STM JOURNALS

Journal of
Mechatronics and Automation

ISSN: 2455-1988
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoMA

https://journals.stmjournals.com/joma

Research JoMA

# AI Based Mock Interview Evaluator

Chethana C. Viraktamath[1,*], K. Shashank[1], Kshama S.D.[1], Madan N.R.[1], Sunil Kumar H.R.[2]

***Abstract***

*AI-based mock interview platform that acts as a bridge between real interviews and preparation. The system evaluates users based on emotions, confidence, and knowledge. Emotions are analyzed using a deep learning CNN algorithm to classify facial expressions. Confidence is also a main feature in interview which is assessed through speech recognition using natural language processing. Knowledge assessment involves taking online tests and answering along with the description. This platform aims to reduce stress and anxiety before actual interviews while enhancing candidate confidence and performance. This innovative project represents a significant advancement in interview preparation, AI technology, and personal development, with the potential to positively impact individuals' career success.*

**Keywords:** Emotion and confidence analysis, artificial intelligence, machine learning, CNN, Interview preparation, knowledge assessment.

## INTRODUCTION

The interview process stands as a crucial gateway to career opportunities across the globe. Every year, millions of individuals participate in interviews for jobs, internships, and academic placements. Studies suggest that a significant portion – estimates range from 40% to 70% – of these interviews result in rejection. This highlights a critical need for improved interview preparation methods. The Mock interview evaluator addresses the critical issues of student interview preparation and job recruitment process by minimizing the risks associated with manual practice and low confidence during interviews. According to the latest data, 98% of people fail interviews. The average job posting receives approximately 250 applications, and most employers typically interview no more than 10 candidates, or 2-3%. Of course, there are many reasons why talks fail. While showing confidence in a job interview is just as important as demonstrating certain skills, confidence is only one part of a successful interview.

***Author for Correspondence**
Chethana C. Viraktamath
E-mail: chethanamath@gmail.com

[1]Student, Department of Computer Science and Engineering, PES Institute of Technology and Management, Shivamogga, Karnataka, India
[2]Assistant Professor, Department of Computer Science and Engineering, PES Institute of Technology and Management, Shivamogga, Karnataka, India

Received Date: May 06, 2024
Accepted Date: May 20, 2024
Published Date: August 23, 2024

**Citation:** Chethana C. Viraktamath, K. Shashank, Kshama S.D., Madan N.R., Sunil Kumar H.R. AI Based Mock Interview Evaluator. Journal of Mechatronics and Automation. 2024; 11(2): 1–5p.

In the broader societal context, interviews are essential for the functioning of businesses, government agencies, educational institutions, and non-profit organizations, as they help in identifying individuals who can contribute to the growth and success of these entities.

Understanding the reasons behind interview failure sheds light on the problem. Common culprits include poor communication skills, lack of confidence, inadequate knowledge of the role or industry, and difficulty managing interview anxiety. These factors can significantly impact a candidate's ability to showcase their true potential. Additionally, the interview format itself presents challenges. Traditional in person interviews can be

 

subjective, with success hinging on factors beyond a candidate's control, such as interviewer bias or a poorly designed interview structure.

The project aims to use the power of artificial intelligence (AI) to solve these problems. We launched AI Mock Interview Evaluator, a new platform designed to bridge the gap between interview preparation and real-life experience. Providing a structured interactive environment and evaluating job candidates based on insight, confidence and knowledge, the platform aims to provide individuals with the tools and insights they need to interview with confidence and success.

## LITERATURE SURVEY

In today's incredibly competitive job market, securing that dream role often comes down to excelling in the crucial interview stage. However, traditional methods of interview preparation may not provide all the necessary tools for success. While practicing with friends or colleagues can be helpful, it may lack objectivity. On the other hand, professional interview coaching can come at a high cost and consume valuable time. Studies have looked at the importance of both verbal and nonverbal cues through three different methods as audio, video, and a questionnaire in determining the capacity of a marketing and business analyst job positions [1]. For detecting, processing and assessing the personality and behavioral changes od the candidate on non-verbal cues [2] is prepared based on machine-learning. Eye movements, head movements, eye and emotions are considered to achieve the above objective.

