BOHR International Journal of Internet of things,
Artificial Intelligence and Machine Learning
2023, Vol. 2, No. 1, pp. 31–36
DOI: 10.54646/bijiam.2023.15
www.bohrpub.com

CONCEPT DEVELOPMENT

Machine learning for autonomous online exam fraud detection: A concept design

Abdul Cader Mohamed Nafrees[1]*, Sahabdeen Aysha Asra[2] and RKAR. Kariapper[3]

[1]South Eastern University of Sri Lanka, Oluvil, Sri Lanka
[2]Department of Information Technology, South Eastern University, Oluvil, Sri Lanka
[3]Department of Information and Communication Technology, South Eastern University, Oluvil, Sri Lanka

***Correspondence:**
Abdul Cader Mohamed Nafrees,
nafrees@seu.ac.lk

Received: 01 August 2023; **Accepted:** 04 August 2023; **Published:** 18 August 2023

E-learning (EL) has emerged as one of the most valuable means for continuing education around the world, especially in the aftermath of the global pandemic that included a variety of obstacles. Real-time online assessments have become a significant concern for educational organizations. Instances of fraudulent behavior during online exams (OEs) have created considerable challenges for exam invigilators, who are unable to identify and remove such dishonest behavior. In response to this significant issue, educational institutions have used a variety of manual procedures to alleviate the situation, but none of these measures have shown to be particularly innovative or effective. The current study presents a novel strategy for detecting fraudulent actions in real time during OEs that uses convolutional neural network (CNN) algorithms and image processing. The development model will be trained using the CK and CK++ datasets. The training procedure will use 80% of the selected dataset, with the remaining 20% used for model testing to confirm the model's efficacy and generalization capacity. This project intends to revolutionize the monitoring and prevention of fraudulent actions during online tests by integrating CNN techniques and image processing. The use of CK and CK++ datasets, as well as an 80–20 split for training and testing, contributes to the study's thorough and rigorous approach. Educational institutions can improve their assessment procedures and maintain the credibility of EL as a credible and equitable way of continuing education by successfully using this unique technique.

Keywords: CNN, E-learning, image processing, machine learning, online exam

1. Introduction

Since the coronavirus disease (COVID) epidemic spreads over the world, E-learning (EL) has become an unstoppable strategy in the education industry, yet EL is not a novel method in the education industry. Also, EL is a significant advance in the field of ICT. Due to the pandemic, various educational organizations use various methods for handling online exams (OEs), such as via ZOOM, planning camera systems around students with safe web pages, and a few others; however, none of these methods prevent students from cheating in exams, and others, such as face-to-face investigation guidance, are ineffective due to numerous practical problems from both students and staff. As a result, developing an experimental solution to check the quality of the test is critical to maintaining the quality of education.

As a result, conducting research to prevent examination fraudulent and teaching actions during OE using ICT is critical. As a result, integrating prior research, this study presents a unique approach based on machine learning (ML) to detect OE fraud by assessing candidates' eye movement, facial expressions, voice recognition, and facial expressions using image processing techniques Certain ML techniques such as selecting features, classification, and others have been used in research to offer a specific strategy/method for online tests.

Machine learning (ML) has proven to be successful in a wide range of tasks. Understanding how support from ML techniques impacts human agency and human performance is critical to realizing the potential of ML for enhancing human judgments. As ML becomes more widely employed in our society to enhance supervision, the ML community must not only construct ML models but also increase our knowledge of how these models are utilized and how people interact with them during the supervisory approach (1).

In the context of the pandemic, e-proctoring technologies have gained popularity, and some academic institutions have employed them to identify fraud in online tests. Lecturers can virtually watch students while they are taking exams utilizing these methods. The professors found that their students cheated on online exams after finding substantial variations in scores between proctored and non-proctored students in this study of computer engineering students (2). Fraud detection in corridor tickets is a challenging task to combine the various photo ingredients in a digital manner. The suggested architecture for disconnected testing uses lobby ticket images to provide security based on their components. Morphological activity is used to process as well as segment the corridor ticket images. The related component naming method is used to identify every fragment (3). It is only logical to employ a convolutional neural network (CNN)-based approach to identify the phony face pictures that arise (4).

The present online exam fraud detection method for combating online cheating is examined in part II. Part III presents the suggested framework, whereas part IV presents the comprehensive results and discussion. The findings are discussed in Section V, which is followed by a summary in Section VI.

