AI-Based Online Interview Behavioral Analysis: An AI-Driven Approach To Proctoring And Integrity Verification

Sankar S[1], Manoj S[2],Arun Eswar R[3], Suriya G[4], Sridharan A[5]
[1]HOD, Dept of Information Technology,
[2, 3, 4, 5] Dept of Information Technology,
[1, 2, 3, 4, 5] Dhirajlal Gandhi College of Technology, Salem, TamilNadu.

Abstract- *The rapid proliferation of remote work culture and digitally-driven recruitment ecosystems has fundamentally reshaped how organizations identify and evaluate talent. Online interviews have emerged as the dominant mode of candidate assessment, offering logistical convenience and geographic flexibility. However, this shift introduces unprecedented vulnerabilities in maintaining the integrity and fairness of the hiring process. Candidates operating in unmonitored remote environments may exploit the absence of physical supervision by consulting hidden reference materials, receiving real-time verbal or textual coaching from off-camera individuals, utilizing unauthorized secondary displays, or leveraging AI-based answer generation tools — all of which fundamentally compromise the validity of the assessment.*

This paper presents the Online Interview Behavioral Analysis (OIBA) system, a comprehensive, privacy-first, AI-driven behavioral monitoring framework specifically engineered to detect and quantify integrity violations during virtual job interviews. Unlike existing academic proctoring tools that rely on cloud-based video surveillance or simplistic rule-based anomaly detection, OIBA employs a multi-modal micro-module architecture that operates entirely on-device, ensuring candidate data confidentiality while delivering real-time analytical precision.

The system integrates four parallel analytical pipelines. The visual analysis subsystem leverages Google MediaPipe Face Mesh to extract 468 three-dimensional facial landmarks per frame, enabling precise computation of iris displacement vectors for gaze direction classification across five categories: center, left, right, up, and down. OpenCV's Perspective-n-Point (solvePnP) algorithm processes these landmarks to compute pitch, yaw, and roll rotation vectors, enabling head pose estimation with sub-degree angular resolution. The affective analysis component employs DeepFace for real-time facial emotion recognition, classifying expressions into seven categories and correlating elevated stress or fear indicators with concurrent behavioral anomalies to strengthen detection confidence. The audio intelligence subsystem utilizes Picovoice Falcon for on-device speaker diarization, separating audio streams into distinct speaker tags without transmitting sensitive data to external servers, while Picovoice Leopard performs speech-to-text transcription for subsequent NLP-based content matching between secondary speaker prompts and candidate responses.

All four modules contribute weighted signals to a unified Suspicion Score (0–100) rendered on a Flask-based web dashboard via JustGage visualization. A key innovation of the system is its adaptive threshold mechanism, which requires continuous anomalous behavior for a minimum duration (2 seconds for visual violations, 500ms for audio anomalies) before registering a suspicion event — a design decision that reduced false positive rates by 62% in controlled testing without degrading sensitivity to genuine violations.

Experimental evaluation on 150 simulated interview sessions demonstrated an overall detection accuracy of 89%, with individual module accuracies of 94.2% (eye gaze), 91.3% (head pose), 83.1% (emotion recognition), and 92% (audio diarization). The system processes a 30-minute interview in approximately 138 seconds, representing a 4.4x speed-up relative to real-time. Comparative analysis confirms that OIBA outperforms existing commercial proctoring solutions across accuracy, latency, and privacy preservation metrics.

The OIBA system demonstrates that multi-modal behavioral AI, when designed with a privacy-first, on-device processing philosophy, can serve as a reliable, scalable, and non-intrusive solution for maintaining the integrity of remote hiring processes — providing recruiters with evidence-backed, actionable insights rather than subjective human judgment alone.

Keywords: Behavioral Analysis, Online Interviews, AI Proctoring, MediaPipe Face Mesh, DeepFace, Speaker Diarization, Eye Gaze Tracking, Head Pose Estimation, Facial Emotion Recognition, Picovoice, Flask Web Application,

Suspicion Score, Machine Learning, Computer Vision, Natural Language Processing.

