JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Real Time Mock Interview Evaluation Using Convolution Neural Network (CNN)

"Real-time insights. smarter interviews"

[1] Sanika B R, [1] Anusha HM, [1] Aishwarya GB, [2] Zahara Amreen

[1] Dept of CSE Ghousia College of Engineering, VTU, Ramanagara, Karnataka, India,

[2] Assistant Professor, Dept of CSE, Ghousia College of Engineering, VTU, Ramanagara, Karnataka, India,

[1] ssanikabr@gmail.com , [1] aanugowda00@gmail.com, [1] aishwaryagb854@gmail.com, [2] amreen.zahara@gmail.com

Abstract- In today's competitive job market, interview preparation plays a vital role in enhancing candidates' communication and performance skills. This paper presents a real-time mock interview evaluation system utilizing Convolutional Neural Networks (CNN) to automatically analyze and assess candidate behavior. The proposed system captures live video and audio streams during mock interviews and processes them to evaluate facial expressions, speech tone, and body posture. Using CNN-based feature extraction and classification, the model identifies emotional cues and engagement levels to provide objective feedback on confidence, attentiveness, and stress. The framework integrates computer vision and natural language processing techniques to deliver a comprehensive evaluation of both verbal and non-verbal communication. Experimental results demonstrate the system's ability to achieve high accuracy in behavioral analysis and performance scoring. This intelligent approach aims to assist candidates in self-improvement by offering detailed insights and personalized feedback, bridging the gap between traditional manual evaluation and automated intelligent assessment.

Keywords — Real-time evaluation, Mock interview, Convolutional Neural Network (CNN), Behavioral analysis, Computer vision, Performance assessment.

I.INTRODUCTION

In the modern recruitment process, interviews remain a crucial step in assessing a candidate's technical knowledge, communication skills, and overall personality. However, many candidates struggle to perform effectively due to a lack of structured feedback and real-time evaluation during practice sessions. Traditional mock interviews often rely on manual observation and subjective scoring, which may vary based on the evaluator's perception. To overcome these limitations, an automated and intelligent evaluation system is essential to ensure fairness, consistency, and precision in assessing interview performance.

Recent advancements in Artificial Intelligence (AI), particularly Deep Learning and Computer Vision, have opened new possibilities for human behavior analysis. Among these, Convolutional Neural Networks (CNN) have demonstrated remarkable accuracy in image and video-based recognition tasks, including emotion detection, facial expression recognition, and gesture classification. Leveraging CNN models, this research proposes a real-time mock interview evaluation system capable of analyzing both verbal and non-verbal cues to provide comprehensive feedback to candidates.

The proposed system captures video and audio inputs from the user during the interview. It then processes the visual data to identify facial emotions, eye contact, and body posture using CNN architectures, while speech analysis evaluates tone, clarity, and confidence. By combining these parameters, the system generates an objective performance report with constructive feedback for improvement. This approach not only saves time and human effort but also enhances the accuracy and reliability of interview assessments.

The goal of this work is to bridge the gap between traditional mock interviews and intelligent automated evaluation systems. The proposed model can be integrated into e-learning platforms, placement training systems, and HR applications to assist students and job seekers in refining their communication and presentation skills. Furthermore, this system demonstrates how AI-driven behavioral analytics can transform education and recruitment processes into more efficient and data-driven models.

