International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 71-74 www.ijaem.net ISSN: 2395-5252

A Comprehensive Survey on Mock Interview Evaluation Systems

Firdouziya M Khadar, Devika G, Nayana Liz Mathews, Parvathy R Prasad, Sulaja Sanal

Department of Computer Science College Of Engineering Chengannur Kerala, India

Date of Submission: 01-12-2024 Date of Acceptance: 10-12-2024

ABSTRACT—Interviews are essential for evaluating candidates' skills, knowledge, and suitability for job roles, while mock interviews serve as valuable tools for preparation. With rapid advancements in artificial intelligence (AI) and machine learning (ML), automated interview evaluation has transformed recruitment by offering more objective, efficient, and comprehensive assessments compared to traditional methods. This survey re- views recent AI-driven methodologies for candidate evaluation in mockinter view settings, focusing on multi-modal techniques such as facial emotion recognition, speech analysis, text mining, and nonverbal behavior analysis. It discusses the effectiveness and limitations of various approaches, providing insights into how AI models assess emotional states, speech confidence, personality traits, and knowledge for comprehensive candidate evaluation.

Index Terms—deep learning, CNN, categorical emotions, NLP, confidence evaluation, web scraping, Speech Emotion Recognition (SER)

I. INTRODUCTION

The rapid advancements in artificial intelligence (AI) and machine learning (ML) have significantly impacted various domains, including recruitment and human resource management. One area that has seen notable progress is auto- mated interview evaluation and candidate assessment. Traditional interview processes can be time-consuming, subjective, and inconsistent, whereas AI-driven solutions offer the potential for a more objective, efficient, and comprehensive evaluation of candidates. The integration of multi-modal AI techniques—incorporating facial emotion recognition, speech analysis, text mining, and nonverbal behavior analysis—has been a promising development in this field. This survey paper reviews recent methodologies employed for candidate evaluation in mock interview settings, providing insights into the effectiveness and limitations of different approaches.

The surveyed methods span a variety of AI techniques, including deep learning, natural language processing (NLP), computer vision, and traditional machine learning. These techniques are often combined to create a more robust evaluation system by leveraging multiple data sources, such as facial expressions, voice tone, speech content, and textual information. The methodologies mentioned in this paper illustrate how various AI models can be utilized to assess emotional states, speech confidence, personality traits, and knowledge, providing a comprehensive assessment of the candidate's performance.

A. Importance of Mock Interviews

1. Self-Assessment of Strengths and Weaknesses: Mock interviews provide an opportunity for candidates to evaluate their own skills and identify areas for improvement. By simulating the interview experience, individuals can better understand their strengths and weaknesses in communication, body language, and overall presentation, allowing for targeted preparation ahead of real interviews.
2. Expert Feedback and Guidance: Participating in mock interviews often involves receiving constructive feedback from experienced professionals or peers. This guidance can help candidates refine their answers, improve their confidence, and learn how to effectively navigate common interview questions and scenarios.
3. Familiarization with the Interview Environment: Mock interviews replicate the real interview setting, which helps reduce anxiety and builds comfort with the process. By becoming accustomed to the structure and dynamics of interviews, candidates are less likely to feel overwhelmed during actual interviews, leading to improved performance and outcomes.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 71-74 www.ijaem.net ISSN: 2395-5252

B. How Current Interview Systems Work

Current interview systems primarily focus on a candidate's knowledge, but they often overlook key elements like behavioral skills, personality traits, and emotional intelligence. Many systems are now shifting to online platforms, especially after the pandemic, with a significant percentage of recruiters using virtual interviews. However, candidates frequently struggle with conveying non-verbal cues like smiles and nods during these interviews, which can affect their performance. To address these issues, AI-driven mock interview systems have been developed that evaluate candidates based on facial expressions and speech patterns, assessing both emotional and confidence levels to enhance overall interview readiness.

II. LITERATUREREVIEW

R.Mandal et al. (2023) [1] proposed a dynamic mock interview system. The methodology integrates facial emotion recognition, speech confidence assessment, and knowledge extraction using AI techniques. For emotion recognition, video frames are preprocessed (cropped, grayscaled, normalized) and passed through a six-layer CNN with four convolutional and two fully connected layers. The model, trained on seven emotion categories, analyzes facial expressions during the interview, adjusting scores accordingly. For confidence analysis, audio data is processed through NLP techniques that evaluate speech clarity, tone, and pitch, categorizing the speaker's confidence level. Simultaneously, web scraping is used to extract key phrases from responses, mapping them to relevant online information to assess the candidate's knowledge. These combined analyses provide a comprehensive evaluation of the candidate's performance in a mock interview setting.

