International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 02 | February - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Video based Interview Assessment System

Abhijit Darade
Dept. of Information Technology
Vidyalankar Institute of Technology
Mumbai, India
abhijit.darade@vit.edu.in

Ayush Kamble
Dept. of Information Technology
Vidyalankar Institute of Technology
Mumbai, India
ayush.kamble@vit.edu.in

Anushka Satav
Dept. of Information Technology
Vidyalankar Institute of Technology
Mumbai, India
anushka.satav@vit.edu.in

Prof. Rasika Ransing
Dept. of Information Technology
Vidyalankar Institute of Technology
Mumbai, India
rasika.ransing@vit.edu.in

***Abstract* — The evolving job interview landscape, driven by technological advancements, has given rise to a pressing challenge in effectively assessing candidates in the era of virtual interviews and remote hiring. To address this, a video-based IAS is introduced, offering a comprehensive solution by integrating innovative features. IAS leverages cutting-edge technologies to analyze audio tones, enhancing the evaluation of candidates' communication skills by assessing confidence and clarity. The system incorporates facial sentiment analysis to gauge emotional expressions, providing insights into interviewee engagement and enthusiasm. IAS streamlines the hiring process by automating candidate comparisons, thereby saving recruiters valuable time. In essence, IAS represents a holistic and efficient approach to modernizing the interviewing and hiring process, revolutionizing candidate evaluation and selection in the digital age.**

Keywords—Interview Assessment System (IAS), Convolutional Neural Networks (CNN), Deep Learning, Machine Learning.

I INTRODUCTION

Among the various methods of candidate assessment, job interviews stand out as the most prevalent. These interviews can take different forms, such as phone interviews, live video sessions, face-to-face meetings, or the more recent asynchronous video interviews. In the case of the latter, candidates log onto a platform, recording their responses to a set of questions curated by the recruiter. The platform then allows multiple recruiters to assess the candidate, engage in discussions, and potentially extend an invitation for an in-person interview. Recruiters opt for these platforms due to the broader candidate pool they provide access to and the accelerated application processing time. Moreover, asynchronous interviews offer candidates the flexibility to conduct interviews at their convenience. However, the increasing volume of asynchronous interviews may pose challenges for recruiters in terms of manageability. The structured nature of asynchronous video interviews, with consistent questions and allotted time per candidate, contributes to heightened predictive validity and diminishes inter-recruiter variability

The process of recruiting the right talent is a complex and crucial undertaking for organizations of all sizes. Recent advancements in technology, coupled with the global shift towards remote work, have introduced a new dimension to the interviewing and candidate assessment process. Traditional face-to-face interviews are progressively being supplanted by video interviews, prompting the need for sophisticated tools and methodologies to effectively evaluate candidates in the digital landscape. The Video-Based Interview Assessment System represents a revolutionary approach to reshape the interviewing and candidate evaluation paradigm. It goes beyond being a mere tool, offering a comprehensive approach that integrates various features to provide a thorough and unbiased assessment of interviewees. In this investigation, we

explore the multifaceted capabilities of Interview Video Analysis, each tailored to address specific aspects of the interview process:

- Audio Tone Analysis: Employing advanced audio processing techniques, Interview Video Analysis analyzes the tone and sentiment of candidates' responses, providing insights into their emotional intelligence and communication skills.
- Grammar Correction: Recognizing the importance of grammatically sound responses, Interview Video Analysis incorporates algorithms for grammar correction, enhancing the clarity and professionalism of candidate responses.
- Fluency Rating: Fluent communication is a crucial factor, and Interview Video Analysis rates candidates based on their ability to express themselves clearly and coherently.
- Facial Sentiment Analysis: Leveraging facial recognition technology, IVA assesses candidates' facial expressions and sentiments during interviews, offering valuable insights into their emotional state and engagement level.
- Ratings/Reviews: Interview Video Analysis enables interviewers and panel members to provide ratings and reviews of candidates, facilitating a structured evaluation process.
- Automated Candidate Comparison: Streamlining the comparison of candidates, Interview Video Analysis aggregates data from various assessment metrics, enabling more informed hiring decisions.
- Email Reminders: Interview Video Analysis offers automated email reminders and scheduling features, ensuring a smooth interview process for both candidates and interviewers.