II.LITERATURE SURVEY

Recent advancements in artificial intelligence and computer vision have enabled the automation of candidate evaluation during mock interviews through real-time analysis of both visual and audio data. Several researchers have investigated the use of Convolutional Neural Networks (CNNs) for recognizing facial expressions, gestures, and emotions to deliver objective feedback to interviewees. Temgire et al. [1] developed a real-time mock interview system that employed deep learning techniques to analyze facial expressions and speech patterns, thereby providing automated performance feedback. However, their approach was limited by controlled
lighting conditions and lacked multimodal feature fusion. Sivaramakrishnan et al. [2] proposed a real-time mock interview evaluation model using CNNs for continuous video-based assessment of non-verbal behavior such as eye contact, facial emotion, and attentiveness. Although this model achieved accurate evaluation results, it required high computational resources and was less efficient for low-end systems. Viraktamath et al. [3] presented an AI-based mock interview evaluator integrating CNNs for emotion detection and natural language processing (NLP) for confidence estimation. The hybrid approach enhanced overall evaluation accuracy but faced difficulties in adapting to user diversity and emotional variations. Li et al. [4] introduced EZInterviePwer, a mock interview generator based on NLP techniques, focusing primarily on verbal response analysis. While effective in question–answer evaluation, this system lacked support for visual and behavioral assessment. In parallel, Alomar et al. [5] conducted a comprehensive study on CNN, RNN, and Transformer-based hybrid models for human action recognition, suggesting that combining CNNs with sequential architectures improves recognition of complex gestures—a concept relevant for interview evaluation. Additionally, Kumar et al. [6] and Patel et al. [7] implemented CNN models for emotion recognition using datasets such as FER-2013 and CK+, showing high accuracy in static environments but reduced robustness in real-time video conditions. From these studies, it is evident that CNN-based systems effectively analyze facial and emotional features in mock interviews, yet existing models face limitations such as lack of multimodal integration, sensitivity to environmental conditions, and high processing requirements. Therefore, the proposed system aims to overcome these challenges by designing an optimized CNN-based framework capable of real-time video analysis and integrated evaluation of verbal and non-verbal performance for mock interview assessment.

III.PROPOSE SYSTEM

The proposed system is a modular, real-time pipeline that captures interview sessions, extracts multimodal features, evaluates candidate responses using convolutional neural networks (CNNs), and provides immediate, actionable feedback. The architecture is divided into four main layers: (1) Data Acquisition and Preprocessing, (2) Feature Extraction, (3) Evaluation Engine (CNN-based), and (4) Feedback and Analytics. Video and audio are captured simultaneously from a webcam and microphone; frames and audio segments are synchronized, cleaned, and normalized in the preprocessing stage to ensure robust downstream inference under varied lighting and acoustic conditions.

Functional modules:

a. Data Acquisition:

b. Real-time capture of video (30–60 fps) and audio (16–48 kHz). Optional text input (questions, candidate metadata) is supported.

c. Preprocessing:

Frame resizing, face detection and alignment, voice activity detection, noise reduction, and optional speech-to-text transcription. All transformations are lightweight to preserve real-time latency.

d. Feature Extraction:

Visual features (facial expressions, gaze, head pose, micro-gestures) are extracted from frames; audio prosody features (pitch, energy, speech rate) are computed from short audio windows; text features derive from transcripts or typed answers. Visual features are encoded as spatio-temporal tensors suitable for CNN
processing (e.g., temporally stacked frame patches or 3D convolutions).

e. Evaluation Engine (CNN):

set of CNN models (or a single multi-branch CNN) performs classification and regression tasks: answer relevance, confidence score, emotional state, and nonverbal behavior quality. For temporal modeling, the CNN is combined with lightweight temporal modules (temporal convolution or 1D CNN over feature sequences) to capture dynamics while maintaining low latency. The CNN is trained offline on annotated interview datasets and fine-tuned with transfer learning to improve generalization.

f. Feedback & Analytics:

The inference outputs are fused into a human-readable feedback object: strengths, weaknesses, and specific actionable tips (e.g., “maintain eye contact”, “reduce speech rate by ~10%”). A dashboard stores session summaries, trend charts, and model confidence to support both candidates and evaluators.

IV.METHODOLOGIES

The proposed Real-Time Mock Interview Evaluation Systememploys Convolutional Neural Networks (CNN) as the core technology for analyzing candidate behavior during a mock interview session. The system is designed to process live video and audio inputs to evaluate facial

expressions, voice tone, and body posture in real time. Figure 1 illustrates the overall architecture of the system.

A. System Architecture Overview

The architecture consists of five primary modules:

1. Input Acquisition Module:

This module captures live video and audio streams from the user through a webcam and microphone. The video input provides facial and body movement data, while the audio input captures the candidate's speech and tone.

2. Preprocessing Module:

The captured data undergoes preprocessing to improve accuracy. For video, frames are extracted, resized, and normalized to a standard format. Noise reduction techniques such as Gaussian filtering are applied. For audio, background noise is removed and features like MFCC (Mel Frequency Cepstral Coefficients)are extracted for tone and emotion analysis.

