5. To design an AI companion that provides real-time support and feedback
 - Create an empathetic, user-friendly virtual assistant capable of delivering coping strategies, motivational messages, and personalized guidance based on the detected mental state.
 - Ensure continuous availability of support, reducing barriers such as stigma and accessibility.
- 6. To promote early intervention and mental health awareness
 - Enable the system to recognize early warning signs of mental health deterioration.
 - Encourage proactive measures before conditions escalate, supporting users in seeking professional help when necessary.
- 7. To ensure scalability and adaptability of the system
 - Design the AI companion to adapt to different user demographics and evolving user behavior.
 - Enable future integration with healthcare platforms, mobile applications, or wearable devices for broader adoption.

II. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

Existing systems for mental health support rely mainly on traditional methods such as in-person therapy, telehealth platforms, and mobile apps like Headspace, Calm, and Moodpath. These platforms provide features such as mood tracking, mindfulness exercises, and CBT-based activities, but they rely heavily on self-reported data, where users manually input their emotions or complete questionnaires.

Some platforms have incorporated emotion detection technology, such as voice tone analysis or facial recognition (e.g., Realeyes, Affectiva). However, these are mostly used in marketing and customer service, with limited adoption in mental health applications.

Other approaches involve quiz-based psychological analysis using diagnostic tools such as PHQ-9 or GAD-7, combined with basic machine learning algorithms. While useful, these methods tend to be static, non-interactive, and not adaptive to real-time emotional changes.

DISADVANTAGES

- Dependence on self-reported data → prone to bias and inaccuracy.
- Lack of real-time interaction → emotions are not monitored continuously.
- Limited personalization → generic responses instead of dynamic, adaptive feedback.
- Narrow scope of AI integration → facial/voice emotion recognition often not combined with text/quiz-based analysis.
- Application gaps → most emotion detection tools are used in business contexts, not specifically for mental health care.

International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

2.2 PROPOSED SYSTEM

The proposed system integrates facial emotion detection using MiniXceptionNet with quiz-based analysis powered by Support Vector Machines (SVM), supported by a user-friendly AI companion interface.

- MiniXceptionNet is used for lightweight, real-time facial emotion detection, capturing emotions such as happiness, sadness, stress, and anxiety from live camera input.
- A quiz-based module dynamically generates personalized questions (based on tools like PHQ-9 and GAD-7), and responses are classified using SVM for mental health assessment.
- Results from both modules are fused to provide personalized feedback, coping strategies, CBT exercises, and mindfulness interventions.
- The system continuously adapts to the user's changing emotional states, offering 24/7 support and reducing barriers like stigma and lack of access to therapy.

ADVANTAGES OF PROPOSED SYSTEM

- Multimodal input (face + quiz/text) → combines objective signals with self-assessment for higher accuracy.
- Real-time emotion detection → continuously monitors user's mood through facial expressions.
- Personalized quizzes and feedback → adapts interventions to individual emotional states.
- Lightweight architecture → MiniXceptionNet ensures efficient real-time processing, even on mobile devices.
- Dynamic adaptability → system updates recommendations based on user's emotional trends over time.
- 24/7 accessibility → provides immediate support, especially for users without access to regular therapy.
- Stigma reduction → offers a private, non-judgmental platform for users hesitant to seek professional help.
- Scalability → can be deployed across web, desktop, and mobile platforms with data security protocols.

2.3 FEASIBILITY STUDY

A feasibility study is carried out to evaluate the practicality, viability, and potential impact of the proposed system before implementation. The study is categorized into social feasibility, technical feasibility, and economic feasibility, ensuring that the system is beneficial to society, technologically achievable, and cost-effective.

3 key considerations involved in the feasibility analysis are

- ◆ ECONOMICAL FEASIBILITY
- ◆ TECHNICAL FEASIBILITY
- ◆ SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

The project is designed to be cost-effective and affordable:

- Development Cost: The majority of tools and frameworks (Python, OpenCV, Scikit-learn, Librosa) are open-source, eliminating licensing expenses.
- Hardware Requirements: A standard computer with a webcam and microphone is sufficient for implementation. No specialized hardware is required.
- Maintenance Cost: Minimal, as updates mainly involve software improvements or dataset expansions.
- Operational Cost: Low, since the system can run locally or be deployed on cloud platforms with affordable hosting plans.
- Benefit-Cost Ratio: The system provides significant benefits, such as mental health support, productivity improvement, and early intervention, compared to its relatively low development and deployment cost.