Throughout this research paper, we thoroughly examine each of these features, highlighting their unique contributions to the interview process. We aim to illustrate how the Video-Based Interview Assessment System has the capacity to improve the efficiency, accuracy, and fairness of candidate evaluations. By the conclusion of this investigation, it will become evident that Interview Video Analysis holds the potential to redefine the strategies organizations employ in talent acquisition, ultimately resulting in more informed hiring decisions and the formation of highly successful teams.

II RELATED WORK

Researchers have shown significant interest in human motion analysis, employing both automated methods and visualization techniques for comprehensive examination. Motion labeling tools have been introduced to facilitate gesture annotations in videos, but their use is often associated with significant time and labor investments. Some scholars have explored automatic methods, while recent studies have delved into the automatic detection and recognition of human motion, with a particular focus on unsupervised approaches [7].

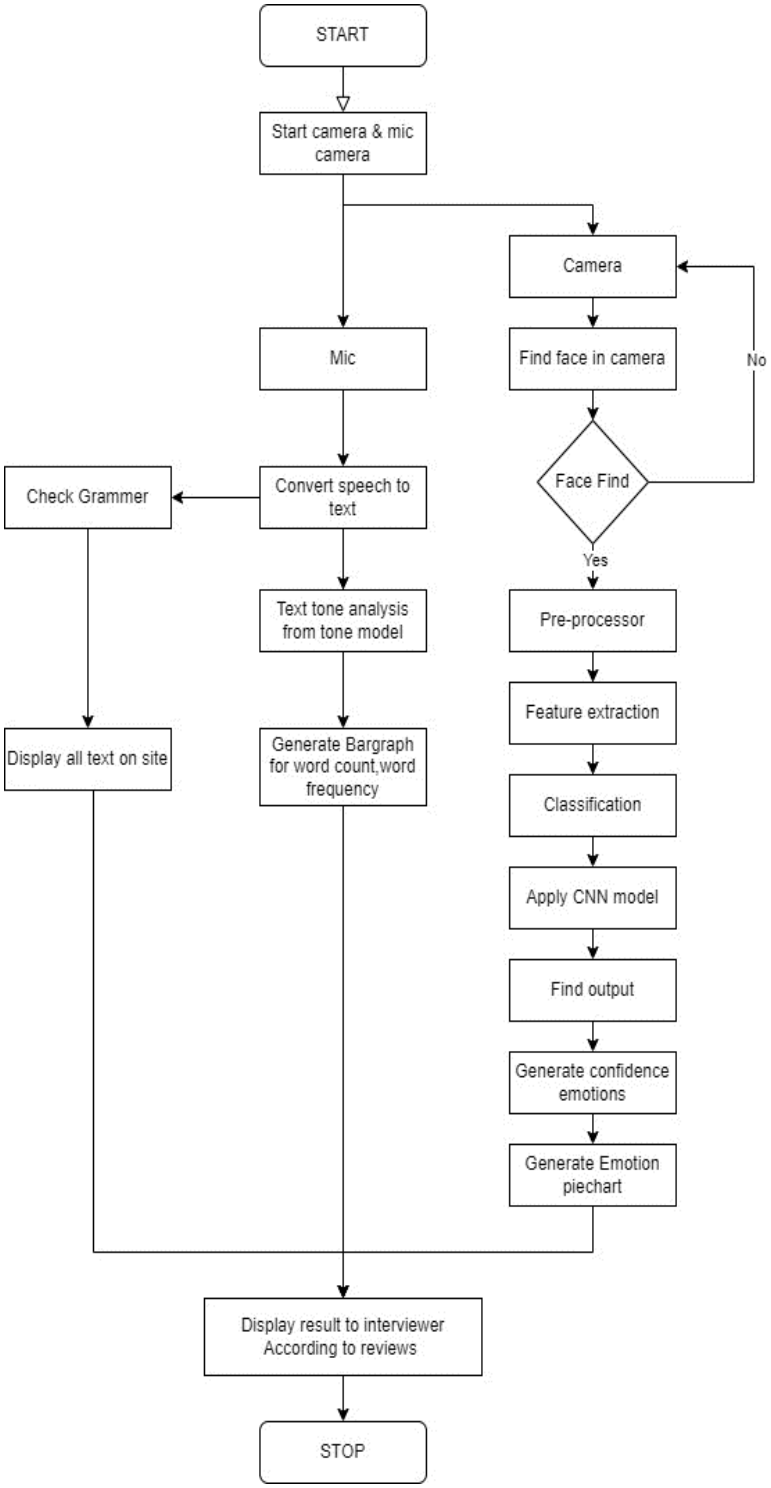
Visualization methods have been extensively utilized to analyze presentation techniques across various dimensions, including text, audio, and facial expressions. For instance, Tanveer [26] conducted an analysis of narrative trajectories by examining transcripts from over 2000 TED talks. This analysis revealed the correlation between narrative trajectories and audience ratings [7].