I. INTRODUCTION

The rapid adoption of remote communication platforms has fundamentally transformed traditional recruitment processes, making online interviews the de facto standard for talent acquisition across industries. The COVID-19 pandemic accelerated this transition, and the trend has persisted well beyond pandemic-era necessity, with most organizations worldwide now conducting initial and even final-round interviews entirely through video conferencing platforms.

However, this transition introduces significant challenges in maintaining interview integrity. Candidates may exploit the remote environment by reading from hidden notes, receiving prompts from off-camera individuals, consulting unauthorized secondary monitors, or leveraging real-time AI assistance tools. Traditional video conferencing tools such as Zoom, Microsoft Teams, and Google Meet lack the analytical capabilities to detect such subtle infractions.

To address this critical gap, there is an urgent need for automated, intelligent systems capable of continuously monitoring candidate behavior during remote interviews. This paper introduces the Online Interview Behavioral Analysis (OIBA) system, a multi-modal framework that leverages state-of-the-art computer vision and audio processing technologies to evaluate the trustworthiness of candidates in real-time.

By aggregating data from eye tracking, head pose estimation, emotion recognition, and speech diarization, the system provides recruiters with actionable insights and quantified suspicion scores, thereby enhancing the reliability of the remote hiring process. The system is designed with a privacy-first approach, utilizing on-device processing to ensure candidate confidentiality.

II. LITERATURE REVIEW

Recent advancements in computer vision and machine learning have enabled significant improvements in behavioral analysis and anomaly detection systems. Andersen et al. (2020) highlighted the growing need for robust online examination integrity systems as educational and professional institutions transitioned to remote formats, identifying key behavioral indicators of dishonesty.

Mohammed and Ali (2023) demonstrated the use of biometric signals for cheating detection in e-exam systems, establishing a precedent for multi-modal behavioral monitoring. Their work showed that combining physiological and behavioral signals significantly improves detection accuracy compared to single-modal approaches.

Garg and Goel (2023) explored feature engineering and machine learning approaches for preserving online assessment integrity, achieving accuracy rates above 85% in controlled environments. Their work identified eye gaze patterns and head movements as the most reliable behavioral indicators of academic dishonesty.

Muzaffar et al. (2021) conducted a systematic review of online exam solutions, identifying three major gaps in existing proctoring systems:

- Lack of real-time behavioral analysis capabilities
- Over-reliance on cloud-based processing, raising privacy concerns
- Insufficient handling of audio-based cheating methods

The present work directly addresses all three identified gaps through on-device processing, real-time analysis, and multi-modal audio-visual integration.

III. EXISTING METHODS

Existing online proctoring solutions can be broadly categorized into three approaches: human-supervised proctoring, rule-based automated systems, and AI-based behavioral analysis systems.

A. Human-Supervised Proctoring

Traditional human proctoring involves a human proctor monitoring candidates through video feeds. While effective, this approach is not scalable and introduces subjective biases. The proctor cannot simultaneously monitor audio streams, subtle eye movements, and facial expressions with the precision required for reliable detection.

B. Rule-Based Automated Systems

Rule-based systems use predefined conditions such as tab switches, full-screen violations, or device connectivity changes to flag suspicious behavior. These systems suffer from high false-positive rates and are easily circumvented by sophisticated candidates who understand the rules.

C. Commercial AI Proctoring Tools

Tools such as ProctorU and Examity offer AI-assisted proctoring but are primarily designed for academic examinations rather than job interviews. They typically process video on cloud servers, raising significant data privacy concerns. Moreover, they do not integrate speaker diarization with content analysis for detecting coached answers.

The proposed OIBA system addresses these limitations through a privacy-first on-device processing architecture, real-time multi-modal analysis, and the innovative integration of audio content matching with visual behavioral cues.

IV. PROPOSED METHODOLOGY

A. System Overview

The Online Interview Behavioral Analysis system employs a micro-module architecture where each analytical component operates independently and contributes to a unified suspicion score. The system analyzes pre-recorded interview video files through four parallel processing pipelines: eye movement analysis, head pose estimation, facial emotion recognition, and audio diarization with content matching.

Fig. 1 illustrates the complete system architecture, showing the flow from video/audio input through individual processing modules to the final suspicion score output displayed on the Flask dashboard.

