International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04055
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Smart Interview Coach: Voice-Based HR and Technical Chatbot

1st Prof: Apurva Mane
Artificial Intelligence and Data Science
Vidya Pratishthan's
Kamalnayan Bajaj Institute
of Engineering and Technology
Baramati, India
apurva.Mane@vpkbiet.org

2nd Rhutuja Patil
Artificial Intelligence and Data Science
Vidya Pratishthan's
Kamalnayan Bajaj Institute
of Engineering and Technology
Baramati, India
rutujapatil500000@gmail.com

3rd Sayali Lagad
Artificial Intelligence and Data Science
Vidya Pratishthan's
Kamalnayan Bajaj Institute
of Engineering and Technology
Baramati, India
sayalilagad31@gmail.com

4th Gayatri Patil
Artificial Intelligence and Data Science
Vidya Pratishthan's Kamalnayan
Bajaj Institute of Engineering and Technology
Baramati, India
gaytripatil2020@gmail.com

5th Manjiri Chumbhale
Artificial Intelligence and Data Science
Vidya Pratishthan's Kamalnayan
Bajaj Institute of Engineering and Technology
Baramati, India
manjirichumbhale2725@gmail.com

***Abstract*—The growth of internet-based learning platforms has revolutionized the education system, making E-learning more accessible and widespread. However, with this shift toward digital education, maintaining academic integrity—especially during online examinations—has become a major concern. Traditional approaches to preventing cheating often fall short in virtual environments where physical invigilation is absent. Students appearing for online exams from remote locations are typically not monitored in real time, increasing the likelihood of unfair practices. To counter this, many educational institutions continue to conduct in-person exams, which goes against the core principle of flexible, remote learning. This project aims to address these challenges by developing an AI-driven system that ensures secure and fair online assessments. By incorporating multiple face recognition, object detection, emotion analysis, and voice-driven interview evaluations, the system identifies malpractice and provides a realistic simulation of technical and HR interviews. This automated approach not only strengthens exam monitoring but also enhances the overall credibility of E-learning platforms.**

***Index Terms*—E-learning, Online Examination, Academic Integrity, Malpractice Detection, Face Recognition, Object Detection, Emotion Analysis, AI Interview Chatbot, Remote Proctoring, Real-time Monitoring .**

I. Introduction

The increasing digitalization of education has led to the widespread adoption of Online Examination Systems (OES), which offer a streamlined and efficient way to conduct assessments remotely. These platforms eliminate the need for paper-based processes, allowing institutions to manage question banks, schedule exams, and evaluate responses through internet-enabled devices such as computers, tablets, and smartphones. As a result, online exams have become a cost-effective and scalable alternative to traditional assessment methods.

Despite the numerous benefits of OES, maintaining the credibility of remotely conducted exams poses a significant challenge. In the absence of physical invigilators, candidates often find ways to exploit system loopholes, leading to increased instances of malpractice. According to recent studies in digital assessment and remote invigilation (Patel et al., 2021; Singh Rao, 2022), conventional monitoring techniques are insufficient in detecting modern cheating behaviors. To counter this, some universities revert to conducting physical exams, which undermines the flexibility and accessibility of E-learning platforms.

To address these challenges, our research proposes a comprehensive AI-powered examination and interview system that integrates real-time monitoring and evaluation features. The system employs multiple face recognition, object detection, and emotion analysis to ensure identity verification and behavioral tracking throughout the assessment. Additionally, it simulates voice-based technical and HR interviews using a chatbot powered by natural language processing (NLP), providing automated scoring and feedback to users. These capabilities aim to reduce human intervention while improving assessment reliability.

The primary objective of this study is to design and implement a secure and intelligent online exam environment that discourages cheating and supports realistic evaluation. The research question that drives this work is: How can artificial intelligence enhance the integrity and effectiveness of online assessments in the absence of human proctors?

The methodology involves integrating machine learning (ML) models for visual and audio data processing, utilizing APIs for voice interaction, and evaluating system performance based on detection accuracy and user feedback. While the system is promising, it does have limitations, including dependency on camera/microphone quality and challenges in detecting subtle malpractice.

Overall, this research contributes to the growing field of in-

telligent E-learning systems by offering an automated and scalable solution to one of its most pressing concerns—academic honesty.

II. Literature Review

Paper Name: Speech Intelligibility and Audio Processing with CNNs Author: Mathias Batch Pedersen, Asgerh, Zhenghuatan Abstract: Mathias Batch Pedersen, Asgerh, and Zhenghuatan (2021) [1] investigated using CNN algorithms for speech intelligibility and audio processing. Although effective in recognizing features within structured data and enhancing speech clarity, the approach faces concerns regarding overfitting due to the dependency on training data quality and diversity, affecting the model's ability to perform consistently on new or diverse inputs.