In the realm of social information processing theory, individuals observe and interpret cues exhibited by others during interactions, such as interviews. Brunswik's lens model illustrates how interviewers use cues to assess an interviewee's personality, establishing a connection between the interviewee's self-assessed personality and the interviewer's perceptual observations. Interviewees convey their apparent personality through distal cues, including facial expressions, gaze, posture, body movement, speaking, and prosody. Conversely, interviewers employ a 'lens' to attribute unobservable personality traits to interviewees through proximal cues, which are behaviors perceived directly by the interviewer. Despite being indirect observable cues, these behaviors contribute to the interviewer's perceptions [27].

Recent progress in the development of toolboxes has significantly enhanced the extraction of features from audio and video streams. Given that asynchronous job interviews are typically conducted through videos, it becomes essential to extract features from each modality (verbal content, audio, and video) frame by frame for constructing a classification model. Audio cues predominantly encompass prosody features (such as fundamental frequency and intensity) and speaking activity indicators (like pauses, silences, and short utterances). Visual cues, on the other hand, are often derived from facial expressions, including facial action units, head rotation and position, and gaze direction. Among these, facial expressions constitute the most commonly extracted visual cues. Moreover, advancements in automatic speech recognition have empowered researchers to leverage the verbal content of candidates. Describing the verbal content involves the use of various features, including lexical statistics (such as the number of words and unique words), dictionaries, topic modelling, bag of words, and more recently, document embedding [28].

III PROPOSED METHODOLOGY

Convolutional Neural Networks (CNNs) represent a prevalent form of deep artificial neural network, particularly well-suited for image processing tasks. In contrast to various other image processing algorithms, CNNs require minimal preprocessing. The architecture of CNNs is inspired by the connectivity pattern observed in the visual cortex of animals. Key components integral to CNNs include a convolutional methodology that dissects the intricate features of the input, and a fully connected layer that utilizes the convolution layer's capabilities to predict optimal definitions. The design of CNNs draws inspiration from the functionality and organization of the cortical area, aiming to replicate neuronal property patterns. Within a CNN, neurons are organized into a 3D structure [2]. This collection of neurons analyzes specific image attributes, with each category of neurons focusing on recognizing distinct parts of the image. Subsequently, the CNN utilizes the predictions from these layers to generate a final output, providing a vector of probability scores that indicate the likelihood of a particular attribute belonging to a specific category [2].



A) Algorithm for Interview Video Analysis

The algorithm for interview video analysis incorporates a comprehensive suite of tools and technologies to enhance the depth and accuracy of the assessment process. Natural Language Processing (NLP) libraries, including NLTK and spaCy, are employed for text extraction, entity recognition, and language comprehension in interview transcriptions. Grammar Checking APIs, such as Grammarly, play a crucial role in ensuring grammatical correctness, thereby elevating the quality of transcriptions and captions. Specialized NLP libraries like VADER and TextBlob contribute to Sentiment Analysis, delving into the emotional dimensions of interview transcripts and providing insights into the conversational context. On the visual front, Computer Vision Libraries like OpenCV are utilized for Facial Sentiment Analysis, extracting emotional cues from facial expressions and enriching the understanding of participants' emotional states. Facial Recognition APIs, such as Microsoft Azure Face API, enhance identity tracking and contribute to security measures. In the realm of audio analysis, tools like pyAudio Analysis scrutinize tones, pitch, and other audio characteristics, unraveling nuances in communication dynamics. Additionally, Email APIs like SendGrid and SMTP seamlessly integrate for efficient communication, delivering timely reminders and notifications related to interviews and events. This comprehensive set of tools collectively amplifies the efficiency and precision of interview video analysis, encompassing both textual and visual dimensions.