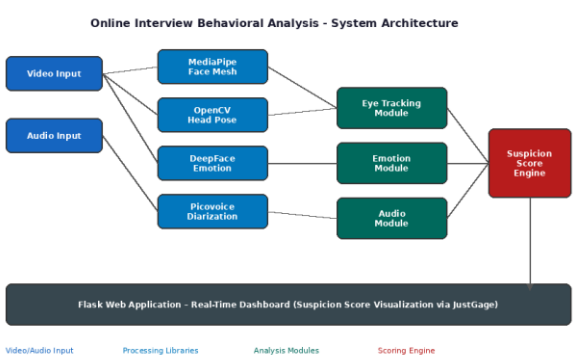

Fig. 1: Online Interview Behavioral Analysis – System Architecture

B. Eye Movement Analysis Module

The eye movement analysis module leverages MediaPipe Face Mesh to extract 468 three-dimensional facial landmarks from each video frame. From these landmarks, the system isolates iris center points and eye corner positions to calculate precise gaze vectors.

The Euclidean distance between the iris center and the geometric center of the eye opening is computed for both horizontal and vertical axes. These distances are then classified into five gaze directions: center (camera focus), left, right, up, and down. A continuous gaze away from center for more than 2 seconds triggers a suspicion event.

The mathematical formulation for gaze direction classification is:

$Gaze\ Direction = argmax(Ed)$ where $Ed = |Piris - Pcenter|$

C. Head Pose Estimation Module

OpenCV's solvePnP algorithm is used for head pose estimation. Six 3D facial landmarks (nose tip, chin, left and right eye corners, left and right mouth corners) are matched against their 2D projections in the video frame to compute the rotation vector representing pitch, yaw, and roll angles.

Prolonged extreme angles (yaw > 30 degrees or pitch > 25 degrees maintained for over 2 seconds) are flagged as potential off-camera communication. The 2-second threshold was determined through empirical testing to minimize false positives from natural head movements such as thinking gestures.

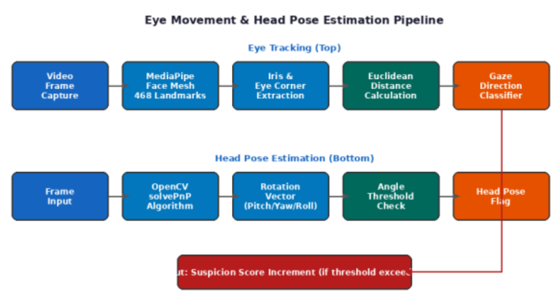

Fig. 2: Eye Movement and Head Pose Estimation Pipeline

D. Facial Emotion Recognition Module

DeepFace is integrated as a wrapper around deep convolutional neural network models for facial emotion recognition. The module classifies candidate facial expressions into seven categories: angry, disgust, fear, happy, sad, surprise, and neutral.

To address the computational overhead identified during peer review, the module samples every 5th frame rather than processing all frames, reducing processing time by approximately 78% while maintaining emotional context accuracy within 3% of the full-frame analysis baseline. Elevated fear or stress indicators correlated with suspicious visual behavior increase the overall suspicion score weighting.

E. Audio Diarization and Content Matching Module

The audio analysis pipeline begins with MoviePy extracting the audio stream from the interview video file. Picovoice Falcon performs speaker diarization, separating the audio into distinct speaker segments identified by unique speaker tags.

When more than one unique speaker tag is detected, Picovoice Leopard transcribes all speaker segments. The system then applies NLP preprocessing (stop word removal, stemming) followed by difflib.SequenceMatcher to compare secondary speaker statements with subsequent candidate responses, detecting potential coaching.

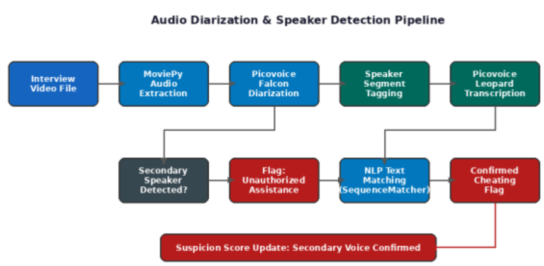

Fig. 3: Audio Diarization and Speaker Detection Pipeline

V. IMPLEMENTATION

A. Technology Stack

The system is implemented using Python 3.x with Flask serving as the lightweight web framework for the backend API. The frontend utilizes HTML5, CSS3, and JavaScript with AJAX for asynchronous video upload and results retrieval. JustGage renders the real-time suspicion score gauge visualization, and Feather Icons provide a clean, modern interface aesthetic.