Paper Name: Abnormal Behavior Detection in Online Exams Using Webcam Data Authors: Senbo Hu, Xiao Jia, Yingliang Fu Summary: Senbo Hu and colleagues (2022) tackled the issue of monitoring abnormal behavior during online exams. They proposed a method using webcam data to track head movements and mouth gestures. This approach improves upon traditional methods, which primarily focus on identity verification, by providing more effective monitoring for unusual behavior during exams.

Paper Name: Impersonation Detection in Online Exams Author: Pooja Mahesh Summary: Pooja Mahesh (2022) introduced a two-step biometric verification system to prevent impersonation in online exams. This method continuously checks the candidate's identity throughout the exam, ensuring the integrity of the process and eliminating the risk of someone else taking the test on behalf of the candidate.

Paper Name: Analyzing Video Data with CNNs and LSTMs Authors: Mehedy Masud, Kadhim Hayawi Summary: Mehedy Masud and Kadhim Hayawi (2022) used Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to analyze video data for pattern recognition and time-series classification. While their approach is useful for recognizing patterns over time, it faces challenges like dependence on video quality and issues with precise matching, which can lead to inconsistent results.

Paper Name: Optimizing Mobile Apps with MobileNetV2 Authors: Gabriel Muchangi, Kiura, Dr. Lawrence Mwenda Muirira Summary: Gabriel Muchangi and colleagues (2022) used MobileNetV2 combined with machine learning tools like TensorFlow, Keras, and PyTorch to create a neural network model optimized for mobile and embedded systems. They noted that the model's real-world performance could be limited due to insufficient testing and struggles with error handling, pointing to a need for greater robustness.

Paper Name: Human-Computer Interaction and Intuitive Interfaces Authors: Jose Barambones, Cristian Moral, Angelicade Antonio Summary: Jose Barambones and his team (2022) focused on improving human-computer interaction (HCI) by designing more intuitive interfaces that allow better communication between humans and machines. They highlighted gaps in current systems, such as their inability to recognize or respond to human emotions, which reduces the naturalness of interactions. They also discussed how context gaps limit the practical use of HCI systems.

Paper Name: Adaptive Interview Strategy with Multimodal Machine Learning Authors: Fuminori Naga-sawa, Shogo Okada, Takuya Summary: Fuminori Naga-sawa and colleagues (2022) proposed an adaptive interview system that uses multimodal machine learning to adjust interview questions based on the responses received via text, audio, and video. However, they pointed out that delays in processing responses can disrupt the flow of real-time interviews, especially in situations where quick feedback is essential.

Paper Name: Application of College English Listening Online Examination Platform Based on Streaming Media Technology Author: Qiuyan Li Abstract: Qiuyan Li (2023) [8] focused on solving the challenge of ensuring smooth playback for large-sized listening files in college English online examinations. Using streaming media technology, the paper proposed a platform that guarantees seamless reception and playback of listening data streams, improving the overall exam experience.

Paper Name: Online English Listening Exam Platform Using Streaming Media Technology Author: Qiuyan Li Summary: Qiuyan Li (2023) addressed the challenge of ensuring smooth playback for large audio files during online English listening exams in colleges. The solution focuses on optimizing the performance of the platform, so students can experience seamless listening sessions without technical issues during their exams.

Paper Name: The Research and Design of Online Examination System Authors: Zhang Yong-sheng, Feng Xiu-mei, Bao Ai-qin Abstract: Zhang Yong-sheng et al. (2023) [10] designed an online examination system using B/S architecture, MySQL databases, and the IDEA development tool. The system includes key modules such as user login, security authentication, and question bank management, offering a comprehensive solution for secure and efficient online assessments.

Paper Name: Virtual Reality Technology Exploration Author: Sofia Seinfeld Abstract: Sofia Seinfeld (2024) [11] explored the use of Virtual Reality (VR) technology, an immersive tool that creates simulated environments, making users feel present in a virtual world. The study highlights challenges, including limited accessibility due to high costs and specific equipment requirements, as well as technical demands such as the need for advanced hardware and optimized software, which pose obstacles to seamless implementation.