B) Performance Analysis of Algorithm

- **Text Sentiment Analysis** (e.g., using VADER sentiment analysis): VADER (Valence Aware Dictionary and sentiment Reasoner) calculates a

Fig1. Flowchart for face and tone/text analysis model

compound sentiment score for a given text. The formula for this compound score is computed as:

  - Compound Sentiment Score = Σ(valence of each word * sentiment booster) / (number of words + alpha)
  - valence of each word' represents the sentiment score of individual words. sentiment booster' accounts for special cases of words with added sentiment impact (e.g., intensifiers or negations).'alpha' is a parameter used to adjust the relative importance of sentiment boosters.

- **Audio Tone Analysis** (e.g., pitch analysis): Pitch analysis often involves complex signal processing techniques. A simplified formula for calculating pitch from audio samples using autocorrelation is as follows:

  - Pitch (in Hertz) = Sampling Rate (in Hz) / Lag at which autocorrelation is maximum
  - 'Sampling Rate' is the rate at which audio samples are recorded. 'Lag at which autocorrelation is maximum' corresponds to the period at which the signal repeats.

- **Fluency Rating Analysis**

  Fluency Rating = (Total Speech Duration - Total Pause Duration) / Total Speech Duration

  - In this formula:
  - Total Speech Duration is the duration of the actual spoken content in seconds. You can calculate this by subtracting the duration of all pauses from the total audio duration.
  - Total Pause Duration is the duration of all pauses in the speech. This can be determined by detecting periods of silence or inactivity in the audio.

IV IMPLEMENTATION AND RESULTS

The implementation phase of our research involves the utilization of Convolutional Neural Networks (CNNs) for the critical task of video-based interview assessment, specifically targeting the categorization of various communication aspects, such as verbal communication, non-verbal cues, and overall presentation. While CNNs are well-regarded for their ability to discern intricate patterns within video data, it is important to acknowledge that the initial model encountered challenges in accurately assessing complete interview performances. Despite these initial assessment discrepancies, we emphasize our unwavering commitment to refining and optimizing the model's performance. Employing a comprehensive approach, we are actively involved in fine-tuning parameters, expanding the training dataset, and exploring advanced CNN architectures to enhance the accuracy of interview assessments. This optimization process not only aims to address current assessment inconsistencies but also strives to improve the model's overall effectiveness in precisely categorizing diverse and intricate aspects of video-based interviews. The iterative refinement underscores our dedication to continually enhance the Video-based Interview Assessment system, ensuring it consistently delivers reliable

and precise results for an enriched user experience in evaluating interview performances. Furthermore, the advancement of our user profiling system is currently in progress, promising a transformative experience for every user. This system is positioned to enable users to curate and save their personalized recommendations, effectively creating a digital repository that aligns with their unique preferences. As we embark on these developments, our focus remains dedicated to delivering a sophisticated and tailored Video-based Interview system. This system not only embraces emerging technologies but also prioritizes user-centric personalization and convenience.

V CONCLUSION

The envisioned Video-Based Interview Assessment System (IAS) represents a groundbreaking approach to transforming traditional job interviews through the integration of advanced technologies. The systematic analysis and meticulous preparation of data have facilitated the development of a precise IAS tailored for video-based assessments. This innovative system incorporates features such as text sentiment analysis and audio tone analysis, providing a comprehensive framework for evaluating candidates. By leveraging these tools, the system aims to enhance the understanding of candidates' communication skills and emotional intelligence. Nevertheless, a rigorous testing phase and seamless integration into the broader IAS infrastructure are imperative to ensure accurate performance assessment. In essence, the Video-Based Interview Assessment System (IAS) holds the potential to revolutionize the landscape of talent acquisition, offering a sophisticated and efficient solution for modernizing the interview and assessment process.

