An Integrated AI System for Exam Proctoring Using YOLOv8, Face Recognition, and Noise Detection

[1]CH. NEVEEN, [2]S. RAMA RAO, [3]N. KRUPARANI

[2] Professor, Department of CSE, MAM Women's Engineering College, Narasaraopet, Palnadu,A.P., India.

[1,3] Asst. Prof., Department of CSE, MAM Women's Engineering College, Narasaraopet, Palnadu,A.P., India.

Abstract-

Online examinations have become increasingly prevalent, demanding robust proctoring solutions to ensure integrity and prevent unauthorized activities. This paper presents an integrated AI-based proctoring system that combines real-time face recognition, object detection using YOLOv8, and audio anomaly detection to enhance the reliability and fairness of online assessments. The proposed system authenticates candidates through webcam-based face recognition using facial encodings, preventing impersonation. Once verified, the system continuously monitors the candidate throughout the exam session.

A YOLOv8 deep learning model is employed to detect the presence of unauthorized objects such as mobile phones, laptops, or multiple persons in the webcam frame. Additionally, Haar cascades are used to track head and eye direction, flagging suspicious behavior such as looking away from the screen. Simultaneously, an audio detection module monitors environmental noise levels to detect conversations or abnormal sounds, triggering real-time alerts.

Detected anomalies activate a visual warning system, sound an alarm, and log the event through automatic screenshot capture and video recording. All sessions and alerts are stored for later review. This multimodal approach enhances exam security and provides scalable, automated proctoring suitable for remote education environments.

I. INTRODUCTION

The rapid adoption of online learning platforms and digital education tools has transformed the way assessments are conducted globally. While virtual classrooms and online examinations offer flexibility and accessibility, they also pose significant challenges in maintaining academic integrity and preventing malpractice during assessments [1], [2]. Traditional human proctoring methods are not only labor-intensive and expensive but also impractical for large-scale or remote testing environments [3]. As a result, there is a growing demand for intelligent, automated proctoring systems capable of real-time monitoring and anomaly detection.

Recent advancements in artificial intelligence (AI), computer vision, and deep learning have opened new possibilities for enhancing online exam security. Techniques such as face recognition, object detection, and activity monitoring are being explored to identify unauthorized activities, including impersonation, the use of mobile phones or laptops, presence of multiple individuals, and distractions during exams [4]–[7]. Various studies have proposed using Convolutional Neural Networks (CNNs), LSTMs, and transformers for recognizing human behavior and suspicious activities in surveillance footage [8]–[10]. However, many of these systems rely on single-modal detection, which limits their effectiveness in complex real-world scenarios.

This research presents an integrated AI-based online exam proctoring framework that combines three core components: YOLOv8 for object detection, face recognition for student verification, and audio analysis for abnormal sound detection. The proposed system performs real-time face matching using pre-registered facial encodings, continuously monitors for forbidden objects like mobile phones or extra persons using YOLOv8, and detects suspicious noise using decibel thresholding techniques. Detected anomalies trigger alarms, capture screenshots, and save video recordings for post-exam review.

By incorporating multi-modal monitoring—visual, audio, and behavioral—this system significantly enhances the reliability and robustness of remote proctoring. The face and eye orientation tracking further adds a behavioral layer to detect inattentiveness or attempts to look away from the screen [11], [12]. The proposed solution is scalable, cost-effective, and easily deployable using basic hardware (webcam and microphone) and open-source frameworks.

In summary, this work addresses key limitations in existing proctoring systems by providing a holistic, AI-driven approach that enhances exam fairness and integrity. It aligns with recent research emphasizing the need for intelligent surveillance in education [13]–[15], while offering a practical and deployable solution for academic institutions, certification bodies, and online learning platforms.

II. LITEARTURE SURVEY

In the era of digital education, maintaining academic honesty during online assessments has become a serious concern. To mitigate these issues, researchers have leveraged artificial intelligence (AI), computer vision, and deep learning techniques to automate proctoring and surveillance systems. Numerous works in recent years have focused on integrating object detection, face recognition, behavioral monitoring, and acoustic analysis to detect suspicious activities in real time.