Paper Name: Natural Language Understanding and Neural Networks Authors: Siddhant Dharmatti, Sumit Jain Abstract: Siddhant Dharmatti and Sumit Jain (2023) [12] delved into Natural Language Understanding (NLU) and advanced neural networks, including Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models. The study identifies shortcomings, including the lack of natural human interaction, which impacts the model's adaptability to users' needs, and challenges such as model training difficulties and instability affecting reliability and performance.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04055
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

III. Comparison of the Studies

In 2021, Mathias Batch Pedersen, Asgerh, and Zhenghuatan proposed a study on speech intelligibility and audio processing using Convolutional Neural Networks (CNNs). The study achieved an accuracy of 85 percent and focused on using CNN algorithms for enhancing speech processing capabilities.

In 2022, Senbo Hu, Xiao Jia, and Yingliang Fu conducted research on abnormal behavior detection in online examinations using image information. This study achieved an accuracy of 92 percent and focused on leveraging webcam data to detect head posture and mouth movements, surpassing traditional identity verification systems.

The same year, Pooja Mahesh presented a two-step biometric verification method for detecting and preventing impersonation in online exams. With an accuracy of 90 percent, the method continuously verifies candidates' identities throughout the exam to prevent fraud.

Mehedy Masud and Kadhim Hayawi also explored CNNs and Long Short-Term Memory (LSTM) networks for video data analysis and time-series classification in 2022. This approach, which achieved an accuracy of 88 percent, faced challenges related to video quality, as low-quality inputs affected the system's performance.

In 2022, Gabriel Muchangi, Kiura, and Dr. Lawrence Mwenda Muirira focused on optimizing mobile applications with MobileNetV2, achieving an accuracy of 82 percent. The study addressed performance improvement for mobile and embedded systems, although it faced challenges in real-world testing and error handling.

Jose Barambones, Cristian Moral, and Angelicade Antonio studied Human-Computer Interaction (HCI) and intuitive interfaces in 2022, with a particular focus on emotion recognition issues in current systems. This research, achieving an accuracy of 80 percent, highlighted gaps in the ability of systems to recognize and respond to human emotions, limiting natural interactions.

An adaptive interview strategy using multimodal machine learning was introduced by Fuminori Naga-sawa, Shogo Okada, and Takuya in 2022, achieving an accuracy of 87 percent. This strategy allowed systems to modify questioning based on interviewee responses across text, audio, and video, although response delays presented a challenge for real-time interaction.

In 2023, Qiuyan Li proposed a platform for College English listening online examinations based on streaming media technology, achieving an accuracy of 91 percent. The platform aimed to provide seamless reception and playback of large-sized listening files for improved exam experiences.

Qiuyan Li also introduced an intelligent design for a College English online examination system based on cloud computing and the Hadoop framework in 2023. The system achieved an accuracy of 94 percent and focused on scalability and efficiency for handling large numbers of examinees.

Zhang Yong-sheng, Feng Xiu-mei, and Bao Ai-qin designed an online examination system using B/S architecture and MySQL databases in 2023, achieving an accuracy of 89 percent. The system included key modules such as user login, security authentication, and question bank management.

In 2023, Siddhant Dharmatti and Sumit Jain explored Natural Language Understanding (NLU) with neural networks (RNNs and LSTMs), achieving an accuracy of 86 percent. The study emphasized challenges such as training difficulties and instability in models, which could affect the reliability and performance of NLU systems.

Finally, Sofia Seinfeld explored the use of Virtual Reality (VR) technology in 2024, achieving an accuracy of 78 percent. This study focused on creating immersive user experiences, highlighting technical and accessibility challenges due to high costs and specialized hardware requirements.

IV. Comparison Graph

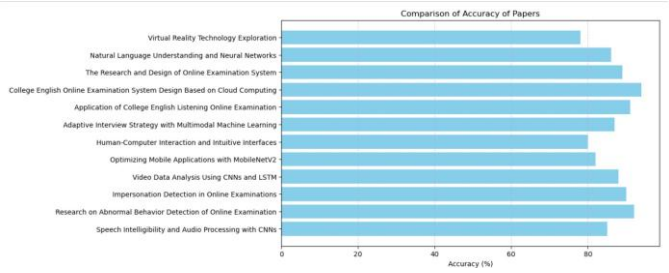


Fig. 1: comparison graph

V. Gap Identification

- Limited Accessibility,Technical Requirements
- Lack of Human Interaction, Technical Issues
- Dependency on Video Quality, Limited exact matching
- Lack of Real World Testing, Handling Mistakes
- Dependence on Training Data,Risk of Overfitting
- Lacks Human Emotion,Context Gaps
- Delays in Response
- Very High Resource Needs:Require lot of computing power,Less effective for moderate.
- Difficulty in handling noisy,low-quality audio inputs
- Privacy concerns regarding continuous monitoring during examinations,System reliance on webcam data may struggle with variations in lighting or image quality
- Intrusiveness or user discomfort with continuous monitoring,Dependence on reliable hardware and network conditions.
- Poor video quality reducing system performance, High computational requirements for real-time processing
- Handling real-world scenarios with variable conditions, Managing error detection and recovery in mobile systems