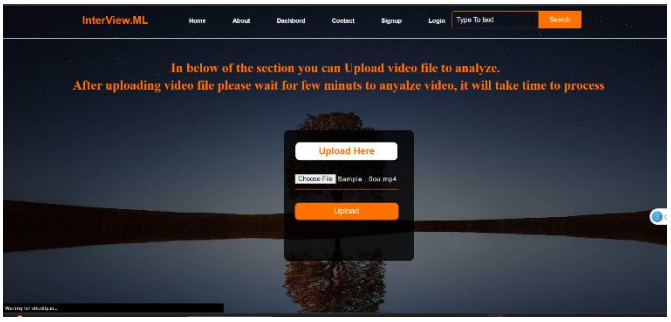


Fig 2. Home Page

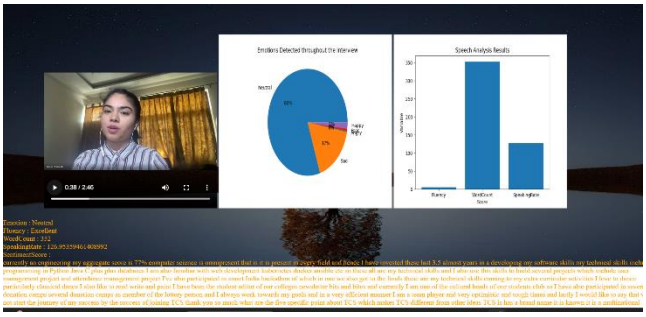


Fig 3. Analysis I

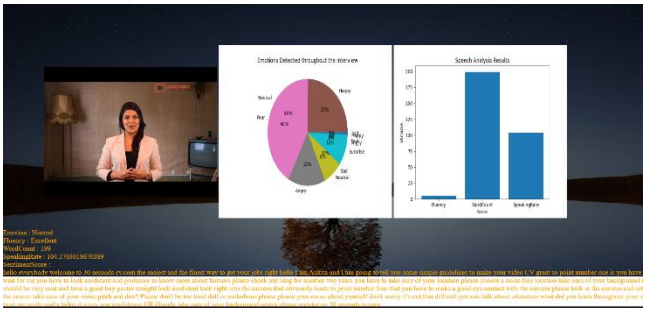


Fig 4. Analysis II

VI FUTURE SCOPE

Looking ahead, our future plans involve a significant enhancement of the system's usability by introducing additional functionalities, including gesture style comparison and gesture recommendation. We aim to embrace multi-modal features and leverage advanced data mining techniques to further elevate the precision and depth of gesture analysis within the Video-Based Interview Assessment System. Additionally, a comprehensive long-term study will be conducted, involving collaboration with more domain experts, to thoroughly evaluate the usability and effectiveness of the system [7]. As part of our commitment to continuous improvement, we are currently validating the accuracy of facial sentiment analysis through the collaboration with external tools and seeking insights from a distinguished professor in PCE. This multifaceted approach underscores our dedication to refining and validating the capabilities of the Video-Based Interview Assessment System, ensuring its continued effectiveness and relevance in the evolving landscape of talent assessment.

REFERENCES

[1] Aparna Joshi, MD, David A. Bloom, MD, Amy Spencer, Kara Gaetke-Udager, MD, Richard H. Cohan, MD 'Video Interviewing: A Review and Recommendations for Implementation in the Era of COVID-19 and Beyond' Volume 27, Issue 9

[2] Hung-Yue Suen, Kuo-En Hung, Chien-Liang Lin, "TensorFlow-Based Automatic Personality Recognition ,Used in Asynchronous Video Interviews" IEEE.

[3] Matthew Krouwel, Kate Jolly, Sheila Greenfield, "Comparing Skype (video calling) and in-person qualitative interview modes in a study of people with irritable bowel syndrome – an exploratory comparative analysis" Springer.

[4] Supriya Anand, Nihar Gupta, Mayesh Mulay, "Personality Recognition & Video Interview Analysis" .

[5] Mandal, R., Lohar, P., Patil, D., Patil, A., & Wagh, S. ," AI -Based mock interview evaluator: An emotion and confidence classifier model" IEEE.