Ding et al. [1] introduced a hybrid approach that combines Vision Transformers with sequence learning for human activity recognition in surveillance videos. This method improved temporal feature extraction, which is essential for detecting prolonged suspicious behavior. Similarly, Selvi et al. [2] proposed an enhanced CNN-based model for recognizing suspicious actions from surveillance feeds, showcasing its effectiveness in real-world security applications. Kamthe and Patil [3] also addressed suspicious activity detection using a combination of motion and appearance-based features for surveillance video analytics.

In the domain of online examinations, Duhaim et al. [4] presented a data mining approach for cheating detection during the COVID-19 pandemic, using behavioral logs and patterns to flag anomalies. Fan et al. [5] explored gesture-based misbehavior detection to identify unnatural hand movements during online assessments, showing how physical cues can be indicative of dishonest behavior. Singh et al. [6] developed a deep learning-based human activity recognition system using convolutional neural networks (CNNs), highlighting its accuracy in real-time environments.

Phyo et al. [7] extended this concept by using skeletal joint tracking with deep learning models to classify human activity, which is highly relevant for detecting unnatural posture or head/eye movements during exams. Mudgal et al. [8] introduced a suspicious action detection method using action attribute modeling, suitable for smart surveillance in restricted areas like examination centers.

Dhulekar et al. [9] focused on motion estimation for surveillance and demonstrated its role in identifying unauthorized movement. Ojha and Sakhare [10] surveyed object tracking techniques in video surveillance, analyzing methods like Kalman filters and mean-shift algorithms, and concluded that combining tracking with object recognition enhances system performance.

Hristov [11] employed a 1D-CNN and LSTM architecture for abnormal activity detection using 3D skeleton data, offering a fine-grained view of body movements. Hristov et al. [12] also proposed a deep learning and SVM hybrid model for human activity recognition, demonstrating improved accuracy in detecting complex actions. Khattar et al. [13] compared several deep learning techniques and emphasized that combining CNNs with RNNs leads to higher reliability in human activity classification tasks.

F and Singh [14] presented a computer vision-based survey on human activity recognition (HAR) systems, challenges, and real-world applications, indicating the increasing demand for AI in smart surveillance. Patel and Shah [15] analyzed supervised learning algorithms in ambient-assisted environments, highlighting their relevance in human-centered activity monitoring systems such as remote healthcare and smart exams.

Collectively, these studies demonstrate the efficacy of AI-powered surveillance systems across various domains. However, many solutions remain single-modal—focused solely on visual cues, face recognition, or sound analysis. The need for a multi-modal, integrated system that simultaneously addresses object detection, identity verification, and acoustic anomaly detection forms the foundation for the system proposed in this research.

III. METHODOLOGY

The proposed system integrates multiple AI-powered components to achieve intelligent and secure online exam monitoring. The overall methodology is structured into five major phases: user registration and verification, object detection, behavioral monitoring, audio anomaly detection, and alert generation with evidence logging. Figure 1 illustrates the high-level system architecture.

User Registration and Face Encoding

In the initial phase, students register by capturing their face through a webcam. The system uses the face_recognition library to detect and extract facial encodings. These encodings are serialized using pickle and stored in a Django-backed database along with student metadata (name, roll number, gender). To ensure accuracy, the system validates that only one face is visible during registration to prevent multi-face registration errors.

Face Verification Before Exam Start

Before starting the exam, the system performs identity verification by comparing the real-time webcam feed's face encoding with stored encodings using face_recognition.compare_faces. If a match is found with acceptable tolerance, the exam session is allowed to proceed; otherwise, it is terminated.

Real-Time Object Detection Using YOLOv8

During the exam, object detection is carried out in real time using the YOLOv8 model from the Ultralytics framework. The model detects specific unauthorized objects such as:

- Cell phones
- Laptops
- Multiple persons

When such objects are detected in the camera feed, bounding boxes are drawn, and the label and confidence scores are shown. The presence of these objects triggers screenshot capture and alerts.