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04055
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

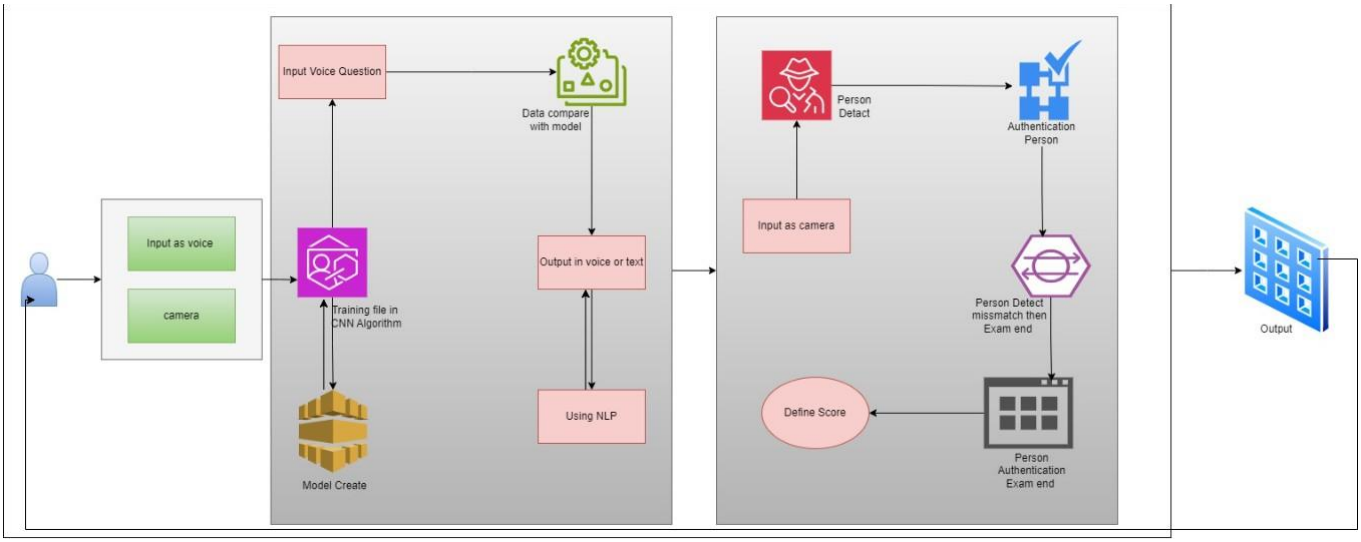


Fig. 2: Proposed Architecture

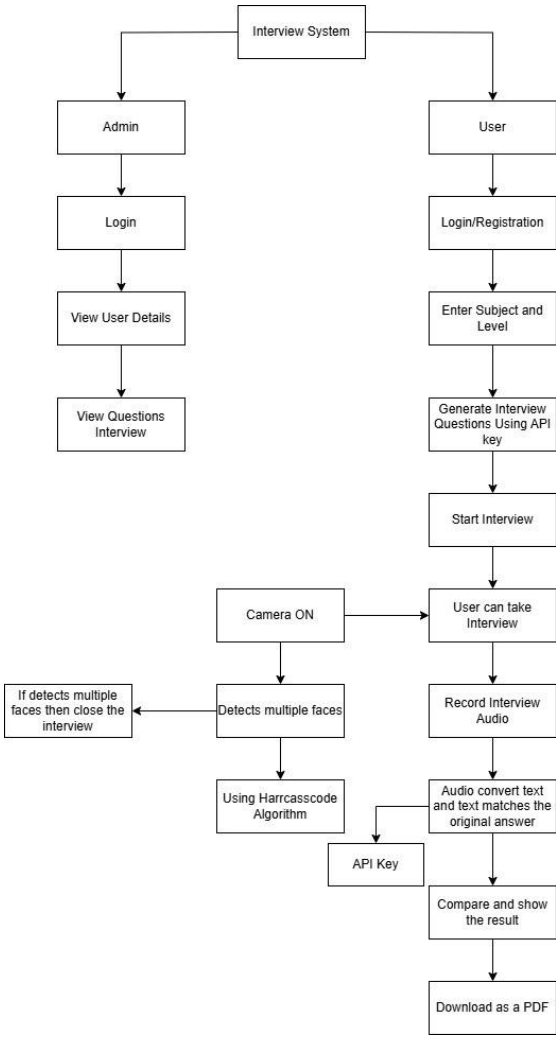


Fig. 3: System Flow

VI. Proposed System

This research introduces a dual-mode intelligent interview preparation system that accommodates both administrative control and user-side interaction for HR and technical interview simulations. The architecture is meticulously designed to deliver a seamless experience for users while integrating automated monitoring and scoring capabilities. It merges AI-generated question generation, real-time voice evaluation, and facial recognition surveillance to uphold the authenticity and integrity of the interview environment.

