2023 International Conference on Frontiers of Information Technology (FIT) | 979-8-3503-9578-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/FIT60620.2023.00032

Enhancing Online Exam Security: Deep Learning Algorithms for Cheating Detection

1st Tanzeela Iqbal
Department of Computer Science
COMSATS University Islamabad,
Sahiwal, Pakistan
tanzilaiqbal8@gmail.com

2nd Tariq Ali
Department of Computer Science
COMSATS University Islamabad,
Sahiwal, Pakistan
tariqali@cuisahiwal.edu.pk

3rdAhmad Shaf
Department of Computer Science
COMSATS University Islamabad,
Sahiwal, Pakistan
ahmadshaf@cuisahiwal.edu.pk

4th Muhammad Shafqat Ali
Department of Computer Science
COMSATS University Islamabad,
Sahiwal, Pakistan
mshafqat448@gmail.com

***Abstract*— Cheating in online exams has become a significant concern in recent years. There are various forms of online exam cheating, including copying answers from external sources. Fear of failure can drive some students to cheat and obtain high grades effortlessly. Many reducing cheating detection methods have limitations in terms of accuracy and time complexity. Even while some clever cheating detection systems have a high detection rate for cheating in videos, they have issue of a significant runtime overhead. Simple procedures with low time complexity tend to have accuracy issues with high false alarm rate. To address these challenges, a unique hybrid deep learning technique based on CNN-BiGRU has been proposed. This approach utilizes recorded videos during an exam to detect cheating instances. The methodology incorporates facial recognition, video analysis, eye tracking and object detection to identify potential cheating behaviors. To evaluate the performance of the proposed approach, 80% of the dataset is used for training the model, while the remaining 20% is employed for testing purpose. Subsequently, accuracy and evaluation matrix are computed. The constructed model achieves an impressive accuracy rate of 92%.**

Keywords—Cheating Detection, Online Exams, Deep Learning, Online Proctor.

I. Introduction

Online tests are excellent for assessing a student's knowledge. The identification of students in online assessments can be deceiving. As a result, the problem of online evaluation fraud is intriguing. For the sake of this study, we define fraud as deceit used in education for personal advantage through fraudulent behaviors. In simple terms, "cheating" is adopting deceptive methods to gain a "higher grade" on an online assessment. Due to the advantages of remote working environments and the Covid-19 pandemic, online tests and job interviews have grown in prominence in recent years. Most corporations and educational institutions utilize these systems for both employment and online assessments [1]. One of the biggest disadvantages of remote examination systems is the inability to conduct tests in a safe environment. Several online assessment apps have been created to meet various technological standards. They all provide the same basic functions, such as giving assessment items, training and evaluation, and grading. These applications store a lot of data to complete their tasks. Starting times, local or distant IP addresses and finishing timings are among the data components kept in the database. Student behavior data, such as frequency of visits, training attempts, and preliminary grades for certain courses, as well as demographics perceptions of the subjects being examined, are all kept. It's critical to make efficient use of this massive amount of data to identify and prevent fraudulent behavior by learning patterns and trends in student behavior.

The online examinations must be invigilated in a protected environment to confirm their authenticity. About 75% of scholars believe that cheating on online exams is simple. Candidates will be supervised by a human proctor, who is commonly used to ensure that tests are administered correctly in traditional settings. On the other side, invigilating is a difficult and time-consuming operation. To avoid pre-planned cheating, it only requires monitoring a person or a group of individuals throughout an exam. In this paper, we present an approach for investigating exam and online interview fraud [2]. The method just demands a recorded video of the candidate during the test. The candidate's condition, the appearance of another person, and the usage of technical gadgets are all detected using the cheating detection pipeline. As a result, we propose a deep learning technique for identifying students (individuals) who cheat in online examinations. The procedure includes face detection, object detection, and tracking systems. To test the pipeline's

2473-7569/23/$31.00 ©2023 IEEE
DOI 10.1109/FIT60620.2023.00032

performance, collect the public filmed dataset. The collection includes both cheating and clean footage. Finally, our pipeline provides a quick and easy way to identify and analyses cheating in online interviews and tests. Our proposed methodology includes facial recognition and video, recognizing sound, timing information, and eye tracking. This will benefit all the universities and schools as well.

