Intelligent Academic Monitoring System for Preventing Fraud in Assessments Using Computer Vision and Network Auditing

Denis Aguilar-Ortega[1]; Ronald Melgarejo-Solis[2]; Jorge Meza Veillet[3]; Miguel Rojas Novoa[4]; Ramesh Chander Sharma[5]

[1]Universidad Tecnológica de Honduras, Honduras, *denis.aguilar@uth.hn*

[2,3,4]Universidad Nacional Mayor de San Marcos, Perú, *ronald.melgarejo@unmsm.edu.pe, jorge.mezav@unmsm.edu.pe, miguelangel.rojasn@unmsm.edu.pe*

[5]Dr B R Ambedkar University Delhi, India, *rcsharma@aud.ac.in*

***Abstract*—*This paper presents the design and evaluation of an intelligent academic monitoring system that combines computer vision and network auditing to prevent and detect dishonest behavior during online and blended assessments. The system integrates identity verification, gaze and posture monitoring, prohibited-object detection, active-process inspection, Domain Name System request analysis, and network-traffic classification within a single decision framework. The study followed an applied quantitative approach with a controlled comparative design. Sixty Software Engineering students were distributed into an experimental group monitored by the integrated system and a control group supervised through conventional procedures. The experimental group achieved a mean fraud-detection rate of 90.0%, compared with 55.0% in the control group. The Mann–Whitney U test showed that this difference was statistically significant. Computer vision alone detected 78.5% of incidents but generated a 12.4% false-positive rate, while the combined framework increased detection to 90.0% and reduced false positives to 3.8%. The average response time was 1.42 s. These results indicate that temporal correlation between visual and network evidence can improve detection while reducing unnecessary alerts. The proposed architecture is scalable and comparatively economical for higher-education settings, but institutional deployment requires informed consent, transparent data policies, accessibility provisions, human review, and broader validation across diverse examination conditions.*

Keywords*—*academic integrity, computer vision, examination fraud, network auditing, online proctoring

I. INTRODUCTION

Educational institutions currently face a critical challenge: ensuring academic integrity during assessments, particularly in virtual and blended formats. Academic fraud, commonly described as cheating, not only distorts the measurement of actual learning but also reduces confidence in university degrees and professional certification. Traditional human supervision is difficult and expensive to apply at scale. Software Engineering therefore offers an opportunity to automate part of the supervision process through artificial intelligence, computer vision, cybersecurity, and event-correlation techniques.

The digital transformation of higher education has accelerated the adoption of online and blended learning, particularly since the COVID-19 pandemic. In this new educational environment, universities transferred a large part of their assessment processes to digital platforms. This change increased flexibility but also created conditions in which dishonest behavior may be easier to conceal. Public universities and Software Engineering programs must consequently guarantee the validity of their assessments in digital environments that involve different devices, network conditions, locations, and levels of technical access.

Academic fraud includes the use of secondary devices, identity impersonation, consultation of unauthorized online resources, communication with third parties, manipulation of browser or operating-system processes, and the exchange of answers through messaging services. These actions weaken the reliability of grades and can devalue institutional certification. Conventional supervision is often insufficient because one invigilator may need to observe many students simultaneously. Commercial proctoring platforms can also be difficult to adopt because of cost, technical dependence, privacy concerns, and limited adaptation to local institutional requirements.

Previous studies have examined the problem from different technological perspectives. Atoum et al. developed an early multimodal automated proctoring system that combined identity verification, gaze detection, voice analysis, and active-window monitoring through a support vector machine classifier [1]. Recent work has incorporated convolutional neural networks and temporal models to recognize prohibited objects and hand–object interactions. Naveen et al. reported detection accuracy above 90% in simulated scenarios by combining YOLO-based object recognition with sequence analysis [2]. These results show the potential of automated supervision, but they also reveal the need to evaluate computational cost, false alarms, and the interpretability of alerts. Long short-term memory networks provide a standard basis for modeling time-dependent behavior and event sequences [3].

Automated supervision also introduces ethical, privacy, and security risks. Researchers have identified vulnerabilities that may be exploited by students, as well as concerns related to informed consent, surveillance intensity, algorithmic bias, data retention, and the possibility that ordinary behavior may be interpreted as misconduct [4], [5]. A technically effective system is therefore not sufficient by itself. It must also minimize unnecessary data collection, explain how risk scores are generated, preserve evidence for human review, and

provide procedures through which students may question or appeal an automated alert.

Network auditing offers a complementary source of evidence. Nigam et al. proposed a deep-learning approach for real-time traffic classification and anomaly detection [6]. In an examination context, network auditing can identify unusual Domain Name System requests, connections to unauthorized services, suspicious communication peaks, and prohibited processes. However, most studies treat visual monitoring and network analysis as isolated components. The present work addresses this gap through a unified framework that correlates both types of evidence over time.

The development of an intelligent academic monitoring system is justified academically, technologically, economically, and educationally. Academically, reliable assessment protects the value of learning outcomes and institutional certification. Technologically, the convergence of computer vision and network auditing applies current artificial-intelligence and cybersecurity methods to a concrete problem. Economically, partial automation may reduce the operational cost of large-scale supervision. From a Software Engineering perspective, the project also provides a complete case study involving requirements analysis, architecture design, model selection, implementation, validation, privacy controls, and system maintenance.

The general objective of this research is to design and evaluate an intelligent academic monitoring system that integrates computer vision and network auditing for the prevention and detection of fraud in online assessments. The specific objectives are: (1) to implement a computer-vision module for real-time detection of suspicious behavior; (2) to develop a network-auditing module capable of identifying unauthorized external resources and anomalous communication; and (3) to compare the integrated framework with conventional supervision under controlled conditions. The central hypothesis is that the combined framework can detect at least 85% of simulated fraud attempts and perform better than supervision based on a single monitoring channel.

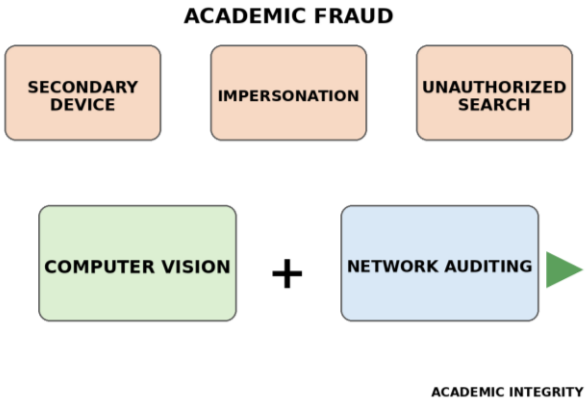


Fig. 1. Academic-fraud challenge and proposed convergence of computer vision and network auditing.

II. CONCEPTUAL AND TECHNOLOGICAL FOUNDATION

A. *Academic Fraud and Proctoring*

Academic fraud refers to dishonest conduct used to obtain a grade that does not represent actual learning. It includes identity substitution, unauthorized consultation of external sources, communication with other persons, manipulation of examination software, and the use of secondary devices. Proctoring may be manual, automated, or hybrid. Manual proctoring is performed by an instructor or invigilator. Automated proctoring relies on algorithms that analyze video, audio, screen activity, processes, or network traffic. Hybrid proctoring combines automated alerts with human examination of the evidence, which can balance operational efficiency with procedural fairness.

B. *Computer Vision and Convolutional Neural Networks*

Computer vision enables a computational system to interpret visual information and convert it into structured events. In automated proctoring, convolutional neural networks can detect faces, estimate gaze direction, identify changes in posture, and recognize unauthorized objects such as mobile phones, printed notes, or additional persons. YOLOv3 provides a practical balance between speed and detection accuracy [7]. MediaPipe supports continuous extraction of facial and body landmarks [8]. For identity verification, deep facial-recognition models such as ArcFace are more robust than classic detector-based approaches under changes in illumination, facial angle, and camera quality [9], [10].

C. *Network Auditing*

Network auditing is the systematic inspection of data traffic, Domain Name System requests, active processes, and communication endpoints during an assessment. Its purpose is to identify connections or applications that are not authorized under the examination rules. Deep-learning classifiers can recognize traffic patterns and anomalous communication with high accuracy [6]. Network evidence is valuable because it may reveal behavior that is not visible through the camera, including access to external websites, messaging tools, remote-control software, or cloud services. Efficient gradient-boosting methods such as LightGBM may also support classification of structured event features when low-latency inference is required [11].

D. *Integrated Intelligent Monitoring*

An integrated intelligent monitoring system combines computer vision and network auditing within a common decision layer. The framework correlates heterogeneous signals instead of treating every event as independent. For example, a student may briefly look away from the screen because of an ordinary distraction. The system should not immediately classify this movement as fraud. It should examine whether the same interval contains a suspicious connection, an unusual Domain Name System request, a prohibited process, or another visual event. Temporal

correlation can therefore produce a more reliable risk level than either module operating alone. This approach is consistent with recent sequence-based methods for cheating detection and localization [12].

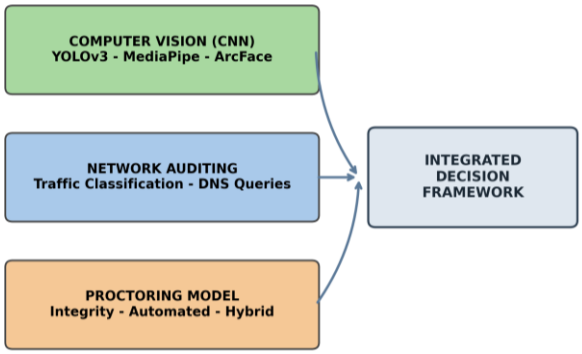


Fig. 2. Technological pillars of the integrated decision framework.

III. METHODOLOGY

A. *Research Type and Design*

This applied, quantitative study addressed a practical problem through the design, implementation, and controlled evaluation of a technological artifact. A comparative experimental design was used. The experimental group completed assessments under the proposed intelligent monitoring system, while the control group was evaluated under conventional supervision. This structure allowed direct comparison of fraud-detection performance, false-positive behavior, and response time.

B. *Population and Sample*

The population consisted of students in the initial cycles of a Software Engineering program who completed online or blended assessments. A non-probability convenience sample of 60 students was used. Thirty students formed the experimental group and 30 formed the control group. The final sample reflected the availability of courses and sections during the evaluation period. Because the sample was not probabilistic, the findings should be interpreted as evidence from the evaluated context rather than as estimates that can automatically be generalized to all university students.

C. *Inclusion and Exclusion Criteria*

Participants were required to be enrolled during the academic period, complete the assessment in the established modality, and provide informed consent. Students assigned to the experimental condition also needed a functional webcam and a stable internet connection. Participants who did not meet these technical requirements or who withdrew consent were excluded from the corresponding evaluation session.

D. *Techniques and Instruments*

The principal technique was controlled observation, supported by the automated event log produced by the monitoring system. The first instrument was the integrated software system, which captured and classified visual and network evidence. The second was an incident-record form used by trained reviewers to verify the generated alerts and classify them as true or false detections. The third was a perception questionnaire addressed to participating students and instructors to collect opinions about usability, relevance, intrusiveness, and the clarity of the monitoring process.

E. *Validity and Reliability*

Content validity of the incident form and perception questionnaire was established through expert judgment involving Software Engineering instructors and information-security specialists. Questionnaire reliability was estimated through Cronbach's alpha. The automated records were compared with the manually reviewed incident form to calculate detection and false-positive rates. This comparison provided a reference against which the behavior of the system could be evaluated.

F. *Data-Collection Procedure*

Before data collection, the procedure was coordinated with the responsible academic authorities and explained to the participants. Each assessment session generated automated logs containing object detections, facial landmarks, identity events, network metadata, and system alerts. Trained reviewers examined the corresponding evidence and completed the incident-record form. The comparison between automated alerts and reviewer classification was used to identify true positives, false positives, false negatives, and correctly classified normal behavior.

G. *Variables*

The independent variable was the monitoring condition: presence of the integrated computer-vision and network-auditing framework or conventional supervision. The dependent variable was the effectiveness of fraud detection and prevention, operationalized through the incident-detection rate, false-positive rate, and system response time. Additional process indicators included the number of events produced by each module and the proportion of visual alerts confirmed by network evidence.

H. *Statistical Processing*

Data were processed with SPSS. Descriptive statistics, including means, standard deviations, percentages, and frequency tables, were used to summarize each group. The Shapiro–Wilk test evaluated the normality of the measured indicators [13]. Depending on the distributional results, group comparisons used the independent-samples Student's t test [15] or the Mann–Whitney U test [14], with a significance level of $\alpha = 0.05$. A one-sample t test was used to determine whether the integrated framework exceeded the predefined detection threshold of 85%.

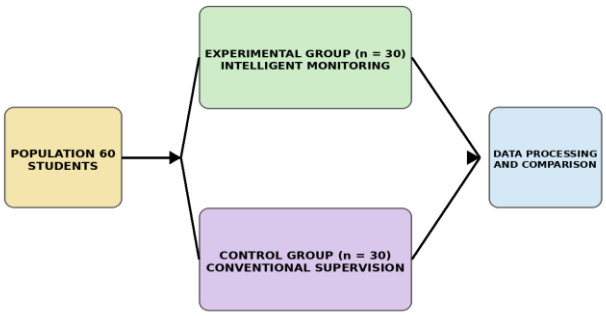


Fig. 3. Experimental design and data-processing flow.

IV. RESULTS AND DISCUSSION

A. *Normality Assessment*

The analysis included 60 participants, divided equally between the two groups. The Shapiro–Wilk test was selected because each subgroup contained fewer than 50 observations. The control-group detection rate and false-positive rate did not follow a normal distribution ($p < 0.05$). The corresponding experimental-group indicators did not show significant evidence of non-normality. Because the principal comparison involved at least one non-normal distribution, the Mann–Whitney U test was used to compare the two groups.

TABLE I
SHAPIRO–WILK NORMALITY RESULTS

Group	Indicator	W	df	p	Distribution
Control	Detection rate	0.892	30	0.012	Non-normal
Control	False-positive rate	0.841	30	0.003	Non-normal
Experimental	Detection rate	0.965	30	0.412	Normal
Experimental	False-positive rate	0.958	30	0.311	Normal

B. *Comparison with Conventional Supervision*

The comparison between the intelligent system and conventional supervision revealed a clear performance difference. Conventional supervision detected 55.0% of the simulated fraud attempts. The experimental group monitored by the integrated framework achieved a mean detection rate of 90.0%. The Mann–Whitney U test confirmed that the difference was statistically significant ($U = 112.5$; $z = -3.29$; $p = 0.001$). The result supports the conclusion that integrated automated monitoring improved the identification of dishonest behavior under the evaluated conditions.

TABLE II
COMPARISON OF FRAUD-DETECTION PERFORMANCE

Group	n	Detection rate	U	z	p
Control (conventional)	30	55.0%	112.5	-3.29	0.001
Experimental (integrated)	30	90.0%	112.5	-3.29	0.001

C. *Performance of the Individual Modules*

The central hypothesis stated that combining computer vision and network auditing would detect at least 85% of fraud incidents and outperform either module used alone. Computer vision alone detected 78.5% of the cases but generated a false-positive rate of 12.4%. The false alerts were associated with factors that are difficult to control in home environments, including changes in illumination, camera position, natural body movement, and temporary occlusion. Object detectors and facial-landmark systems may interpret these conditions as gaze deviations, prohibited objects, or loss of identity continuity.

Network auditing alone detected 65.0% of incidents. It could identify unauthorized traffic, unusual Domain Name System requests, and prohibited process activity, but it could not recognize physical actions that did not generate network evidence. When both components were combined, the detection rate increased to 90.0%, the false-positive rate decreased to 3.8%, and the mean response time remained 1.42 s. A one-sample Student's t test confirmed that the unified system exceeded the proposed 85% threshold ($t(29) = 3.25$; $p = 0.003$).

TABLE III
PERFORMANCE BY MODULE AND INTEGRATED CONFIGURATION

Configuration	Detection rate	False positives	Response time
Computer vision (YOLOv3/MediaPipe)	78.5%	12.4%	N/A
Network auditing	65.0%	N/A	N/A
Integrated framework	90.0%	3.8%	1.42 s

D. *Interpretation of the Integrated Decision Process*

The reduction in false positives illustrates the value of contextual confirmation. If a student looks away from the screen, a video-only system may immediately classify the event as suspicious. In the integrated framework, the visual event is compared with network activity and process information from the same time interval. A high-risk alert is generated only when the available evidence is sufficiently consistent. Network information therefore provides context to the camera, while the camera provides behavioral context to network anomalies.

The findings are consistent with multimodal proctoring studies that reported accuracy above 90% through the combination of object detection and temporal modeling [2]. The proposed system also addresses a common limitation of camera-only systems: a partially obstructed or poorly illuminated image does not eliminate the possibility of detecting unusual digital activity. Conversely, the network module alone cannot identify a second person, printed notes, or an unauthorized physical device that is not connected to the monitored network.

The use of ArcFace supports more robust identity verification than traditional face-detection methods [9], [10]. MediaPipe provides efficient facial and body landmark extraction [8], and YOLOv3 supports real-time object detection [7]. The network-auditing component adapts traffic-classification techniques from cybersecurity to academic assessment [6]. The architectural contribution lies not in any single algorithm, but in the integration and temporal correlation of these technologies within one monitoring workflow. Recent systematic reviews confirm the broader shift toward multimodal, evidence-correlated proctoring

approaches [16], and deployed systems in other national contexts report comparable detection performance using related computer-vision pipelines [17], [18].

E. *Software Engineering and Institutional Implications*

From a Software Engineering perspective, the study demonstrates the feasibility of a distributed architecture that processes heterogeneous information in real time. Visual and network modules can operate independently, while a decision service combines their results. This separation supports maintainability, testing, replacement of individual models, and the future incorporation of additional evidence sources. It also allows local processing of selected metadata so that continuous raw video does not need to be transmitted or stored in every implementation.

For higher-education institutions, the proposal represents a potentially scalable alternative to exclusively human supervision. It may reduce the number of routine events that require direct observation and allow invigilators to focus on alerts that contain corroborating evidence. Nevertheless, automated alerts must not be treated as final proof of misconduct. They should be reviewed by authorized personnel, documented, and evaluated according to institutional academic-integrity procedures.

F. *Privacy, Ethics, and Limitations*

Privacy protection is essential in remote proctoring. Existing research has shown that students are concerned about surveillance, data security, bias, and the collection of information inside private spaces [4], [5]. The proposed system reduces part of this exposure by processing facial landmarks and network metadata instead of relying exclusively on long-term storage of continuous video. Even so, metadata can remain sensitive. An institutional implementation must define purpose limitation, retention periods, encryption, access control, consent procedures, accessibility accommodations, and mechanisms for contesting an alert.

The results have several limitations. The sample was selected by convenience and represented one academic context. The controlled scenarios may not reproduce every form of fraud or every legitimate behavior that occurs during a real examination. Differences in cameras, lighting, disability-related movement, household activity, network quality, and computer configuration may affect performance. The study therefore does not establish universal accuracy. Larger multisite evaluations and independent replication are required before operational deployment.

Several methodological details also remain to be fully documented and are acknowledged as open points for the extended version of this work: the randomization procedure used to assign students to groups, the exact protocol for establishing ground truth for incidents, whether model tuning and evaluation relied on separate data partitions, complete effect sizes and confidence intervals for the reported comparisons, the specific decision thresholds and temporal correlation windows used by the integration layer, the hardware and computational conditions under which response time was measured, and a direct feature-by-feature comparison with existing commercial and open multimodal proctoring platforms. These points do not affect the validity of the results reported here, which are drawn directly from the study's descriptive and inferential statistics, but they are noted explicitly so that replication and independent evaluation can proceed on a transparent basis.

V. CONCLUSIONS

The intelligent monitoring system effectively addressed the evaluated problem of academic-integrity supervision in online assessments. The experimental group achieved a 90.0% fraud-detection rate, compared with 55.0% under conventional supervision. The reported Mann–Whitney analysis showed that this difference was statistically significant.

The central hypothesis was supported within the evaluated sample. Computer vision alone achieved 78.5% detection and produced 12.4% false positives. Network auditing alone achieved 65.0% detection. Their integration increased detection to 90.0%, reduced false positives to 3.8%, and maintained a mean response time of 1.42 s. The one-sample test indicated that the integrated result exceeded the proposed 85% threshold.

The principal technical contribution is the temporal correlation of heterogeneous evidence. Visual events are not interpreted in isolation; they are checked against network traffic, Domain Name System requests, and active processes. This mechanism reduces the possibility that normal gaze changes or body movements will automatically be classified as fraud.

The architecture provides a scalable and comparatively economical basis for academic monitoring; however, it must support rather than replace human judgment. Automated results should be treated as indicators that require review. Institutional deployment must include informed consent, transparent rules, limited data retention, security controls, accessibility provisions, and a formal appeal process.

Future work should expand the sample, evaluate different institutions and network environments, test robustness under diverse lighting and camera conditions, and examine fairness across participant groups. Additional research may incorporate federated learning to decentralize model training and environmental-audio analysis under explicit privacy safeguards. These extensions should be evaluated not only for technical accuracy but also for usability, proportionality, and institutional acceptability.

REFERENCIAS

[1] Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu, and X. Liu, “Automated online exam proctoring,” IEEE Transactions on Multimedia, vol. 19, no. 7, pp. 1609–1624, 2017.

[2] A. K. Naveen, B. Singla, R. Wankhade, S. M., R. S., and R. M. R. Guddeti, “AutoOEP: A multi-modal framework for online exam proctoring,” arXiv:2509.10887, 2025.

[3] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[4] B. Burgess, A. Ginsberg, E. W. Felten, and S. Cohney, “Watching the watchers: Bias and vulnerability in remote proctoring software,” arXiv:2205.03009, 2022.
[5] D. G. Balash, D. Kim, D. Shaibekova, R. A. Fainchtein, M. Sherr, and A. J. Aviv, “Examining the examiners: Students’ privacy and security perceptions of online proctoring services,” in Proc. Symposium on Usable Privacy and Security, 2021.
[6] S. Nigam, A. Singh, and A. K. Misra, “Efficient real-time network traffic classification and anomaly detection using deep learning,” International Journal of Communication Systems, vol. 35, no. 5, Art. no. e5063, 2022.
[7] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” arXiv:1804.02767, 2018.
[8] C. Lugaresi et al., “MediaPipe: A framework for building perception pipelines,” arXiv:1906.08172, 2019.
[9] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive angular margin loss for deep face recognition,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4690–4699.
[10] P. Viola and M. Jones, “Robust real-time face detection,” International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.
[11] G. Ke et al., “LightGBM: A highly efficient gradient boosting decision tree,” in Proc. Advances in Neural Information Processing Systems, 2017, pp. 3149–3157.
[12] Y. Liu, J. Ren, J. Xu, X. Bai, R. Kaur, and F. Xia, “Multiple instance learning for cheating detection and localization in online examinations,” IEEE Transactions on Cognitive and Developmental Systems, vol. 16, no. 4, pp. 1315–1326, 2024.
[13] S. S. Shapiro and M. B. Wilk, “An analysis of variance test for normality (complete samples),” Biometrika, vol. 52, no. 3/4, pp. 591–611, 1965.
[14] H. B. Mann and D. R. Whitney, “On a test of whether one of two random variables is stochastically larger than the other,” Annals of Mathematical Statistics, vol. 18, no. 1, pp. 50–60, 1947.
[15] Student [W. S. Gosset], “The probable error of a mean,” Biometrika, vol. 6, no. 1, pp. 1–25, 1908.
[16] M. Malhotra and I. Chhabra, “Ensuring academic integrity through automated online exam proctoring: A decade-long systematic review,” Discover Education, vol. 5, no. 1, 2026.
[17] A. Nurpeisova, A. Shaushenova, Z. Mutalova, M. Ongarbayeva, S. Niyazbekova, A. Bekenova, L. Zhumaliyeva, and S. Zhumasseitova, “Research on the development of a proctoring system for conducting online exams in Kazakhstan,” Computation, vol. 11, no. 6, art. 120, 2023.
[18] P. Verma, N. Malhotra, R. Suri, and R. Kumar, “Automated smart artificial intelligence-based proctoring system using deep learning,” Soft Computing, vol. 28, no. 4, pp. 3479–3489, 2024.