Emotion recognition is a very complex and a difficult job to achieve with very high accuracy. Number of researches [3] done to analyze the emotions of the candidate deeply for detecting the emotions. Still there does not exist a perfect feature set that characterize the emotion correctly. This is because of the factors like different speakers, sentences, speaking styles, speaking rates and different languages.

In the research [4], it is planned to discuss the monitoring situation and the selection of the deep learning method. First, a corpus of interviews was collected from 12 participants and recorded along with their spoken words and actions. Next, short-term memory and neural networks are used to predict the speech state, and deep learning is used to study the relationship between speech and action. Finally, options are used to create questions for the interview. A search method [5, 6] provides different types of inferences to obtain visual, audio, and textual information from a video call. The multimodal features include head pose, gaze direction, action units, several acoustic features, and text cues. Both text and audio cues yielded satisfactory results in predicting hiring recommendation scores. Another study collected prosodic, lexical, and facial features (smile intensity and head pose) of each interview. These features were then used to predict behavioral traits and generate recommendation scores.

Speech and speaker recognition is an important and widely researched area in digital signal processing. Feature extraction using MFCC is the most widely used method as it remains robust and effective, but not without limitations. Practically it lacked due to the cepstral features' sensitivity towards background and channel noise. Studies shown in make it clear how noise affects the efficiency of the process. Many techniques were presented to refine the process, and these techniques can be broadly classified into three types: filter-based compensation, noise model based compensation and empirical compensation [7].

Automatic emotional recognition (AER) [8] can be done in two ways: through speech or face. In human-computer interaction, voice, facial expressions, and gestures are the main targets of emotion recognition. Words are considered a powerful way to express thoughts and feelings. In recent years, many studies have been conducted on using speech to understand people's emotions. Facial expression extraction is an important part of FER. Facial changes caused by different facial expressions are typically extracted using appearance-based methods or geometry-based methods. Appearance-based features describe facial textures resulting from expression, such as wrinkles in

 

appearance-based FER, facial features are extracted by applying image filters, such as the Gabor wavelets filter, local binary patterns (LBP) filter, and histogram of oriented gradient (HOG) filter to the whole face or to specific regions [9, 10].

## SYSTEM DESIGN

The system architecture illustrates a process designed to analyze live video inputs for facial emotion detection, confidence evaluation and to assess knowledge through tests (Figure 1). Here's a stepwise description of the system architecture:

1. *Live video input:* The foundation of the system, it receives a live video feed which is then directed into two separate analysis streams:
   i. *Video cue:* This stream processes visual data to extract features using the Histogram of Oriented Gradients (HOG) method, known for object detection tasks, particularly facial detection.
   ii. *Audio cue:* In parallel, audio data from the video is analyzed using Mel-Frequency Cepstral Coefficients (MFCC), a feature widely used for voice recognition and sound analysis.
2. *Facial emotion detection:* The features extracted from the video cue are used in this module to detect and interpret the emotional state exhibited by the faces in the video. The emotion recognition is performed on the individuals captured in the video input.
3. *Confidence detection:* Complementary to the Facial Emotion Detection, this module interprets cues from the audio (potentially tone or pitch of the voice) to determine the confidence level of the person speaking.
4. *Knowledge assessment module:* This module allows the user to take the test using their credentials. It is then followed by “Take Test” step where users are then prompted to take a test, presumably to assess their knowledge on a particular subject. This seemingly outputs the results of the test, accompanied by comprehensive explanations or descriptions, which could provide feedback based on the user's performance and detected emotional state.