Thus, the system is economically feasible, making it practical for deployment in institutions such as colleges, workplaces, and even individual use.

International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

TECHNICAL FEASIBILITY

The system leverages proven technologies and frameworks to ensure effective functioning:

- Facial Expression Recognition: Implemented using OpenCV for preprocessing and SVM classifiers for emotion detection.
- Text-Based Sentiment Analysis: Performed using Random Forest algorithms to classify text input into mental health states.
- Audio Emotion Detection: Uses MFCC (Mel-Frequency Cepstral Coefficients) for feature extraction and SVM for classification.
- Decision Fusion: Combines results from face, text, and audio for a more reliable outcome.
- Hardware/Software Requirements: The model can run efficiently on mid-level systems with Python libraries (NumPy, Scikit-learn, OpenCV, Librosa) and does not require expensive high-end infrastructure.
- Scalability: The framework is designed to integrate additional modalities like physiological signals in the future.

Hence, the system is technically feasible, as all required tools and algorithms are readily available, lightweight, and compatible with current hardware/software environments.

SOCIAL FEASIBILITY

The project addresses one of the most pressing social issues of the modern era—mental health challenges such as stress, anxiety, and depression. Despite the growing awareness, stigma and lack of accessibility still prevent many individuals from seeking professional help. The proposed AI companion offers:

- Accessibility: Provides 24/7 support, making mental health assistance available even in remote or underserved areas.
- Confidentiality: Ensures that users can express their emotions privately without fear of judgment or social stigma.
- Awareness and Early Intervention: Encourages individuals to recognize mental health symptoms early, preventing escalation.
- Impact on Society: By assisting students, working professionals, and vulnerable groups, the system helps promote productivity, emotional stability, and overall well-being.

Thus, the system is socially acceptable and likely to gain positive acceptance among individuals who are hesitant to approach traditional therapy at the initial stage.

III. SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE

The system architecture for emotion detection typically starts with data acquisition, where inputs such as text, speech, facial expressions, or physiological signals (e.g., heart rate) are collected from users. This raw data is then passed through a preprocessing stage to remove noise, normalize formats, and prepare it for analysis. Feature extraction follows, where relevant emotional cues—such as tone of voice, facial landmarks, or sentiment-bearing words—are derived. These features are then fed into a classification model (e.g., CNN, RNN, or transformer-based models), which is trained to recognize and categorize emotions like happiness, anger, sadness, or surprise. The prediction results are passed to a visualization or response layer, which may trigger adaptive feedback, such as chatbot responses or alerts in mental health applications. A feedback loop is integrated to improve the model over time using user interactions or corrections, ensuring the system evolves with context and accuracy. Figure 3.1 Represents the System Architecture.

