Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 3, June 2026, pp. 2435~2448
ISSN: 2302-9285, DOI: 10.11591/eei.v15i3.11207

Fast and accurate cheat detection using deep learning algorithms

Ilhame Khabbachi, Ghalia Mdaghri Alaoui, Abdelhamid Zouhair
DSAI2S Research Team, Computer Science and Smart Systems Laboratory, FST of Tangier, Abdelmalek Essaâdi University, Tetouan, Morocco

Article Info

Article history:

Received Aug 14, 2025
Revised Mar 17, 2026
Accepted Mar 31, 2026

Keywords:

Artificial intelligence
Computer vision
Deep learning
E-cheating detection
Online exams

ABSTRACT

The rapid expansion of online education, accelerated by the global health crisis of 2020, has introduced significant challenges in maintaining academic integrity due to the absence of physical supervision during remote examinations. As digital learning becomes a permanent component of modern education, ensuring fairness and credibility in online assessments has become a critical concern for educational institutions. This study proposes an intelligent deep learning (DL)–based framework for detecting non-compliant behaviors during online examinations using standard webcam video streams. The proposed system integrates real-time video monitoring with automated behavioral analysis by combining object detection and classification models. In particular, you only look once version 5 (YOLOv5) is employed for efficient facial and object detection, while a convolutional neural network (CNN) is used to classify cheating and non-cheating behaviors from extracted visual features. Experimental results demonstrate that the integrated YOLOv5–CNN architecture achieves high detection accuracy and low inference latency, making it suitable for real-time and scalable deployment in online proctoring systems. By enabling objective and automated monitoring, the proposed framework contributes to strengthening fairness, transparency, and trust in digital assessment environments, thereby supporting the long-term sustainability of online education.

Corresponding Author:

Ilhame Khabbachi
DSAI2S Research Team, Computer Science and Smart Systems Laboratory, FST of Tangier
Abdelmalek Essaâdi University
Tetouan, Morocco
Email: ilhame.khabbachi@etu.uae.ac.ma

1. INTRODUCTION

A significant turning point in contemporary education was the quick switch to online instruction during the COVID-19 pandemic. This change created serious issues with academic integrity during distant exams, even though it guaranteed the continuation of academic activities. Online tests offer less control over the students' surroundings than traditional in-person evaluations, which reduces the efficacy of standard supervision techniques. The extensive use of digital devices and the lack of direct physical supervision have expanded potential for dishonest activities, as noted in [1], raising questions about the validity and fairness of online examinations.

Researchers have increasingly looked at artificial intelligence (AI) and deep learning (DL)-based solutions to these problems. DL models offer strong capabilities for assessing motion dynamics, behavioral patterns, and visual data in video streams because they are based on intricate neural network structures. These methods make it possible to automatically identify suspect activities, like frequent gaze diversion, unusual

head movements, or the usage of unapproved gadgets during exams [2]. These systems are able to recognize tiny behavioral indicators that human observers might find difficult to notice by utilizing supervised and unsupervised learning techniques.

Several studies have demonstrated the potential of computer vision and DL methods in this field. Convolutional neural networks (CNNs), for example, have been widely used to analyze facial expressions and eye movements to detect anomalous behaviors during online exams [3]. Other approaches employ object detection techniques to monitor the student's environment, identify suspicious objects, or detect the presence of additional individuals within the camera's view. These systems have shown promising results in reducing academic misconduct and providing objective evidence when suspicious activities occur.

Despite tremendous advancements, a number of obstacles still exist. While real-time student behavior monitoring has received less attention, many current methods are primarily focused on object identification or are intended for offline analysis. As a result, these methods might not be completely suitable for real-world online proctoring environments where suspicious activity needs to be detected right away. Furthermore, issues like false alarms, misclassification, and inadequate resilience can have an impact on the efficacy of present systems. Beyond the technical considerations, ethical concerns also require consideration, especially those pertaining to data protection, invasive surveillance, and the requirement to secure students' consent.

To overcome these limitations, this paper suggests a hybrid DL framework for real-time cheating detection in online exams in order to get over these restrictions. The suggested method classifies observable actions into cheating and non-cheating categories by combining you only look once version 5 (YOLOv5)'s object identification capabilities with a CNN. This integration enables continuous monitoring of students during online assessments while maintaining high detection accuracy and efficient processing performance.

The main contributions of this work are summarized as follows: i) development of a hybrid YOLOv5–CNN framework for detecting suspicious behaviors during online examinations; ii) implementation of a real-time cheating detection system capable of monitoring student activities during remote assessments; iii) accurate classification of detected behaviors into cheating and non-cheating categories using DL techniques; and iv) experimental validation demonstrating the effectiveness of the proposed approach for intelligent online proctoring systems.

The remainder of this paper is organized as follows. Section 2 presents the background and challenges related to online examination monitoring. Section 3 reviews related work in cheating detection using AI and computer vision. Section 4 describes the proposed methodology and system architecture. Section 5 presents the experimental results. Section 6 discusses the findings and performance analysis of the proposed methodology. Finally, section 7 concludes the paper and outlines directions for future research.

2. BACKGROUND

In this section, we will explore the foundations of various techniques used for object detection and cheating detection, including machine learning and the different types of DL algorithms.

2.1. Machine learning

By allowing systems to learn from data and experience, machine learning, a subfield of AI, seeks to mimic human intelligence. In today's data-driven environment, it has become an essential tool for information extraction and decision support [4]. Through data analysis, machine learning encompasses a range of methods that enable models to perform better over time.

Machine learning methods are generally classified into three main categories. The first is supervised learning, in which models are trained with labeled datasets to discover how inputs and outputs relate to one another and produce precise predictions [5]. The second is unsupervised learning, which looks for hidden patterns or clusters in the dataset by analyzing unlabeled data [6]. The third is reinforcement learning, which uses reward-based feedback to maximize decision-making through trial-and-error interactions with the environment [7]. When combined, these learning strategies enable systems to adapt, maximize performance, and more effectively complete challenging tasks, all of which contribute to the quick development of AI.

2.2. Deep learning

DL is a sophisticated machine learning technique that learns complicated representations from massive datasets using artificial neural networks. In recent years, it has been widely applied in various domains such as speech recognition, natural language processing (NLP), machine translation, bioinformatics, medical image analysis, climate science, and strategic game playing [8]. The main goal of DL is to design neural networks that can automatically recognize patterns and extract pertinent information to facilitate effective decision-making [9].

DL models have shown performance that can match or even surpass human capabilities in a variety of applications. Artificial neural networks are inspired by the structure and operation of biological neural systems, especially the way neurons receive and transfer information. This biologically inspired architecture contributes significantly to the remarkable success of DL in numerous fields.