Behavioral Monitoring (Head and Eye Movement)

The system uses Haar cascade classifiers to detect head orientation and eye movement. The x-coordinate of detected features is compared against the center of the frame to determine whether the student is looking away or showing signs of distraction. If excessive deviation is found, a warning is displayed (“look properly” or “look at me”) and treated as a potential anomaly.

Noise Detection Using Audio Monitoring

A separate thread monitors background noise using the sounddevice library. The audio amplitude is measured in real-time. If the volume exceeds a predefined threshold (e.g., 5 dB), the system identifies it as an abnormal sound, triggers an alert, and saves a screenshot with a “noise” label.

Alert Generation and Logging

When any abnormal activity is detected (unauthorized object, noise, eye/head deviation, multiple persons), the system:

- Plays a platform-specific alarm sound
- Saves a timestamped screenshot
- Stores video recordings of the full session
- Logs alerts for administrator review

This multi-modal detection ensures a robust and tamper-proof online exam monitoring experience.

IV. SYSTEM ARCHITECTURE

The system architecture is presented in fig.1.

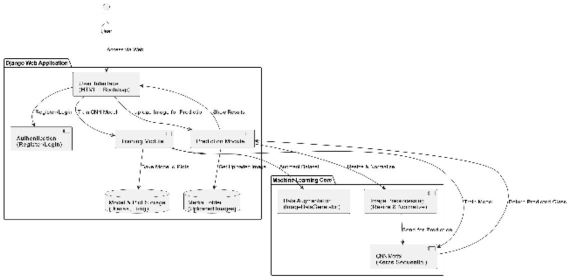

User
Access via Web
Django Web Application
User Interface (HTML + Bootstrap)
Register/Login
Train CNN Model
Upload Image for Prediction
Show Results
Authentication (Register/Login)
Training Module
Prediction Module
Save Model & Plots
Get Uploaded Image
Augment Dataset
Resize & Normalize
Model & Plot Storage (.keras, .png)
Media Folder (Uploaded Images)
Machine Learning Core
Data Augmentation (ImageDataGenerator)
Image Preprocessing (Resize & Normalize)
Train Model
Return Predicted Class
Send for Prediction
CNN Model (Keras Sequential)

Fig.1. System architecture

The system architecture of the proposed AI-based online exam proctoring solution integrates multiple components to ensure secure, automated supervision of remote assessments. On the client side, the student interacts with a web interface that captures real-time webcam and microphone input. The webcam feed is used for both face recognition and continuous monitoring, while the microphone captures audio signals for abnormal noise detection. These inputs are sent to the Django-based backend server, where several modules work in coordination. The Face Recognition Module extracts facial encodings and compares them against stored entries in the Face Encoding Database to authenticate the student before allowing access to the exam. Once verified, the Start Exam module initiates real-time surveillance using three parallel systems: YOLOv8-based object detection, behavior analysis using head and eye tracking, and an audio anomaly detection module. The YOLOv8 Object Detection model identifies unauthorized items like cell phones, laptops, or the presence of multiple people, while the Behavior Monitoring Module flags suspicious activity like frequent gaze shifts or head turns. Simultaneously, the Noise Detection Module triggers alerts if ambient noise exceeds a threshold, suggesting possible verbal communication or disturbances. All abnormal events are handled by the Alert Handler, which activates alarms, captures screenshots, and logs recordings for later review. These are stored in dedicated Screenshot and Video Storage modules. Finally, the admin accesses the Admin Dashboard to review user registrations, screenshots, and video logs, enabling manual validation and post-exam inspection. This architecture ensures a multi-modal, AI-driven framework for effective remote exam invigilation.

V. IMPLEMENTATION

The implementation of the proposed system, *ExamGuard*, integrates cutting-edge AI technologies to create a reliable and intelligent online exam proctoring solution. It comprises multiple modules including user registration, face recognition-based login, AI-powered monitoring, violation detection, and result-based termination. Each component is developed using Python, Django, OpenCV, face recognition APIs, YOLOv8 for object detection, and sound analysis libraries.