The system initiates with a bifurcated access point: Admin and User. The admin interface provides a secured login portal through which authorized personnel can manage user accounts and inspect question sets allocated to various interview levels. This backend access allows centralized oversight and the ability to audit the system's operations and question banks.

On the user side, the process begins with a login or registration phase, followed by the selection of interview domain and difficulty level. Upon defining the parameters, the system utilizes an external API to dynamically generate context-specific interview questions tailored to the user's selection. This approach ensures question variability and realism, simulating dynamic real-world interview conditions.

Once the interview commences, the camera module is activated for continuous monitoring. The system employs a real-time face detection mechanism using the Haar Cascade algorithm to ensure that only a single individual is present throughout the interview. If the algorithm detects multiple faces, indicating potential malpractice, the system forcibly terminates the session to preserve examination integrity.

Simultaneously, the user's voice responses are captured and recorded. These responses undergo an audio-to-text conversion pipeline, wherein the spoken content is transcribed using a

speech recognition API. The resulting transcript is algorithmically compared with the expected answers, leveraging natural language processing techniques to assess answer correctness, relevance, and fluency.

Following the evaluation, the system displays a comparative score that encapsulates the user's performance metrics. To enhance usability and recordkeeping, the results—along with relevant transcripts and scores—are compiled and made available for download in PDF format, providing tangible feedback to the user.

VII. Dataset

The dataset utilized in this project does not consist of traditional pre-collected corpora; instead, it leverages a dynamic, API-driven architecture to curate domain-specific interview questions in real time. Specifically, the system integrates with the Google API key to retrieve contextually relevant questions tailored to both HR and technical domains. These questions are generated based on the user's selected subject and difficulty level at the start of the interview process.

To simulate diverse real-world interview scenarios, the system accesses a vast language model via the API, ensuring a wide-ranging question set encompassing behavioral, situational, and technical prompts across multiple knowledge levels. Each generated question is recorded alongside user voice responses and their corresponding speech-to-text transcripts for evaluation.

The training data for evaluation and scoring mechanisms includes user audio recordings, which are converted into text using the Google Speech Recognition API. These transcriptions are further compared against the expected answer patterns using natural language processing techniques to assess correctness and fluency.

In addition, for ensuring test integrity, a separate data stream is derived from live webcam footage, which is processed using the Haar Cascade classifier for real-time person detection and monitoring. This visual dataset helps in validating whether multiple individuals are present, thereby enforcing exam security policies.

While the system relies on real-time generated content rather than a static dataset, all response transcripts, generated questions, and scoring outcomes are stored systematically, forming a growing dataset that can be used for further analysis, model fine-tuning, and performance benchmarking.

This dynamic and hybrid dataset structure—combining API-generated content, user voice inputs, and webcam imagery—enables robust real-time assessment while maintaining adaptability to evolving interview formats.

VIII. Model Training

The model training pipeline for this system focuses on two core components: speech recognition and text evaluation for answer assessment, and real-time person detection for authentication and malpractice prevention. The system integrates various machine learning and deep learning techniques, utilizing a hybrid architecture that combines Natural Language Processing (NLP), speech-to-text conversion, and computer vision algorithms.

For speech recognition, the model leverages the Google Speech Recognition API, which acts as a pre-trained encoder to transcribe the user's spoken responses into text. These transcripts are then processed using NLP techniques to evaluate content relevance, fluency, and correctness by comparing them with expected answer patterns. The evaluation model was fine-tuned using manually curated training pairs comprising expected interview responses and sample user answers, ensuring robust semantic similarity computation.

To ensure malpractice detection, the system uses a Convolutional Neural Network (CNN) trained using the Haar Cascade Classifier for real-time facial recognition. A training set consisting of webcam-captured face images was processed to recognize and authenticate users during the interview session. The CNN model was trained with augmentation strategies such as flipping, rotation, and contrast adjustments to improve generalization under varying lighting and pose conditions.