1.1. Literature review

The research offered a framework for conducting online exams in a safe manner in order to prevent test difficulties and obstacles (5). Similarly, another research recommended an OEL system for conducting tests from anywhere at any time. This ensures the examination's validity. (6). A group of students suggested using an ML technique to conduct OEs, which helps to decrease examination fraud by evaluating every student's activities during the test period (7).

Another study found that using live and artificial intelligence (AI)-based remote monitoring considerably boosts the amount of security that OEs have against student test fraud (8, 9). To prevent examination infractions, the authors (10) recommend using low-cost keystroke analysis, voice recognition, facial recognition, iris reader, and fingerprint reader as well as using technologies like Blackboard Learning and Lockdown Browser (11).

Table 1 summarizes some other techniques used in fraud detection in online examination too.

According to the table, the main technologies or algorithms used for the examination fraud detection and support to identify fake faces are LSTM, KDE, CNN, ALEXNET, OpenPose, MATLAB, GAN, KNN, and SVM. Most of the research papers include CNN as the technology for their system. Usually, the uses of main technologies in each paper do not differ much. The main uses of the technologies are face recognition, authentication, and face extraction. But MATLAB technology is very helpful for deaf invigilators and visually impaired and also detects cheating automatically.

The VGG-16 Net is used to categorize six aberrant behavior-related poses using the picture within the generated bounding box for each distinguished human subject. While assessing the behavior of a testing set, the six-dimensional output vector of VGG-16 Net is passed to LSTM to detect unauthorized behaviors that arise instantly (17). A study found that both the crowd-sourced and ML approaches lower the likelihood of exam cheating (18). Similarly, tools such as computer vision for facial recognition and ML may be utilized to detect student participation throughout online learning sessions (19). The spatiality reduction approach, in which a starting dataset is separated as well as decreased to further manageable teams, may include feature extraction (20). The most crucial element of those big data sets is the large number of variables required. A large amount of computer capacity is needed to process these variables. As a consequence, by choosing and merging variables into alternatives, feature extraction assists in the induction of the most effective feature from enormous data sets. These qualities are simple to construct, yet they can correctly and uniquely describe the data collection in issue (3).

The main aim of this concept proposal was to look at a novel method for detecting probable cheating scenarios utilizing the ML approach. Keeping academic integrity is one of the most difficult tasks. The online test is an important part of EL. Student's tests in EL are completed virtually, with no physical invigilators present. Because of the easiness with which students might cheat on OE, EL colleges rely on an examination technique in which students take a face-to-face examination in a physical venue defined on the organization's facilities under supervised conditions.

To train a KNN classifier to distinguish between a spoof and a real face, image quality measurements and color texturing are utilized. To generate the final feature representation, texture and quality data are obtained and connected. After that, the characteristics are put into a KNN classifier. KNN classifies an unknown sample by returning the label that occurs most frequently between the nearest training distance K and the testing dataset. Both the unknown and training samples are measured using Euclidean distance. KNN is more precise and faster to perform. SVM (supervised learning model) is a data analysis and pattern recognition model that is used for classification and regression analysis. SVM is a method that

TABLE 1 | ML technologies used for OE fraud detection.

References	Used techniques	Technology used for	Accuracy	Limitation
Kamalov et al. (12)	Long short-term memory networks (LSTM), KDE (Kernel density estimation)-based outlier detection	KDE - Recognize outliers in a dataset. LSTM - Cope with sequence data efficiently by enabling prior outputs to be used as inputs.	High level of accuracy	-
Sharma et al. (13)	CNN	Detection of fraud during e-exams.	-	-
Nishchal et al. (14)	OpenPose, ALEXNET model, CNN	OpenPose is used to identify posture; Model ALEXNET is used to identify types of fraud	77.8% of accuracy	The camera should be placed at a suitable height to detect all the persons
Asadullah et al. (15)	MATLAB	Detection of cheating automatically	The likelihood of false acceptance (FAR) is 1%, while the probability of false rejection (FRR) is 3%.	Unusual behaviors such as whispering and visualizing
Mo et al. (4)	Generative adversarial network (GAN), CNN	GAN - generate realistic looking faces, CNN - identify fake face images	99.4% of accuracy	-
Pranoto et al. (16)	KNN, SVM	Authentication and recognition process	95% ± 0.2 accuracies, 84% accuracy when the head is angled approximately 45%, 94% accuracy for small detected faces, 75% accuracy with little occlusion on the face, and 20% accuracy with major occlusion.	-

uses the structural risk minimization (SRM) concept to choose the optimum hyperplane (21). Since technological advancement has dramatically improved, trying to capture irregular aspects relevant to unethical activities requires an examination of the quantity of variations that can be investigated concurrently. This can be settled by expanding the number of neurons and/or layers at the cost of broader online activities of increasing sophistication neural networks (22).