International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

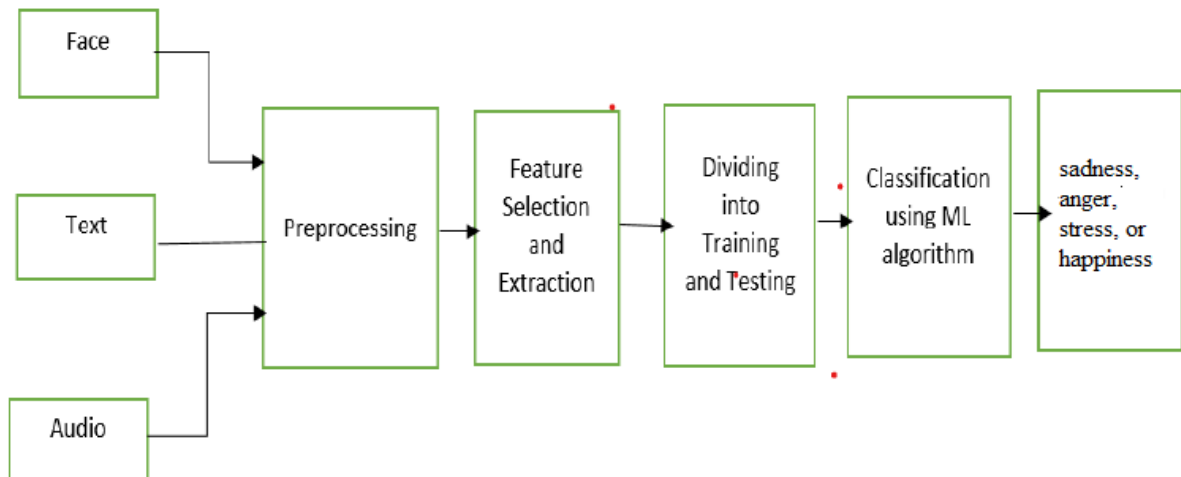
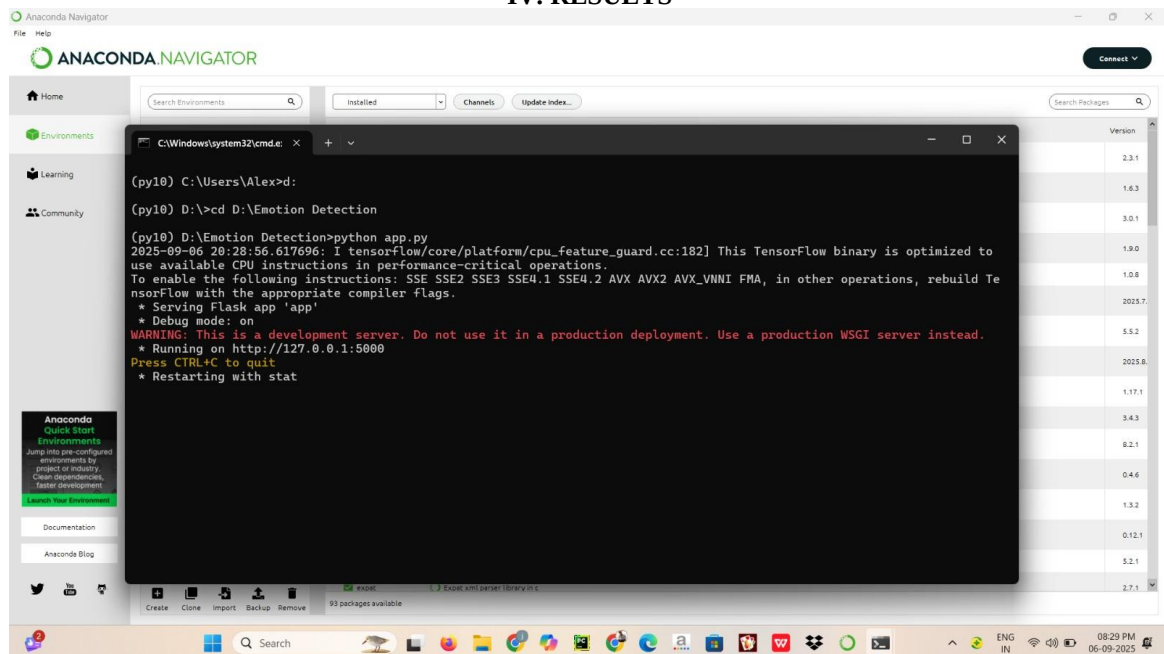


Figure:3.1 System Architecture

IV. RESULTS

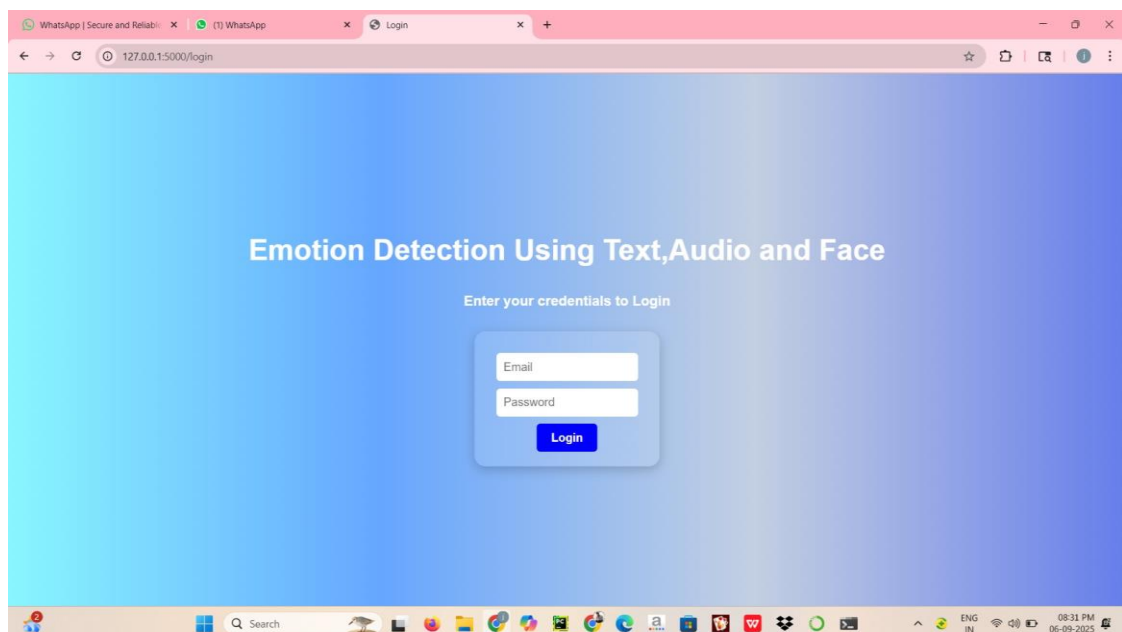
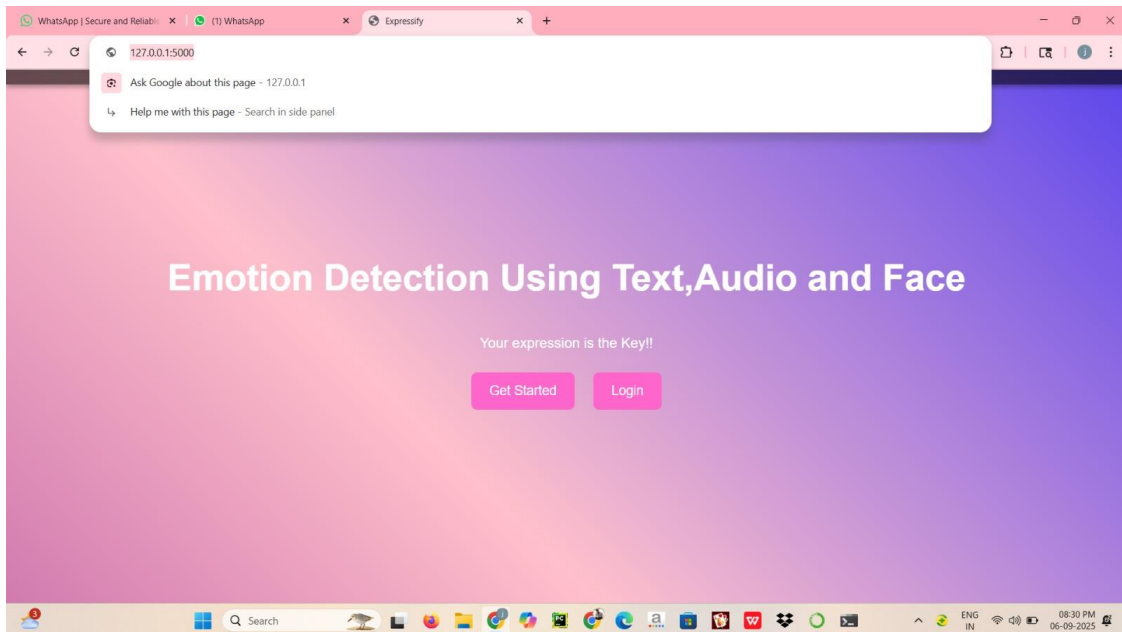


International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

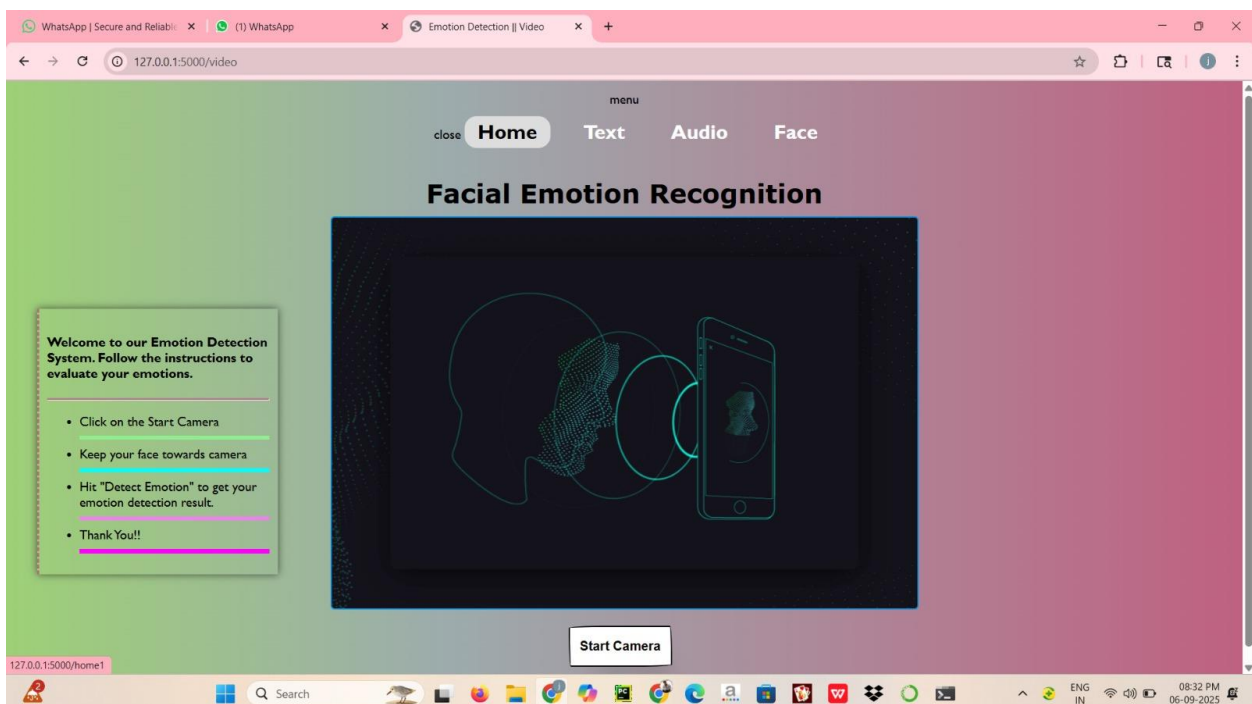
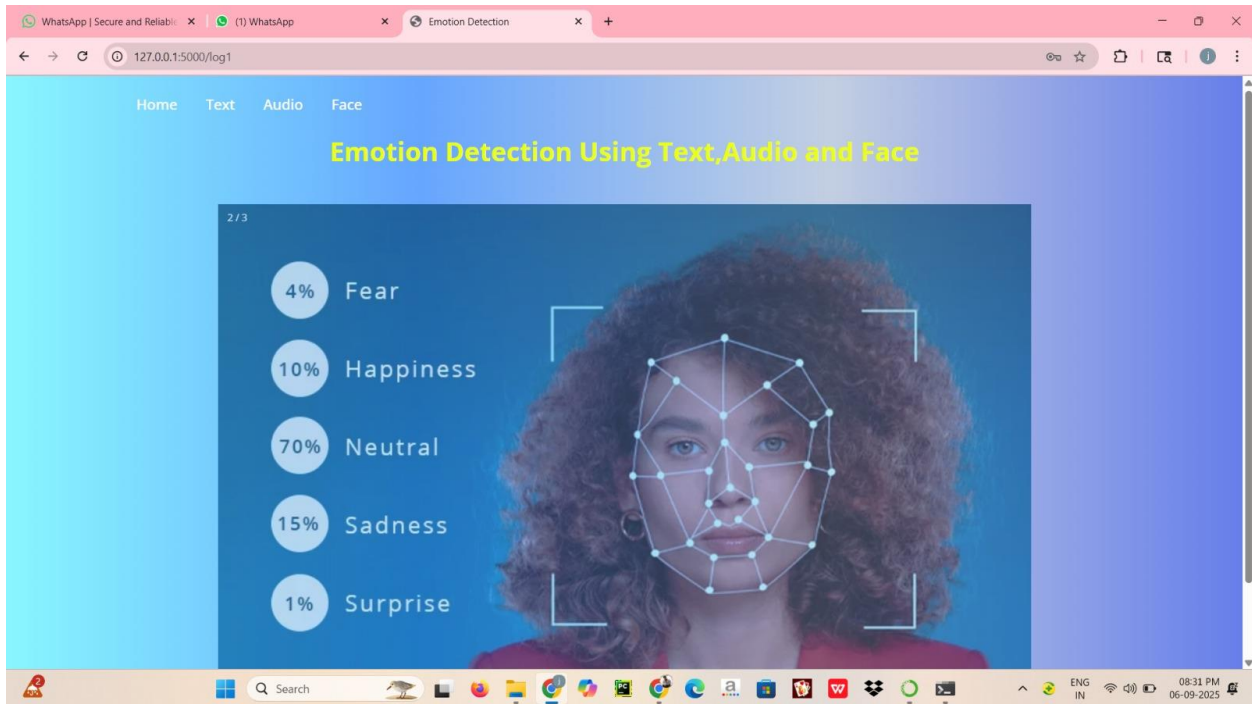


International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

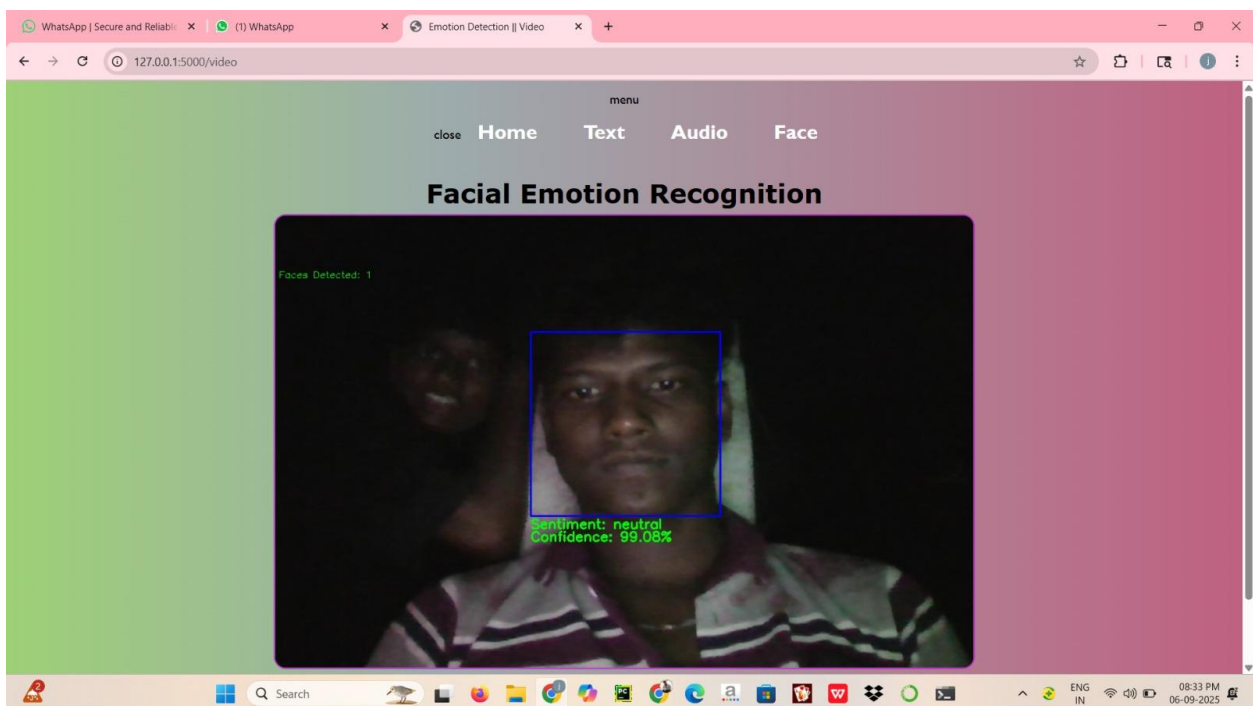
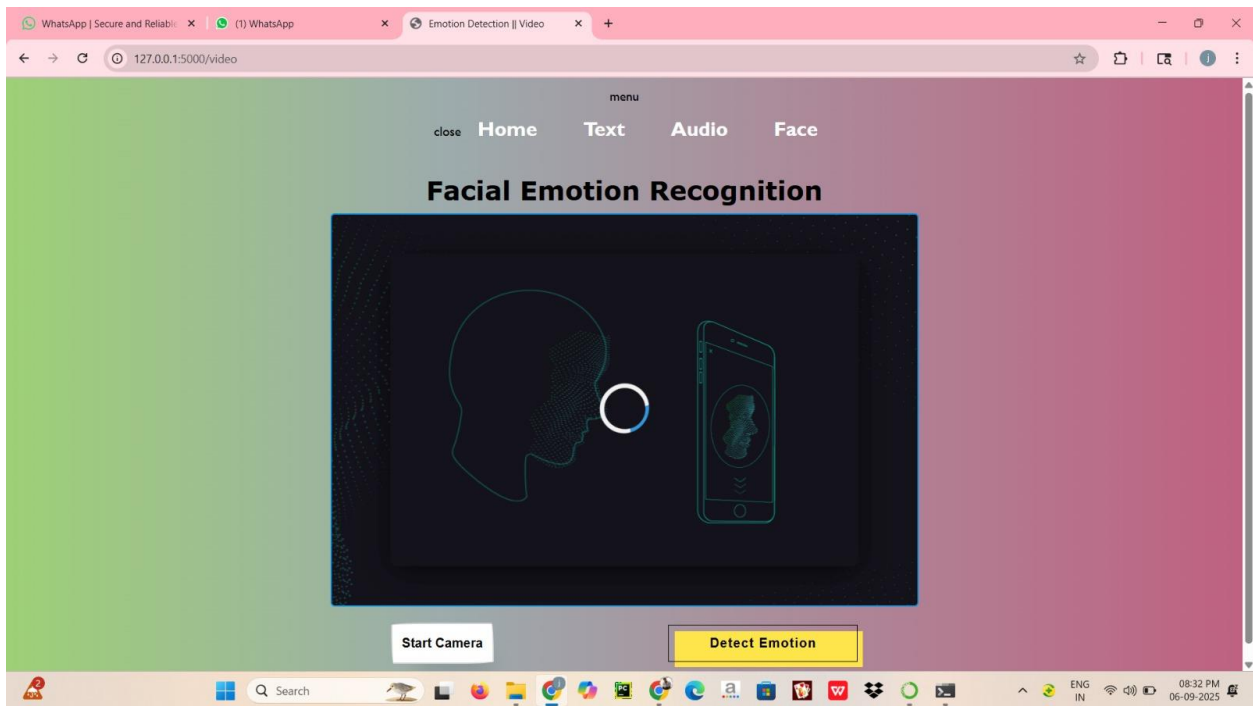