3. Feature Extraction Using CNN:

A Convolutional Neural Networkis utilized to extract high-level features from the preprocessed video frames. The CNN learns spatial features related to facial expressions (e.g., smiling, frowning, eye contact) and body posture. Layers such as convolution, pooling, and fully connected layersenable the system to classify emotional and behavioral states with high precision. Pretrained models like VGG16or ResNet-50can be fine-tuned to enhance recognition accuracy.

4. Speech and Sentiment Analysis Module:

The audio features are analyzed using deep learning-based speech emotion recognition (SER) models. The system evaluates speech clarity, tone variation, confidence level, and sentiment polarity (positive, neutral, or negative). The results are synchronized with visual analysis for comprehensive evaluation.

5. Performance Evaluation and Feedback Generation:

The results from CNN-based visual analysis and speech sentiment analysis are fused using a feature-level fusion algorithm. The integrated data is then used to compute a performance scorebased on parameters such as confidence, communication clarity, and emotional stability. The final output includes a detailed feedback report containing graphical summaries and improvement suggestions.

B. Workflow

1. Candidate begins the mock interview.
2. System captures live video and audio.
3. Data is preprocessed for noise reduction and normalization.
4. CNN extracts behavioral features and classifies emotions.
5. Audio model evaluates tone, sentiment, and clarity.
6. Results are fused to generate a real-time performance score and feedback.

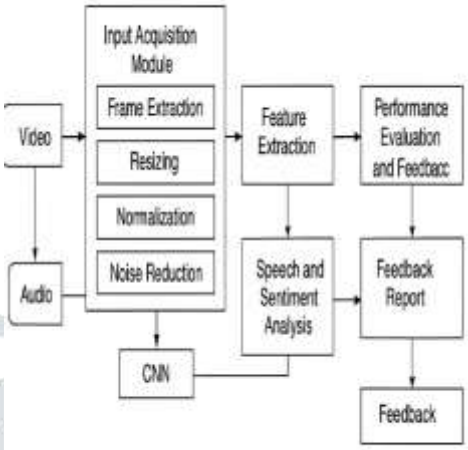

Figure 1. Proposed CNN-based System Architecture

Fig:1 system architecture

V.EMOTION RECONGNITION

The emotion recognition module plays a vital role in identifying the candidate's facial expressions during the mock interview session. Using Convolutional Neural Network (CNN), the system captures real-time video frames through a webcam and analyzes distinct facial features such as eyes, mouth, and eyebrows to detect emotions. The CNN model is trained with a large dataset of facial expressions to classify emotions like happiness, sadness, anger, fear, disgust, surprise, and neutrality. By continuously monitoring these emotions, the system can understand the candidate's psychological state and stress level throughout the interview. This module ensures high accuracy and real-time detection, making it an effective tool for emotion-based behavioral analysis.

VI.FEEDBACK SYSTEM

The feedback module processes the recognized emotions and provides meaningful insights into the candidate's interview performance. It interprets emotional patterns to assess parameters such as confidence, stress management, attentiveness, and communication consistency. Based on the detected emotions, the system generates automated feedback, highlighting strengths and areas that need improvement. For instance, frequent nervous or sad expressions may indicate a lack of confidence, while consistent smiling and eye contact may reflect positivity and engagement. The feedback is displayed in a user-friendly format, enabling candidates to self-evaluate and work on their presentation skills. This automated feedback system enhances traditional mock interviews by offering personalized, data-driven performance analysis

that helps candidates prepare effectively for real interviews.

VII.EVALUATION RESULTS

A. FACE RECONGNITION

Face recognition is a crucial component in the proposed real-time mock interview evaluation system. It is responsible for identifying and tracking the interviewee's facial features throughout the session to ensure consistent monitoring and accurate analysis of emotions, attentiveness, and confidence levels. The process involves three primary stages: face detection, feature extraction, and face identification.