2. Methodology

2.1 Problem statement

As per the past segment, several digital solutions and manual methods have been presented as well as employed to prevent OEs from engaging in illegal behavior and scamming, but no secured specific development, technique, or method has yet been offered for executing protected automatic OEs. This results in unjust exam outcomes for students and also challenges in examination monitoring for staff; it also contributes to the majority of educational sectors delaying final tests, lengthening the course time period. As a result, in the future, it will be important to create secured practical solutions for OEs that will assist in avoiding all the OE difficulties described previously.

Moreover, the mentioned practical reasons from both students and staff of educational organizations motivated this research study:

- Completing the course/study within the scheduled school program
- Preserving the performance of EL as if it were face-to-face learning
- Minimizing supervisors' anxiety levels during OEs
- Maintaining impartial exam results to ensure student sincerity throughout OEs.
- Important technological developments for EL practical problems.

Most of the foregoing arguments made us to do a study in the subject of EL on "Development of a unique machine-learning algorithms for the automated monitoring of online examination fraudulent activities."

2.2. Research questions (RQ) and Research objectives (RO)

As per prior study findings, OEs are the most difficult activities in the EL framework for active monitoring to prevent examination fraud. However, until now, no effective method has been devised or supplied to address this issue. As a result, this work provides a viable strategy for filling this gap employing image processing approaches using ML techniques.

This model will be trained using the provided dataset and evaluated using real-time collected photos to corroborate the study's importance. Moreover, the same technique may be used to perform OEs while using any EL tool (ZOOM, VLE, etc.), as shown in **Table 2**.

TABLE 2 | RQs and ROs for the proposed study.

S. No	Research question	Research objectives
1	What are the best ML algorithms for detecting learners' activity during OE?	Examine and evaluate existing research for the most advanced ML approaches for voice recognition, eye movement, and facial recognition as well as their performance.
2	How can I use ML methods to create an app that captures students' activity through webcam?	Build an ML-based technology to identify students' questionable actions as well as alert supervisors during OE.
3	How can the correctness of the produced ML system be verified?	To check the correctness of the built system, train it with accessible picture data sets as well as test it in a real-time context.

The suggested model will be developed using ML techniques, and it will assess students' facial recognition (fear/nervousness, sorrow, and happiness), head pose, and eye movement during OEs to detect exam cheating as well as fraudulent activities. As a result, the CK and CK + databases will be used to train the model, and real data gathered from the camera will be used to evaluate the model to determine the system's correctness. In addition, if a graphics processing unit (GPU) is not really available, an i7 higher graphic general-purpose computer will be utilized to build the overall model.

Argumentation and pre-processing of the proposed system, design as well as implementation using ML approaches, training the model using CK and CK + datasets, and eventually testing the entire system with real data will all be included in this research article.

1. Pre-processing and Argumentation

This procedure will considerably increase the accuracy of the produced model's findings. As a result, every image in the training dataset and testing dataset is shrunk to the same X * X pixel size. The model's errors will be estimated using a sample size of 64 or 32. Furthermore, argumentation techniques like shear, horizontal flip, and zoom will be employed for the training data sets to decrease erroneous predictions.

1. Design and Implementation

As once preceding steps have been performed, these photos will be sent to the ML layers for additional processing.

1. Model Training

For the model training procedure, an equal number of photos from each type of face expression, head stance, and eye movements will be used. The model overall accuracy will be determined by the greatest accuracy. In addition, Anaconda Navigator will be utilized for the desktop user interface, and the generated model will be trained using TensorFlow open library software.

1. Model Testing

Just after the model has been effectively trained, this process will be carried out with real data sets collected during OEs. The below diagram depicts how the trained ML model will recognize as well as inform the supervisors.

2.3. Expected outcomes

The ML-based model will assist, identify, and warn supervisors about students' examination fraudulent or cheating behaviors during online tests, as stated and illustrated above. With this technology, a trained model will identify students' expression in real-time collected photographs and alert the supervisors if the expression is suspected of being cheating or fraudulent. Furthermore, it has been determined to publish research publications in reputable indexing journals and IEEE or Springer conferences.

3. Results and discussion

Researchers frequently offer various strategies to address the problems raised by online tests. Features such as examinee variation suspicious behavior identification, overall system safety, question paper production, and so on are very significant in online examinations. According to the previous papers, there are numerous ML technologies and algorithms that are used in the concept of autonomous monitoring of online examination to identify examination fraudulence. Most of the papers confess that CNN, LSTM, KDE, ALEXNET, OpenPose, MATLAB, GAN, KNN, and SVM are the suitable ML techniques. Also, they indicate the accuracy and limitations as well. Almost every article's techniques show a high level of accuracy. Meanwhile, some research papers discuss the limitations. The main limitations are that covering large areas with a camera is harder and the system could not identify unusual behaviors (e.g., whispering).

We proposed a concept or solution to fill the gap that we indicated previously. Especially, the proposed model would measure students' facial recognition (fear/nervousness, grief, and happiness), head attitude, and eye movement to detect exam cheating and fraudulent activities using ML approaches. As a consequence, the CK and CK + databases will be utilized to train the model, and real data from the camera will be used to test the model and determine whether the system is right. In addition, if a graphics processing unit (GPU) is not accessible, the total model will be built on an

i7 higher visual general-purpose computer. As a result, the training and testing datasets' images were all downsized to the identical X * X pixel size. A sample size of 64 or 32 will be used to estimate the model's errors. Additionally, for the training data sets, argumentation techniques such as shear, horizontal flip, and zoom will be used to reduce erroneous predictions. An equal number of photographs from each sort of facial expression, head stance, and eye movements will be employed in the model training phase. The greatest accuracy will decide the model's accuracy rate. The desktop user interface will be Anaconda Navigator, and the resultant model will be trained using the Tensorflow open library software. This research article will cover the rationale and pre-processing of the proposed system as well as the design and implementation of the system using ML techniques, training the model with CK and CK + datasets, and finally testing the system with actual data.

Furthermore, training the model using CK and CK + datasets is the most suitable and user-friendly among other ML approaches. In addition, the webcam will capture the student's activity and process the image by sending or verifying within the trained dataset. The processed picture is then provided to an ML model to train it to alert the examination about the students' expression during the online assessment. So, there is identification of the correctness or testing in a real-time context of the previous or recommended article. They mention the accuracy level is high. But the article does not mention the numeral accuracy level.

4. Conclusion

During critical events such as war, natural disasters, and pandemics, EL has shown promise. As a result, during the last three decades, numerous techniques and EL systems have been created with the aim of appropriately carrying and enhancing EL. Using ICT to avoid examination fraud and instructional activities during OE is critical in this context.

According to prior study findings, OEs are the most difficult activities in the EL process of active monitoring to prevent examination fraud. However, until now, no effective method has been devised or supplied to address this issue. As a result, this work provides a viable strategy for filling this gap employing image processing approaches using ML techniques. This model will be trained using the provided data and evaluated using real-time collected photos to corroborate the research's importance. Furthermore, the same technique may be used to perform OEs while using any EL tool (ZOOM, VLE, etc.).

Furthermore, among other ML techniques, CNN is the most recommended because of its accuracy and processing timing. ML approaches that train the model using CK and CK + datasets are the most suited and user-friendly. In the proposed system, the webcam will record the student's behavior and submit or verify the image inside the training dataset. Then the processing image is sent to train the model via an ML model to notify to the examination about the expression of the students during the online examination. So, the preceding or suggested article can determine the accuracy or test it in a real-time environment. They claim that the degree of accuracy is great. However, the numerical accuracy level was not specified in the paper.

Therefore, during major disasters such as wars, natural catastrophes, and pandemics over the last three decades, EL has demonstrated its potential. Many approaches and systems for improving EL have been created, including ICT-based methods to prevent examination fraud during OEs. However, successfully monitoring and eliminating fraud in OEs continues to be difficult. This work presents a way for monitoring students' behavior during OEs by utilizing ML and visual processing techniques, notably CNN. To confirm its importance, the model will be trained and tested using real-time data. Despite the fact that the proposed system appears to be accurate, exact numerical findings were not supplied. The analysis admits the possibility of some essential contributions being overlooked due to title mismatches.

There are some limitations in this study. In this context, it is possible that we originally overlooked a few significant studies whose titles do not correspond to the study's real contributions or contents.

Author contributions

AN: concept desining, literature review, and methodology. SA: introduction, abstract, and conclusion. RK: overall reading and supervision. All authors contributed to the article and approved the submitted version.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Lai V, Tan C. On human predictions with explanations and predictions of machine learning models: A case study on deception detection. *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency, FAT* 2019*. New York, NY: (2019). p. 29–38. doi: 10.1145/3287560.3287590
2. Balderas A, Palomo-duarte M, Caballero-hernández JA, Rodriguez-M, Dodero JM. Learning analytics to detect evidence of fraudulent behaviour in online examinations. *Int J Interact Multimedia Artific Intellig.* (2022) 7:241–9. doi: 10.9781/ijimai.2021.10.007
3. Chhabile SS, Varma AK. *Automated fraud detection in examination from hall ticket. 03, 1393–1396*. Navi Mumbai: Projectwale (2021).

