DOI: http://dx.doi.org/10.26483/ijarcs.v17i4.7539

Volume 17, No. 4, July-August 2026

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A COMPREHENSIVE SURVEY OF AI-BASED ONLINE EXAMINATION MONITORING AND BEHAVIOURAL PERFORMANCE ANALYSIS SYSTEMS

Alan C B, Ansha Anilkumar E, Nandana P V, Pooja T K, Aruna A S
B. Tech Information Technology Graduates,
College of Engineering Vadakara
Kerala, India

Abstract: The rapid adoption of online examination systems in educational institutions has introduced significant challenges in ensuring academic integrity, reliable student monitoring, and meaningful performance evaluation. Traditional online proctoring approaches primarily rely on webcam surveillance and screen recording, which are often insufficient to capture subtle behavioural patterns such as student engagement, attention, and cognitive difficulty during examinations. To address these limitations, recent research has explored the use of artificial intelligence and computer vision techniques for automated exam monitoring and behaviour analysis. This survey presents a comprehensive review of existing studies on AI-based online examination systems, focusing on key components such as face recognition for identity verification, mouse behaviour and heatmap analysis for interaction assessment, facial emotion recognition for affective state detection, head pose estimation for attention monitoring, and object detection for exam integrity enforcement. Furthermore, the survey examines multimodal data fusion techniques used to integrate heterogeneous behavioural signals for holistic performance evaluation. By critically analysing the methodologies, algorithms, strengths, and limitations of prior research, this survey identifies existing research gaps and establishes the conceptual foundation for the development of an integrated smart examination platform capable of real-time monitoring and post-exam behavioural performance analysis.

Keywords: Online Examination Systems, AI-Based Proctoring, Mouse Heatmap Analysis, Facial Emotion Recognition, Head Pose Estimation, Object Detection, Multimodal Data Fusion, Behavioural Performance Analysis, Academic Integrity

I. INTRODUCTION

An examination system is a structured mechanism used by educational institutions and professional organizations to evaluate a learner's knowledge, skills, and understanding of a subject. Traditionally, examinations were conducted in physical classrooms under the supervision of invigilators, where question papers were distributed manually and answer scripts were evaluated offline. While such manual examination systems ensured controlled environments and direct supervision, they required significant logistical arrangements, infrastructure, and human resources. Conducting large-scale examinations involved venue allocation, physical monitoring, transportation of question papers, and manual result processing, which often made the system time-consuming and resource-intensive.

With the rapid growth of digital education, e-learning platforms, and remote learning environments, institutions have increasingly shifted from manual examination systems to online examination systems. Online exams enable flexible scheduling, remote participation, automated evaluation, and scalability for large numbers of candidates. They are widely used in universities, competitive examinations, recruitment assessments, certification programs, and corporate training environments. The COVID-19 pandemic further accelerated this transition by necessitating remote assessment mechanisms. However, although online examination systems provide convenience and scalability, they introduce new challenges related to identity verification, exam integrity, student authentication, and monitoring of candidate behaviour. Conventional online proctoring systems primarily rely on webcam surveillance, screen recording, and browser lockdown mechanisms to detect malpractice. These systems attempt to replicate physical invigilation in a virtual setting. While such approaches can detect visible irregularities such as the presence of additional persons or unauthorized devices, they often fail to capture deeper behavioural patterns such as student engagement, attention levels, hesitation, and cognitive difficulty. As a result, most existing systems focus predominantly on cheating detection rather than comprehensive behavioural analysis, limiting the pedagogical insights that can be derived from examination data.

Recent advancements in artificial intelligence (AI)[1], machine learning, and computer vision have enabled the development of more intelligent examination monitoring frameworks. Techniques such as face recognition for identity verification, head pose estimation for attention tracking, facial emotion recognition for affective state analysis, mouse movement and heatmap analytics for interaction assessment, and object detection models for identifying prohibited items have been explored in recent research. Furthermore, multimodal data fusion techniques allow the integration of visual, behavioural, and interaction-based signals to provide a more holistic understanding of student performance beyond final scores alone.

In this context, the objective of this survey is to review and analyse existing AI-based online examination monitoring systems, examine their methodologies, datasets, evaluation strategies, and limitations, and identify key research gaps. By synthesizing current advancements and highlighting persistent challenges, this study establishes the conceptual foundation for developing an integrated smart examination platform capable of real-time monitoring as well as post-exam behavioural performance analysis.

II. RELATED WORKS

A. AI-Based Online Examination Proctoring Systems.

Niharika G. N and Nayak S. N[2] presented an AI-based Online Examination Proctoring System aimed at enhancing academic integrity in remote assessment environments. Early developments in AI-driven online examination monitoring focused primarily on automated visual surveillance mechanisms. One of the foundational contributions in this domain was presented by Atoum et al.[3], who explored automated exam proctoring using computer vision techniques for detecting abnormal head movements, gaze deviations, and environmental anomalies. Their work demonstrated the feasibility of replacing traditional invigilation with intelligent visual monitoring systems, thereby establishing a technical basis for subsequent research in AI-based proctoring frameworks.

The primary objective of their work was to automate the invigilation process by reducing dependence on manual supervision during online examinations. The proposed framework operates on real-time webcam video streams captured during examination sessions rather than on publicly available benchmark datasets. The input data consist of live facial images, head movements, eye gaze behaviour, and environmental activity monitored through the student's camera and microphone.

The system architecture is divided into monitoring and analysis components. During the examination, the monitoring module activates the webcam and continuously captures visual frames. Image preprocessing techniques such as grayscale conversion, noise removal, and segmentation are applied to enhance image clarity. Feature extraction is performed using the Gray Level Co-occurrence Matrix (GLCM)[4] to analyse spatial pixel relationships. For behavioural detection and classification, the system integrates Convolutional Neural Networks (CNN)[5],[6], Haar Cascade classifiers for face detection, and the YOLO v3[7] algorithm for object detection, particularly for identifying mobile phone usage and the presence of additional persons. The system monitors eye movements, head pose deviations, mouth movements, face spoofing attempts, and suspicious environmental activities. When irregular behaviour is detected, alerts are generated and logged in spreadsheet format for subsequent review by human proctors.

The evaluation of the proposed system is primarily demonstrated through functional testing and implementation screenshots rather than large-scale experimental validation. Although the system successfully identifies multiple forms of suspicious activity under controlled testing conditions, the study does not report detailed quantitative performance metrics such as accuracy, precision, recall, F1-score, or Area Under Curve (AUC)[8]. Furthermore, no standardized dataset benchmarking or comparative performance analysis is provided.

While the framework contributes to automated real-time proctoring using computer vision techniques, it exhibits certain limitations. The approach relies heavily on vision-based monitoring without incorporating multimodal behavioural indicators such as mouse dynamics or keystroke patterns. The absence of structured quantitative evaluation limits its generalizability, and post-examination behavioural performance analytics are not considered. Additionally, issues related to scalability and privacy preservation are not extensively addressed within the study.

Several subsequent studies extended vision-based proctoring architectures by incorporating additional behavioural monitoring strategies. Sharma and Tripathi [9] proposed a remote examination monitoring framework that combined facial detection, anomaly tracking, and system-level activity monitoring to strengthen examination security. Similarly, Awaghade et al. [10] introduced an AI-assisted examination platform designed to automatically detect suspicious activities through integrated computer vision models, aiming to minimize reliance on manual supervision.

Biometric authentication has also been investigated as a core component of exam integrity systems. Hadian et al.[11] developed an online examination framework emphasizing face detection and verification to ensure that the registered candidate remains continuously authenticated throughout the examination session. These approaches collectively demonstrate the growing emphasis on identity validation and real-time behavioural surveillance in AI-based proctoring research.

B. Systematic Reviews of Online Examination Solutions.

Muzaffar et al.[12] conducted a comprehensive systematic literature review focusing specifically on online examination solutions within E-learning environments.

The primary objective of their study was to analyse recent developments in online examination systems and to identify major features, underlying development approaches, datasets, tools, and global research trends between January 2016 and July 2020. Unlike experimental studies that implement a single proctoring model, this work synthesizes findings from multiple research contributions to provide a structured overview of the domain.

The authors followed a formal Systematic Literature Review methodology. Six major scientific databases IEEE, Elsevier, Springer, ACM, Wiley, and Taylor & Francis were explored using predefined inclusion and exclusion criteria. After a multi-stage filtering process involving title screening, abstract review, general analysis, and detailed evaluation, 53 studies were selected for final analysis. Rather than relying on a single dataset, the study examined datasets used across the selected works, including video-based, image-based, and text-based datasets for verification, abnormal behavior detection,question bank generation, and evaluation. Public benchmark datasets as well as privately constructed datasets were reported in the surveyed literature.

Methodologically, the review categorized existing online examination solutions into three primary groups: biometric-based systems, software application-based systems, and general hybrid approaches. The study further analysed underlying development strategies such as Machine Learning, AI, Formal Methods, and traditional programming approaches. Frequently adopted techniques identified in the reviewed studies included CNN for face recognition, head pose estimation algorithms, Natural Language Processing (NLP)[13] for automated evaluation, genetic algorithms for question bank optimization, and rule-based application monitoring for abnormal behaviour detection.

The evaluation component of the surveyed works varied considerably. Many studies reported accuracy, precision, recall, F1-score, false acceptance rate (FAR)[14], and false rejection rate (FRR)[14] for authentication and abnormal behaviour detection models. Other works validated their systems through controlled experiments, student participation trials, surveys, or prototype implementations rather than standardized benchmarking. The review highlighted that question bank generation and evaluation (20 studies) and verification and abnormal behaviour detection (19 studies) were the most frequently targeted features during the analysed period.

Despite its broad coverage, the study identified several limitations within the existing research landscape. A significant portion of proposed tools lacked public availability, limiting reproducibility and further validation. Many solutions relied on small or non-standardized datasets, restricting generalizability. Additionally, several systems focused on isolated features such as authentication or question generation without integrating multimodal behavioural analytics. The review also emphasized disparities in adoption across countries, noting that financial constraints, network infrastructure, hardware requirements, and implementation complexity influence practical deployment.

Overall, Muzaffar et al. (2021) provide a structured analytical overview of online examination research trends and highlight the necessity for integrated, scalable, and cost-aware smart examination platforms capable of combining multiple features within a unified framework.

C. Context-Aware and Continuous Authentication in Online Learning.

Ryu et al.[15] conducted a structured survey focusing on context-aware continuous implicit authentication mechanisms in online learning environments.

The study aimed to examine how unobtrusive and adaptive authentication approaches can strengthen academic integrity while maintaining a seamless user experience in digital learning platforms.

The authors employed a Systematic Literature Review methodology. Using predefined search strings related to continuous and context-aware authentication, an initial corpus of 1,766 publications was identified. After applying strict inclusion and exclusion criteria, 17 relevant studies were selected for detailed analysis. The selected works were examined across four analytical dimensions: authenticators used, authentication process design, machine learning techniques adopted, and contextual information integration. The reviewed systems were further categorized based on application scope, including online learning platforms, online examinations, and virtual laboratory environments.

As a survey-based study, the paper did not implement a single dataset but instead analysed datasets used in the reviewed research. The findings revealed that most experimental systems relied on privately collected biometric datasets, often involving small participant groups ranging from fewer than 10 to approximately 30 users. A few large-scale projects included datasets exceeding several hundred or thousand participants. Public benchmark datasets for face recognition and keystroke dynamics were occasionally used, but their availability was limited for context-aware continuous authentication evaluation. Many experiments were conducted under semi-controlled conditions, manipulating variables such as lighting, typing tasks, or user posture to simulate contextual variations.

In terms of methodology, the survey identified face recognition and keystroke dynamics as the most frequently applied authenticators due to their passive data acquisition capability through standard webcams and keyboards. The authentication process was divided into identity assurance (verifying the enrolled student at login and during sessions) and authorship assurance (ensuring that the same individual completes academic tasks). Various machine learning techniques were reported across the reviewed studies, including CNN for facial recognition, object detection frameworks for monitoring, classical classifiers such as Naïve Bayes and Decision Trees for keystroke analysis, and ensemble-based multimodal fusion strategies. Contextual information such as environmental lighting, background noise, typing behaviour, and device interaction patterns was incorporated in several systems to improve robustness.

Evaluation approaches across the surveyed works were predominantly quantitative. Common performance metrics included accuracy, precision, recall, FAR, FRR, and Equal Error Rate (EER)[16]. Some systems additionally measured cheating detection efficiency and fraud identification performance. However, the survey highlighted inconsistencies in evaluation methodologies, making direct comparison between systems difficult. Only a limited number of studies incorporated user-centered evaluation measures such as usability, user acceptance, or system disturbance assessment. The authors identified several limitations within the existing research landscape. Most systems concentrated primarily on online exam scenarios rather than the entire learning lifecycle. Authenticator diversity was limited, with heavy reliance on facial and keystroke biometrics. Template ageing and long-term biometric variation were insufficiently addressed. Large-scale validation in real-world educational environments remained scarce. Furthermore, privacy concerns and vulnerability to presentation attacks were not comprehensively analysed in many works. The study concluded by emphasizing the need for scalable, adaptive, and privacy aware continuous authentication frameworks for online learning ecosystems.

D. Mouse Behavior and Heatmap-Based Learning Analytics

Khaesawad et al.[17] proposed a machine learning-based framework for classifying students' Programming Logic Understanding Levels (PLUL) using mouse-tracking heatmap representations generated from interactions within a block-based coding platform.

The primary objective of the study was to support instructors in identifying students with lower levels of programming comprehension during introductory programming courses by analysing behavioural interaction data rather than relying solely on traditional assessment scores.

The dataset used in the study was collected from 50 first-year university students enrolled in a software engineering program. Data collection was conducted under controlled experimental conditions to reduce variability due to hardware and environmental factors. The dataset consisted of two components: (1) PLUL test scores derived from a structured 30-question assessment covering sequential execution, conditional statements, iteration, and function usage; and (2) time-series mouse-tracking data captured while students completed assigned tasks on a block coding learning platform. The PLUL scores were grouped into low and high categories using the K-means clustering algorithm, which generated centroid values distinguishing the two performance groups.

Methodologically, the raw mouse interaction data including cursor coordinates, click events, and duration of mouse stops were transformed into three distinct heatmap images representing mouse movement, mouse click activity, and mouse duration time. These images were resized and combined in different configurations to generate multiple heatmap representations. The heatmaps were converted into feature matrices, and Principal Component Analysis (PCA) was applied to reduce dimensionality while retaining at least 95% of the variance. Five supervised machine learning classifiers Support Vector Machine (SVM)[18], k-Nearest Neighbours (k-NN)[19], Decision Tree[20], Random Forest[21], and Multi-Layer Perceptron (MLP)[22] were trained using these features. Model performance was evaluated using Leave-One-Out Cross-Validation (LOOCV)[23] to address the limited sample size.

The experimental results demonstrated that the Decision Tree classifier applied to a combined mouse movement and mouse duration heatmap representation achieved the highest performance. The reported metrics included an accuracy of 76.00%, F1-score of 77.36%, sensitivity of 75.93%, and specificity of 76.09%. These findings indicate that mouse interaction patterns can provide meaningful indicators of students' conceptual understanding during programming tasks. The study also observed that combining movement and duration features produced more discriminative representations than click-based heatmaps alone.

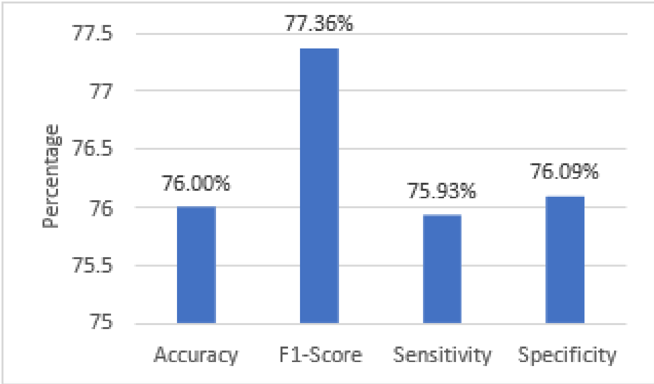

Figure 1. Performance metrics of the Decision Tree classifier using combined mouse movement and duration heatmap representations (adapted from [17]).

Despite its contributions, the study presents certain limitations. The dataset size was relatively small, consisting of only 100 samples (pre- and post-course data from 50 students), which restricts generalizability. Deep learning models such as CNN were not employed due to concerns regarding overfitting with limited data. Additionally, the framework focuses exclusively on behavioural interaction analytics and does not incorporate biometric modalities such as facial recognition or head pose estimation. The applicability of the approach beyond block-based coding environments was not extensively validated, and large-scale real-world deployment remains to be explored.

Overall, this work demonstrates the feasibility of using mouse-tracking heatmap analytics combined with traditional machine learning algorithms to estimate programming comprehension levels, highlighting the potential of behavioural interaction data in educational assessment contexts.

E. Vision-Based AI Proctoring Systems

Satre et al.[24] proposed an AI-based Online Exam Proctoring System designed to enhance academic integrity in remote assessment environments.

Beyond purely visual monitoring, multimodal audio-visual surveillance mechanisms have also been explored. Prathish [25] proposed an examination monitoring system that integrates facial analysis with environmental audio monitoring to identify potential irregularities such as external assistance or suspicious communication. By combining visual and auditory cues, such systems attempt to enhance detection robustness compared to single-modality approaches.

The primary objective of the study was to develop a web-based automated examination framework capable of detecting and flagging suspicious behaviour during online exams without the need for physical examination center's.

The system operates using real-time webcam video streams captured during the examination session rather than a standardized benchmark dataset. The input data primarily consist of live facial images, head movements, environmental visuals, and object presence within the camera frame. The authors focused on monitoring behaviours such as lip movement, mobile phone detection, detection of additional persons in the frame, and abnormal head orientation. Methodologically, the framework integrates computer vision and deep learning techniques. The YOLO algorithm is employed for object detection, enabling identification of unauthorized items such as mobile phones and additional individuals within the examination environment. Face detection and recognition are implemented using OpenCV[26]-based approaches, including CNN and deep learning-based face encoding models. The system also incorporates anomaly detection mechanisms to identify unusual head movements and suspicious behavioural patterns. An alert mechanism is integrated into the architecture to notify instructors in real time when potential misconduct is detected. The system architecture includes user authentication, background application monitoring, exam engine integration, and teacher-side monitoring dashboards, as illustrated in the block diagram provided in the implementation section of the paper.

The evaluation of the system is primarily demonstrated through implementation results and functional testing rather than large-scale quantitative benchmarking. The results section presents example scenarios where the system successfully detects head movement deviations, mobile phone usage, and user verification during exam attempts (as shown in the result figures). However, the study does not provide detailed statistical performance metrics such as accuracy, precision, recall, or F1-score. Additionally, no standardized dataset validation or comparative experimental analysis is reported.

Although the proposed system contributes a practical implementation of AI-driven proctoring with real-time monitoring and alert generation, several limitations are evident. The framework relies heavily on vision-based monitoring and does not incorporate behavioural interaction analytics such as mouse tracking or keystroke dynamics. Quantitative evaluation[27] using large-scale datasets is absent, limiting generalizability. Furthermore, privacy implications and computational scalability in large institutional deployments are not extensively discussed. The system also focuses primarily on real-time anomaly detection without incorporating post-examination behavioural performance analysis.

Overall, the work by Satre et al. demonstrates the feasibility of integrating YOLO-based object detection and face recognition algorithms within a web-based proctoring framework, contributing to the advancement of automated remote examination monitoring systems.

III. COMPARATIVE ANALYSIS OF SURVEYED SYSTEMS

To provide a clearer understanding of methodological differences, dataset characteristics, evaluation strategies, and research gaps among the surveyed studies, a comparative analysis is presented in Table 1. The comparison highlights variations in data sources, machine learning techniques, reported performance metrics, and identified limitations.

Table 1. Comparative analysis of surveyed systems.

Method	Dataset Type	Methodology	Evaluation Metric	Key Limitation
AI-Based Online Examination Proctoring Systems[2]	Real-Time Webcam Data	CNN, Haar Cascade, YOLOv3, GLCM	Functional validation (no detailed statistical metrics)	Vision-Only approach; no multimodal fusion.
Systematic Reviews of Online Examination Solutions[12]	53 Selected studies (SLR review)	Systematic Literature Review	Accuracy, FAR, FRR (reported across studies)	No unified Integrated framework
Context- Aware and Continuous Authentication in Online Learning[15]	17 Selected Studies	Continuous authentication using CNN & keystroke dynamics	Accuracy, EER, Precision, Recall	Limited large-scale real-world validation
Mouse Behaviour and Heatmap- Based Learning Analytics[17]	50 Students (mouse-tracking dataset)	PCA + SVM, k-NN, Decision Tree, MLP	Accuracy-76%, F1 -77.36%, Sensitivity-75.93%	Small dataset; no Biometric integration
Vision-Based AI Proctoring Systems[24]	Real- Time webcam video	YOLO-based object detection, OpenCV face recognition	Functional testing (no benchmark metrics)	Lack of Quantitative benchmarking

Figure 2 presents a structured taxonomy of AI-based online examination monitoring systems identified in the surveyed literature. The classification groups existing approaches into vision-based proctoring, continuous authentication, behavioural interaction analytics, and multimodal integrated systems. This taxonomy highlights the methodological progression from single-modality surveillance frameworks toward comprehensive multimodal behavioural analysis architectures.

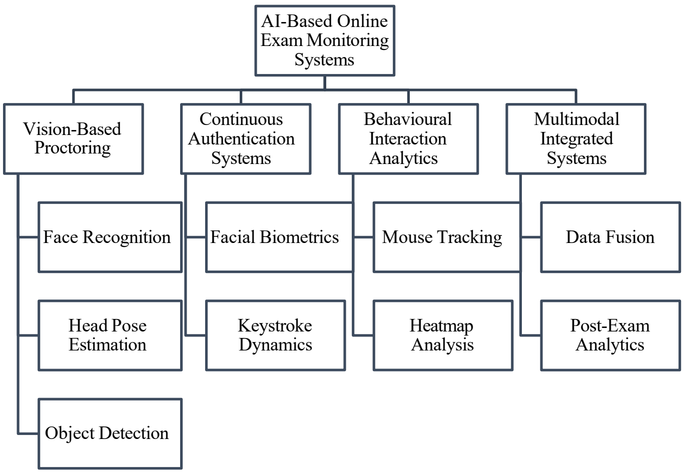

Figure 2. Taxonomy of AI-Based Online Examination Monitoring Systems

IV. LIMITATIONS AND ENHANCEMENTS IN EXISTING SYSTEMS

A comprehensive analysis of the surveyed literature reveals that although significant progress has been made in AI-based online examination monitoring, several technical, methodological, and practical limitations persist across existing systems.

One of the most prominent limitations is the over-reliance on vision-based monitoring techniques. The majority of existing proctoring frameworks primarily depend on webcam-based face detection, object detection, and head pose estimation. While these approaches are effective in identifying visible misconduct such as mobile phone usage or the presence of additional persons, they do not adequately capture deeper cognitive and behavioural indicators such as hesitation patterns, engagement levels, or problem-solving difficulty. This narrow focus restricts the analytical depth of current proctoring systems.

Another observed limitation is the lack of multimodal data integration. Most surveyed systems operate using a single modality either facial biometrics, object detection, or interaction logs without fusing multiple behavioural signals into a unified analytical framework. Even in cases where multiple detection mechanisms are present, they often function independently rather than through structured data fusion. As a result, contextual understanding of student behaviour remains fragmented and incomplete.

A further challenge lies in limited dataset size and lack of standardized benchmarking. Several experimental studies rely on small, institution-specific datasets collected under controlled conditions. The absence of publicly available large-scale datasets for online examination behaviour reduces reproducibility and restricts performance comparison across different systems. Moreover, many implementations report functional validation rather than comprehensive statistical evaluation using standardized metrics such as precision, recall, F1-score, AUC, and EER.

Privacy and ethical considerations also remain insufficiently addressed in many frameworks. Continuous video monitoring raises concerns regarding data storage, user consent, biometric misuse, and surveillance anxiety. Few studies propose privacy-preserving mechanisms such as edge processing, data anonymization, or federated learning architectures.

Scalability is another critical issue. Real-time deep learning models, particularly those based on CNN and object detection algorithms, demand substantial computational resources. This may limit deployment feasibility in institutions with constrained infrastructure or in environments with unstable internet connectivity.

In addition, most systems emphasize real-time cheating detection but overlook post-examination behavioural analytics. Current frameworks typically generate alerts during the examination process but do not provide comprehensive performance dashboards analysing engagement trends, time allocation per question, confusion indicators, or behavioural patterns correlated with academic outcomes. Consequently, the pedagogical value of collected behavioural data remains underutilized.

V. POTENTIAL ENHANCEMENTS AND FUTURE RESEARCH DIRECTIONS

To address these limitations, future online examination systems should adopt a multimodal integrated architecture[28] that combines visual monitoring, behavioural interaction analytics, and contextual data within a unified analytical model. Data fusion techniques can enhance reliability by correlating facial cues, mouse dynamics, head movement patterns, and object detection results.

The integration of behavioural analytics, such as mouse heatmap analysis and interaction timing metrics, can provide deeper insights into student engagement and cognitive load. Unlike purely vision-based methods, behavioural interaction signals offer a less intrusive yet informative mechanism for assessing attention and problem-solving strategies.

Standardized evaluation protocols and larger datasets should be developed to improve reproducibility and comparability across systems. Publicly available benchmark datasets for online examination behaviour would significantly strengthen research validation.

To address privacy concerns, future systems should explore privacy-aware AI mechanisms, including on-device processing, encrypted storage[29], consent-driven data handling policies, and federated learning models that reduce centralized data collection.

Scalability can be improved through lightweight model architectures and hybrid cloud-edge[30],[31] processing strategies that distribute computational load efficiently.

Finally, online examination systems should evolve beyond surveillance tools and transform into intelligent assessment ecosystems. By incorporating post-exam analytics dashboards, performance profiling, engagement scoring, and personalized feedback generation, AI-driven platforms can support both academic integrity and pedagogical improvement.

VI. CONCLUSION

The continuous growth of digital education has intensified the need for secure and intelligent online examination mechanisms capable of ensuring academic integrity while supporting meaningful assessment practices. This review critically analysed recent research contributions in AI-driven examination monitoring, encompassing vision-based proctoring systems, continuous authentication models, mouse interaction analytics, and emerging multimodal behavioural analysis frameworks. The surveyed studies demonstrate that AI techniques including CNN, object detection algorithms, biometric verification methods, and classical machine learning classifiers have significantly enhanced the automation of remote invigilation processes.

Despite these advancements, the literature reveals notable research gaps. A considerable proportion of existing systems depend predominantly on webcam-based surveillance, which limits behavioural interpretation to visible anomalies and increases the risk of misclassification. Comprehensive multimodal integration remains underexplored, and temporal modelling of user behaviour across examination sessions is rarely implemented. Furthermore, many frameworks lack interpretability, standardized benchmarking protocols, and structured post-examination analytics capable of generating pedagogically valuable insights. Practical concerns related to data privacy, computational scalability, and ethical AI deployment further complicate large-scale adoption.

Addressing these challenges requires a shift toward more holistic system architectures. Future online examination platforms should incorporate multimodal data fusion strategies that combine biometric, behavioural, and contextual signals within unified analytical models. Temporal deep learning approaches such as Long Short-Term Memory (LSTM) networks and transformer-based architectures can improve sequential behaviour understanding. The integration of explainable AI techniques will enhance transparency and trust in automated decision-making processes. In addition, privacy-preserving solutions including federated learning and edge-based AI processing must be prioritized to ensure responsible data handling.

In summary, while AI-enabled examination monitoring technologies have evolved considerably, the next phase of development must emphasize integration, interpretability, scalability, and pedagogical relevance. By aligning advanced AI methodologies with ethical and educational considerations, future systems can move beyond mere surveillance and evolve into comprehensive intelligent assessment ecosystems that support both examination security and improved learning outcomes. The integration of multimodal behavioural analytics with privacy-aware AI architectures represents the next evolutionary step in intelligent digital assessment ecosystems.

VII. DECLARATIONS

All authors declare that they have no conflicts of interest.

VIII. REFERENCES

[1] Valavanidis, A. (2023). Artificial intelligence (ai) applications. *Department of Chemistry, National and Kapodistrian University of Athens, University Campus Zografou*, *15784*.

[2] N. G. Niharika and S. N. Nayak, "Artificial Intelligence Based Online Examination Proctoring System," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 11, no. IX, pp. 569–575, Sep. 2023.

[3] Y. Atoum, L. Chen, A. Liu, S. D. Hsu, and X. Liu, "Automated Online Exam Proctoring," *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1609–1624, Jul. 2017.

[4] Pi, P., & Lima, D. (2021). Gray level co-occurrence matrix and extreme learning machine for Covid-19 diagnosis. *International Journal of Cognitive Computing in Engineering*, *2*, 93-103.

[5] Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H. I. K., Frisky, A. Z. K., & ul Haq, Q. M. (2023).Understanding of convolutional neural network (cnn): A review. *International Journal of Robotics and Control Systems*, *2*(4), 739-748.

[6] Aruna, A. S., Babu, K. R., & Deepthi, K. (2025). A deep drug prediction framework for viral infectious diseases using an optimizer-based ensemble of convolutional neural network: COVID-19 as a case study. Molecular Diversity, 29(3), 2473-2487.

[7] Alsanad, H. R., Sadik, A. Z., Ucan, O. N., Ilyas, M., & Bayat, O. (2022). A-V3 based real-time drone detection algorithm. *Multimedia tools and applications*, *81*(18), 26185-26198.

[8] Turner, J. R. (2020). Area under the curve (AUC). In *Encyclopedia of Behavioral Medicine* (pp. 146-146). Cham: Springer International Publishing.

[9] P. Sharma and U. Tripathi, "AI-Based Remote Examination Monitoring System," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, pp. 311–318, 2021.

[10] D. Awaghade, T. Bombe, T. Deshmukh, and K. Takawane, "Automated Online Examination System with AI-Based Monitoring," *International Journal of Computer Applications*, vol. 176, no. 29, pp. 1–6, 2020.

[11] H. S. G. Hadian, Y. Bandung, and A. A. Putra, "Online Examination System Using Face Detection and Verification," *Proc. Int. Conf. Information Technology Systems and Innovation*, pp. 1–6, 2019.

[12] A. W. Muzaffar, M. Tahir, M. W. Anwar, Q. Chaudry, S. R. Mir, and Y. Rasheed, "A Systematic Review of Online Exams Solutions in E-Learning: Techniques, Tools, and Global Adoption," *IEEE Access*, vol. 9, pp. 32689–32712, 2021, doi: 10.1109/ACCESS.2021.3060192.

[13] Fanni, S. C., Febi, M., Aghakhanyan, G., & Neri, E. (2023). Natural language processing. In *Introduction to artificial intelligence* (pp. 87-99). Cham: Springer International Publishing.

[14] Lim, C. Y., Markus, C., & Loh, T. P. (2021). Precision verification: effect of experiment design on false

acceptance and false rejection rates. *American Journal of Clinical Pathology*, *156*(6), 1058-1067.

[15] R. Ryu, S. Yeom, D. Herbert, and J. Dermoudy, "A Comprehensive Survey of Context-Aware Continuous Implicit Authentication in Online Learning Environments," *IEEE Access*, vol. 11, pp. 24561–24585, 2023, doi: 10.1109/ACCESS.2023.3253484.

[16] Zhang, L., Wang, X., Cooper, E., Evans, N., & Yamagishi, J. (2023). Range-based equal error rate for spoof localization. *arXiv preprint arXiv:2305.17739*.

[17] A. Khaesawad, Y. Fukuchi, V. Yem, and N. Nishiuchi, "Machine Learning-Based Classification of Programming Logic Understanding Levels by Mouse-Tracking Heatmaps," *IEEE Access*, vol. 13, pp. 89905–89918, 2025, doi: 10.1109/ACCESS.2025.3571050.

[18] Valkenborg, D., Rousseau, A. J., Geubbelmans, M., & Burzykowski, T. (2023). Support vector machines. *American journal of orthodontics and dentofacial orthopedics*, *164*(5), 754-757.

[19] Acito, F. (2023). k nearest neighbors. In *Predictive Analytics with KNIME: Analytics for Citizen Data Scientists* (pp. 209-227). Cham: Springer Nature Switzerland.

[20] Priyanka, & Kumar, D. (2020). Decision tree classifier: a detailed survey. *International Journal of Information and Decision Sciences*, *12*(3), 246-269.

[21] Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69-79.

[22] Kruse, R., Mostaghim, S., Borgelt, C., Braune, C., & Steinbrecher, M. (2022). Multi-layer perceptrons. In *Computational intelligence: a methodological introduction* (pp. 53-124). Cham: Springer International Publishing.

[23] Sreedharan, R., Prajapati, J., Engineer, P., & Prajapati, D. (2023). Leave-one-out cross-validation in machine learning. In *Ethical issues in AI for bioinformatics and chemoinformatics* (pp. 56-71). CRC Press.

[24] S. Satre, S. Patil, T. Mane, V. Molawade, T. Gawand, and A. Mishra, "Online Exam Proctoring System Based on Artificial Intelligence," *Proc. IEEE Int. Conf.*, 2023.

[25] S. Prathish, "Audio-Visual Based Online Exam Proctoring System," *International Journal of Engineering Research & Technology*, vol. 9, no. 6, pp. 120–125, 2020.

[26] A Mishra, S., Verma, V., Akhtar, N., Chaturvedi, S., & Perwej, Y. (2022). An intelligent motion detection using OpenCV. *International Journal of Scientific Research in Science, Engineering, and Technology*, *9*(2), 51-63.

[27] A Lim, W. M. (2025). What is quantitative research? An overview and guidelines. *Australasian Marketing Journal*, *33*(3), 325-348.

[28] A. Vieira, S., Pinaya, W. H. L., Garcia-Dias, R., & Mechelli, A. (2020). Multimodal integration. In *Machine Learning* (pp. 283-305). Academic Press.

[29] Manthiramoorthy, C., & Khan, K. M. S. (2024). Comparing several encrypted cloud storage platforms. *International Journal of Mathematics, Statistics, and Computer Science*, *1*, 44-62.

[30] Pal, S., Jhanjhi, N. Z., Abdulbaqi, A. S., Akila, D., Almazroi, A. A., & Alsubaei, F. S. (2023). A hybrid edge-cloud system for networking service components optimization using the internet of things. Electronics, 12(3), 649.

[31] Liang, Q., Shenoy, P., & Irwin, D. (2020, October). Ai on the edge: Characterizing ai-based IOT applications using specialized edge architectures. In IEEE International *Symposium on Workload Characterization (IISWC)*

