IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153129

Multi Model Emotion Aware Conventional Chatbot Using Facial Expression and Text Sentiment Fusion

Mrs. K. Tejaswi[1], Ch. Pushpa Manasa[2], D. Hima Sravanthi[3], B. Pravallika[4], G. Girishma[5]

Assistant Professor, Department of Information Technology, KKR & KSR Institute of Technology and Sciences Guntur, India[1]

Student, Department of Information Technology, KKR & KSR Institute of Technology and Sciences Guntur, India[2-5]

Abstract: Understanding or analyzing the human emotion is difficult for smart, adaptive human-computer interaction systems. This project proposed a multi-modal emotion-aware conversational chatbot that recognizes emotions from the users by the help of facial expression, text-based sentiment analysis from speech-to-text conversion and also written text. Facial emotions are extracted using computer vision techniques. Such as eye movement, mouth shape, and eyebrow position, based on all these it captures the emotion state. while speech is converted to text, and also by direct text using natural language processing (NLP) methods. The emotional information obtained from both modalities is combined using a weighted fusion mechanism to clarify the user's overall emotional state. Based on this, the chatbot generates emotionally appropriate and relevant context- aware responses. The proposed system adapts human-centered interactions and shows its utility for mental health support. Using OpenCV facial expression recognition, VGG16 convolutional neural networks, automatic speech recognition (speech-to-text), VADER sentiment analysis, and weighted multimodal fusion. Experimental results show that the performance of emotion recognition information response relevance compared to single- modality systems, leading to the that multimodal emotion fusion greatly improves the effectiveness and empathy of conversational AI systems.

Index Terms: Multimodal Emotion Analysis, Emotion-Aware Chatbot, Facial Expression Recognition, Sentiment Analysis, OpenCV, VGG16, Natural Language Processing, VADER

I. INTRODUCTION

Even the AI get more developed in the recent years, also included with human and computer interaction or chatbot etc. There is more lack of understanding according to chatbot it also becomes some critical in hard situations. However, humans also communicate through their faces and tone of voice, with such communication difficult to capture through text and making emotion recognition less accurate. The project investigates a multi-modal emotion-aware conversational chatbot, using textual analysis and facial analysis, to help overcome this limit.

Expression recognition, spoken input is converted to text and then processed using natural language processing, before analyzing the corresponding facial expressions using computer vision as camera. The system could thus be improved by integrating visual and textual emotional signals to detect user simultaneous emotion.

With such a structure, the chatbot is able to, both produce answers to the user and respond to the user's context and also emotional state. Therefore, conversations with the chatbot are more relaxed and natural as human, and can be used in various fields such as education and training, mental health, and smart virtual agents. Overall, this shows the importance of multi-modal emotion recognition to the development of more human-like conversational agents.

II. LITERATURE REVIEW

Although emotion recognition has been the main focus of most of the existing studies (e.g., facial emotion recognition, speech emotion recognition, and text-based sentiment analysis), multimodal emotion recognition systems generally perform unimodal methods. However, most multimodal emotion recognition research has been mainly focused on the recognition task and not been expanded to include emotion- aware conversational responses similarly [1]. Chatbot emotion detection and real-time facial analysis have been explored for mental health monitoring, but are

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026
DOI: 10.17148/IJARCCE.2026.153129

limited in works related to selective disorders, and lack integration of speech-to-text sentiment analysis and expression analysis. Papers including surveys of OpenCV and sentiment analysis have identified importance in mental health detection, but they do not present real-time approaches or conversational frameworks in real up to now [3].

The Other works employed deep learning techniques to detect facial expressions from video feeds with high accuracy and in real time. However, these works only rely on visual information and do not consider speech-based or text-based sentiment, rendering them less adaptable in conversational agents [4], [8]. Emotion aware and real time facial expression recognition agents have been developed to respond according to facial expressions, but they do not work on speech-based sentiment analysis, nor provide context-aware and empathetic responses [5]. While multimodal emotion recognition systems have been used in limited application domains such as business negotiations, these models are not easily adapted to open- domain conversation chatbots [6]. Likewise, while multimodal speech and facial emotion detection models achieve high detection accuracy, they do not done for dialog context and emotion-based response generation [7]. However, existing systems are purely voice-based and also do not include facial expression analysis and multimodal fusion, making them week [9].

III. METHODOLOGY

The formation and creation of a multimodal emotion-aware conversational chatbot is achieved by the use of systematic, modular strategy that by combines facial expression recognition, speech-to-text conversion, and text evaluation. The comprehensive approach is designed to achieve accurate emotion detection and meaningful chatbot responses by combining multiple sources of gained results of emotional information

A. *Obtaining User Input*

The process gets starts by collecting input from the user through three phases as facial expressions, speech, and text. A camera is used to capture real time facial images, while a microphone records spoken input and convert it into text. Users can also directly type text messages. This multi way of input approaches allows the system to capture and gather the information from the user simultaneously.