User Registration Module

The implementation begins with the user registration interface, where candidates input their personal details such as name, roll number, and gender, and capture their facial image using the webcam. This image is encoded into a numerical representation using the face_recognition library and securely stored in the database for future verification.

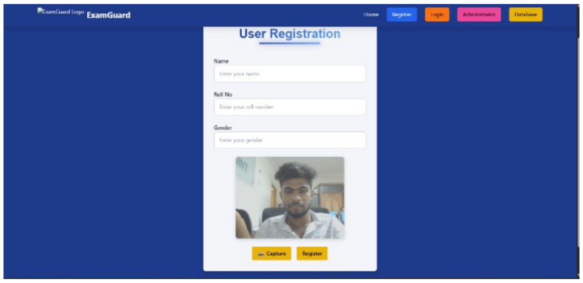
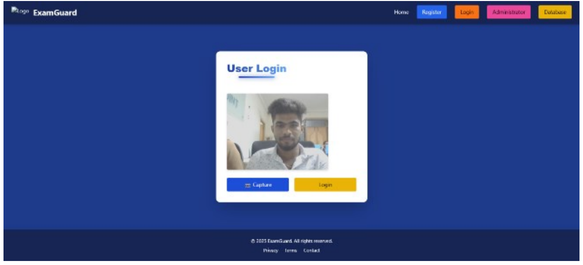


Fig. 2 – User Registration Page

This figure shows the registration interface where the candidate enters information and captures a facial image using the webcam.

User Login and Authentication

The login module captures a fresh webcam image of the user and compares it against the previously stored face encoding to verify identity. Only upon successful facial match does the system permit the user to access the examination. This ensures the authenticity of the examinee and prevents impersonation.



© 2025 ExamGuard. All rights reserved.

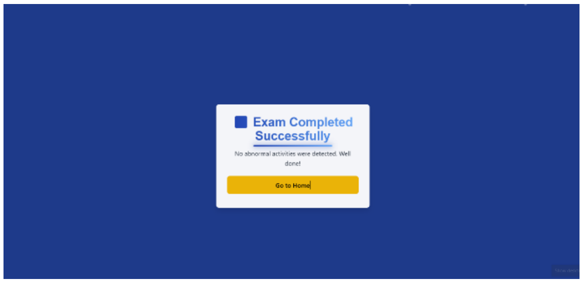
Fig. 3 – User Login with Webcam Face Verification
This screen captures a new image and compares it with stored face data for biometric authentication.

AI-Based Proctoring and Monitoring

During the exam, the system activates multiple monitoring components in parallel. The YOLOv8 model continuously detects unauthorized objects such as mobile phones or extra persons in the frame. Simultaneously, Haar cascades ensure that the user's face and eyes remain visible and focused on the screen. An audio stream listener checks for surrounding noise levels to catch possible verbal cheating.

Exam Completion

If the user adheres to all exam rules, the exam is marked as completed successfully and the system displays a confirmation message.

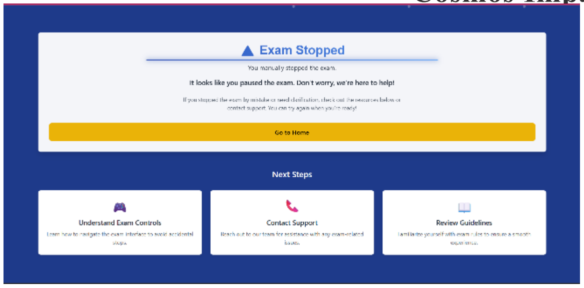


No abnormal activities were detected. Well done!

Fig. 4 – Exam Completed Successfully Screen
This figure confirms that the exam session concluded without any violations or manual interruptions.

Manual Termination by User or Admin

If the exam is ended manually by either the user or the administrator (due to emergencies or technical errors), a different screen is shown to indicate the termination.



▲ Exam Stopped

You manually stopped the exam.