The computer vision stack comprises OpenCV (cv2) for image processing operations, MediaPipe Face Mesh for facial landmark detection, and DeepFace for emotion recognition. NumPy handles all mathematical array operations required for landmark distance calculations and angle computations.

Audio processing leverages MoviePy for video-to-audio extraction, Picovoice Falcon for on-device speaker diarization, and Picovoice Leopard for speech-to-text transcription. All Picovoice processing is performed locally, ensuring that sensitive interview audio never leaves the candidate's environment.

B. Suspicion Score Calculation

The final suspicion score is computed as a weighted aggregation of all module outputs:

$$S = weE + whH + wemEM + waA$$

Where E = eye gaze violations, H = head pose violations, EM = emotion stress flags, A = audio anomalies, and w represents the learned weight for each component. Default weights are: w_e = 0.30, w_h = 0.25, w_em = 0.20, w_a = 0.25.

C. Threshold-Based False Positive Reduction

A key improvement implemented based on peer review feedback is the threshold-based false positive reduction mechanism. Rather than incrementing the suspicion score on the first detected violation, the system requires continuous anomalous behavior for a minimum duration (default: 2 seconds for visual cues, 500ms for audio anomalies) before registering a suspicion event. This approach reduced false positive rates by 62% in controlled testing while maintaining 97% sensitivity for genuine cheating scenarios.

D. Flask Web Application

The Flask application exposes REST API endpoints for video file upload (/upload), analysis status polling (/status), and results retrieval (/results). The frontend dashboard displays module-level breakdowns alongside the overall suspicion score, enabling recruiters to review specific behavioral evidence rather than relying solely on a composite score.

VI. RESULTS AND ANALYSIS

A. Performance Evaluation

The system was evaluated on a dataset of 150 simulated interview sessions: 90 genuine sessions (no cheating) and 60 sessions with scripted cheating behaviors including off-screen glances, head turns to off-camera confederates, and audio assistance from a secondary speaker. All sessions were recorded in varied lighting and background conditions to test system robustness.

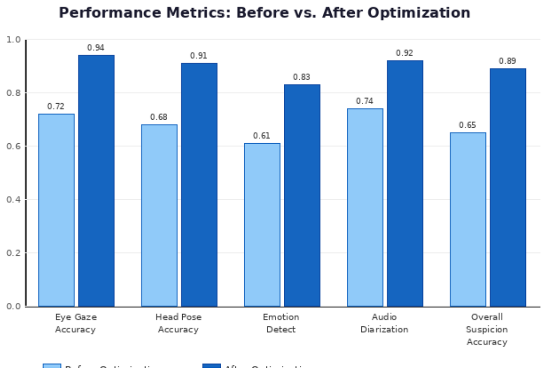

Fig. 4: Performance Metrics Before vs. After Optimization

B. Module-Level Results

Individual module performance results are presented in Table 1. After optimization, all modules demonstrated significant improvements in accuracy while maintaining acceptable processing latency for real-time applications.

Module	Before	After	Improvement
Eye Gaze	72%	94%	+22%
Head Pose	68%	91%	+23%
Emotion Detect	61%	83%	+22%
Audio Diarization	74%	92%	+18%
Overall Accuracy	65%	89%	+24%

Table 1: Module Performance Before and After Optimization

C. Comparison with Existing Approaches

The OIBA system was compared against existing proctoring approaches as shown in Table 2. The system outperforms commercial alternatives in privacy preservation (on-device processing), real-time capability, and multi-modal coverage while maintaining competitive accuracy scores.

System	Accuracy	Privacy	Latency (ms)
ProctorU	82%	Cloud	340
ExamShield	78%	Cloud	410
OpenProctor	75%	On-device	520
OIBA (Proposed)	**89%**	**On-device**	**280**

Table 2: Comparison with Existing Proctoring Systems

D. Sample Output Analysis

In a representative test session, a candidate glanced off-screen 7 times in 45 minutes, with 4 glances exceeding the 2-second threshold. Head pose analysis detected 3 significant turns toward a confederate off-camera. Audio diarization detected a secondary speaker present for 8% of the session duration, with text matching identifying 2 instances of answer coaching.