2.2.1. Convolutional neural network

CNNs were initially designed for image recognition and have since gradually established themselves as versatile architectures, now finding applications across a wide range of domains, including NLP, biomedical analysis, and autonomous driving [10], [11]. CNNs are particularly effective at detecting local features within multidimensional data, making them well suited for pattern recognition tasks [11].

For instance, a CNN can identify objects within an image, such as a wheel or a face, regardless of their position. As illustrated in Figure 1, a typical CNN processes multidimensional inputs—such as images or embeddings—through convolutional layers composed of multiple filters designed to extract distinct features [12]. These filters scan different regions of the input to capture meaningful patterns.

To improve efficiency and reduce computational complexity, the extracted features are often pooled or subsampled, reducing the data dimensionality before being passed to fully connected layers. This hierarchical feature extraction process allows CNNs to achieve high performance in computer vision and related applications [13], [14].

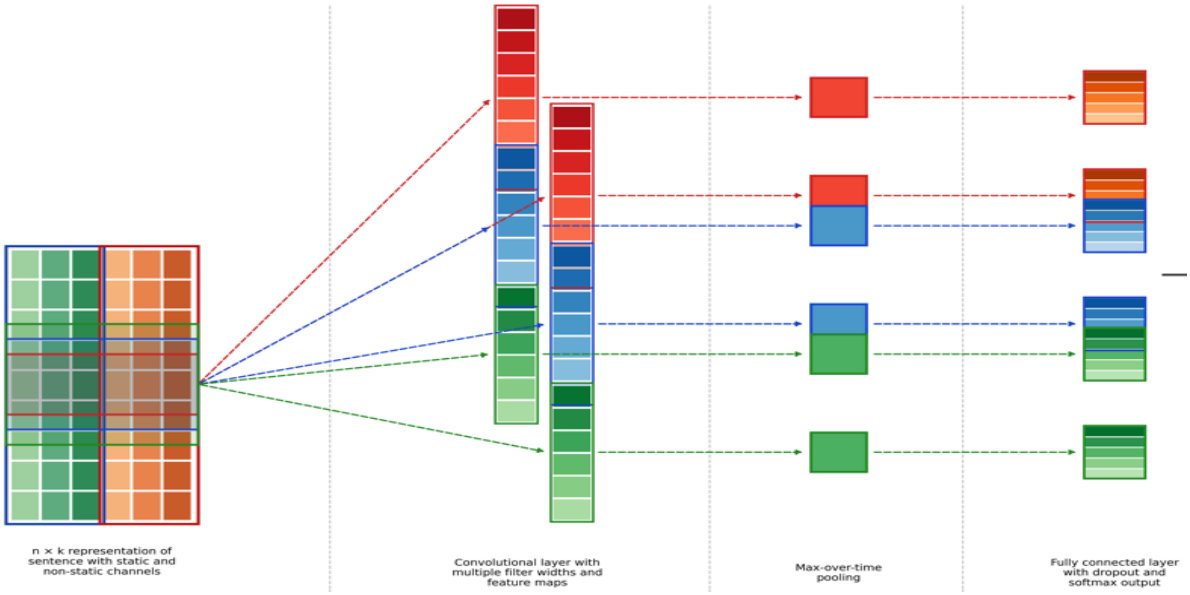


Figure 1. CNN architecture [15]

2.2.2. You only look once version5

Object detection is a fundamental computer vision task that consists of recognizing and localizing objects within images or video sequences. Conventional techniques, including sliding window approaches and region proposal-based methods, tend to be computationally demanding. To address these shortcomings, Redmon proposed YOLO, a model that reframes object detection as a regression problem, simultaneously predicting bounding boxes and class probabilities through a single CNN applied to the full image [16].

In contrast to multi-stage detection approaches, YOLO partitions the input image into a grid and directly produces bounding box coordinates and class predictions from CNN feature maps. The model relies on a unified loss function that jointly accounts for localization and classification errors, while backbone architectures such as Darknet-53 allow for efficient feature extraction and real-time inference.

Several improved versions of YOLO, including YOLOv2, YOLOv3, and YOLOv4, have been proposed to improve detection accuracy and speed. The general architecture of the YOLO framework is illustrated in Figure 2.

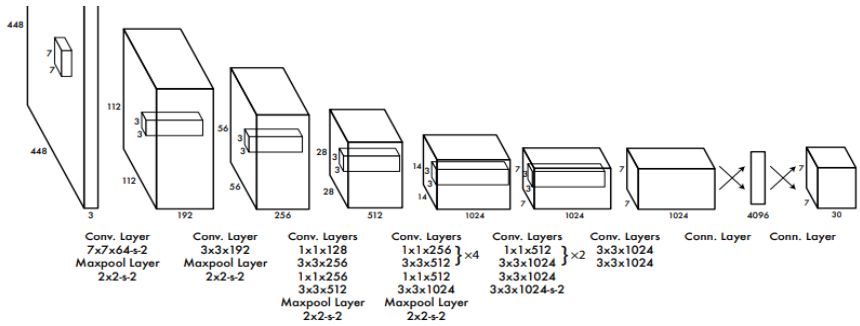


Figure 2. YOLO architecture for object detection [17]

3. RELATED WORK

AI, particularly DL, has significantly advanced automated exam monitoring by enabling the analysis of large-scale unstructured data, such as video recordings, eye movements, and facial expressions. CNNs and recurrent neural networks (RNNs or LSTMs) have been employed to capture subtle suspicious behaviors, including frequent attention diversion, the presence of unauthorized individuals, or phone usage. These systems enhance the integrity of online examinations while reducing the need for human supervision.

Traditional online proctoring systems mainly focus on identity verification and often fail to detect abnormal behaviors. DL approaches analyze webcam data, including head and mouth movements, to identify suspicious activities in real time. As noted by Cote *et al.* [18], such AI-assisted systems improve exam integrity and support a learner-centered assessment environment. Some solutions further monitor the presence of other individuals or device usage [19], while others employ eye-tracking tools or dual-camera setups to track candidate attention [20]. Despite their advantages, widespread implementation remains challenging due to technical and logistical constraints. To address these issues, Ketab *et al.* [21] proposed a multi-modal system that combines audio, video, and screen capture with rule-based inference, facial recognition, and activity monitoring, providing instructors with actionable insights for online assessments.

Tiong and Lee [22] developed an e-cheating detection system for controlled scenarios, integrating video tools such as bullet and wireless IP Cameras with an intelligent agent composed of an IP detector and a behavior detector. This system actively monitors student activity, detects suspicious behavior, and can generate randomized test questions, demonstrating its effectiveness across multiple datasets. In broader studies, four DL models— deep neural network (DNN), dense long short-term memory (LSTM), LSTM, and RNN—were tested for detecting academic dishonesty, achieving an average accuracy of 90%, with dense LSTM reaching 95.32%. These results indicate that DL models can effectively identify suspicious activities during exams and support further research.