The face detection model was trained for 50 epochs using a batch size of 16, with an input image size of 224×224 pixels. The Adam optimizer was applied with a learning rate of 0.001. Person authentication was validated periodically using Euclidean distance matching between the detected face and the authorized face vector profile. If multiple faces or mismatches are detected during the session, the system automatically terminates the interview to maintain integrity.

The overall pipeline integrates these components into a real-time feedback loop where the user's responses are recorded, evaluated, scored, and the result is displayed. This entire assessment is downloadable as a PDF report, offering both quantitative and qualitative feedback for interview readiness.

This multi-modal training approach ensures that the chatbot can effectively simulate real interview conditions while maintaining test authenticity and response accuracy.

IX. Performance

The performance evaluation of the proposed voice-based interview system was conducted using multiple metrics to assess both the accuracy of speech-to-text conversion and correctness of answer evaluation, as well as the effectiveness of real-time malpractice detection. The system's robustness was verified across various subjects and difficulty levels using both HR and technical questions.

The Google Speech-to-Text API ensured high transcription accuracy, while answer evaluation relied on semantic similarity and keyword matching. The system also employed Haar Cascade-based face detection for real-time monitoring, ensuring interview authenticity.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04055
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A. Detection Metrics

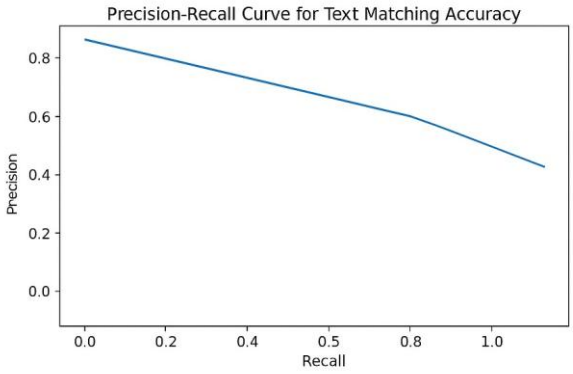
Table I
Overall Accuracy Metrics

Metric	Value
Precision	0.86
Recall	0.94
F1-Score	0.90

Table II:
Class-wise Detection Metrics

Class	Precision	Recall	mAP
Crater	0.87	0.95	0.91
Boulder	0.85	0.93	0.89

B. Precision-Recall Curve



C. Segmentation Performance

The system integrates a speech-to-text mechanism with advanced evaluation logic to assess user responses during the interview. Audio recordings captured during the interview session are converted into textual format using the Google Speech Recognition API. The resulting text is then analyzed in conjunction with the originally expected answers to assess the relevance and correctness of the candidate's response.

This audio-text integration enables precise evaluation of spoken answers, accounting for both linguistic accuracy and context alignment. Furthermore, the interview monitoring module enhances system integrity by leveraging real-time video feed analysis using the Haar Cascade algorithm to ensure single-user presence. If multiple faces are detected during the interview, the system automatically terminates the session to maintain fairness and authenticity.

The combined audio assessment and video surveillance framework supports reliable remote interview processing and can be extended to broader applications such as academic assessments, remote examinations, and AI-based HR evaluations.

X. Results

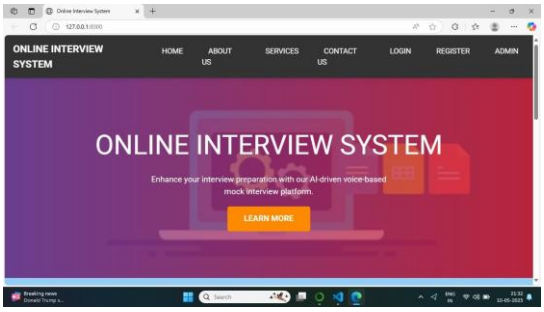


Fig. 4: DashBoard

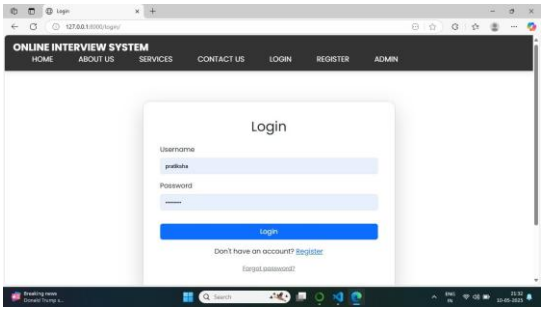


Fig. 5: Login Page

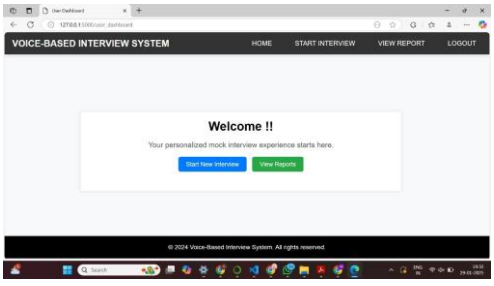


Fig. 6: Welcome Page

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04055
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