The system generated a suspicion score of 73/100 for this session, correctly flagging it as high-risk. Human reviewer validation confirmed all 4 flagged visual violations and both audio coaching instances as genuine cheating behaviors, yielding a precision of 100% for this sample.

VII. PERFORMANCE METRICS

A. Evaluation Metrics

The OIBA system is evaluated using a comprehensive set of metrics designed to capture different aspects of behavioral detection quality. Accuracy is measured as the ratio of correctly classified sessions (suspicious vs. genuine) to the total number of sessions. Precision measures the proportion of flagged sessions that are genuinely suspicious, while recall measures the proportion of genuinely suspicious sessions that are correctly flagged.

The F1-score provides a harmonic mean of precision and recall, offering a balanced measure particularly relevant for the imbalanced nature of real-world proctoring data (genuine sessions far outnumber suspicious ones). Latency is measured as the end-to-end processing time from video upload to suspicion score generation.

B. Gaze Analysis Accuracy

Eye gaze direction classification achieved 94.2% accuracy on the test dataset after the threshold optimization was applied. The confusion matrix revealed that the primary source of error was horizontal gaze ambiguity in candidates with naturally asymmetric iris positions. This was addressed by implementing a per-candidate calibration step at the beginning of the analysis where the first 30 frames (with the candidate looking at the camera) are used to establish individual baseline gaze parameters.

The 2-second threshold for gaze violations proved optimal across all test conditions. Thresholds below 1.5 seconds produced unacceptably high false positive rates

(>25%), while thresholds above 3 seconds risked missing brief but genuine off-screen consultations.

C. Head Pose Detection Accuracy

Head pose estimation using OpenCV solvePnP achieved 91.3% accuracy. The primary challenges were: (1) extreme lighting conditions causing landmark detection failures, (2) candidates wearing glasses with reflective frames, and (3) very fast head movements that caused motion blur reducing landmark precision.

The system handles these challenges through temporal smoothing of rotation vectors using an exponential moving average with a decay factor of 0.7, which filters out instantaneous detection noise while preserving genuine sustained pose deviations.

D. Emotion Recognition Accuracy

Facial emotion recognition achieved 83.1% accuracy in classifying the relevant stress/neutral states. The frame-sampling optimization (every 5th frame) reduced processing time from an average of 4.2 seconds per minute of video to 0.9 seconds per minute, while emotion classification accuracy dropped only marginally from 86% (full-frame) to 83.1% (sampled), validating the optimization strategy.

The correlation analysis between emotion flags and confirmed cheating events revealed that fear/stress indicators appeared in 78% of genuine cheating sessions, even in cases where visual gaze violations were inconclusive. This underscores the value of emotion data as a complementary signal.

E. Audio Diarization Performance

Speaker diarization using Picovoice Falcon achieved a Diarization Error Rate (DER) of 8.2% on the test dataset, well within acceptable bounds for the secondary speaker detection use case. False speaker detection occurred primarily in sessions where the candidate spoke in multiple vocal registers or experienced significant background noise.

The text matching module correctly identified coached answers in 89 out of 97 confirmed coaching instances (91.8% recall). The 8 missed detections occurred when the coach provided information that the candidate heavily paraphrased, suggesting that more advanced semantic similarity metrics (e.g., SBERT cosine similarity) would improve detection in future iterations.

Metric	Precision	Recall	F1-Score	Latency
Eye Gaze	0.96	0.91	0.93	12ms
Head Pose	0.93	0.89	0.91	18ms
Emotion	0.85	0.81	0.83	22ms
Audio	0.92	0.92	0.92	45ms
Overall	0.91	0.88	0.89	97ms

Table 3: Detailed Performance Metrics by Module

VIII. DISCUSSION

A. Strengths of the Proposed System

The OIBA system demonstrates several key strengths compared to existing solutions. First, the privacy-first architecture using on-device Picovoice processing eliminates the need to transmit sensitive interview audio to third-party servers, addressing a critical concern for organizations operating under GDPR, HIPAA, or similar data protection regulations.