B. *Processing Facial Expression*

Captured facial images are first pre-processed to improve recognition accuracy. This includes face detection, resizing, and normalization of detected face to remove noise and background variations to get clarity. The processed facial data is then passed to a trained facial expression recognition model, which analyses facial features such as eye movement, mouth shape, and eyebrow position. Based on all these features, at last the model classifies the user's facial expression into predefined emotional categories.

C. *Speech-to-Text Conversion*

When the user provides the voice input, the recorded audio signal is processed by the speech recognition module. Then the audio gets cleaned to reduce background noise and then converted into text format using automatic speech recognition techniques. This step enables spoken communication to be made the same way as typed input for further emotional analysis.

D. *Analyzing Text*

The text obtained either from the direct user input or from speech-to-text conversion is analyzed using natural language processing methods. The system examines the word usage, sentence structure, and contextual meaning to determine the emotion expressed by the user this is used to analyze. The sentiment is classified into emotional states positive, negative, or neutral, along with associated emotional information where applicable this helps to identify the emotion.

E. *Emotion Integration Approach*

To improve or to develop the reliability of the system, the generated results from the facial expression and the text analysis that are combined using an emotion fusion mechanism that compares the results of the two modalities and adjusts their contribution. If the results of one modality are unclear, a decision is made using the other modality. This helps pre- vent misinterpretation of stimuli and inappropriate emotional response.

F. *Emotion-Based Response Generation*

Once the final emotion gets identified, the chatbot generates a response based on the analyzed emotion. The response logic ensures that replies that are empathetic, supportive, and contextually related. This makes the interaction more natural and emotionally intelligent compared to traditional chatbots.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153129

G. ***Output Delivery***

The generated response is finally displayed to the user in the form of text and can also be delivered as speech. The system then waits for the next user input, allowing continuous and simultaneous interaction.

IV. RESULT

The system presents a multimodal, emotion-aware conversational chatbot that can understand and respond to user emotions by combining facial expressions with text-based sentiment from speech. Facial expressions are captured using computer vision, while spoken input is converted to text and analyzed using natural language processing techniques. By mixing information from both sources, using a weighted fusion method to understand how the user's feels and allows the chatbot to respond appropriately to the situation and show understanding of emotions. Using techniques like OpenCV facial recognition, VGG16 neural networks, speech-to-text, VADER sentiment analysis, and multimodal fusion, can improve sys- tem better understands emotions and responds appropriately, helping in mental health support, customer service and more effective for various applications.

Overall, the results show that combining emotional inputs improve chatbot performance, makes the chatbot smarter, and more reliable for real-world tasks

V. DISCUSSION

Using different methods to understand the user emotions can improve the performance of chatbot. By combining facial expressions, voice input, and text, the system can understand user emotions clearly, even when messages are unclear. Facial expressions show different emotions clearly, while voice input makes interaction easier and more natural. Text based emotion detection works for basic feelings, but not suites well for sarcasm.

When these emotions are combined together, the chatbot can give better and more understanding replies than older systems. Even though there are few limitations, combining different emotion inputs improves human-computer interaction much better.

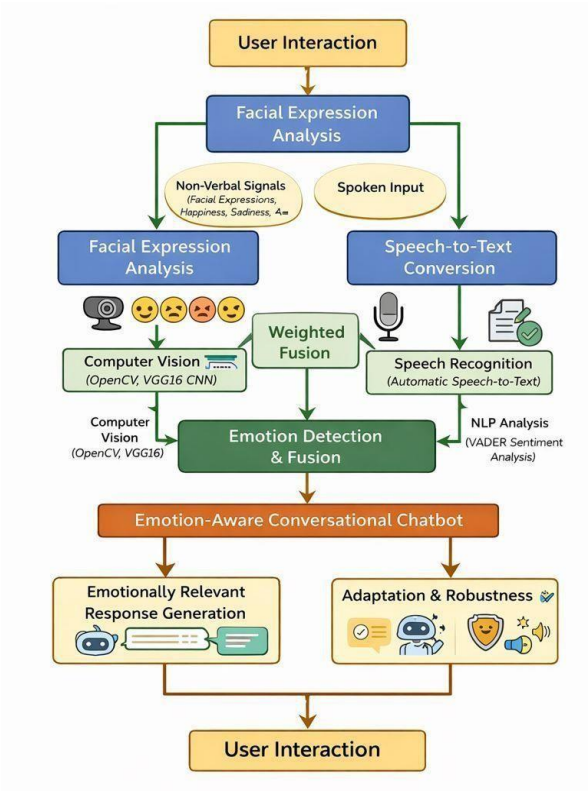

Fig. 1. System architecture of the multimodal emotion-aware conversational chatbot

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153129

Using speech-to-text made communication feel easier, user- friendly and more natural. Users found speaking easier instead of typing that makes the system easier and comfortable to use. The text was converted well, but background noise sometimes creates small errors. Even though there were some issues, voice input helps to improve user experience better. Using both facial emotions and text feelings can give better results than using them separately. If one method failed, the other one can be used to decide the emotion. This helped the chatbot to respond more accurate and efficiently.