It looks like you paused the exam. Don't worry, we're here to help!

If you stopped the exam by mistake or need clarification, check out the resources below or contact support. You can try again when you're ready!

Go to Home

Next Steps

Understand Exam Controls
Learn how to navigate the exam interface to avoid accidental stops.

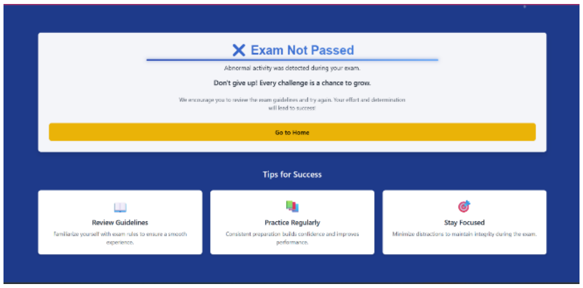
Contact Support
Reach out to our team for assistance with any exam-related issues.

Review Guidelines
Familiarize yourself with exam rules to ensure a smooth experience.

Fig. 5 – Exam Stopped Manually Screen
This image shows a system status update when the exam is halted intentionally before completion.

Violation Detection and Failure

In cases where the system detects a violation—like multiple faces, unauthorized devices, or loud noises—the exam is automatically stopped. The system alerts the user with a "not passed" status and saves the evidence for administrative review.



✕ Exam Not Passed

Abnormal activity was detected during your exam.

Don't give up! Every challenge is a chance to grow.

We encourage you to review the exam guidelines and try again. Your effort and determination will lead to success!

Go to Home

Tips for Success

Review Guidelines
Familiarize yourself with exam rules to ensure a smooth experience.

Practice Regularly
Consistent preparation builds confidence and improves performance.

Stay Focused
Minimize distractions to maintain integrity during the exam.

Fig. 6 – Exam Not Passed Due to Rule Violation
This screen notifies the user that the exam was terminated because of suspicious activity or detected anomalies.

Data Storage and Backend Integration

All captured data—including screenshots, video recordings, and face encodings—are stored on the server. Screenshots are timestamped and saved under the static/screenshots/ directory, while full session recordings go to static/recordings/. Face encodings are stored in binary format using pickle, enabling quick authentication during login.

The backend is handled using Django's ORM, and threading is employed to run real-time tasks (video, audio, YOLO, etc.) simultaneously without performance lags.

VI. CONCLUSION

The proposed AI-driven exam proctoring system offers a robust and automated solution to ensure integrity in online assessments. By combining **YOLOv8-based object detection**, **face recognition authentication**, and **audio-based anomaly detection**, the system effectively identifies unauthorized behaviors such as multiple persons, use of mobile phones or laptops, and suspicious noise during the examination process.

The implementation demonstrates real-time monitoring capabilities, with automatic screenshot capture and video recording during detected violations. The use of Haar cascades further enhances the system's precision in tracking eye and head movement, discouraging off-screen glances and ensuring user attention remains focused on the test interface.

This project overcomes the limitations of traditional manual invigilation, especially in remote or large-scale online exam environments. Its modular architecture allows for future enhancements such as gaze tracking, browser lockdown, and integration with learning management systems (LMS). Moreover, the system is lightweight and deployable using open-source tools, making it accessible for educational institutions with limited resources.

In conclusion, this integrated AI-based proctoring tool promotes fairness and transparency in online examinations, providing both scalability and reliability for modern digital education systems.

REFERENCES

[1] B. Y. Ding, A. Hussain, T. Hussain, W. Ullah, and S. W. Baik, "Vision Transformer and Deep Sequence Learning for Human Activity Recognition in Surveillance Videos," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 3454167, 2022.[Online].Available: https://doi.org/10.1155/2022/3454167

[2] E. Selvi et al., "Suspicious Actions Detection System Using Enhanced CNN and Surveillance Video," *Electronics*, vol. 11, no. 24, Article ID 4210, 2022.[Online].Available: https://doi.org/10.3390/electronics11244210

[3] U. M. Kamthe and C. G. Patil, "Suspicious Activity Recognition in Video Surveillance System," in *Proc. 2018 Fourth Int. Conf. Comput. Commun. Control Automat. (ICCUBEA)*, 2018.