II. LITERATURES SURVEY

Online proctoring systems and their characteristics are compared in depth in these studies. The present challenges in online assessments are addressed in this paper, which are particularly pertinent during the Covid-19 pandemic. E-learning with a mobile twists [3]. Autonomous detection systems are often integrated with human proctors and well-defined workflows in certain methods, as demonstrated in [4] present a user verification pipeline that includes dataset collection and training processes. The paper [5] aims to investigate various methods employed by students to cheat in online tests, focusing on constant authentication and online proctoring. They have also created an e-exam management system that is designed to identify and stop cheating on online tests. The technology includes an eye-tracking gadget and a fingerprint reader authenticator to keep an eye on students during the test. The face of the candidate was recognized, and feature points were extracted to determine the subject's head position [6].

Student behavior toward cheating in online tests is influenced by emotional elements, contextual factors at schools, and teaching approaches used by teachers, according to [7]. The development of courses that were compatible with a range of mobile phone models, brands, and platforms was motivated by the arrival of portable devices. Tests should receive more attention if we want to improve mobile learning credentials. A major problem with mobile tests is the identification and verification of students before and during the exam session. [8] A standard username and password as well as biometric iris recognition to verify student identity before and during the mobile test are two solutions available to assure student identification in the absence of a proctor. A novel paradigm for authenticating students is given, which includes two types of authentications. Even the most well-behaved and well-educated students might engage in cheating during online exams under certain circumstances. To avoid such scenarios, author recommended adopting webcam-based proctoring to identify potential cheaters. Using this method, however, necessitates regular inspection and monitoring of the webcam's recorded images, which is like the old manner of proctoring in paper-based examinations. Speed of typing, interruptions, inconsistent typing, and other factors could be used to identify dishonest students.

For identity verification in [9], the student combines traditional login techniques such a username and password with biometric identification like iris recognition. Additionally, the suggested approach randomly checks for students during the exam to avoid impersonality. A strategy was outlined to eliminate the need for a physical proctor during exams by utilizing a multi-modal system. To achieve this, webcams and active window capture were employed to capture audio and video data, which was then processed by an intelligent rule-based inference engine.

The system uses various factors, such as yaw angle fluctuations, audio presence, and active window capture, to detect any misbehavior [10]. Later, author stressed the concept of using the sort of interaction between the computer and the person as a tool for detecting those who cheat during eExams [11]. The well-rounded inference system can assist instructors in monitoring students during online exams. As part of the system's planned features, capacity was built to perform feature point extraction and yaw angle detection [12]. The project's focus is on a solitary situation with a solitary learner. New tools for recording video such as the system can also be created using Bullet Cameras and Wireless IP Cameras [13]. Another variation of this strategy involves pairing pre-emptive systems with human proctors.

Researcher attempted to build a paperless assessment system for the SQL language on both a practical and theoretical level. The researcher's project, Smart Online Exam Proctoring Assist for Cheating Detection, was completed in [14]. If face-to-face education must be discontinued due to a forced incident, like the most recent Covid-19 outbreak, this style of assessment has become even more prevalent and necessary. In [15], present extensive pipelines containing voice detection, gaze estimation, head-pose estimation, and other aspects in multimodal techniques. On the contrary, this kind of expansive pipeline typically requires more significant computational capacity. When it comes to online exams, the proctoring procedures used provide a significant issue for the research community. Additionally, there are automated interview systems available, as described in [15] that utilize online coding platforms to evaluate candidates. While these systems offer plagiarism detection, they may not be able to detect multiple individuals or the use of electronic devices during the assessment. The researchers conducted two experiments with students for text-based evaluation and computerized examination to better understand the writing styles, habits, and tactics they use during the exams [16].

The discipline of e-learning has grown exponentially over the past ten years due to the Internet's fast spread and technical improvements. Exam cheating, however, is a widespread problem that exists everywhere, irrespective of the degree of progress. By analyzing two key parameters, namely the total time spent away from the screen and the frequency of screen disengagement, they can determine whether an examinee is engaged in cheating behavior or not. Multiple cheater analysis paradigms are combined into multi-modal procedures that use both visual and auditory clues. evaluates a candidate's hiring potential in light of these traits and psychological aspects [17].

Later, an author proposed an enhanced system to address e-exam and e-evaluation security concerns [18]. Another study looked at the interactions between students and teachers to see if they were willing to cheat [19]. This paper [20] made a similar endeavor to study students' attitudes on e-exams to assess their undergraduate pupils. Students' displeasure with the questions being asked, the format of the questions, the timings, and other issues were cited as some of the obstacles. When opposed to on-site face-to-face tests, conducting online exams presents a far larger problem in terms of maintaining academic integrity [21]. There is no human proctor that to oversee the examinee, which increases the likelihood that they may cheat. Numerous online exams proctoring systems are used by educational institutions all over the world to prevent the chance of cheating. These instruments' most common technique is to record the examinee's video and audio throughout the duration of the test.