[6] Chou, Y.-C., Wongso, F. R., Chao, C.-Y., & Yu, H.-Y., "An AI Mock-interview Platform for Interview Performance Analysis" IEEE.

[7] Zeng, H., Wang, X., Wang, Y., Wu, A., Pong, T.-C., & Qu, H., "GestureLens: Visual Analysis of Gestures in Presentation Videos" IEEE.

[8] Brown, J. A., & Vincent, V., "Video-based structured interviews to assess managerial candidates" Emerald Insight,

[9] Pandey, R., Chaudhari, D., Bhawani, S., Pawar, O., & Barve, S. "Interview Bot with Automatic Question Generation and Answer Evaluation" IEEE.

[10] Artiran, S., Ravisankar, R., Luo, S., Chukoskie, L., & Cosman, P. , "Measuring Social Modulation of Gaze in Autism Spectrum Condition With Virtual Reality Interviews" IEEE.

[11] Gorman, C. Allen, Gamble, Jason S., Robinson, Jim, "An investigation into the validity of asynchronous web-based video employment-interview ratings." Apa.org.

[12] Smith, A., Johnson, B., & Brown, C, "Automated Analysis of Facial Expressions in Job Interviews"

[13] Lee, S., Kim, J., & Park, H. , "Predicting Interview Performance using Machine Learning"

[14] Garcia, R., Martinez, M., Perez, J., "Automated Assessment of Live Job Interviews Using Machine Learning and Natural Language Processing"

[15] Zhang, Q., Li, W., & Wang, Z., "Detecting Fumbling Behavior in Interviews using Deep Learning"

[16] Lee, S., Kim, H., Park, J., "Real-Time Analysis of Live Job Interviews Using Multimodal Deep Learning"

[17] Zhang, Y., Liu, Y., Li, X., "Real-Time Emotion Recognition in Live Job Interviews Using Deep Neural Networks"

[18] Ahmad Shahvaroughi, Hadi Bahrami, Javad Hatami, Rui Paulo., "The cognitive interview: comparing face-to-face and video-mediated interviews"

[19] Yi-Chi Chou; Felicia R. Wongso; Chun-Yen Chao; Han-Yen Yu, "An AI Mock-interview Platform for Interview Performance Analysis"

[20] Haipeng Zeng; Xingbo Wang; Yong Wang; Aoyu Wu; Ting-Chuen Pong; Huamin Qu., "GestureLens: Visual Analysis of Gestures in Presentation Videos"

[21] J. Allen Brown Vinod Vincent, "Video-based structured interviews to assess managerial candidates"

[22] Raj Pandey; Divya Chaudhari; Sahil Bhawani; Omkar Pawar; Sunita Barve, "Interview Bot with Automatic Question Generation and Answer Evaluation"

[23] Saygin Artiran; Raghav Ravisankar; Sarah Luo; Leanne Chukoskie; Pamela Cosman. "Measuring Social Modulation of Gaze in Autism Spectrum Condition With Virtual Reality Interviews"

[24] Keyur Patel; Dev Mehta; Chinmay Mistry; Rajesh Gupta; Sudeep Tanwar; Neeraj Kumar; Mamoun Alazab, 'Facial Sentiment Analysis Using AI Techniques: State-of-the-Art, Taxonomies, and Challenges'

[25] Orhan Emre Aksoy1 , Selda Güney 'Sentiment Analysis from Face Expressions Based on Image Processing Using Deep Learning Methods' 2022, Vol. 8, Issue 4.

[26] M. I. Tanveer, S. Samrose, R. A. Baten, and M. E. Hoque, "Awe the audience: How the narrative trajectories affect audience perception in public speaking," in Proceedings of the CHI Conference on Human Factors in Computing Systems,

[27] A. Vinciarelli and G. Mohammadi, ''A survey of personality computing,''IEEE Trans. Affect. Comput., vol. 5, no. 3, pp. 273–291,

[28] Leo Hemamou, Ghazi Felhi, Vincent Vandenbussche, Jean-Claude Martin, Chloe Clavel "HireNet: A Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews"