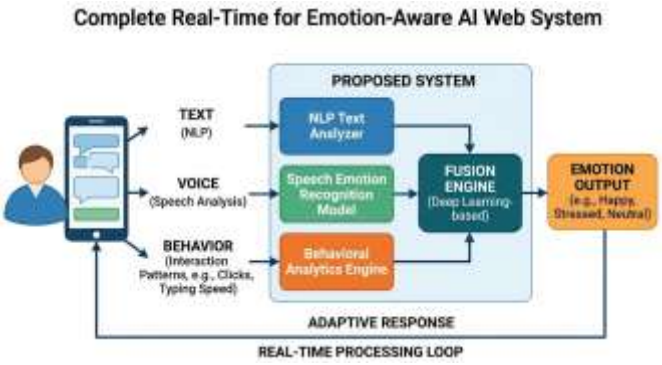
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Real-Time Emotion-Aware AI Web Systems Using Multimodal Inputs (Text, Voice, and Behavioral Signals)

Abhishek Thakur Barhi[1] , Asst. Prof. Mohmmad Asif[2]

[1] *B.Tech Computer Science and Engineering , Parul institute of Technology, Vadodara, Gujarat India*
[2] *B.Tech Computer Science and Engineering , Parul institute of Technology, Vadodara, Gujarat India*

Abstract -Emotion recognition is a critical component of next-generation intelligent systems, enabling more natural and adaptive human–computer interaction. This paper presents the design and development of a Real-Time Emotion-Aware AI Web System that integrates multimodal inputs including text, voice, and user behavior. The proposed system leverages Natural Language Processing (NLP), speech emotion recognition, and behavioral analytics to detect emotional states in real time. A fusion-based deep learning approach is used to combine features from different modalities for improved accuracy. Experimental results demonstrate that multimodal systems outperform unimodal approaches in emotion detection accuracy and responsiveness, making them suitable for applications such as mental health monitoring, customer support, and adaptive user interfaces.



Keywords

Emotion Recognition, Multimodal AI, NLP, Speech Analysis, Behavioral Analytics, Human-Computer Interaction

I. Introduction

Human emotions play a crucial role in communication, decision-making, and interaction. Traditional web systems lack emotional intelligence, resulting in limited user engagement. Recent advancements in Artificial Intelligence have enabled systems to detect and respond to human emotions.

Multimodal emotion recognition, which combines text, voice, and behavioral signals, has emerged as a powerful approach for improving accuracy and robustness. Unlike single-modal systems, multimodal systems capture diverse emotional cues, leading to better performance.

This research proposes a real-time web-based system that integrates multiple input modalities to detect user emotions dynamically and adapt system responses accordingly.

II. Literature Review

Recent studies highlight the importance of multimodal emotion recognition:

- Emotion recognition systems perform better when combining audio, text, and visual/behavioral data instead of relying on a single modality.
- Deep learning models such as CNN, RNN, and transformers are widely used for extracting emotional features.
- Speech-based emotion recognition relies on acoustic features like pitch, tone, and energy patterns.
- Multimodal systems improve accuracy by fusing complementary information from different sources.

However, most existing systems focus on offline processing and lack real-time web integration with behavioral signals.

III. Problem Statement

Existing systems:

- Use single-modal emotion detection
- Lack real-time processing
- Ignore behavioral signals (clicks, typing speed, interaction patterns)
- Provide limited adaptability in web environments

This research addresses these limitations by developing a real-time multimodal emotion-aware web system.

IV. Objectives

- To design a multimodal emotion detection system
- To integrate text, voice, and behavioral inputs
- To implement real-time emotion recognition
- To evaluate system accuracy and performance

V. Proposed System Architecture

A. System Overview

The system consists of:

1. **Frontend Layer**
   - Web interface for user interaction
   - Captures text, voice, and behavioral inputs
2. **Backend Layer**
   - API server (Flask/Django)
   - Handles processing and communication
3. **AI Processing Module**
   - NLP model for text sentiment analysis
   - Speech emotion recognition model
   - Behavioral analysis module
4. **Fusion Engine**
   - Combines outputs from all modalities
   - Uses deep learning-based fusion techniques
5. **Database**
   - Stores user interaction and emotion data

B. Architecture Flow

User Input → Data Capture → AI Models → Fusion Engine → Emotion Output → Adaptive Response

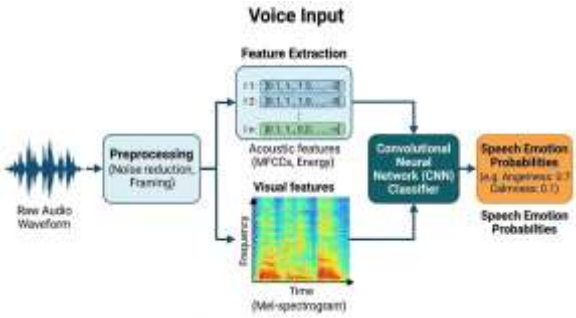
VI. Methodology

A. Data Collection

- Text input from chat/messages
- Voice input from microphone
- Behavioral data (typing speed, mouse movement, session time)

B. Feature Extraction

- Text: Sentiment scores using NLP models
- Voice: Acoustic features (pitch, MFCC, tone)
- Behavior: Interaction metrics

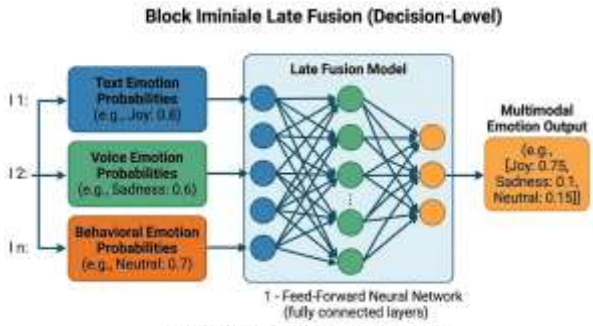


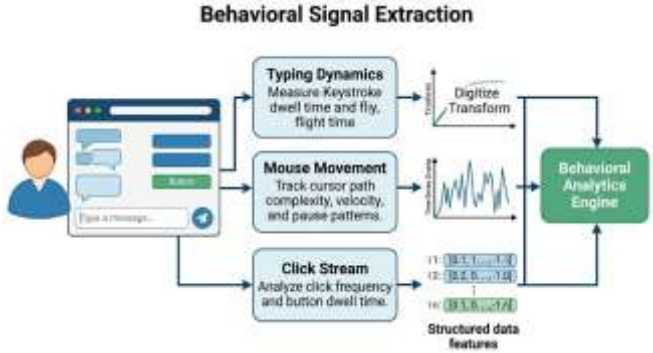
C. Model Implementation

- NLP models (BERT/GPT-based)
- Speech models (CNN-RNN)
- Behavioral ML models

D. Multimodal Fusion

- Early fusion (feature-level)
- Late fusion (decision-level)
- Hybrid fusion for improved accuracy





VII. Implementation

A. AI Modules

- Text sentiment analyzer
- Speech emotion recognition system
- Behavioral tracking system

B. Web Integration

- Real-time processing using APIs
- WebSocket-based communication for live updates

C. Tools & Technologies

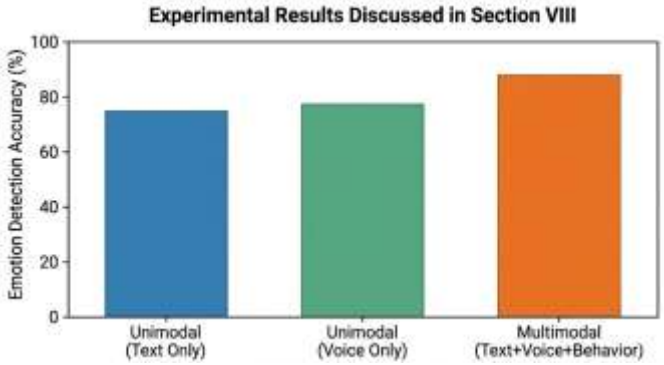
- Python
- Flask/Django
- React.js
- TensorFlow/PyTorch
- OpenAI / Gemini API

VIII. Results and Discussion

A. Performance Comparison

Model Type	Accuracy
Text Only	75%
Voice Only	78%
Multimodal	88%

Multimodal systems outperform unimodal systems due to the integration of complementary emotional cues.



B. Observations

- Behavioral data significantly improves accuracy
- Real-time processing enhances user engagement
- Fusion models provide robust predictions

IX. Advantages

- High accuracy through multimodal fusion
- Real-time emotion detection
- Improved user experience
- Adaptive system behavior

X. Limitations

- High computational cost
- Privacy concerns (voice & behavior tracking)
- Need for large datasets

XI. Future Scope

- Integration with wearable devices
- Advanced emotion prediction models
- Personalized AI assistants
- Ethical AI frameworks for privacy

XII. Conclusion

This paper presents a Real-Time Emotion-Aware AI Web System that integrates text, voice, and behavioral inputs for accurate emotion detection. The system demonstrates improved performance compared to traditional approaches and highlights the importance of multimodal AI in future web technologies. Emotion-aware systems will play a key role in enhancing human-computer interaction and intelligent automation.

References

[1] P. Praveenkumar et al., "Multimodal Emotion Recognition Using Deep Learning," *ICAISDA*, 2026.

[2] C. Wu et al., "Multimodal Emotion Recognition in Conversations," *EMNLP*, 2025.

[3] S. Chamishka et al., "Voice-Based Emotion Detection Using RNN," *Multimedia Tools*, 2022.

[4] S. Byun et al., "Multi-Modal Emotion Recognition Using Speech and Text," *Applied Sciences*, 2021.

[5] "Deep Learning-Based Multimodal Emotion Recognition Review," *Expert Systems with Applications*, 2023.