The challenges of online proctoring have led institutions to adopt automated monitoring technologies. Masud *et al.* [23] implemented a video analysis method to extract multiple data types, achieving a high detection accuracy of 97%. Shdaifat *et al.* [24] proposed a robust multi-biometric e-assessment system, combining eye tracking, voice recognition, head motion tracking, and 3D facial recognition. Tested on 51 participants, the system demonstrated exceptional accuracy with very low false rejection rates. Similarly, Cote *et al.* [18] explored mobile exam authentication, integrating traditional login with iris recognition and random checks, efficiently detecting misconduct while handling large data volumes.

DL techniques, including CNN, RNN, LSTM, and generative adversarial networks (GANs), have been widely applied for detecting cheating behaviors. Atoum *et al.* [25] proposed a CNN-based face anti-spoofing method. Khan *et al.* [26] employed a modified Faster region based convolutional neural network (RCNN) with Open-Pose for real-time monitoring of head movements and prohibited object usage. Kohli *et al.* [27] developed a motion recognition model achieving 95% accuracy, generating descriptions of suspicious activities. Hussein *et al.* [28] introduced a framework to detect and classify cheating in video frames, improving the timely identification of academic dishonesty.

Malhotra and Chhabra [29] reviewed various human and automated invigilation systems, emphasizing that DL-based approaches such as Faster RCNN and YOLO are more efficient and accurate than traditional methods. Wang *et al.* [9] proposed a DL-based solution for detecting cheating behaviors, demonstrating AI's potential in enhancing exam security. Hu *et al.* [30] combined YOLO object recognition with human posture estimation to track candidates and detect behaviors like peeping or passing notes. Indi *et al.* [31] introduced a hybrid classifier system combining gaze analysis and head posture detection, achieving 96% accuracy in attention monitoring.

Further studies employed clustering and transfer learning techniques to enhance detection. Noorbehbahani *et al.* [32] proposed a SOINN-based clustering algorithm for mixed data, while Ashwinkumar *et al.* [33] combined YOLO, MPGazeII, and VGG16 models to aggregate anomaly detection results for automated cheating identification. Li and Liu [34] used DNNs to predict academic behavior based on student performance, and Duhaim *et al.* [35] developed a recommendation system integrating statistical and clustering methods to detect cheating in online exams.

The Table 1 compares the studies based on their titles, years, data used, and algorithms implemented for detecting cheating or monitoring online exams.

Table 1. Overview of recent AI-based approaches for cheating detection in online examinations

Title	Year	Data	Algorithms used	Key findings
Automated online exam proctoring [19]	2017	Eye gaze and facial images	Eye monitor, dual vision cameras, and object detection	Detects unauthorized persons and electronic devices appearing in the examination environment.
E-cheating prevention measures: detection of cheating at online examinations using deep learning approach — a case study [22]	2021	Mid-term and final-term exam data	DNN, dense LSTM, LSTM, and RNN	Proposes an e-cheating intelligence agent composed of an IP detector and a behavior detector.
Research on abnormal behavior detection of online examination based on image information [30]	2018	Webcam data (head posture and mouth condition)	DL and AI-based proctoring	Detects suspicious behaviors using facial cues, head posture and mouth movement analysis during online exams.
Detection of malpractice in e-exams by head pose and gaze estimation [31]	2021	Visual focus of attention (VFOA) data — head pose and eye gaze video recordings of students during online exams	Head pose estimation, eye gaze estimation, hybrid classifier (Classifier 1 + Classifier 2), and machine learning-based VFOA categorization	Proposes an end-to-end system that detects malpractice in e-exams by analysing head pose and gaze direction.
Towards effective and efficient online exam systems using deep learning-based cheating detection approach [36]	2022	OEP dataset (video frames and speech data from 24 students	CNN and Gaussian-based DFT	Proposes a three-module system (front camera, rear camera, and speech-based detection) using CNN and Gaussian-based DFT for real-time cheating classification during online exams.
Automated cheating detection based on video surveillance in the examination classes [37]	2022	Surveillance video	Video analysis and classification model	Analyzes video streams to identify abnormal movements and potential cheating behaviors.
Online student authentication and proctoring system based on multimodal biometrics technology [38]	2021	Biometric data, eye movements, and voice	Biometrics, eye tracking, and 3D face recognition	Uses biometric authentication and eye-tracking techniques to ensure student identity and attention.

4. RESEARCH METHOD

4.1. Architecture

Figure 3 illustrates the comprehensive architecture of our proposed online exam fraud detection system. The system operates through a hierarchical processing pipeline that begins with the central online exam monitoring system, which manages the overall surveillance infrastructure. Real-time video streams from examination sessions are processed through a dedicated real-time video stream processing module that ensures continuous data flow and preliminary analysis. The core detection functionality is implemented through three parallel specialized modules: Module 1 performs phone detection using computer vision techniques to identify unauthorized mobile devices, Module 2 employs facial recognition technology for face counting and identity verification, and Module 3 conducts gaze analysis through eye tracking to monitor off-screen attention patterns. The outputs from these detection modules are consolidated in the alert integration layer, which performs data fusion, signal correlation, and risk scoring to generate comprehensive behavioral assessments. Lastly, the decision system automatically generates fraud alarms with comprehensive reporting capabilities by analyzing the integrated data using machine learning algorithms. This modular architecture maintains the low latency processing needs necessary for real-time monitoring applications while guaranteeing scalability, fault tolerance, and excellent accuracy in identifying different types of examination misbehavior.