D. Y. Dissanayake et al. (2021) [2] proposed a system that detects personality traits through facial and head movements, enabling nonverbal cue recognition in online interactions, offering continuous behavioral and group analysis. The methodology for detecting smiles and their genuineness using CNN architectures, along with personality trait prediction based on nonverbal cues, can be summarized as follows. Smile detection is performed using a LeNet-based CNN model, which classifies smiles as genuine or posed. To further analyze smile genuineness, a ResNet-18 combined with Conv LSTM is applied to identify discriminative features of spontaneous versus posed smiles over consecutive frames. For personality traits prediction, nonverbal cues such as smiles, gaze, and head movements are analyzed, and a Random Forest (RF) model is employed to classify traits like agreeableness, openness, and extraversion based on these cues. This integrated approach leverages deep learning and machine learning to assess both emotional and behavioral characteristics of individuals during interviews or other assessments.

Y. -C. Chou et al. (2020) [3] proposed a system that help job seekers to match with suitable positions at large job fairs. It analyzes applicants' resumes to assess their competitiveness and personality traits, recommends relevant job opportunities, and provides companies with lists of potential candidates. The methodology involves using machine learning and text mining to analyze resumes for job applications. The system processes resumes to assess the candidate's competitiveness in areas like skills, education, and experience. It then applies the DISC model, which categorizes personality traits into dominance, influence, steadiness, and compliance, to evaluate the applicant's soft skills. Based on these analyses, the system generates job recommendations by matching the resume content with available job vacancies, providing candidates with a list of suitable roles. This approach enhances the job-seeking process by aligning candidate profiles with job opportunities.

Taiba Majid Wani et al. (2021) [4] introduces a Speech Emotion Recognition(SER) system that utilizes strong feature extraction methods, including Mel-Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstral Coefficients (LPCC),to analyze speech data that has been preprocessed for emotion detection. The methodology involves reviewing various Speech Emotion Recognition (SER) systems to evaluate the effectiveness of different feature extraction and classification methods. The approach includes analyzing techniques used to extract features from speech signals, such as prosodic, spectral, voice quality, and Teager Energy Operator (TEO) features, which are crucial for capturing emotional content. Additionally, the methodology assesses various classification methods, including traditional models like GMM, HMM, and SVM, as well as deep learning approaches such as CNN, RNN, and DNN, to determine the most effective techniques for accurately recognizing emotions in speech.

Mingzhe Li et al. (2023) [5] proposed a mock interview generator namely EZ Interviewer for job-seekers, generating interviews based on resumes, job requirements, and previous responses. The methodology for the mock interview generator

includes collecting online interview data to create a comprehensive question database and analyzing candidate resumes to extract key details like skills and experiences. Using this information, the system generates customized interview questions tailored to each candidate's background and the job description, ensuring contextual coherence. The quality of the questions is evaluated through BLEU scores and human assessments, with ongoing feedback incorporated for iterative improvements. This process allows the generator to adapt to changing job market trends and enhance the interview experience.

Ch. Sri Latha et al. (2023) [6] proposed an automated interview evaluation system using SBERT model, to evaluate the user's responses and provide them with feedback on their performance. An automated interview evaluation system uses a multi-strategy operation that begins with recording the candidate responses in to text via speech recognition. Next, the text is pre-processed by cleaning and deducing structure from contextual meaning. Answers are checked for relevance and quality using natural language processing (NLP) techniques. Questions are delivered in audio format, and answers to queries come from the system, where evaluations of responses have been performed using Sentence-BERT to measures mantic similarity between potential answers and correct ones. It offers a way to operation alize feedback so candidates know specifically what needs improvement.

Samaneh Madanian et al. (2023) [7] proposed a system that aims to enhance Speaker-Independent Speech Emotion Recognition(SER) by systematically addressing common challenges and leveraging effective techniques across three critical stages: data pre-processing, feature extraction, and emotion classification. The method for processing speech signals commences with pre-processing where the speech is framed and windows to convert noises in to fixed segments. Speech data is converted from the time to frequency domain using the Fast Fourier Transform (FFT), which helps to extract key features such as pitch and energy. Discrete Wavelet Transform (DWT) — combining both time and frequency parts giving you higher resolution for high and low-frequency details. Feature Extraction: Important elements like Mel-Frequency Cepstral Coefficients (MFCC), pitch, and Intensity are then identified during feature extraction often performed by tools such as Open SMILE. Finally, there are machine learning algorithms such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Recurrent Neural Networks which classify the features to differentiate emotional states in speech. This integrated approach efficiently manages the complexities of speech signal analysis.