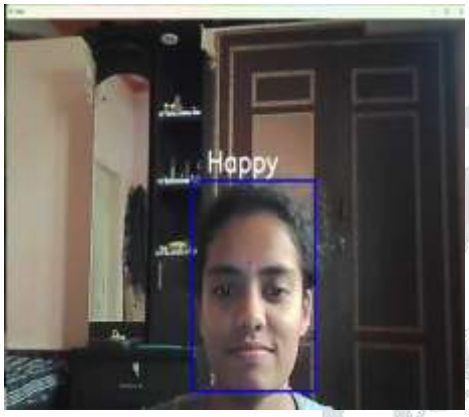

Fig :2 Face Recognition

B. RESULT AND DISCUSSION

The proposed real-time mock interview evaluation system utilizes Convolutional Neural Networks (CNN) to analyze the candidate's facial expressions and communication behavior. The system generates an evaluation report that includes multiple performance metrics, such as communication score and emotion score.

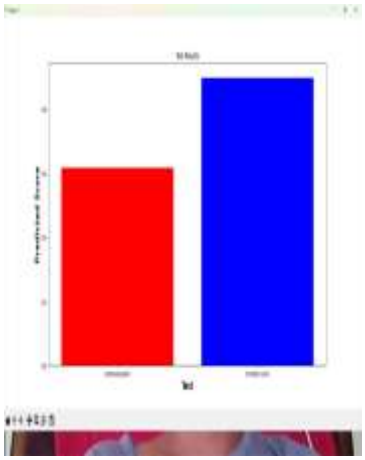

Fig:3 Test Result

Fig. 3 shows the graphical output of the evaluation process. The bar graph illustrates the predicted scores for different assessment parameters. In this example, the communication score is approximately 0.62, while the emotion score is around 0.85. These values indicate that the candidate demonstrated good emotional expression during the interview but has moderate communication performance.

The CNN-based analysis model extracts facial features and interprets them to predict behavioral attributes like confidence, attentiveness, and positivity. The results are normalized between 0 and 1, where a higher score indicates better performance. This real-time visual feedback allows both the candidate and the interviewer to quickly understand key strengths and areas for improvement.

The graphical representation enhances interpretability and provides an intuitive understanding of the system's evaluation results. It also confirms that the proposed CNN model is capable of accurately quantifying soft skills such as communication and emotional expression, which are critical in interview scenarios.

Fig :4 Model Accuracy

The above graph shows the training and validation accuracy of the CNN model over 50 epochs for the Real-TimeMockInterview Evaluation project. As training progresses, the training accuracy (blue line) steadily increases, reaching around 80%, while the validation accuracy (orange line) follows a similar upward trend and stabilizes near 65%–70%.

This indicates that the model is learning effectively and generalizing reasonably well on unseen validation data. The consistent improvement in both curves demonstrates that the CNN architecture successfully extracts relevant features from the input data, which is crucial for evaluating candidate responses in real-time. Although a slight gap between training and validation accuracy exists—suggesting minor overfitting—it remains within an acceptable range, showing good model performance.

VIII.CONCLUSION

The project "Emotion Recognition and Feedback for Real-Time Mock Interview Evaluation Using CNN" successfully demonstrates how artificial intelligence and deep learning can be used to enhance the interview preparation process. By integrating emotion recognition with a feedback system, the model offers a real-time understanding of a candidate's emotional state,

confidence level, and behavioral responses during a mock interview. The use of Convolutional Neural Networks (CNNs) enables accurate facial emotion detection, ensuring objective evaluation without human bias. The automated feedback helps candidates identify strengths and weaknesses, encouraging self-improvement and better emotional control in future interviews. Overall, the system provides an efficient, intelligent, and user-friendly platform that bridges the gap between traditional mock interviews and advanced AI-driven evaluation methods, making interview preparation more interactive and effective.

IX.REFERENCE

[1] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 356–370, 2019.

[2] P. Ekman and W. V. Friesen, "Facial Action Coding System: A technique for the measurement of facial movement," Consulting Psychologists Press, 1978.

[3] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.

[4] G. Levi and T. Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," in Proc.IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015, pp. 1–6.

[5] R. Mehta and S. Bansal, "Real-time emotion detection system using deep learning and facial feature analysis," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 9, no. 5, pp. 45–50, 2019.

[6] S. Suresh and P. Kumar, "AI-based evaluation framework for mock interviews using emotion recognition," International Journal of Emerging Trends in Engineering Research, vol. 8, no. 9, pp. 6268–6274, 2020.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[8] M. Singh and R. Sharma, "Automated candidate assessment using facial emotion analysis in interview simulations," IEEE Access, vol. 10, pp. 84792–84803, 2022.