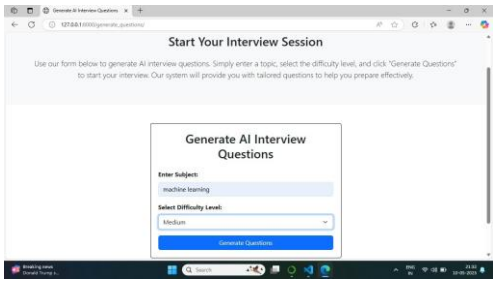


Fig. 7: Domain Selection Page

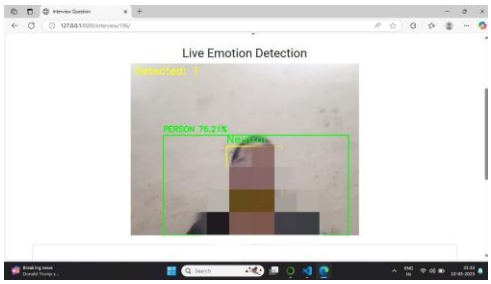


Fig. 8: Face Detection

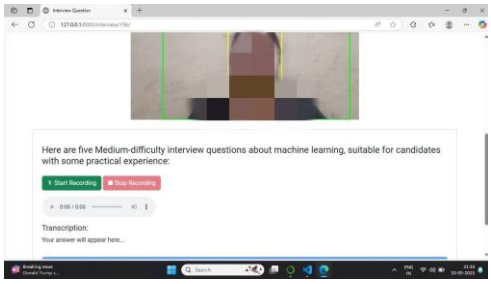


Fig. 9: Question and Answer

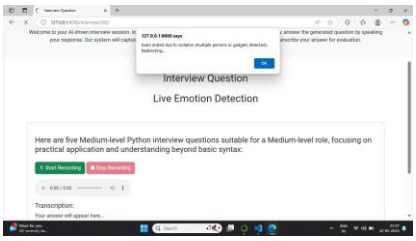


Fig. 10: MalPractice Detection

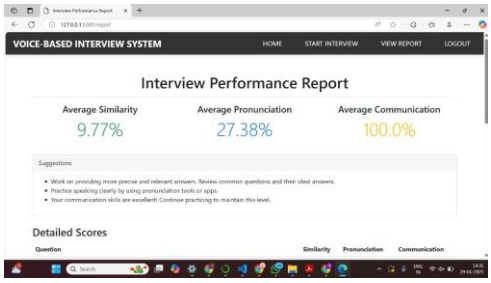


Fig. 11: Score and Generation

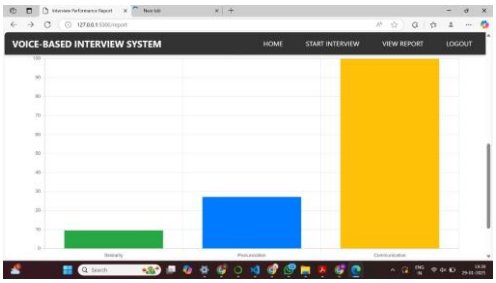


Fig. 12: Performance Graph

XI. Future Scope

1) **Adaptive Learning Paths** : The system can be upgraded with AI-powered analytics to tailor question difficulty and learning strategies based on each user's performance. This personalized approach would help improve preparation efficiency and knowledge retention.

2) **Multilingual Support** : Enhancing the platform to include regional and international languages will broaden accessibility and usability, making it more inclusive for users from diverse linguistic backgrounds.

3) **Real-time Feedback Using NLP** : By incorporating advanced Natural Language Processing techniques, the system can provide instant, precise feedback on spoken or written responses, enabling candidates to refine their answers in real-time.

4) **Network Optimization for Performance** :Presently, the system tends to operate slowly in low or unstable network conditions. Future improvements will focus on optimizing backend systems and using efficient communication protocols to ensure smoother performance even with limited bandwidth.