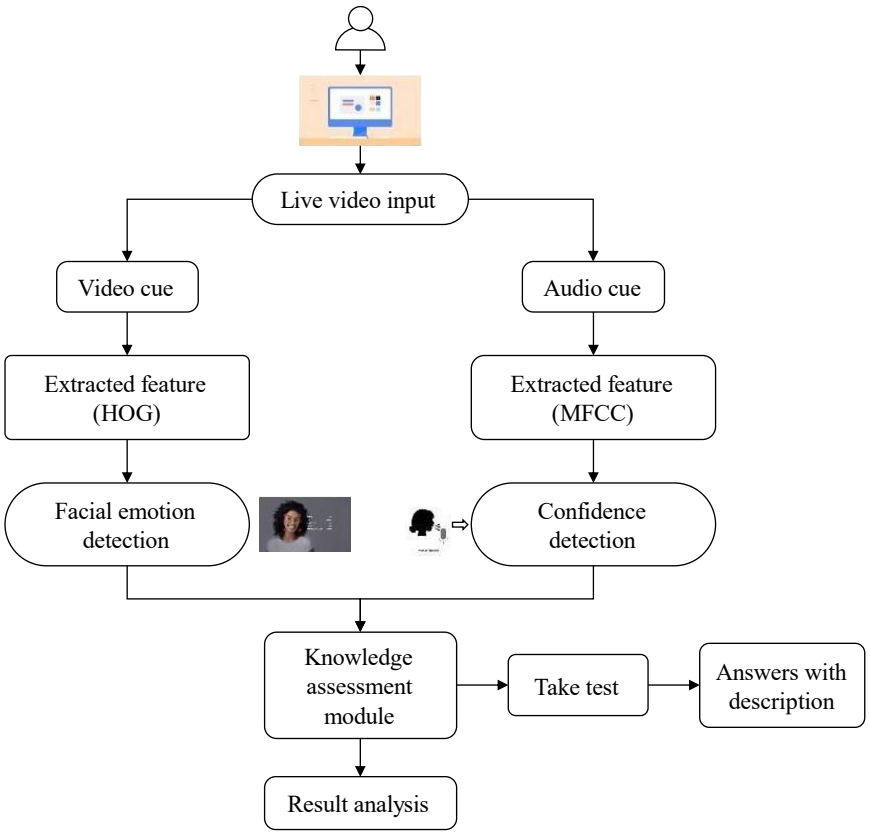


**Figure 1.** System architecture.

 

## METHODOLOGY

The User log-in to the website and click start for the video interview. The System starts with the live video input from the user. The live video data is used as input which is approximately 5–10 seconds. These video data inputs are then distributed for emotion and voice analysis (Figure 2).

The Video Cue enables to extract HOG features from input and detect facial emotion. Total 8 emotions of face has been used for emotion analysis such as Neutral, Happy, Sad, Surprise, Disgust, Fear, Anger and Contempt. Deep Learning CNN algorithms are used to detect the facial expressions.

The Audio Cue enables MFCC (Mel-Frequency Cepstral Coefficients) to extract features from input and detect voice emotion. Same emotions are used for the speech emotion analysis and deep learning CNN models are used. The Pitch and loudness are the two main features extracted using MFCC. The Confidence evaluation of the speaker is decided based on the combination of the speech emotions. If the detected emotions are loud, happy, neutral, then confidence seems to be high. Otherwise for other emotions, it seems to be low confidence.

After successful video and audio feature extraction, a pop-up with the message as “Take Test” appears on the screen. The user will be forwarded to the question-answer form after clicking on the button. Along with the answer, the user must provide the best descriptions for those answers which makes their knowledge to be calculated. This is called Knowledge Assessment module of the project. Technical and Non-technical questions are updated on the quiz form to be answered by the candidate attending interview.

The Result analysis is the last part of the project where results appear in three sections that is facial emotion, confidence evaluated and the knowledge assessed.

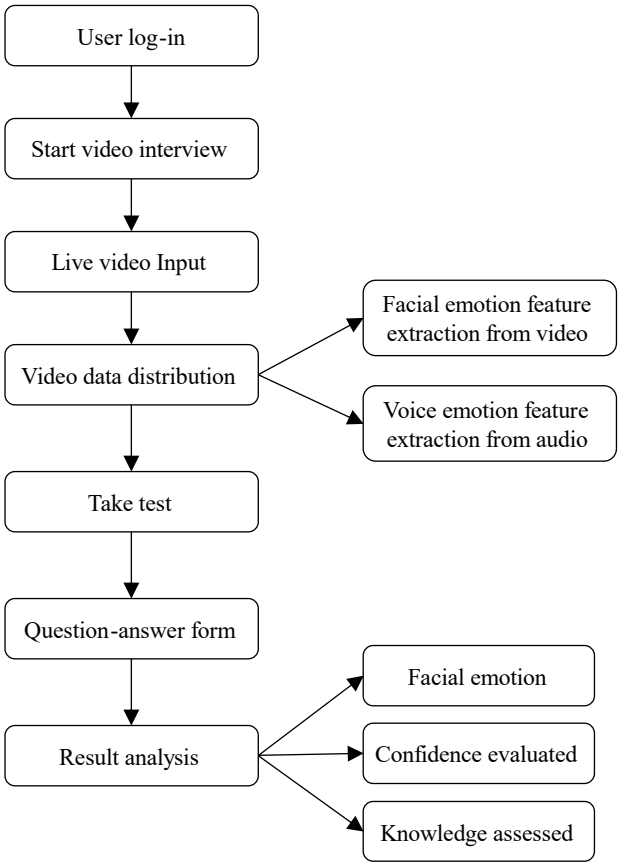


**Figure 2.** Flow chart of the evaluator

 

## CONCLUSION

This paper presented an artificial intelligence (AI) powered mock interview system that uses audio and video analysis to evaluate a user's knowledge and communication abilities. Utilizing deep learning for both voice analysis (using MFCC features) and facial emotion detection (using HOG features), the system calculates user confidence in conjunction with their performance on a knowledge assessment module. This comprehensive assessment provides insightful information about interview readiness. This data offers a thorough analysis of the user's interview performance, including with user answers to a knowledge assessment module comprising both technical and non-technical questions. With this technique, interview preparation is made easy and scalable, giving users the chance to practice and get feedback in an authentic environment.

Interview practice is made easy with the system's scalable and convenient platform. Future research might look into adding elements like body language analysis or the ability to generate interviewer feedback. Further research could improve the existing confidence estimation by taking into account variables like speech fluency and hesitations, as it is dependent on fundamental emotion recognition. All in all, this system offers a promising method for interview preparation, enabling users to enhance their communication abilities and obtain feedback based on data.

## REFERENCES

1. L. S. Nguyen, D. Frauendorfer, M. S. Mast, and D. Gatica-Perez, “Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior,” IEEE transactions on multimedia International Conference.
2. Dulmini Yashodha Dissanayake, Venuri Amalya, Raveen Dissanayaka, Lahiru Lakshan, Pradeepa Samarasinghe, Madhuka Nadeeshani, Prasad Samarasinghe, "AI-based Behavioural Analyser for Interviews/Viva ", IEEE 2021.
3. M.S. Sinith, Aswathi E., Deepa T. M., Shameema C. P and Shiny Rajan, "Emotion Recognition from Audio Signals using Support Vector Machine", December 2015.
4. Ming-Hsiang Su, Kun-Yi Huang, Tsung-Hsien Yang, Kuan-Jung Lai and Chung-Hsien Wu, "Dialog State Tracking and Action Selection Using Deep Learning Mechanism for Interview Coaching", IEEE 2016.
5. L. Chen, R. Zhao, C. W. Leong, B. Lehman, G. Feng and M. E. Hoque, "Automated video interview judgment on a large-sized corpus collected online", 2017.
6. I. Naim, M. I. Tanveer, D. Gildea and M. E. Hoque, "Automated Analysis and prediction of Job Interview Performance", IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 191-204, 2018.
7. N. Dhananjaya and B. Yegnanarayana, Correlation-Based Similarity Between Signals for Speaker Verification with Limited Amount of Speech Data, Chennai 600 036, India: Indian Institute of Technology Madras.
8. F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier and B. Weiss, "A database of German emotional speech", Proc. Interspeech, 2005.
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 886–893.2005.
10. De la Torre, F., Chu, W.S., Xiong, X., Vicente, F., Ding, X., Cohn, J.F.: Intraface. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp 1–8. FG.2015.7163082 (2015).

 