[4] A. M. Duhaim, S. O. Al-Mamory, and M. S. Mahdi, "Cheating Detection in Online Exams during Covid-19 Pandemic Using Data Mining Techniques," *Webology*, vol. 19, no. 3, 2022.

[5] Z. Fan, J. Xu, W. Liu, and W. Cheng, "Gesture-based misbehavior detection in online examination," in *Proc. 2016 11th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Nagoya, Japan, 2016, pp. 234–238, doi: 10.1109/ICCSE.2016.7581586.

[6] P. Singh, I. Jindal, P. Panwar, H. Sirohi, and P. Kaushik, "Human Activity Recognition Using Deep Learning," in *Proc. 2022 1st Int. Conf. Informatics (ICI)*, Noida, India, 2022, pp. 75–79, doi: 10.1109/ICI53355.2022.9786890.

[7] C. N. Phyo, T. T. Zin, and P. Tin, "Deep Learning for Recognizing Human Activities Using Motions of Skeletal Joints," *IEEE Trans. Consum. Electron.*, vol. 65, no. 2, pp. 243–252, May 2019, doi: 10.1109/TCE.2019.2908986.

[8] M. Mudgal, D. Punj, and A. Pillai, "Suspicious Action Detection in Intelligent Surveillance System Using Action Attribute Modelling," *J. Web Eng.*, vol. 20, no. 1, pp. 129–146, Jan. 2021, doi: 10.13052/jwe1540-9589.2017.

[9] P. A. Dhulekar, S. T. Gandhe, A. Shewale, S. Sonawane, and V. Yelmame, "Motion estimation for human activity surveillance," in *Proc. 2017 Int. Conf. Emerg. Trends Innov. ICT (ICEI)*, Pune, India, 2017, pp. 82–85, doi: 10.1109/ETIICT.2017.7977015.

[10] S. Ojha and S. Sakhare, "Image processing techniques for object tracking in video surveillance—A survey," in *Proc. 2015 Int. Conf. Pervasive Comput. (ICPC)*, Pune, India, 2015, pp. 1–6, doi: 10.1109/PERVASIVE.2015.7087180.

[11] P. Hristov, "Real-time Abnormal Human Activity Detection Using 1DCNN-LSTM for 3D Skeleton Data," in *Proc. 2021 12th Nat. Conf. Int. Participation (ELECTRONICA)*, Sofia, Bulgaria, 2021,pp.1–4,doi: 10.1109/ELECTRONICA52725.2021.9513696.

[12] P. Hristov, A. Manolova, and O. Boumbarov, "Deep Learning and SVM-Based Method for Human Activity Recognition with Skeleton Data," in *Proc. 2020 28th Nat. Conf. Int. Participation (TELECOM)*, Sofia, Bulgaria, 2020, pp. 49–52, doi: 10.1109/TELECOM50385.2020.9299541.

[13] L. Khattar, C. Kapoor, and G. Aggarwal, "Analysis of Human Activity Recognition using Deep Learning," in *Proc. 2021 11th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Noida, India, 2021,pp.100–104,doi: 10.1109/Confluence51648.2021.9377114.

[14] A. M. F and S. Singh, "Computer Vision-based Survey on Human Activity Recognition System, Challenges and Applications," in *Proc. 2021 3rd Int. Conf. Signal Process. Commun. (ICSPC)*, Coimbatore, India, 2021, pp. 110–114, doi: 10.1109/ICSPC51351.2021.9451736.

[15] A. D. Patel and J. H. Shah, "Performance Analysis of Supervised Machine Learning Algorithms to Recognize Human Activity in Ambient Assisted Living Environment," in *Proc. 2019 IEEE 16th India Council Int. Conf. (INDICON)*, Rajkot, India,2019,pp.1–4,doi: 10.1109/INDICON47234.2019.9030353.