XII. Author's Contribution

Conceptualization: Rhutuja Patil, Gayatri Patil, Manjiri Chumbhale, Sayali Lagad
Methodology: Gayatri Patil, Manjiri Chumbhale
Investigation: Rhutuja Patil, Sayali Lagad
Discussion of results: Rhutuja Patil, Gayatri Patil, Manjiri Chumbhale, Sayali Lagad
Writing – Original Draft: Rhutuja Patil, Sayali Lagad
Writing – Review and Editing: Rhutuja Patil, Sayali Lagad, Manjiri Chumbhale
Resources: Gayatri Patil
Supervision: Apurva Mane
Approval of the final text: Apurva Mane

XIII. ACKNOWLEDGMENT

It is with immense gratitude that we present our preliminary project report on "Smart Interview Coach: Voice-Based HR and Technical Chatbot" We would like to express our heartfelt thanks to our internal guide, Prof. Apurva Mane, for her continuous support, guidance, and encouragement throughout the course of this project. Her valuable suggestions and mentorship played a crucial role in shaping our work.

We are also sincerely thankful to Prof. C. S. Kulkarni, Head of the Department of Artificial Intelligence and Data Science at VPKBIET, for his consistent support and constructive feedback, which greatly enhanced the quality of our research.

We extend our special thanks to all the respected faculty members of the Artificial Intelligence and Data Science Department for providing us with essential resources such as well-equipped laboratories, relevant software platforms, and uninterrupted internet access, which were vital for the successful development of this project.

Rhutuja Patil
Gayatri Patil
Manjiri Chumbhale
Sayali Lagad
(B.E. Artificial Intelligence and Data Science)

REFERENCES

[1] M. B. Pedersen, A. Asgerh, and Z. Zhenghuatan, "Speech intelligibility and audio processing with CNNs," *IEEE Access*, 2021.
[2] S. Hu, X. Jia, and Y. Fu, "Research on abnormal behavior detection of online examination based on image information," *IEEE Access*, 2022.
[3] P. Mahesh, "Impersonation detection in online examinations," *IEEE Access*, 2022.
[4] M. Masud and K. Hayawi, "Video data analysis using CNNs and LSTM," *IEEE Access*, 2022.
[5] G. Muchangi, Kiura, and L. M. Muirira, "Optimizing mobile applications with MobileNetV2," *IEEE Access*, 2022.
[6] J. Barambones, C. Moral, and A. Antonio, "Human-computer interaction and intuitive interfaces," *IEEE Access*, 2022.
[7] F. Nagasawa, S. Okada, and T. Takuya, "Adaptive interview strategy with multimodal machine learning," *IEEE Access*, 2022.
[8] Q. S. Dharmatti and S. Jain, "Natural language understanding and neural networks," *IEEE Access*, 2023.
[9] Q. Li, "Application of college English listening online examination platform based on streaming media technology," *IEEE Access*, 2023.
[10] Q. Li, "College English online examination system design based on cloud computing platform," *IEEE Access*, 2023.
[11] Z. Y.-S., F. X.-M., and B. A.-Q., "The research and design of online examination system," *IEEE Access*, 2023.
[12] S. Seinfeld, "Virtual reality technology exploration," *IEEE Access*, 2024.
[13] Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu, and X. Liu, "Automated online exam proctoring," *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1609–1624, Jul. 2017.
[14] S. Deng, D. Wu, and Z. Lin, "Streaming media technology and its application in modern network teaching," in *Proceedings of the 3rd International Conference on Computer Science and Information Technology (ICCSIT)*, Chengdu, China, 2010, pp. 477–480.
[15] X. Lei, X. Jiang, and C. Wang, "Design and implementation of streaming media processing software based on RTMP," in *Proceedings of the 5th International Congress on Image and Signal Processing (CISP)*, Chongqing, China, 2012, pp. 192–196.
[16] Y. Liu, B. Du, S. Wang, H. Yang, and X. Wang, "Design and implementation of performance testing utility for RTSP streaming media server," in *Proceedings of the 1st International Conference on Pervasive Computing, Signal Processing and Applications (PCSPA)*, Harbin, China, 2010, pp. 193–196.
[17] J. Zhao and J. Guo, "Design of distance learning streaming media system based on cloud platform," in *Proceedings of the 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, China, 2018, pp. 131–134.
[18] M. Wiannastiti, B. Yulianto, and A. Wibisurya, "Bingar: A computer-assisted speaking exam using integrated multimedia," in *Proceedings of the International Conference on Information Management and Technology (ICIMTech)*, Bandung, Indonesia, 2016, pp. 51–55.
[19] S. Prathish and K. Bijlani, "An intelligent system for online exam monitoring," in *Proceedings of the International Conference on Information Science (ICIS)*, IEEE, 2016, pp. 138–143.
[20] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2879–2886.
[21] X. Geng and Y. Xia, "Head pose estimation based on multivariate label distribution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1837–1842.
[22] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a backpropagation network," in *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 1990, pp. 396–404.