International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

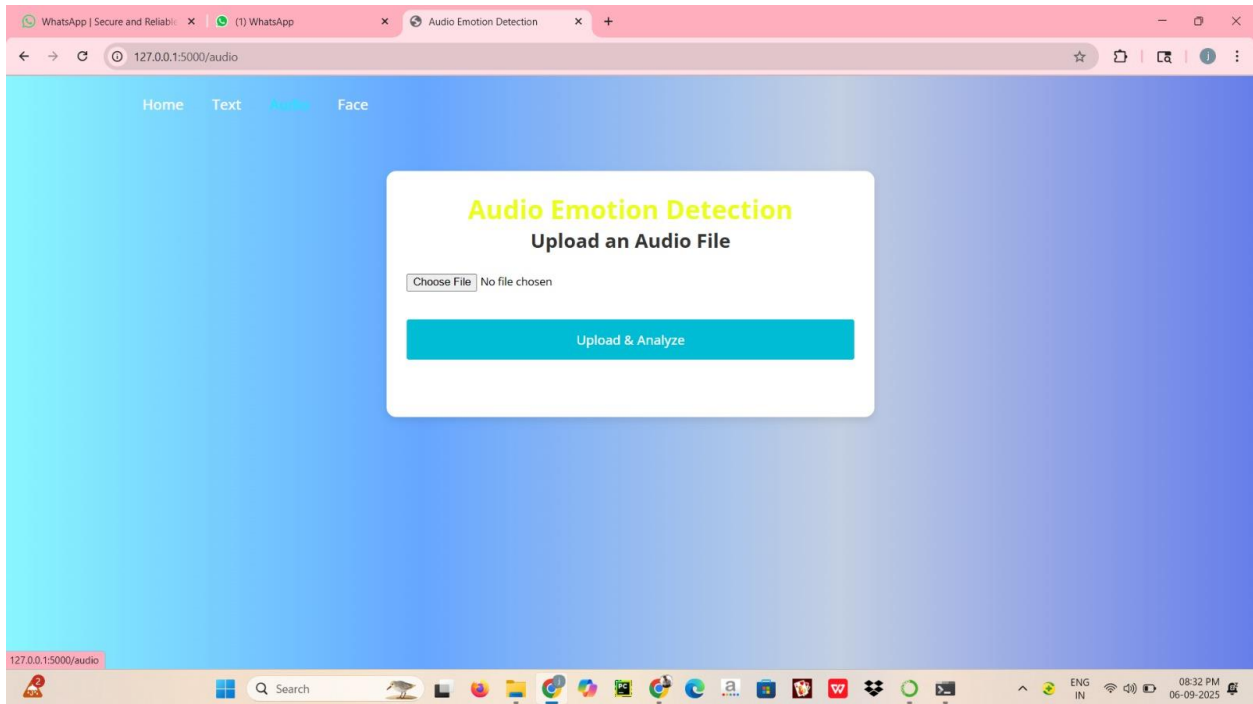


International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017



V. CONCLUSION

The proposed system, Mental Health Identification from Face, Text, and Voice, demonstrates the potential of artificial intelligence in addressing one of society's critical challenges—mental health detection and support. By leveraging multimodal inputs such as facial expressions (OpenCV + SVM), text sentiment analysis (Random Forest), and audio emotion recognition (MFCC + SVM), the system provides a more comprehensive and accurate evaluation of an individual's psychological state compared to single-modality approaches.

The AI companion offers real-time support, early intervention, and privacy, making it a socially acceptable and accessible solution, especially for individuals reluctant to seek professional help. Furthermore, the system's modular architecture and reliance on open-source tools make it technically and economically feasible for large-scale deployment.

Overall, this project contributes toward reducing the barriers of stigma and accessibility in mental health care, acting as a supplementary tool for emotional well-being and resilience, while complementing traditional therapy and counseling practices.

VI. FUTURE ENHANCEMENT

While the system achieves promising results, several improvements can be made in future work:

- Deep Learning Integration – Use advanced models such as CNNs, LSTMs, or Transformers for improved accuracy in face, text, and voice analysis.
- Multilingual Support – Extend text and speech recognition to handle multiple languages and dialects for broader accessibility.
- Mobile/IoT Deployment – Develop a lightweight mobile application or IoT-based mental health monitoring system for continuous tracking.

International Journal of Innovative Research in Science, Engineering and Technology

(An UGC Approved, High Impact Factor, Monthly, Peer Reviewed Journal)

UGC Approval No: 48027

Vol. 6, Issue 2, February 2017

- Integration with Wearables – Incorporate physiological signals such as heart rate, EEG, or galvanic skin response for more reliable detection.
- Personalized Recommendations – Implement reinforcement learning to adapt coping strategies and exercises based on user progress.
- Cloud and Chatbot Integration – Deploy the system on cloud platforms with chatbot interfaces for scalability and real-time interaction.
- Dataset Expansion – Use larger and more diverse datasets to improve system robustness across age groups, cultures, and environments.

REFERENCES

1. S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," PLoS ONE, vol. 13, no. 5, pp. 1–35, May 2018.
2. P. Ekman and W. V. Friesen, "Facial Action Coding System (FACS): A technique for the measurement of facial action," Consulting Psychologists Press, Palo Alto, CA, 1978.
3. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
4. S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," Information Fusion, vol. 37, pp. 98–125, Sept. 2017.
5. A. Dhalla, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using facial expression and speech," in Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), 2012, pp. 806–810.
6. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, May 2015.
7. J. Han, E. Friesen, and C. Darwin, "Multimodal approaches to emotion recognition: A survey," IEEE Trans. Affective Computing, vol. 11, no. 1, pp. 17–37, Jan.–Mar. 2020.
8. 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, Mar. 2011.
9. C. C. Aggarwal, Machine Learning for Text, Springer, 2018.
10. H. Abadi et al., "Deep affective neural networks for multimodal emotion recognition," in Proc. ACM Int. Conf. Multimodal Interaction (ICMI), 2015, pp. 483–490.