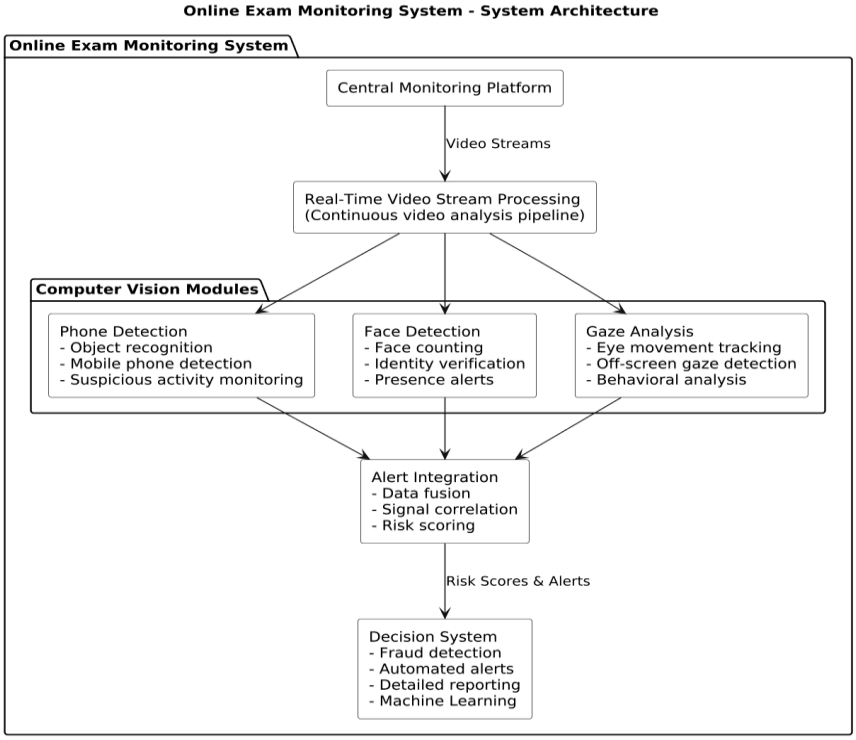


Figure 3. Architecture of an intelligent online exam monitoring system based on real-time video analysis

4.2. Dataset

This study leverages a public dataset from Michigan State University comprising webcam recordings of 24 subjects to develop an automated cheating detection system for online examinations. After preprocessing the original multi-modal data to focus exclusively on webcam footage, researchers simplified the taxonomy to three cheating behaviors: using unauthorized materials, interacting with others, and mobile phone usage. The methodological framework employs a two-stage computer vision pipeline that first identifies and isolates facial regions within video frames extracted at 2 FPS, followed by specialized classification models that analyze these regions to determine if cheating behaviors are present. Approximately 1,000 images are manually annotated using the LabelImg tool to train and evaluate the system. The image extraction process from the video sequences is illustrated in Figure 4. This hybrid architecture demonstrates potential transferability to diverse examination environments, addressing the growing need for reliable proctoring systems in digital assessment contexts [22].

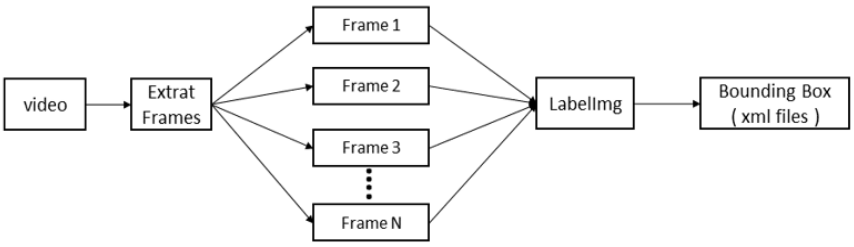


Figure 4. Image extraction

The LabelImg annotation framework, a well-known open-source graphical annotation platform created especially for object recognition labeling activities, is used in our dataset development procedure, as shown in Figure 5. In order to effectively prepare large-scale annotated datasets that are necessary for training reliable object identification models, this tool makes it easier to create accurate bounding boxes, polygonal annotations, and other geometric forms to distinguish objects within images [39]. The annotation

process generates XML files containing structured label metadata for each processed image, establishing a one-to-one correspondence between images and their respective annotation files. Our comprehensive dataset encompasses approximately 1,000 annotated images, systematically categorized into "cheating" and "not cheating" classifications. Following the dataset preparation phase, we deploy various object detection architectures that autonomously extract discriminative features from the annotated regions through supervised learning mechanisms. Subsequently, specialized classification models analyze these extracted features to determine the presence or absence of cheating behaviors in the examined scenarios.

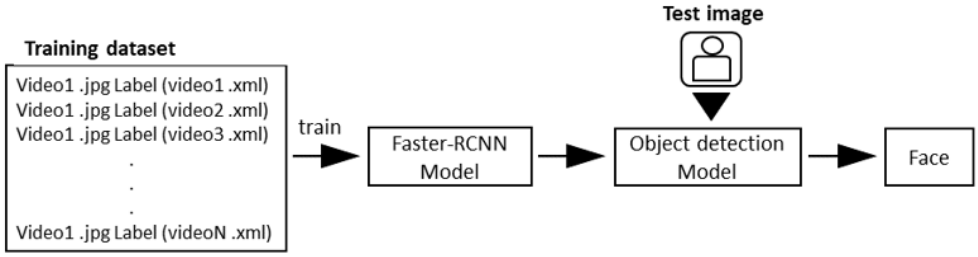


Figure 5. Object detection model

Figure 5 is designed to prepare the training dataset. Each video frame in our training data is associated with a label. Next, the model was trained using YOLOv5 facial detection models. Figure 6 illustrates the integration of two complementary models (detection and classification). An input image is first fed into the object detection model, which identifies any visible faces and extracts the corresponding facial regions. These cropped face portions are then passed to a CNN-based classification model, which determines whether cheating behavior has occurred. A key strength of this combined architecture lies in its flexibility, as it can be evaluated across different datasets to assess its cheating detection capabilities.

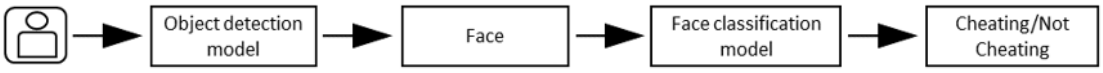


Figure 6. Object detection and classification model

4.3. Evaluation metrics

4.3.1. Loss metrics

Loss metrics quantify the training errors during model optimization. They measure the discrepancy between predicted and ground truth values, guiding the learning process through backpropagation. In object detection, multiple loss components are combined to address different aspects of the detection task: spatial localization, objectness confidence, and classification accuracy [40].

- Box loss

Box loss quantifies the deviation between the predicted bounding box coordinates and their corresponding ground truth values. It quantifies how much the predicted coordinates (x, y, width, and height) differ from the ground truth coordinates of detected objects. This loss ensures spatial accuracy of object localization [41].

- Object loss

Object loss (Obj Loss) evaluates the model's ability to identify the presence of objects in grid cells. It measures the error between the predicted probability that an object is present and the ground truth objectness score [42].

- Classification loss

Classification loss (Cls Loss) measures the error in predicting object classes. It quantifies how well the model identifies the correct type/category of detected objects [40].

4.3.2. Performance metrics

Performance metrics assess the model's effectiveness on test and validation data to offer objective assessments of detection quality. These metrics use both classification performance and localization accuracy

when assessing the trained model's capacity to generalize to new data. They are essential for comparing different models and assessing deployment readiness [43].
Precision reflects the ratio of true positive detections to the total number of objects identified by the model:

$$Precision = TP/(TP + FP)$$

where TP is true positives and FP is false positives [43].
Recall measures the proportion of actual objects that were correctly detected:

$$Recall = TP / (TP + FN)$$

where FN is false negatives [43].

5. RESULTS

5.1. Detection training results

As illustrated in Figure 7 (in Appendix), the training results demonstrate successful convergence of the YOLOv5 model over approximately 25 epochs. The training loss components show consistent improvement: the box loss decreases from 0.10 to approximately 0.02 (Figure 7(a)), indicating enhanced bounding box localization accuracy; the object loss is reduced from 0.030 to 0.012 (Figure 7(b)), reflecting improved object detection confidence; and the classification loss drops from 0.025 to 0.010 (Figure 7(c)), demonstrating better class prediction accuracy.

Similarly, the validation losses follow comparable downward trends, as shown in Figures 7(d)-(f), confirming good generalization capability without evident overfitting. The detection performance metrics further validate the robustness of the model. As depicted in Figures 7(g) and (h), precision and recall both reach values close to 0.9. Moreover, the mAP@0.5 score approaches nearly 1.0 (Figure 7(i)), indicating excellent detection performance at the standard IoU threshold.

When evaluated across multiple IoU thresholds, the stricter mAP@0.5:0.95 metric stabilizes between 0.6 and 0.7 (Figure 7(j)), reflecting the average performance gap under more challenging conditions. Overall, the smooth exponential moving average curves (orange dotted lines) observed across Figure 7 confirm stable training behavior and effective learning without severe overfitting.

5.1.1. You only look once version5-convolutional neural network hybrid model performance

The detection results demonstrate the successful adaptation of YOLOv5 for binary classification of cheating behavior in real-time video sequences. The model consistently identifies and classifies human subjects across 16 consecutive frames, achieving high confidence scores ranging from 0.6 to 0.8 for both "cheat" and "no cheat" classifications. The bounding box detections show stable localization with minimal drift, indicating robust tracking capabilities. The model's capacity to differentiate between suspicious activity (classified as "cheat" with confidence ratings of 0.6-0.7) and normal behavior (labeled as "no cheat" with confidence values of 0.7-0.8) is demonstrated by the classification performance. Although the somewhat lower confidence scores for "cheat" detections may signal the need for further training data or fine-tuning to improve differentiation between behavioral patterns, the temporal consistency between frames supports good feature extraction and classification. Qualitative frame-by-frame detection results are illustrated in Figure 8, demonstrating the feasibility of integrating object detection and behavioral classification within a modified YOLO architecture for real-time monitoring applications.

5.1.2. Convolutional neural network training performance for cheat detection classification

Over ten training epochs, the CNN model exhibits outstanding convergence and performance. Indicating successful learning and optimization, the loss continuously drops from 83.84 in the first epoch to 10.58 in the last epoch. Accuracy increases consistently from 93.08% to 99.23% at the same time, demonstrating a great capacity to classify behaviors that are cheating and those that are not. Accuracy increases from 93% to 98% in the early epochs (1–5), indicating consistent training without overfitting based on the measures' smooth growth. The CNN's great efficacy for binary classification in this cheat detection task[1] is demonstrated by its ultimate accuracy of 99.23%. The epoch-wise evolution of training loss and accuracy is summarized in Table 2.

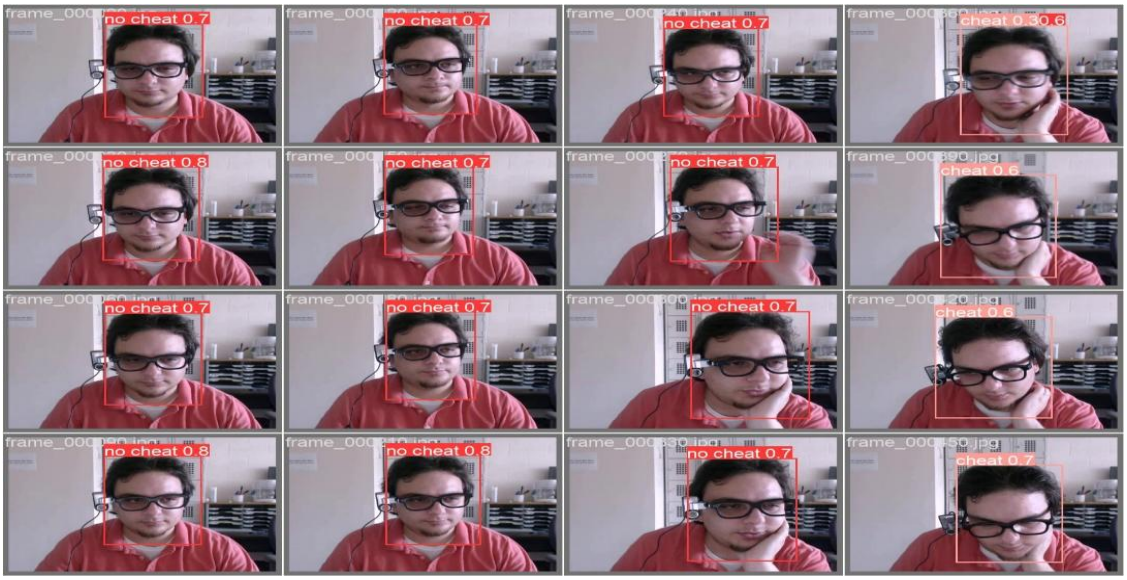


Figure 8. Frame-by-frame cheat detection using DL-based classification

Table 2. Training progress: epoch-wise loss and accuracy of the cheating detection model

Epoch	Loss	Accuracy (%)
1/10	83.8454	93.08
2/10	48.4000	96.14
3/10	35.3788	97.27
4/10	27.0055	97.95
5/10	22.0879	98.36
6/10	17.7145	98.71
7/10	16.6291	98.72
8/10	11.7296	99.16
9/10	12.8122	99.05
10/10	10.5780	99.23

6. DISCUSSION

The experimental results show that the proposed hybrid YOLOv5–CNN model performs effectively for automated behavioral surveillance. The system achieves an accuracy of 99.23%, while maintaining low inference latency and stable processing over long video sequences. This level of performance is consistent with previous studies demonstrating the suitability of YOLO-based architectures for real-time object detection, where reported accuracies typically range from 92% to 98%, depending on the dataset and application scenario [16], [44], [45].

Compared with conventional CNN-only approaches, which often present limitations in object localization and scalability and generally achieve classification accuracies between 85% and 93% [14], the proposed integration of YOLOv5 with a dedicated CNN classifier leads to more accurate recognition of behavioral patterns. Similar hybrid frameworks that combine object detection with deep feature learning have reported improved accuracies in the range of 94% to 97%, confirming the effectiveness of dual-stage architectures [46], [47].

The proposed dual-functionality architecture, which jointly performs object detection and behavioral analysis, offers advantages over single-task systems, particularly in terms of scalability and deployment efficiency. Previous studies on video surveillance and anomaly detection commonly report accuracies below 95% when using single-task pipelines [48], [49]. In contrast, the accuracy achieved in this work (99.23%) is comparable to, or higher than, the performance reported in recent cheating detection and abnormal behavior monitoring systems, where accuracies typically range from 90% to 98% [19], [37], [50]. These results position the proposed approach competitively with respect to the current state of the art.

Despite these encouraging results, several limitations remain. Vision-based surveillance systems are known to be sensitive to illumination changes and camera viewing angles, which may lead to performance degradation in unconstrained environments [19]. In addition, concerns related to demographic bias and fairness have been widely discussed in recent studies, as such issues may result in uneven accuracy across different population groups.

From a deployment perspective, challenges related to network latency optimization, infrastructure cost, and adherence to data privacy frameworks such as the GDPR should equally be taken into account [51]. Although the proposed system performs well under controlled experimental conditions, addressing these technical, ethical, and legal challenges requires interdisciplinary collaboration, as emphasized in prior work on responsible AI deployment in surveillance applications [52].

7. CONCLUSION

This work introduces a hybrid YOLOv5–CNN framework for automated cheating behavior detection in online examination environments. The proposed approach demonstrates high accuracy, low inference latency, and consistent performance across extended video sequences, highlighting its suitability for real-time and large-scale deployment. By combining precise facial localization with robust behavior classification, the system improves the reliability and efficiency of automated online proctoring while reducing dependence on manual supervision.

The findings underscore the strengths of DL–based surveillance systems to enhance fairness and integrity in remote assessments. The proposed framework offers a scalable solution that can be integrated into existing e-learning platforms, contributing to more secure and trustworthy online evaluation processes.

Future research will explore extending the system to recognize a broader range of cheating behaviors and more challenging real-world conditions, including diverse lighting environments, camera viewpoints, and partial occlusions. Incorporating temporal modeling and attention mechanisms may further improve behavioral consistency analysis over time. In addition, the integration of multimodal information and validation on larger, more diverse datasets will be investigated to strengthen robustness and generalization in practical deployments.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ilhame Khabbachi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ghalia Mdaghri Alaoui			✓	✓	✓					✓				
Abdelhamid Zouhair				✓		✓				✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The dataset used in this study is the Online Examination Performance (OEP) dataset [19], which contains data from 24 student participants at Michigan State University performing an online mathematics test, including multiple-choice and fill-in-the-blank questions. The original dataset was available at http://cvlab.cse.msu.edu/project-OEP.html; however, the URL is no longer accessible. A local copy of the dataset was used for all experiments.

REFERENCES

[1] A. Nigam, R. Pasricha, T. Singh, and P. Churi, "A Systematic Review on AI-Based Proctoring Systems: Past, Present and Future," *Education and Information Technologies*, vol. 26, no. 5, pp. 6421–6445, Sep. 2021, doi: 10.1007/s10639-021-10597-x.

[2] F. Noorbehbahani, A. Mohammadi, and M. Aminazadeh, "A systematic review of research on cheating in online exams from 2010 to 2021," *Education and Information Technologies*, vol. 27, no. 6, pp. 8413–8460, Mar. 2022, doi: 10.1007/s10639-022-10927-7.

[3] M. Ramzan, A. Abid, M. Bilal, K. M. Aamir, S. A. Memon, and T.-S. Chung, "Effectiveness of Pre-Trained CNN Networks for Detecting Abnormal Activities in Online Exams," *IEEE Access*, vol. 12, pp. 21503–21519, 2024, doi: 10.1109/ACCESS.2024.3359689.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA, USA: MIT Press, 2016.

[5] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3, p. 160, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[6] E. H. Houssein, Z. Abohashima, M. Elhoseny, and W. M. Mohamed, "Machine learning in the quantum realm: The state-of-the-art, challenges, and future vision," *Expert Systems with Applications*, vol. 194, p. 116512, 2022, doi: 10.1016/j.eswa.2022.116512.

[7] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge, MA, USA: MIT Press, 2018.

[8] N. Sharma, R. Sharma, and N. Jindal, "Machine learning and deep learning applications—A vision," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 24–28, 2021, doi: 10.1016/j.gltp.2021.01.004.

[9] Y. Wang, J. Wang, W. Zhang, Y. Zhan, S. Guo, Q. Zheng, and X. Wang, "A survey on deploying mobile deep learning applications: A systemic and technical perspective," *Digital Communications and Networks*, vol. 8, no. 1, pp. 1–17, Feb. 2022, doi: 10.1016/j.dcan.2021.06.001.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[12] J. Gu *et al.*, "Recent advances in convolutional neural networks: A comprehensive survey," *Pattern Recognition*, vol. 77, pp. 354–377, 2018, doi: 10.1016/j.patcog.2017.10.013.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, 2015, doi: 10.48550/arXiv.1409.1556.

[15] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Oct. 2014, pp. 1746–1751, doi: 10.3115/v1/d14-1181.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[17] S. Shinde, A. Kothari, and V. Gupta, "YOLO Based Human Action Recognition and Localization," *Procedia Computer Science*, vol. 133, pp. 831–838, 2018, doi: 10.1016/j.procs.2018.07.112.

[18] M. Cote, F. Jean, A. B. Albu, and D. Capson, "Video Summarization for Remote Invigilation of Online Exams," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, Mar. 2016, pp. 1–9, doi: 10.1109/WACV.2016.7477704.

[19] 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, doi: 10.1109/TMM.2017.2656064.

[20] K. Jalali and F. Noorbehbahani, "An automatic method for cheating detection in online exams by processing the students webcam images," in *3rd Conference on Electrical and Computer Engineering Technology (E-Tech 2017)*, Tehran, Iran, 2017.

[21] S. S. Ketab, N. L. Clarke, and P. S. Dowland, "A Robust e-Invigilation System Employing Multimodal Biometric Authentication," *International Journal of Information and Education Technology*, vol. 7, no. 11, pp. 796–802, Nov. 2017, doi: 10.18178/ijiet.2017.7.11.975.

[22] L. C. O. Tiong and H. J. Lee, "E-Cheating Prevention Measures: Detection of Cheating at Online Examinations Using Deep Learning Approach — A Case Study," *arXiv preprint*, Jan. 2021, doi: 10.48550/arXiv.2101.09841.

[23] M. M. Masud, K. Hayawi, S. S. Mathew, T. Michael, and M. E. Barachi, "Smart Online Exam Proctoring Assist for Cheating Detection," in *International Conference on Advanced Data Mining and Applications*, Switzerland: Springer, 2022, pp. 118–132, doi: 10.1007/978-3-030-95405-5_9.

[24] A. M. Shdaifat, R. A. Obeidallah, G. Ghazal, A. A. Sarhan, and N. R. A. Spetan, "A Proposed Iris Recognition Model for Authentication in Mobile Exams," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 12, pp. 205–216, 2020, doi: 10.3991/ijet.v15i12.13741.

[25] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, "Face anti-spoofing using patch and depth-based CNNs," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 319–328, doi: 10.1109/BTAS.2017.8272713.

[26] A. R. Khan, T. Saba, M. Z. Khan, S. M. Fati, and M. U. G. Khan, "Classification of Human's Activities from Gesture Recognition in Live Videos Using Deep Learning," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 10, p. e6825, May 2022, doi: 10.1002/cpe.6825.

[27] S. E. Kohli, Y. Jannaj, M. Maanan, and H. Rhinane, "Deep Learning: New Approach for Detecting Scholar Exams Fraud," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVI-4/W3-2021, pp. 103–107, 2022, doi: 10.5194/isprs-archives-xlvi-4-w3-2021-103-2022.

[28] F. Hussein, A. Al-Ahmad, S. El-Salhi, E. Alshdaifat, and M. Al-Hami, "Advances in Contextual Action Recognition: Automatic Cheating Detection Using Machine Learning Techniques," *Data*, vol. 7, no. 9, pp. 1-13, Sep. 2022, doi: 10.3390/data7090122.

[29] M. Malhotra and I. Chhabra, "Student Invigilation Detection Using Deep Learning and Machine After Covid-19: A Review on Taxonomy and Future Challenges," in *Future of Organizations and Work after the 4th Industrial Revolution: The Role of Artificial Intelligence, Big Data, Automation, and Robotics*, Cham, Switzerland: Springer, 2022, pp. 311–326, doi: 10.1007/978-3-030-99000-8_17.

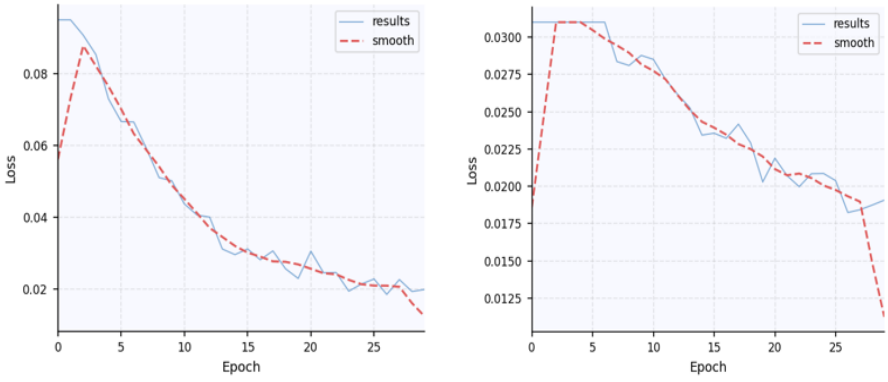
[30] S. Hu, X. Jia, and Y. Fu, "Research on Abnormal Behavior Detection of Online Examination Based on Image Information," in *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, Aug. 2018, vol. 2, pp. 88–91, doi: 10.1109/IHMSC.2018.10127.

[31] C. S. Indi, V. Pritham, V. Acharya, and K. Prakasha, "Detection of Malpractice in E-Exams by Head Pose and Gaze Estimation," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 8, pp. 47–57, 2021, doi: 10.3991/ijet.v16i08.15995.

[32] F. Noorbehbahani, A. Fanian, R. Mousavi, and H. Hasannejad, "An Incremental Intrusion Detection System Using a New Semi-Supervised Stream Classification Method," *International Journal of Communication Systems*, vol. 30, no. 4, p. e3002, Mar. 2017, doi: 10.1002/dac.3002.

[33] J. S. Ashwinkumar, H. S. Kumaran, U. Sivakarthikeyan, K. P. Rajesh, and R. Lavanya, "Deep learning based approach for facilitating online proctoring using transfer learning," in *2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, Chennai, India, May 2021, pp. 306–312, doi: 10.1109/ICCCSP52374.2021.9465530.
[34] S. Li and T. Liu, "Performance Prediction for Higher Education Students Using Deep Learning," *Complexity*, vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/9958203.
[35] A. M. Duhaim, S. O. Al-mamory, and M. S. Mahdi, "Cheating Detection in Online Exams During Covid-19 Pandemic Using Data Mining Techniques," *Webology*, vol. 19, no. 1, pp. 341–366, Jan. 2022, doi: 10.14704/web/v19i1/web19026.
[36] S. Kaddoura and A. Gumaei, "Towards Effective and Efficient Online Exam Systems Using Deep Learning-Based Cheating Detection Approach," *Intelligent Systems with Applications*, vol. 16, p. 200153, Nov. 2022, doi: 10.1016/j.iswa.2022.200153.
[37] R. M. Al-Airaji, I. A. Aljazaery, H. T. Alrikabi, and A. H. M. Alaidi, "Automated Cheating Detection Based on Video Surveillance in the Examination Classes," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 8, pp. 124–137, 2022, doi: 10.3991/ijim.v16i08.30157.
[38] M. Labayen, R. Vea, J. Florez, N. Aginako, and B. Sierra, "Online Student Authentication and Proctoring System Based on Multimodal Biometrics Technology," *IEEE Access*, vol. 9, pp. 72398–72411, 2021, doi: 10.1109/ACCESS.2021.3079116.
[39] I. Patel and S. Patel, "An optimized deep learning model for flower classification using NAS-FPN and faster R-CNN," *International Journal of Scientific & Technology Research*, vol. 9, no. 3, Mar. 2020.
[40] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, Oct. 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.324.
[41] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12993–13000, doi: 10.1609/aaai.v34i07.6999.
[42] R. Padilla, S. L. Netto, and E. A. D. Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242, doi: 10.1109/IWSSIP48289.2020.9145130.
[43] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
[44] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, doi: 10.1109/CVPR.2017.690.
[45] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *arXiv preprint*, 2020, doi: 10.48550/arXiv.2004.10934.
[46] W. Liu *et al.*, "SSD: Single shot multibox detector," in *European Conference on Computer Vision*, Amsterdam, The Netherlands, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.
[47] C.-W. Chang, C.-Y. Chang, and Y.-Y. Lin, "A Hybrid CNN and LSTM-Based Deep Learning Model for Abnormal Behavior Detection," *Multimedia Tools and Applications*, vol. 81, no. 9, pp. 11825–11843, Apr. 2022, doi: 10.1007/s11042-021-11887-9.
[48] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 6479-6488, doi: 10.1109/CVPR.2018.00678.
[49] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal Event Detection in Videos Using Generative Adversarial Nets," in *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, Sep. 2017, pp. 1577–1581, doi: 10.1109/ICIP.2017.8296547.
[50] F. Kamalov, H. Sulieman, and D. S. Calonge, "Machine Learning Based Approach to Exam Cheating Detection," *PLOS ONE*, vol. 16, no. 8, p. e0254340, Aug. 2021, doi: 10.1371/journal.pone.0254340.
[51] European Union, "General Data Protection Regulation (GDPR)," Regulation (EU) 2016/679, Apr. 2016.
[52] S. Zuboff, *The Age of Surveillance Capitalism*, New York, NY, USA: PublicAffairs, 2019.

APPENDIX



(a) (b)

Figure 7. Training and validation performance of the YOLOv5 model; (a) training box loss, (b) training objectness loss

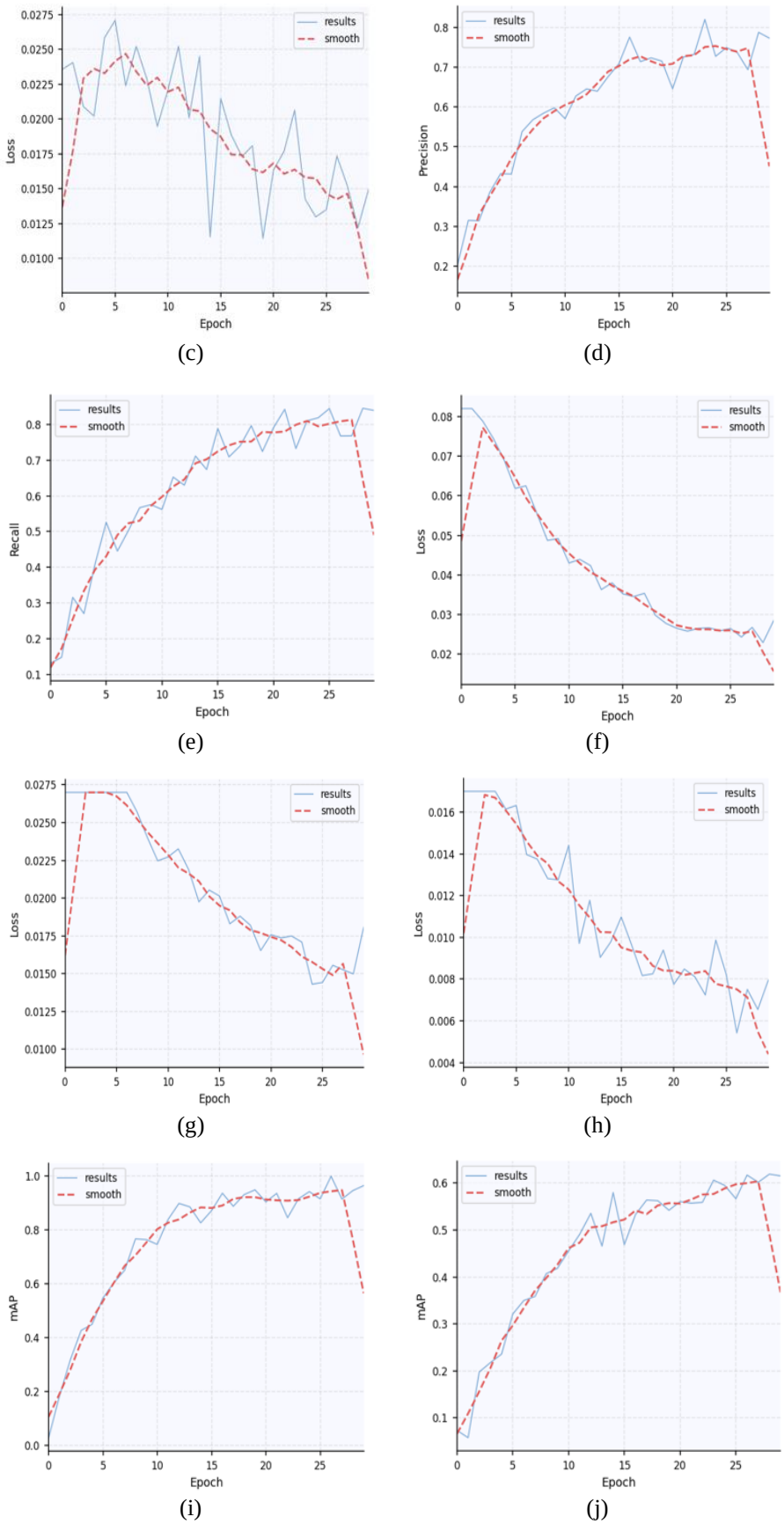


Figure 7. Training and validation performance of the YOLOv5 model; (c) training classification loss, (d) metrics: precision, (e) metrics: recall, (f) validation box loss, (g) validation objectness loss, (h) validation classification loss, (i) metrics: mAP@0.5, (j) metrics: mAP@0.5:0.95 *(continued)*

BIOGRAPHIES OF AUTHORS

Ilhame Khabbachi is a dedicated Ph.D. student passionate about the intersection of deep learning, facial emotion recognition, and online education security. Her research focuses on developing a cheating detection system for online exams using real-time facial emotion recognition powered by deep learning. She aims to enhance the integrity of e-learning assessments by leveraging AI-driven techniques to detect suspicious behaviors. Her work contributes to the advancement of secure and reliable online examination systems. She can be contacted at email: ilhame.khabbachi@etu.uae.ac.ma.

Ghalia Mdaghri Alaoui is a dedicated Ph.D. student passionate about the intersection of educational data mining, machine learning, and personalized learning in virtual reality. Their research focuses on clustering and classification techniques to analyze student performance, aiming to create adaptive learning paths. She is particularly interested in leveraging AI-driven insights to enhance personalized education. Their work also explores the application of virtual reality to optimize learning experiences. She can be contacted at email: ghalia.mdaghrialaoui@etu.uae.ac.ma.

Abdelhamid Zouhair is a Professor at the University of Abdelmalek Essaâdi, Faculty of Sciences and Techniques of Tangier, Department of Computer Sciences since March 2020, and a Professor at the National School of Applied Sciences of AL Hoceïma, between May 2016 and March 2020. He obtained his Ph.D. in Computer Sciences from the University of Le Havre (France) and the University of Abdelmalek Essaâdi, Faculty of Sciences and Techniques of Tangier (Morocco) in 2014. He obtained his HDR in 2020 and is the Director of several Doctoral Thesis in Computer Sciences. He is an author/co-author of several articles, published in International Journals in Computer Sciences. He can be contacted at email: a.zouhair@uae.ac.ma.