Sheradha Jauhari et al. (2022) [8] proposed a system that aims to develop a Virtual Mock Interview Assistant, a chatbot designed to help students prepare for job interviews by providing instant, conversational support. The methodology for conducting mock interviews using dialogue and sentiment analysis involves several key components. Speech-to-Text Conversion (ASR) transcribes spoken responses into text for further analysis, while Facial Emotion Recognition (FER) using CNN, detects emotions like happiness or stress by analyzing facial expressions to assess the candidate's emotional reactions. Gesture Recognition with 3D-CNN captures gestures such as hand movements and posture, providing insights into non- verbal communication. The system also performs Dialogue and Sentiment Analysis, evaluating emotional tone (e.g., positive, neutral, or negative) in responses and offering feedback on how the candidate handled specific questions. Based on the analysis of speech, facial expressions, and gestures, the system generates Human-like Feedback that mimics constructive criticism from a real interviewer, helping users identify strengths and areas for improvement. This integrated approach simulates real interview scenarios, giving users comprehensive feedback to enhance their preparation.

Hailun Lian et al. (2023) [9] proposed a system that uses deep learning techniques to enhance Multimodal Emotion Recognition (MER). MER is the framework that combines speech, text, and facial data for emotion recognition, where deep learning methods like convolutional neural networks (CNNs) are used to convert face images into features. This merges these data types by increasing the accuracy of detecting emotions, providing a more comprehensive and precise insight into emotional expressions and behaviors.

Rohan Patil et al. (2021) [10] proposed a software frame- work that allows virtual recruiters to adapt to users' behavior during mock interviews by recognizing social cues. These cues, relevant to job interview settings, were identified through studies involving real jobseekers. Facial expression recognition was performed using Convolutional Neural Networks (CNNs), achieving a strong accuracy of 88.56% after effective training. AI models based on TensorFlow were also employed to identify personality traits with precision ranging from 90.9% to 97.4%, particularly focusing on the "Big Five" traits—openness, conscientiousness, extraversion, agreeableness, and neuroticism. These

deep learning techniques enable real-time feedback, enhancing the analysis of both personality and behavior during user interactions.

III. CONCLUSION

The survey highlights the progress made in AI-driven emotion and confidence recognition systems, particularly for mock interview applications. Deep learning models, such as Convolutional Neural Networks(CNN) for facial recognition and advanced Speech Emotion Recognition (SER) techniques, have significantly improved the ability to evaluate human behavior and emotions (Training accuracy: 80.75%, Testing accuracy: 76.34%). While traditional machine learning approaches offer basic insights, deep learning methods have shown superior performance, especially in speaker-independent scenarios, despite ongoing challenges with dataset diversity, multimodal data fusion, and generalization. Of all the methods discussed, deep learning-based SER models employing hybrid fusion techniques are the most effective, demonstrating higher accuracy in emotion recognition tasks. To further enhance these systems, future work should focus on expanding datasets, improving the fusion of multimodal inputs, and boosting generalization to ensure the practical applicability of these systems in real-world interview environments.

REFERENCES

[1] AR. Mandal, P. Lohar, et al., ”AI -Based mock interview evaluator: An emotion and confidence classifier model,”2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore,India,2023,pp.521-526

[2] D. Y. Dissanayake et al., ”AI-based Behavioural Analyser for Inter-views/Viva,” 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS), Kandy, Sri Lanka, 2021, pp. 277-282

[3] Y.-C.Chou, H.-Y.Yu,”Based on the application of AI technology in resume analysis and job recommendation,” 2020 IEEE International Conference on Computational Electromagnetics (ICCEM), Singapore,2020, pp. 291-296.

[4] Taiba Majid Wani, Teddy Surya Gunawan, et al., ”A Comprehensive Review of Speech Emotion Recognition Systems, ”IEEE Access, vol.9, pp. 47795-47814,2021

[5] Mingzhe Li, Xiuying Chen, et al.,”EZ Interviewer: To Improve Job Inter-view Performance with Mock Interview Generator,” ACM International Conference on Web Search and Data Mining(WSDM 2023)

[6] Ch. Sri Latha, N. Krishna Kalyan, et al.,” Automated interview Evaluation,” E3S Web of Conferences, 010 (2023) ICMPC 2023

[7] Samaneh Madanian, Talen Chen, et al.,” Speech Emotion Recognition using Machine Learning — A Systematic Review,” Elsevier-Intelligent Systems with Applications 20 (2023) 200266

[8] Sheradha Jauhari, Chetan Aggarwal, et al., ”Virtual Mock Interview Assistant (Video Bot-based),” EasyChair Preprint(2022).

[9] HailunLian, Cheng Lu, et al.,”A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face,” MDPI-Entropy, 2023, 25, 1440.

[10] Rohan Patil, Akash Butte, etal.,”Real Time Mock Interview using Deep Learning,” International Journal of Engineering Research & Technology(IJERT),ISSN: 2278-0181 Vol. 10 Issue 05, May-2021