According to the author [22], the proposed system does not require any additional components and establishes continuous authentication by employing randomly generated questions from current data. Poor financial circumstances may also lead them to cheat on online tests [23].

III. PROPOSED METHODOLOGY

Our proposed framework means to identify predetermined cheating exercises in a fully programmed manner by relying on captured video evidence of the test-taker or - as we call-up-and-comer. Aside from a webcam or other compatible camera, the framework doesn't require any further hardware. Additionally, there isn't a verification phase when the applicant sends some printed and visible data in advance. In Fig. 1, system architecture of the proposed model is shown. The architecture consists of the following steps: Video pre-processing: This step cleans up the video and removes any noise or artifacts. It also involves resizing the video or adjusting the brightness and contrast. Candidate's face detection: This step uses a face detector to identify any potential faces in the video. Face recognition: This step extracts a unique faceprint from each candidate face and compares it to a database of known faces. If a match is found, the system can identify the person in the video. Analyzing results for other: object detection, another person device detection: This step looks for other objects in the video, such as other people or devices. This information can be used to make a more informed decision about the person's identity. Decision: The system makes a decision about whether or not the person in the video is known. If the system is unable to make a clear decision, it will process the whole video.

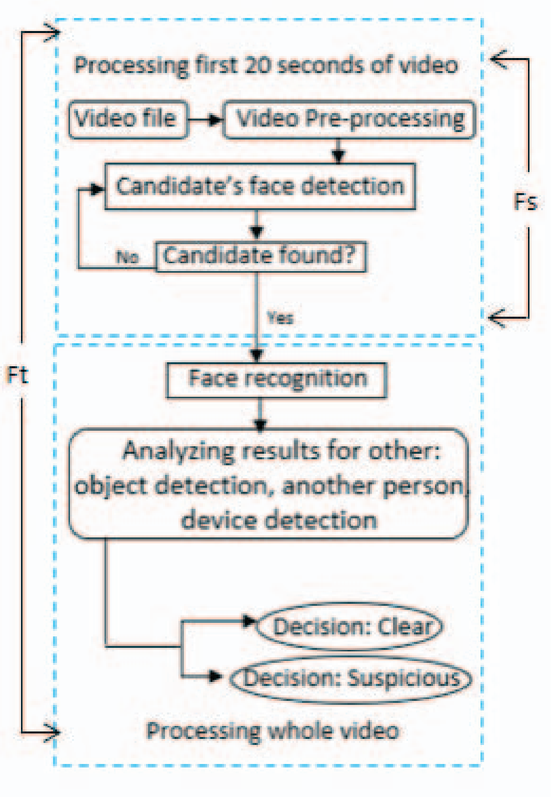

Processing first 20 seconds of video
Video file
Video Pre-processing
Candidate's face detection
No
Candidate found?
Yes
Fs
Ft
Face recognition
Analyzing results for other: object detection, another person device detection
Decision: Clear
Decision: Suspicious
Processing whole video

Fig 1. Proposed cheating detection pipeline

A. Dataset

We have taken openly accessible dataset for our proposed work [5]. A total of 24 participants, all of whom were enrolled at Michigan State University, took part in the data collection process. Among them, 15 participants played the role of actors simulating exam cheaters. They were instructed to engage in cheating behaviors during the session, with no specific guidelines on what cheating methods to employ or how to execute them. It's worth noting that these actors sometimes exhibited behavior that appeared artificial. In order to capture authentic exam scenarios, nine actual students were asked to take a genuine exam, with their scores being recorded. It was assumed that these students were unlikely to cheat in the controlled data collection environment. To induce cheating behaviors, the proctor interacted with them by talking, approaching them, or offering items such as books, thus creating a more realistic exam setting. This combination of two distinct participant groups contributed to the database's diversity, incorporating a wide range of cheating techniques and replicating the genuine engagement experienced during real exams. These videos incorporate having someone else in the exam climate other than the applicant, the candidate using book, using a cell phone or PC apparently, or some blend of these events all through the video.