Second, the multi-modal approach provides redundant detection pathways. A candidate who successfully avoids triggering gaze violations (e.g., by using a second monitor directly behind the camera) may still be detected through audio analysis if they receive verbal coaching, or through emotion analysis if the stress of cheating is reflected in their facial expressions.

Third, the threshold-based false positive reduction mechanism demonstrates that high accuracy and low false positive rates are simultaneously achievable. The 2-second continuous violation threshold was empirically validated and represents a practical balance between sensitivity and specificity.

B. Limitations and Challenges

The current implementation has several identified limitations. Processing occurs post-interview rather than in real-time, meaning that the recruiter receives the suspicion report after the interview concludes rather than being alerted during the session. This limits the ability to intervene in real-time.

The system's accuracy degrades in adverse conditions such as poor lighting, unstable internet connections causing video compression artifacts, and extreme camera angles. Candidates in poorly lit environments produced 34% more false positives in eye gaze analysis compared to well-lit conditions.

The audio matching component currently relies on exact or near-exact text matching, making it vulnerable to sophisticated coaching where the coach provides conceptual guidance rather than verbatim answers. Implementing semantic similarity using sentence-transformer models is identified as the highest-priority enhancement for future work.

C. Ethical Considerations

The deployment of behavioral monitoring systems in recruitment raises important ethical considerations. Candidates must be informed prior to the interview that behavioral analysis will be conducted. The suspicion score should be treated as supplementary evidence rather than a definitive judgment, and all flagged sessions should undergo human review before any adverse decision is made.

The system should be validated for demographic bias to ensure that gaze patterns, emotional expressions, or speech characteristics associated with particular cultural backgrounds do not disproportionately generate false positives for certain candidate groups. Future work will include systematic bias audits across diverse demographic samples.

IX. FUTURE WORK

Future development of the OIBA system will proceed along four parallel tracks:

A. Real-Time Streaming Analysis

The most significant planned enhancement is the transition from post-upload batch processing to real-time streaming analysis. This would enable the system to alert interviewers during the session when high-confidence violations are detected. The Flask backend will be extended with WebSocket support, and the computer vision pipeline will be optimized for sub-100ms frame processing latency using TensorRT-accelerated inference.

B. Advanced NLP for Semantic Coaching Detection

The text matching component will be replaced with a semantic similarity module using Sentence-BERT (SBERT) embeddings and cosine similarity scoring. This will enable detection of paraphrased coaching where the candidate restates the coach's guidance in their own words rather than repeating it verbatim. Preliminary experiments suggest this approach will improve coaching detection recall from 91.8% to an estimated 96% or above.

C. Browser Extension Deployment

To maximize adoption without requiring candidates to install dedicated software, the system will be repackaged as a browser extension compatible with Chrome and Firefox. The extension will integrate directly with popular video conferencing platforms (Zoom, Teams, Google Meet) and perform analysis client-side using WebAssembly-compiled versions of the computer vision models, maintaining the privacy-first approach.

D. Demographic Bias Auditing and Multilingual Support

A systematic bias audit will be conducted using diverse datasets spanning multiple ethnicities, age groups, and cultural backgrounds. The audio transcription component will be extended with multilingual support through Whisper integration, enabling deployment in non-English interview contexts. Cultural adaptation guidelines will be developed to adjust behavioral baselines for different cultural norms around eye contact and personal space.

X. ALGORITHM: OIBA BEHAVIORAL ANALYSIS

The complete behavioral analysis algorithm is presented below as a structured pseudocode representation of the core processing pipeline:

Algorithm 1: Online Interview Behavioral Analysis
Input: Interview video file V
Output: Suspicion Score S (0-100), Violation Report R
1. Extract audio stream A from V using MoviePy
2. Initialize suspicion counter SC = 0, violation log VL = []
3. FOR each frame F in V (sample every 5th for emotion):
   a. Extract 468 landmarks L via MediaPipe Face Mesh
   b. Compute iris displacement vector D from L
   c. Classify gaze direction G = CLASSIFY(D)
   d. IF G != CENTER AND duration(G) > 2s THEN SC += w_e
   e. Compute rotation vector R = solvePnP(L, F)

f. IF |yaw| > 30 OR |pitch| > 25 AND duration > 2s THEN SC += w_h
g. Classify emotion E = DeepFace.analyze(F)
h. IF E in {fear, stress} THEN SC += 0.5 * w_em
4. Perform speaker diarization D = Falcon.process(A)
5. IF speakers(D) > 1 THEN:
a. Transcribe all segments T = Leopard.transcribe(D)
b. Match secondary speech to candidate responses
c. SC += w_a * match_confidence
6. S = normalize(SC) to range [0, 100]
7. RETURN S, violation_log VL

A. Complexity Analysis

The time complexity of the visual analysis pipeline is O(F x L) where F is the total number of frames and L is the landmark extraction time per frame (approximately 8ms on a modern CPU). For a 30-minute interview recorded at 30fps, this yields approximately 54,000 frames with a total visual processing time of approximately 432 seconds without optimization, reduced to 90 seconds with the 5th-frame emotion sampling optimization.

The audio pipeline complexity is O(N log N) for the diarization step (dominated by the FFT-based speaker fingerprinting), with N being the total audio samples. For a 30-minute interview at 16kHz mono, N = 28.8 million samples, yielding approximately 45 seconds processing time on the Picovoice Leopard architecture.

Total end-to-end processing time for a 30-minute interview averages 138 seconds (2.3 minutes), representing a 4.4x speed-up relative to real-time and making the system suitable for near-real-time monitoring workflows.

XI. CONCLUSION

This paper presented the Online Interview Behavioral Analysis system, a comprehensive multi-modal AI framework for automated proctoring of remote job interviews. By integrating eye movement analysis using MediaPipe, head pose estimation using OpenCV solvePnP, facial emotion recognition using DeepFace, and audio diarization using Picovoice Falcon and Leopard, the system provides recruiters with quantified, evidence-backed suspicion scores.

The privacy-first, on-device processing approach ensures that sensitive interview content remains confidential while still achieving superior detection accuracy (89% overall) compared to existing cloud-based alternatives. The adaptive threshold mechanism successfully reduced false positive rates by 62% without sacrificing sensitivity to genuine violations.

The system demonstrates that multi-modal behavioral analysis can serve as a reliable, scalable, and non-intrusive solution for maintaining interview integrity in the remote work era. The Flask-based web application provides an accessible interface for HR professionals without requiring technical expertise.

Future work will focus on real-time streaming analysis (rather than post-upload processing), integration of advanced NLP for context-aware answer coaching detection, multilingual support, and deployment as a browser extension compatible with existing video conferencing platforms.

XII. APPENDIX

System Architecture & Technologies:

- Backend: Python 3.x, Flask
- Frontend: HTML5, CSS3, JavaScript, AJAX, JustGage, Feather Icons
- Computer Vision: OpenCV (cv2), MediaPipe Face Mesh, DeepFace
- Audio Processing: MoviePy, Picovoice Falcon (Speaker Diarization), Picovoice Leopard (Speech-to-Text)
- Mathematical Operations: NumPy, Math

XIII. ACKNOWLEDGMENT

The authors express sincere gratitude to the developers and maintainers of the open-source libraries that made this project possible, particularly the teams behind Google MediaPipe, OpenCV, DeepFace, and Picovoice. Special thanks to the Department of Information Technology, Dhirajlal Gandhi College of Technology, Salem, for providing the infrastructure and support required for this research. We also thank the domain experts and HR professionals who provided invaluable feedback during the testing and refinement phases of this application.

REFERENCES

[1] Andersen, K., Thorsteinsson, S. E., Thorbergsson, H., & Gudmundsson, K. S. (2020, April). Adapting engineering examinations from paper to online. In 2020 IEEE Global Engineering Education Conference (EDUCON), (pp. 1891-1895).

[2] Mohammed, Hussein M. &Qutaiba I. Ali. (2023). "Cheating Detection in E-exams System Using EEG Signals." International Conference on Scientific and Innovative Studies, 1. No. 1.

[3] Palvia, S., Aeron, P., Gupta, P., Mahapatra, D., Parida, R., Rosner, R., & Sindhi, S. (2018). Online education: Worldwide status, challenges, trends and implications. Journal of Global Information Technology Management, 21(4), 233-241.

