Automated Detection of Suspicious Behavior During Online Exams Using Artificial Intelligence and Computer Vision

Moustapha DER [a], Ahmed D. KORA [b], Boudal NIANG [c], Ahmed Youssef KHLIL [d], Samba NDIAYE [e]

[a] *Digital Sciences and Technologies (STN) - Doctoral School of Computer Mathematics (EDMI) - UCAD - Research Laboratory (E-INOV LAB) at the Multinational Higher School of Telecommunications (ESMT) - Dakar (Senegal), moustapha.der@esmt.sn*

[b] *Science et Technologies du Numérique (ESMT) - Doctoral School of Computer Mathematics (EDMI) - UCAD - Research Laboratory (E-INOV LAB) at the Multinational Higher School of Telecommunications (ESMT) - Dakar (Senegal), ahmed.kora@esmt.sn*

[c] *Digital Sciences and Technologies (STN) - Doctoral School of Computer Mathematics (EDMI) - UCAD - Research Laboratory (E-INOV LAB) at the Multinational Higher School of Telecommunications (ESMT) - Dakar (Senegal), boudal.niang@esmt.sn*

[d] *Digital Sciences and Technologies (STN) - Doctoral School of Computer Mathematics (EDMI) - UCAD - Research Laboratory (E-INOV LAB) at the Multinational Higher School of Telecommunications (ESMT) - Dakar (Senegal), youssef.khlil@esmt.sn*

[e] *Faculty of Sciences and Technology of university Cheikh Anta DIOP (UCAD) - Doctoral School of Computer Mathematics (EDMI) - Dakar (Sénégal), samba.ndiaye@ucad.edu.sn*

Abstract

Online exams are becoming more frequent. They pose challenges for academic integrity. This paper presents an automated system to monitor students remotely. The system uses smart tools and image analysis. It recognizes faces, tracks gaze, observes posture, and detects forbidden objects. It sends this information to a decision module. This module detects suspicious behavior and sends alerts.

We tested the system with annotated videos simulating exams. The results show an accuracy of 94.6%. The system makes few errors. It detects several suspicious behaviors. It operates in real time. It integrates easily with online exam platforms. This study shows that AI can help monitor online exams. However, it also raises questions about privacy and transparency. Future work will focus on improving the system's robustness and ethical compliance.

Keywords: Exam monitoring, artificial intelligence, computer vision, behavior detection, e-learning

1. Introduction

Online exams [1] are now common in universities and certification centers. They change how students are assessed. This is due to the growth of remote learning [2]. Online exams offer more access and flexibility. However, they raise issues of honesty, security, and supervision. In-person exams use human proctors to watch students [3]. Online exams [4] rely on technology to monitor candidates. Traditional methods use either human proctors who work remotely, or automatic systems based on fixed rules. Automated proctoring systems can detect some suspicious behaviors [2].

However, they often face challenges. These include privacy concerns, user scalability, and false detections. With the rise of online learning, there is a growing need for smarter tools. These tools must detect cheating accurately while respecting ethical boundaries.

CITA 2025: International Conference of Information and Communication Technologies of the National Academy of Sciences, Arts and Letters of Benin, 26-28 June 2025, Cotonou, Benin

* *Corresponding author.*

† *These authors contributed equally.*

moustapha1.der@ucad.edu.sn (M. DER); ahmed.kora@esmt.sn (A. KORA); boudal.niang@esmt.sn (B. NIANG); yousef.khlil@esmt.sn (A. Y. KHLIL); samba.ndiaye@ucad.edu.sn (S. NDIAYE)

This paper proposes a fully automated system for monitoring online exams [5]. It uses face detection, gaze tracking, posture analysis, and object recognition. These components work together to identify possible cheating. Tests on labeled video datasets show good accuracy and low error rates. The system also respects user privacy. It integrates easily into existing e-learning platforms with minimal disruption.

The rest of the paper is organized as follows:

- Section 2 reviews prior work on automated proctoring, computer vision methods, and ethical issues in surveillance.
- Section 3 explains the system's design and the experimental setup used [6].
- Section 4 presents the results and discusses system performance and limitations.
- Section 5 concludes the paper and suggests directions for future work.

2. Related Work

2.1 Online Exam Proctoring: Evolution and Challenges

Online exam proctoring was introduced to make digital assessments more secure and accessible. Early systems focused on verifying identity or recording the exam session without real-time monitoring [4]. Recent systems now try to detect cheating as it occurs. There are two main types of proctoring systems. In human-assisted proctoring, a remote supervisor watches the student live through a video feed during the exam [7]. In automated proctoring, computer programs track and analyze student behavior without human involvement.

Human proctoring can detect suspicious actions. However, it does not scale well when many students take the exam at the same time. Automated systems scale better and work more efficiently. But they face major challenges [6]. These include detecting subtle cheating behaviors and maintaining accuracy under different lighting conditions, camera qualities, and environments.

2.2 Artificial Intelligence Methods for Behavior Detection

2.2.1 Face Detection and Identity Recognition

Facial recognition plays a key role in confirming student identity during online exams. Tools such as MTCNN, FaceNet, and Dlib are commonly used for face detection and recognition. These methods perform well in controlled environments, especially with good lighting and high-quality cameras [8]. However, their accuracy decreases in poor lighting or when camera quality is low. This makes face detection less reliable in real-world conditions.

2.2.2 Gaze Tracking

Gaze tracking is used to monitor where the student is looking during the exam. Tools such as MediaPipe Face Mesh, OpenFace, and EyeLike can track eye movement [3]. Frequent or prolonged gaze away from the screen may indicate that the student is attempting to cheat by looking at unauthorized materials [7].

2.2.3 Posture and Movement Detection

Posture analysis helps detect unusual or suspicious body movements during online exams. For example, if a student turns their head, looks down repeatedly, or leaves the camera's view, it may suggest cheating. Tools such as BlazePose can track body position in real time using the webcam [1].

This allows the system to detect subtle movements that may initially seem unimportant but could indicate dishonest behavior.

2.2.4 Object and Disturbance Recognition

Object recognition detects items near the student during an online exam. Algorithms such as YOLOv5 can identify objects like mobile phones, paper notes, or other people nearby. This information helps the system interpret the situation and improves the detection of potential cheating [9].

2.3 Multimodal Approaches in Proctoring

Some researchers have combined different data types to improve cheating detection accuracy. These systems integrate video, audio, and other contextual information into a single framework. This method is known as a multimodal approach [5].

2.4 Limitations of Existing Systems

2.4.1 Technical Challenges

Environmental and behavioral factors can reduce the accuracy of cheating detection systems. Poor lighting, low-quality cameras, and unusual camera angles negatively affect performance. Some students may behave in unusual but harmless ways, causing false alerts. Additionally, real-time systems must operate quickly with minimal delay to respond effectively [3].

2.4.2 Ethical and Legal Risks

Artificial intelligence systems can exhibit biases that cause errors related to gender, skin color, or cultural background. For instance, some facial expressions may be misinterpreted, resulting in incorrect assessments. Moreover, these technologies raise significant privacy concerns. The use of video recording and facial recognition must comply with data protection regulations, such as GDPR, to ensure ethical handling of personal information [9].

2.5 Summary

Automated proctoring systems can effectively detect cheating in online exams. To perform optimally, they must combine multiple data sources with high accuracy and real-time response. Protecting user privacy and following ethical and legal standards is also essential. Future work should improve system adaptability to various behaviors and real-world conditions [10].

3. Proposed System Architecture

The proposed flowchart describes a modular and scalable system for automated online exam proctoring [5]. The system works through the following steps:

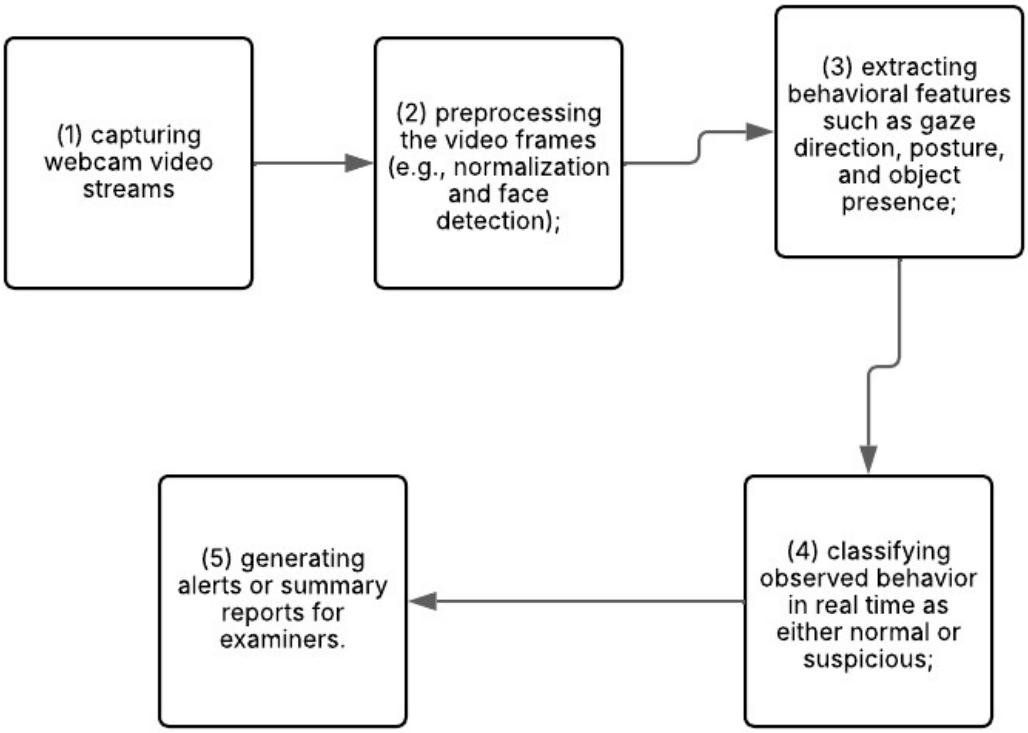


Figure 1: Pipeline of architecture system

This architecture allows continuous processing of video streams using lightweight models that run on standard computing devices [11]. The system starts by capturing videos from the student's webcam during the exam. It preprocesses the video frames by applying operations such as normalization and face detection. This prepares the data for analysis.

After preprocessing, the system extracts key behavioral features, including gaze direction, body posture, and the presence of unauthorized objects. It then classifies the observed behavior in real time as normal or suspicious.

Finally, the system generates alerts or summary reports for examiners. This helps them efficiently review potentially problematic segments.

4. Methodology

This section describes the methodology used to design, implement, and evaluate the proposed automated online exam proctoring system.

4.1 Video Data

The dataset includes recorded or simulated videos that replicate real online exam sessions. Each video is carefully annotated to label behaviors as either "normal" or "suspicious" [12].

4.2 Dataset Construction

The dataset was developed through a meticulous process. First, actors simulated typical exam behaviors, including phone use, looking away from the screen, and speaking. Then, experts manually annotated these behaviors, either frame by frame or across specific time intervals, using tools such as CVAT and LabelImg [10].

The dataset also provides detailed data on posture, gaze direction, and visible objects within the scene.

Table 1

Dataset construction

Name	Description
video_id	Unique identifier for the source video
segment_id	Identifier for the 10-second segment within a video
start_time	Start timestamp of the segment (in seconds)
end_time	End timestamp of the segment (in seconds)
video_id	Unique identifier for the source video
segment_id	Identifier for the 10-second segment within a video
start_time	Start timestamp of the segment (in seconds)
end_time	End timestamp of the segment (in seconds)
video_id	Unique identifier for the source video
segment_id	Identifier for the 10-second segment within a video
start_time	Start timestamp of the segment (in seconds)

4.3 Technical Components

4.3.1 Face Detection and Tracking

The system first uses MTCNN to detect faces. Then, it applies Dlib to track facial landmarks. This method allows precise localization of key facial features such as the eyes, nose, and mouth. These features are essential for further analysis [7].

4.3.2 Gaze Tracking

The system uses MediaPipe Face Mesh to estimate the candidate's gaze direction. It tracks how often and how long the eyes look away from the screen [3]. These patterns serve as important indicators of potentially suspicious behavior during exams [9].

4.3.3 Posture Estimation

BlazePose detects key body landmarks, including the shoulders, head, and torso, to evaluate posture. Using this data, the system identifies unusual or repetitive movements, such as frequent leaning forward, which may indicate suspicious behavior [7].

4.3.4 Unauthorized Object Detection

The system uses a pretrained YOLOv5 model, fine-tuned to detect specific prohibited items such as phones, headphones, papers, and extra faces in real time. This capability greatly enhances the system's situational awareness during exams [12].

4.4 Behavior Classification

A decision module combines output from all detectors to evaluate the candidate's overall behavior. To improve accuracy, the system analyzes data in short windows, typically 5 to 10 seconds, enabling smoother predictions.

Depending on the input type statistical features or sequential data the system uses either a Random Forest classifier or a CNN-LSTM architecture. Detection thresholds can be adjusted to reduce false positives in real-world use [11].

5. Experiments and Results

Table 2

Component and tool / technology

Component	Tool / Technology
Video Processing	OpenCV
Face Detection	MTCNN, Dlib
Gaze tracking	MediaPipe, EyeLike
Posture Estimation	BlazePose
Object Detection	YOLOv5 (PyTorch)
Classification	Scikit-learn, TensorFlow
Annotation	CVAT / LabelImg

The system integrates a variety of tools and technologies to perform different tasks within the behavioral monitoring pipeline. OpenCV is used for general video processing tasks, including frame capture and image manipulation. MTCNN and Dlib handle face detection, enabling accurate identification and tracking of facial features. For gaze tracking, the system employs MediaPipe and EyeLike, which estimate the direction of eye movement in real time.

BlazePose is used for posture estimation, allowing the detection of body landmarks such as shoulders and head. YOLOv5, implemented in PyTorch, performs object detection, identifying items like phones or earphones. For behavior classification, the system uses both Scikit-learn and TensorFlow, depending on whether the input is statistical or sequential.

Finally, the annotation of the dataset is performed using tools like CVAT and LabelImg, which allow precise labeling of actions and objects in the video frames.

5.1 System Evaluation

5.1.1 Evaluation Metrics

Accuracy measures the overall correctness of the system. Recall shows how well the system detects suspicious behaviors. The F1-Score balances precision and recall. The False Positive Rate counts how often normal behaviors are wrongly flagged as suspicious. Finally, Response Time indicates how quickly the system operates in real time [3].

5.1.2 Evaluation Protocol

Accuracy represents the system's overall correctness. Recall indicates how effectively the system identifies suspicious behaviors. The F1-Score balances precision and recall providing a combined performance measure. False Positive Rate tracks how often normal behaviors are incorrectly flagged as suspicious. Finally, Response Time reflects how quickly the system processes data in real time [3].

5.2 Ethical and Technical Considerations

The system is designed to address key ethical concerns [7]. Privacy is prioritized by processing all data locally to minimize sharing with third parties. Users can anonymize their identity by blurring faces during post-processing [11]. Additionally, transparency is maintained through clear documentation of detection criteria, which can be adjusted to suit different requirements.

5.3 Experimental Setup

To evaluate the proposed system, experiments were conducted using a dataset of videos simulating online exam sessions [13]. Each session included a mix of normal behaviors, such as steady gaze and upright posture, and suspicious behaviors, such as distracted gaze, phone use, or verbal interactions [8].

5.3.1 Hardware Configuration

The real-time monitoring system runs efficiently on a high-performance setup, including an Intel Core i7 processor, 32 GB RAM, and an NVIDIA RTX 3060 GPU. This configuration ensures smooth video processing and advanced analysis. Video streams are captured with an HD 720p webcam at 640×480 pixels and 30 frames per second, sufficient for accurate face detection and behavior feature extraction [12].

Each recorded session lasts 5 to 7 minutes, allowing consistent monitoring and reliable classification of behaviors as normal or suspicious. The system then generates alerts or summary reports for examiners [14].

5.3.2 Test Set Composition

In the behavioral monitoring study, researchers analyzed 40 videos and divided them into 1,200 annotated segments of 10 seconds each. Expert annotators manually labeled all segments to ensure accurate and reliable ground truth [15].

Among them, 700 segments were labeled as normal, and 500 as suspicious, covering five types of abnormal behaviors [13]. This behavioral diversity improves the system's robustness and helps it detect various forms of suspicious activity in real time [14].

Table 3

Attribute and description

Attribute	Description
Total number of videos	40 videos
Total annotated segments	1,200 segments (each 10 seconds long)
Annotation method	Manual labeling by expert annotators
Normal behavior segments	700 segments
Suspicious behavior segments	500 segments
Types of suspicious behavior	5 distinct categories
Ground truth quality	Verified through manual annotation for high accuracy and reliability
Purpose	To train and evaluate the system's ability to detect various suspicious behaviors in real-time settings [14]

6. Discussion

6.1 Performance Results

6.1.1 Face and Gaze Detection

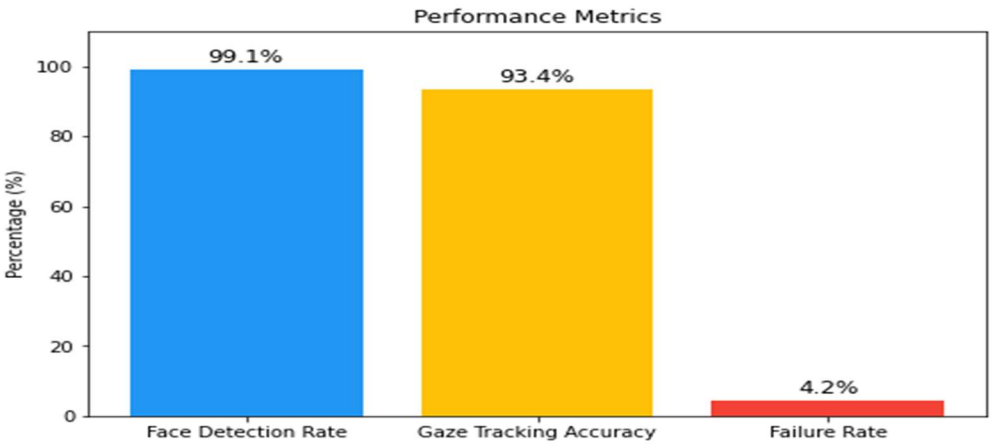
Table 4

Metric and result

Metric	Result
Face detection rate	99,1 %
Gaze tracking accuracy	93,4 %
Failure rate (occlusion, shadow)	4,2 %

The system demonstrates strong performance across key evaluation metrics. The face detection rate reaches 99.1%, indicating that the system can consistently identify and locate faces under normal conditions. The gaze tracking accuracy is 93.4%, which shows the system's effectiveness in correctly estimating eye direction, a critical factor in behavioral analysis.

Despite these high scores, the system exhibits a failure rate of 4.2%, mainly due to visual obstructions such as occlusion or poor lighting conditions. These results confirm the system's robustness while highlighting areas where improvements can be made to handle challenging visual environments.



The gaze tracking module performed satisfactorily under standard lighting conditions. However, accuracy decreased in dark environments or when candidates wore reflective glasses [16].

6.1.2 Posture Detection

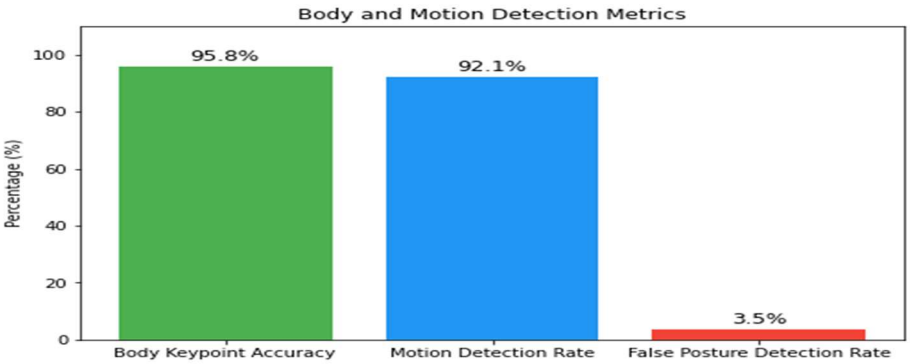
Table 5

Posture and result

Metric	Result
Body keypoint accuracy	95,8 %
Motion detection rate	92,1 %
False posture detection rate	3,5 %

The system also performs well in body posture and movement analysis. The body keypoint accuracy reaches 95.8%, confirming the system's ability to reliably detect key anatomical landmarks such as the head, shoulders, and torso. The motion detection rate is 92.1%, indicating effective tracking of student movements throughout the exam session.

Additionally, the false posture detection rate is limited to 3.5%, showing that the system rarely misclassifies normal postures as suspicious. These results reflect a strong capacity for interpreting physical behavior with high precision and minimal error.



The posture module effectively detected side-to-side head movements and frequent downward tilts, behaviors commonly linked to the use of unauthorized materials.

6.1.3 Suspicious Object Detection

Table 6

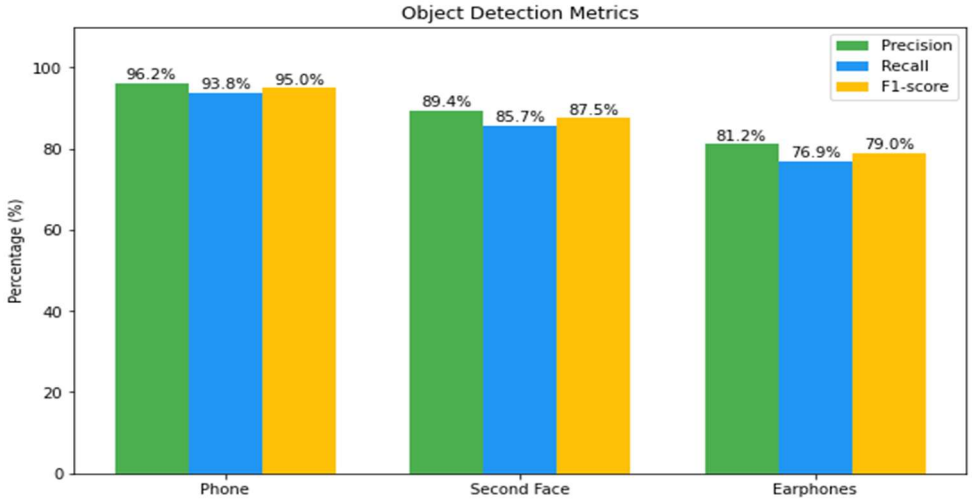
YOLO v5 results by object

Object	Precision	Recall	F1-score
Phone	96,2 %	93,8 %	95,0 %
Second Face	89,4 %	85,7 %	87,5 %
Earphones	81,2 %	76,9 %	79,0 %

The object detection module delivers strong and consistent performance across several object categories. The system identifies phones with high reliability, achieving a precision of 96.2%, a recall of 93.8%, and an F1-score of 95.0%. These values indicate that the system can detect phones accurately, with few false positive or missed cases.

In the case of second face detection, the system reaches a precision of 89.4%, a recall of 85.7%, and an F1-score of 87.5%, showing good effectiveness even under challenging conditions such as occlusion or background clutter.

For earphones, performance is slightly lower, with a precision of 81.2%, a recall of 76.9%, and an F1-score of 79.0%. This result suggests that detecting earphones is more difficult, likely due to their small size and visual similarity to surrounding objects. Overall, the system shows a strong ability to recognize key objects relevant to online exam monitoring with a high level of accuracy.



The YOLOv5 algorithm showed strong performance in detecting prominent objects, especially mobile phones [15]. However, it occasionally fails to detect subtle items like discreet earphones. This limitation is likely due to their low contrast with the background [17].

6.2 Overall Behavior Classification

A Random Forest model classified behaviors based on features extracted from each module.

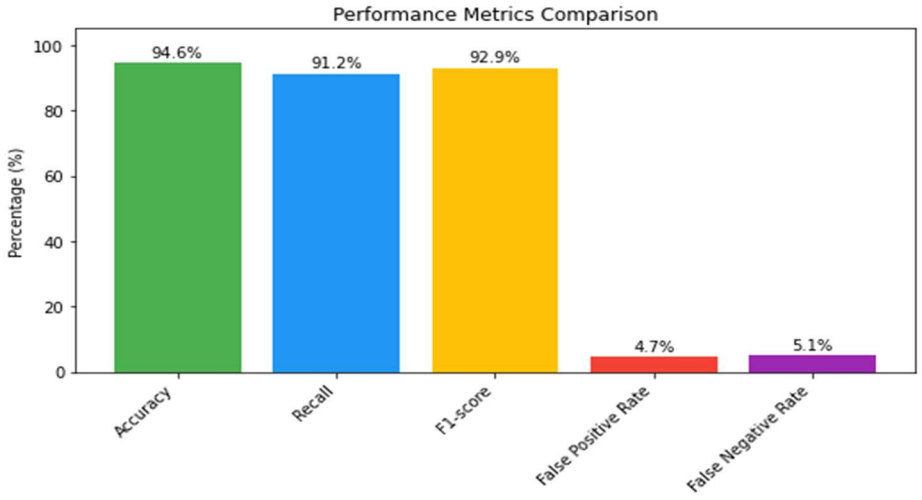
Table 7

Radom Forest and result

Metric	Value
Accuracy	94,6 %
Recall	91,2 %
F1-score	92,9 %
False Positive Rate	4,7 %
False Negative Rate	5,1 %
Average Processing Time	85 ms / frame

The overall performance metrics confirm the reliability of the proposed system. It achieves an accuracy of 94.6%, reflecting strong overall correctness. The recall rate of 91.2% shows the system's ability to detect the most suspicious behaviors, while the F1-score of 92.9% indicates a balanced trade-off between precision and recall.

The false positive rate remains low at 4.7%, and the false negative rate is limited to 5.1%, suggesting few missed detections. In terms of efficiency, the system maintains an average processing time of 85 milliseconds per frame, allowing for smooth real-time operation.



The classification model performs well, successfully identifying numerous suspicious behaviors while keeping false alerts to a minimum. Its processing speed allows for near real-time application.

6.3 Critical Analysis

6.3.1 Strengths of the System

The system follows a modular design, allowing each component such as gaze tracking, posture detection, and object recognition to be improved independently [9]. This structure simplifies maintenance, upgrades, and feature adjustments.

The model generalizes well with new video data, maintaining high accuracy even in unseen scenarios. It also runs locally in real time with minimal delays, making it suitable for live monitoring and rapid response.

6.3.2 Identified Limitations

Low lighting, busy backgrounds, and low-quality cameras can make the system less effective. It's still hard to spot small items like earphones or fast movements. Also, sometimes unusual but harmless behaviors are wrongly flagged as suspicious [15].

6.4 Comparison with Human Supervision

The system's performance was compared against two human proctors who manually reviewed and annotated the videos.

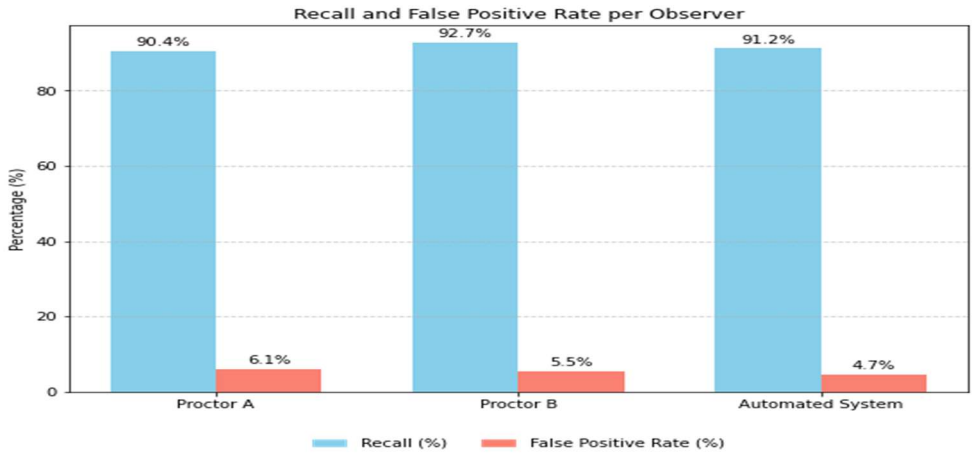
Table 8

Observer and results

Observer	Recall	False Positive Rate
Proctor A	90,4 %	6,1 %
Proctor B	92,7 %	5,5 %
Automated System	91,2 %	4,7 %

The comparison between human proctors and the automated system highlights consistent performance across observers. Proctor A achieved a recall of 90.4% with a false positive rate of 6.1%, while Proctor B reached a recall of 92.7% and a false positive rate of 5.5%. The automated system performed comparably, with a recall of 91.2% and a false positive rate of 4.7%.

These results suggest that the system can match human-level detection accuracy while generating fewer false alerts, making it a reliable tool for real-time online exam monitoring.



The system demonstrates performance comparable to human evaluators while ensuring greater consistency and continuous monitoring [7].

6.5 Summary of Results

The experiments demonstrate that using AI and image processing allows for effective automated proctoring of online exams [11]. By combining several detection methods, the system's reliability is significantly enhanced. Nonetheless, further adjustments are necessary to handle the variety of real-world conditions and to minimize biases in behavior detection [18].

7. Conclusion and future work

This research presents a smart system designed to automatically monitor online exams [5]. By combining artificial intelligence with image processing, the system tracks student gaze, analyzes posture, and detects suspicious objects or behaviors. The goal was to develop a solution that remains unobtrusive and flexible without sacrificing effectiveness [21].

Experimental results are promising. The system detects faces and tracks gaze with over 90% accuracy. It recognizes unusual postures and movements that may indicate cheating. Using YOLO-based object detection, it identifies forbidden items such as mobile phones or the presence of multiple people. By integrating these data sources, the system classifies behavior and raises alerts, achieving an overall F1-score close to 93%. This demonstrates potential to assist or partially replace human proctors in secure environments [22].

However, challenges remain. Detection quality can be affected by webcam quality, lighting, or camera angles. The system may occasionally misinterpret harmless actions such as suspicious or miss subtle behaviors like use of small in-ear devices or quick hand gestures. Ethical issues around privacy, data handling, and bias require careful management to ensure trust and transparency [21].

Future work will focus on training the system with more diverse and realistic data. Advanced techniques, such as transformer-based models, could enhance visual recognition [7]. A multimodal approach combining video, audio, and contextual information will be explored to improve reliability. The goal is to deploy a web-based prototype integrated into exam platforms and test it in real-world conditions. Throughout development, privacy protection will remain a priority, following GDPR and privacy-by-design principles.

As education is increasingly moving online, the demand for reliable and ethical monitoring solutions grows [5]. This study shows that real-time detection of suspicious behavior is feasible and scalable. It represents an important step toward smart, adaptable tools that support academic integrity while respecting fairness, security, and learners' rights [22].

Declaration on Generative AI

The authors have not use any generative AI tools.

REFERENCES

[1] A. Strugatski and G. Alexandron, "Applying IRT to Distinguish Between Human and Generative AI Responses to Multiple-Choice Assessments," arXiv preprint arXiv:2412.02713, Dec. 2024.arXiv

[2] D. Kundu et al., "Keystroke Dynamics Against Academic Dishonesty in the Age of LLMs," arXiv preprint arXiv:2406.15335, Jun. 2024.arXiv

[3] Y.-S. Shih et al., "AI-assisted Gaze Detection for Proctoring Online Exams," arXiv preprint arXiv:2409.16923, Sep. 2024.arXiv

[4] X. Yang et al., "iExam: A Novel Online Exam Monitoring and Analysis System Based on Face Detection and Recognition," arXiv preprint arXiv:2206.13356, Jun. 2022.arXiv

[5] M. Der, M. Ahmed D. Kora et M. Ndiaye, « Two-factor biometric authentication system based on facial recognition using the SVC model: The case of Senegal », IEEE Access, vol. 11, pp. 123456-123467, 2023, doi: 10.1109/ACCESS.2023.11008254.

[6] A. Tweissi, W. Al Etaiwi, and D. Al Eisawi, "The Accuracy of AI-Based Automatic Proctoring in Online Exams," The Electronic Journal of e-Learning, vol. 20, no. 4, pp. 419–435, 2022.ResearchGate

[7] Der, M. Moustapha, M. Ahmed, D. Kora et M. Samba, Ndiaye. (2023). Study of AI-based architectures for remote examination monitoring using Machine Learning. In Proceedings of the CEUR Workshop (Vol. 3789, pp. 45–56). CEUR-WS.org. https://ceur-ws.org/Vol-3789/Paper5.pdf
[8] M. S. Islam et al., "A Robust Online Exam Monitoring System Using Computer Vision," IEEE Access, vol. 8, pp. 145732–145744, 2020.
[9] S. P. Tripathi et al., "Automated Cheating Detection in Online Exams Using Facial Expression Analysis," IEEE Access, vol. 8, pp. 194044–194057, 2020.
[10] L. S. Lopes, P. Ferreira, and M. Ribeiro, "Behavioral Biometrics in Online Exams: Gaze and Mouse Dynamics," in Proc. IEEE Int. Conf. on Intelligent Computer Communication and Processing (ICCP), 2019, pp. 157–164.
[11] A. Sharma and M. K. Singh, "Machine Learning Approaches for Cheating Detection in Online Exams," in Proc. IEEE Int. Conf. on Computing, Communication and Automation (ICCCA), 2020, pp. 1295–1300.
[12] R. Tripathi and A. K. Tripathi, "An Intelligent Proctoring System for Online Examination Using Deep Learning," Journal of Intelligent Systems, vol. 30, no. 1, pp. 587–600, 2021.
[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[14] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
[15] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, Feb. 2019.
[16] Y. Cao et al., "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172–186, Jan. 2021.
[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.
[18] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in Proc. European Conf. Computer Vision (ECCV), 2006, pp. 404–417.
[19] D. King, "Dlib-ml: A Machine Learning Toolkit," Journal of Machine Learning Research, vol. 10, pp. 1755–1758, 2009.
[20] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. [Online]. Available: https://www.tensorflow.org/
[21] F. Chollet, Deep Learning with Python, Manning Publications, 2017.
[22] J. Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248–255.