4. Mo H, Chen B, Luo W. Fake faces identification via convolutional neural network. IH and MMSec 2018. *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*. New York NY: (2018). p. 43–7. doi: 10.1145/3206004.3206009
5. Fahd K, Parvin S, Venkatraman S, Di Serio A, de Souza-Daw A, Overmars A, et al. An innovative framework for online examination in a higher education setting: a response to the COVID-19 crisis. *J Inst Res South East Asia* (2021) 19:93–118.
6. Tripathi AM, Kasana R, Bhandari R, Vashishtha N. Online examination system. *Lecture Notes Netw Syst.* (2022) 286:709–17. doi: 10.1007/978-981-16-4016-2_67
7. Sinha P, Dileshwari, Yadav A. Remote proctored theory and objective online examination. *Int J Adv Netw Applic.* (2020) 11:4494–500. doi: 10.35444/ijana.2020.11068
8. Karim MN, Kaminsky SE, Behrend TS. Cheating, reactions, and performance in remotely proctored testing: an exploratory experimental study. *J Bus Psychol.* (2014) 29:555–72.
9. LaFlair GT, Settles B. *Duolingo English Test: Technical Manual. Duolingo Research Report*. Pittsburgh, PA: Duolingo Inc (2019).
10. Berkey D, Halfond J. *Cheating, Student Authentication and Proctoring in Online Programs*. Boston, MA: New England Journal of Higher Education (2015).
11. Ali L, Al-Dmour NAHH. The shift to online assessment due to covid-19: An empirical study of university students, behaviour and performance, in the region of UAE. *Int J Inform Educ Technol.* (2021) 11:220–8. doi: 10.18178/ijiet.2021.11.5.1515
12. Kamalov F, Sulieman H, Calonge DS. Machine learning based approach to exam cheating detection. *PLoS One.* (2021) 16:e0254340. doi: 10.1371/journal.pone.0254340
13. Sharma NK, Gautam DK, Rathore S, Khan MR. CNN implementation for detect cheating in online exams during COVID-19 pandemic: a CVRU perspective. *Mater Today Proc.* (2021):doi: 10.1016/j.matpr.2021.05.490
14. Nishchal J, Reddy S, Navya PN. Automated cheating detection in exams using posture and emotion analysis. *Proceedings of CONECCT 2020 - 6th IEEE International Conference on Electronics, Computing and Communication Technologies*. Bangalore: (2020). doi: 10.1109/CONECCT50063.2020.9198691
15. Asadullah M, Nisar S. An automated technique for cheating detection. *Proceedings of the 2016 6th International Conference on Innovative Computing Technology (INTECH), 2016*. Dublin (2017). p. 251–5. doi: 10.1109/INTECH.2016.7845069
16. Pranoto H, Kusumawardani O. Real-time triplet loss embedding face recognition for authentication student attendance records system framework. *JOIV Int J Inform Vis.* (2021) 5:150–5.
17. Ko KE, Sim KB. Deep convolutional framework for abnormal behavior detection in a smart surveillance system. *Eng Applic Artif Intellig.* (2018) 67:226–34. doi: 10.1016/j.engappai.2017.10.001
18. Alenezi HS, Faisal MH. Utilizing crowdsourcing and machine learning in education: literature review. *Educ Inform Technol.* (2020) 25:2971–86. doi: 10.1007/S10639-020-10102-W
19. Dewan MAA, Murshed M, Lin F. Engagement detection in online learning: a review. *Smart Learn Environ.* (2019) 6:1–20. doi: 10.1186/s40561-018-0080-z
20. Bora A, Sharma S. A review on video summarization approaches: recent advances and directions. *Proceedings of the IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking, ICACCCN*. Greater Noida: (2018). p. 601–6. doi: 10.1109/ICACCCN.2018.8748574
21. Rahmad C, Arai K, Asmara RA, Ekojono E, Putra DRH. Comparison of geometric features and color features for face recognition. *Int J Intellig Eng Syst.* (2021) 14:541–51. doi: 10.22266/IJIES2021.0228.50
22. Sravani PVG, Sikndar S. Fraud identification: fraud monetary detection with aid of human behaviour appraisal examination. *Int J Adv Res Sci Technol.* (2021) 11:181–8.