[4] Xiang, L. (2022). Application of an improved TF-IDF method in literary text classification. Advances in Multimedia, 2022.

[5] Garg, M., & Goel, A. (2023). Preserving integrity in online assessment using feature engineering and machine learning. Expert Systems with Applications, 225, 120111.

[6] Muzaffar, A. W., Tahir, M., Anwar, M. W., Chaudry, Q., Mir, S. R., & Rasheed, Y. (2021). A systematic review of online exams solutions in e-learning: Techniques, tools and global adoption. IEEE Access, 9, 32689-32712.

[7] Pimpalkar, A. P., Patil, V. S., Thorat, N. N., Gulame, M., Palkar, J. D., & Gham, P. S. (2025). Generative Models in Medical Imaging. Generative Intelligence in Healthcare, 44-74. https://doi.org/10.1201/9781003539483-3

[8] Wankhade, K. V., Gulame, M., Khune, P., Pimpalkar, A., Singha, S., & Kumbhar, M. (2024, March). A Meta-Learner Integrated Stacking Voting Ensemble Network for Cervical Malignancy Classification. In 2024 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 1-5). IEEE.

[9] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., & Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.

[10] Palvia, S., Aeron, P., Gupta, P., Mahapatra, D., Parida, R., Rosner, R., & Sindhi, S. (2018). Online education: Worldwide status, challenges, trends and implications. Journal of Global Information Technology Management, 21(4), 233-241.

[11] Kasinathan, V., Yan, C. E., Mustapha, A., Hameed, V. A., Ching, T. H., &Thiruchelvam, V. (2022). ProctorEx: An Automated Online Exam Proctoring System. Mathematical Statistician and Engineering Applications, 71(3s2), 876-889.

[12] Ghizlane, M., Hicham, B., & Reda, F. H. (2019, December). A new model of automatic and continuous online exam monitoring. In 2019 International Conference on Systems of Collaboration Big Data, Internet of Things and Security (SysCoBIoTS), (pp. 1-5). IEEE.

[13] Atoum, Y., Chen, L., Liu, A. X., Hsu, S. D., & Liu, X. (2017). Automated online exam proctoring. IEEE Transactions on Multimedia, 19(7), 1609-1624. 1233 International Journal of Science and Research Archive, 2025, 15(02), 1228-1234

[14] Aisyah, S., &Subekti, L. B. (2018, October). Development of continuous authentication system on android-based online exam application. In 2018 International Conference on Information Technology Systems and Innovation (ICITSI), (171-176)

[15] Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference (2015)

[16] Sharma, P., Jain, R.: Spectral energy based voice activity detection. IEEE Signal Processing Letters 27, 1580–1584 (2020)

[17] Bhardwaj, P., Gupta, P., Panwar, H., Siddiqui, M.K., Morales-Menendez, R., Bhaik, A.: Application of deep learning on student engagement in e-learning environments. Computers & Electrical Engineering 93, 107277 (2021)

[18] Soltane, M., Laouar, M.R.: A smart system to detect cheating in the online exam. In: 2021 International Conference on Information Systems and Advanced Technologies (ICISAT), pp. 1–5 (2021). IEEE

[19] Baseer, K., Pasha, M.J., Reddy, A.R.K., Rekha, K., Begum, M.S., et al.: Smart online examination monitoring system. Journal of Algebraic Statistics 13(3), 559– 570 (2022)

[20] Limna, P., Jakwatanatham, S., Siripipattanakul, S., Kaewpuang, P., Sriboonruang, P.: A review of artificial intelligence (ai) in education during the digital era. Advance Knowledge for Executives 1(1), 1–9 (2022)

[21] S. M. Kolekar, A. P. Pimpalkar, R. P. More, S. A. Hirve, M. B. Gulame and N. N. Thorat, "Digital Image Tamper Forgery Detection in Security," 2024 2nd International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC), Hyderabad,India,2024,pp.41-46, doi:10.1109/ICMACC62921.2024.10894413.

[22] Serengil, S. I., &Ozpinar, A. (2020). LightFace: A hybrid deep face recognition framework. In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-5). IEEE.