Compared to older chatbots, this one is more useful and responds very quickly. Even there are few limitations, combining multiple emotion inputs improves how humans interact with computer.

VI. CONCLUSION

A multimodal emotion-aware chatbot was developed in such a way, that easily identifies user emotions using facial expressions, speech-to-text, and text sentiment analysis. By combining all these inputs will helps us to understand emo- tions and provides friendly and understanding responses. Only voice interaction can improve user interactions, but combining different inputs gives better results than a single input. The results show that using multiple emotion inputs can make chatbots more interactive, and able to understand the situation with possible inputs such as voice tone, body movements and live emotion detection.

REFERENCES

[1] S. S. Malik, M. Ilyas, Y. ul Haq, R. Sana, M. S. Razzaq, F. Maqbool, and M. S. Pathan, "Multi-modal emotion detection and sentiment analysis," *IEEE Access*, vol. 13, pp. 59790–59806, 2025, Doi: 10.1109/ACCESS.2025.3552475.

[2] M. Swapna, S. S. Vadana, S. I. Keerthi, and S. N. Kodavath, "Depression detection using Chabot and live video facial analysis," *International Journal for the Multidisciplinary Research*, vol. 7, no. 3, pp. 1–10, May– June 2025.

[3] A. K. Krishna et al., "A survey on mental health state detection using OpenCV and sentimental analysis," *Journal Emerging Technologies and Innovative Research (JETIR)*, vol. 12, no. 7, July 2025.

[4] K. Sukeerth and V. Gowthami, "Face Recognition and Emotion De- tection in the Video Streams," *International Journal of Sciences and Innovation Engineering*, vol. 2, no. 9, pp. 807–811, Sept. 2025, Doi: 10.70849/ijsci.

[5] P. V. Kurhe, P. K. Take, M. S. Palande, S. M. Barode, and P. S. Pachorkar, "Emotion-Aware Conversational Agent with Real-Time Facial Recogni- tion," *International Research Journal of Modernization in Engineering, Technology and Science (IRJMETS)*, vol. 7, no. 11, pp. 5780–5786, Nov. 2025.

[6] L. Zhao, "Design of a Multimodal Emotion Analysis System Based on Speech and Text in Business Negotiations," in *Proceedings of the 2025 IEEE 3rd International Conference on Image Processing and Computer Applications (ICIPCA)*, 2025, pp. 1677–1681.

[7] V. V. Avabratha, S. Rana, S. Narayan, S. Y. Raju, and S. Sahana, "Speech and facial emotion recognition using convolutional neural network and random forest: A multimodal analysis," in *Proc. 2024 Asia Pacific Conf. on Innovation in Technology (APCIT)*, IEEE, 2024.

[8] K. Park, S. Johnson, and M. Taylor, "Face Detection and Emotion Recognition in Real-Time Video Streams Using Optimized Neural Networks," *IEEE Transactions on Multimedia*, vol. 26, pp. 9123–9137, Dec. 2024, Doi: 10.1109/TMM.2024.3571414.

[9] C. V. Chethan, K. S. Greeshma, and K. Y. C. Kiran, "Emotion detection via voice and speech recognition," *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, vol. 4, no. 1, pp. 635–642, Feb. 2024, Doi: 10.48175/IJARSCT-15385.

[10] H. S. Harini and R. Bhuvana, "Multimodal approaches for depression detection: A comprehensive review," *International Journal of Research Publication and Reviews*, vol. 5, no. 1, pp. 5644–5650, Jan. 2024.

[11] T. Guo, W. Zhao, M. Alrashoud, A. Tolba, S. Firmin, and F. Xia, "Multi- modal Educational Data Fusion for Students' Mental Health Detection," *IEEE Access*, vol. 10, pp. 70370–70382, 2022, Doi: 10.1109/ACCESS.2022.3187502.

[12] P. Tiwari and A. D. Darji, "Therapy bot: A multimodal stress/emotion recognition and alleviation system," *International Journal of Computer Applications*, vol. 183, no. 33, pp. 1–8, Oct. 2021.

[13] "Facial Expression Rendering in Medical Training Simulators: Current Status and Future Directions," *IEEE Access*, 2020.

[14] R. W. Picard, "Affective computing," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 11–20, Jan.–Jun. 2010.