B. CNN-BiGRU Model

A model with the name CNN-(Bi-GRU) is proposed whose architecture is shown in Fig. 2. The recorded video content is given to the initial convolution layer as input. The output of the convolutional layer is pooled into a smaller dimension. This output is then used as input in the Bi-GRU layer. The convolution layer from CNN will extract the features automatically and then the Bi-GRU layer will memorize the ordering of those extracted features which helps to learn about the ordering of the events happening in the video.

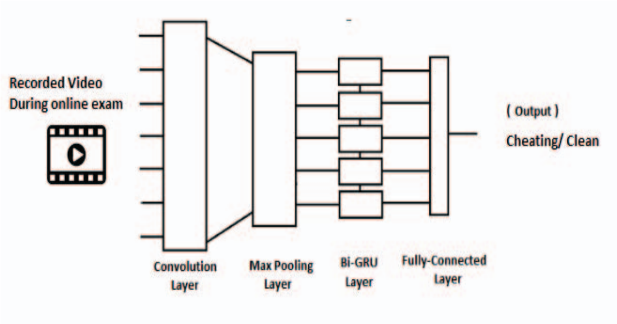


Fig 2. CNN-BiGRU architecture

C. Video Pre-processing

Video pre-handling is basic since it assists resulting phases with run quicker. Additionally, it changes all recordings in an average unraveling design with static goal and fps values. We set our recordings at a consistent 3 fps rate. It might appear to be a piece lower for a standard frame grouping yet separating significant cheating relations between frames is yet sufficient. Additionally, we involved 400-pixel widths for outlines while keeping unique perspective proportions of them.

D. Face detection of Candidate

We utilize HOG-based SVM and CNN-based identifier as core computations for facial recognition. We focus on the applicant's face, record its face encodings for cheating investigation. Currently, only HOG identification is used. We analyze every appearance along the video's edges within the first 20 seconds. A head assistant assesses faces by pointing at the screen. Face photographs outside predefined bounds are not recognized. Encoded face boundaries are saved for valid faces.

E. Face Recognition

During processing the full footage, we examine each edge to check for the candidate's face, add up the faces and bodies, and look for any technological devices. We build a link between the encodings of the enlisted face and the applicant's face for each detected face while exploring the frames. We refer to the enlisted outlines that were taken from the first 20 seconds of the film as Fs, where s is the frame number in the first 20 seconds. Crossed outlines are referred to as Ft, which displays the edge records for the full footage. With the use of a continuous thresholding tool, anomalous face separations are resolved. The frames with someone else are those edges with face removal that go beyond the limit, which is 0.65 in our studies. Afterwards, we connect these frames to each other based on their proximity in the sequence to determine how frequently someone else appeared in the video. On the off chance that those edges associate (assuming they are near one another), each arrangement is considered 1 appearance of someone else.

F. Object Detection

Like face discovery, object not entirely settled through a thresholding component. In any case, in contrast to confront recognition, since we acquire a certainty value from the MobilenetSSD model, we limit these qualities to conclude cheating exercises rather than distance values. For body and gadget recognition, the limits are 0.65 and 0.30 separately in the examinations. These are improved limit values. We collect corpse counts and certainty values from each frame for body finding. The edges allotted are those with a body count of 1, but no candidate's face was discovered, or with a body count of more than 1, but applicant's face was discovered. Like the face discovery part, those named outlines are connected to one another relying upon their distances in the arrangement to conclude how often numerous bodies showed up in the video.

IV. Experimental Results

In the experiments, we initially set the hyperparameters of the deep CNN and Bi-GRU model to their optimal values. Finding the best hyperparameter values in deep learning is challenging, so we started with a wide range of values and progressively refined them based on validation results. For the hybrid model's training optimization, we chose the RMSprop optimizer, which is widely used and effective. We kept the learning rate and rho parameters of RMSprop at their default values. The second fully connected dense output layer consisted of a single neuron that produced a binary output with a probability score. The first fully connected dense hidden layer contained 500 neurons. Additionally, we employed a dropout ratio of 0.5. In the model's convolutional layers, we used a kernel size of two and a maximum pooling size of two. The model underwent training for 20 epochs. To evaluate our proposed approach, we split the dataset into 80% for training and 20% for testing. After the split, we computed accuracy and evaluation metrics. We employed the confusion matrix depicted in Fig. 3 to determine the values of these TP, TN, FP, and FN. Fig. 4 shows sample pictures of the preprocessed dataset. Out of the 200 frames, 85 were classified as suspicious frames and 95 as clean frames, as shown in Fig. 5. The constructed hybrid model achieved an accuracy rate of 92% according to the confusion matrix in the same figure. Our proposed model demonstrated favorable recall, precision, F1-score, and accuracy, as evident from the comparison and the assessment metrics presented in Tab. 1 for the test dataset.

G. *Evaluation metrics*

Three metrics are used to evaluate how well our proposed classifier performs at spotting cheating: precision, recall and accuracy. The suggested algorithm's precision is determined as follows:

$$Precision = \frac{FP}{TN + FP} \tag{1}$$

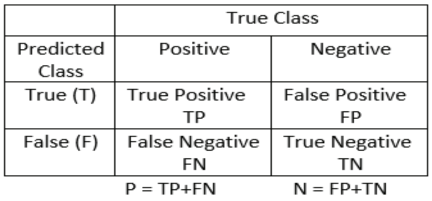
Where FP represent False Positive and TN represent True negative. The following formula is used to compute recall:

$$Recall \ = \frac{FN}{TP + FN} \tag{2}$$

Where FN represents False Negative and TP represents True Positive. The suggested algorithm's Classification Accuracy is calculated as:

$$Prediction\ accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Where TP represent True Positive, TN represents False Negative, FP represents False Positive, and FN represents False Negative.



	True Class	
Predicted Class	Positive	Negative
True (T)	True Positive TP	False Positive FP
False (F)	False Negative FN	True Negative TN
	P = TP+FN	N = FP+TN

Fig 3. Confusion matrix

Where TP is the number of accurately recognised true positives. The term "FP" stands for "false positives," or the quantity of false positives. TN stands for the number of successfully recognised false negatives. The number of false-negative occurrences that were mistakenly identified is represented by symbol FN.

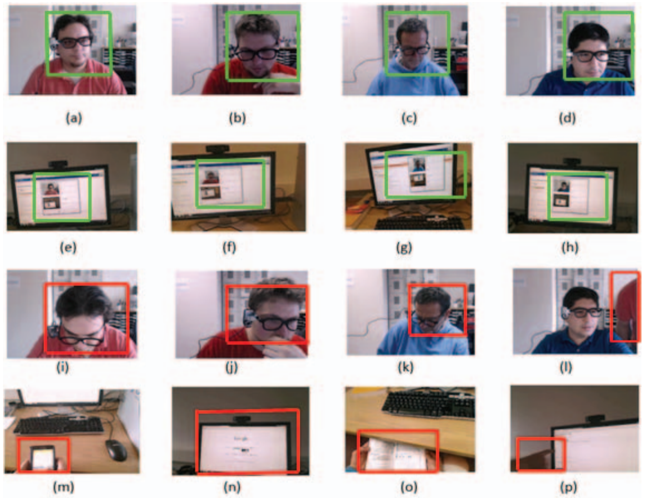


Fig 4. Some pictures extract from recorded video dataset, the pictures classified as "not cheating" are highlighted in green, while those classified as "cheating" are highlighted in red.

TABLE 1. Experimental evaluation of proposed model

Class name	Precision	Recall	F1-score	Accuracy
Yes	0.94	0.89	0.92	0.92
No	0.90	0.95	0.93	0.92

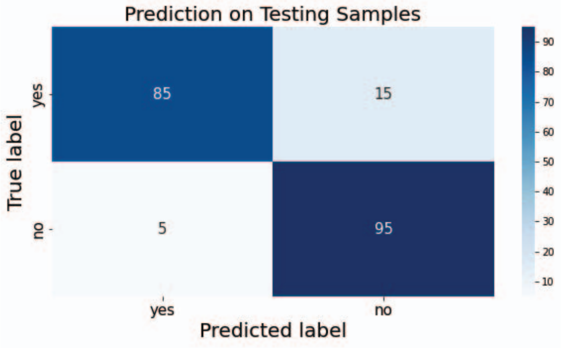


Fig 5. Testing the confusion matrix of the proposed model

V. Conclusion and Future work

The CNN-based method we developed not only facilitates cheating detection but also enables automated execution of online proctoring. Consistent training is essential to enhance the accuracy of correctly classifying the students taking online tests. Additionally, comparing the stored values in the database can contribute to improving face verification. There are several intriguing avenues for expanding on the findings presented in this study. Firstly, the system can be implemented online, allowing its utilization over the internet. Furthermore, incorporating voiceprint as an additional feature in the system can enhance continuous authentication during E-exam sessions. Continuous authentication techniques, such as face recognition, analyze and compare patterns in digital photographs to accurately identify individuals.

Another potential enhancement is the utilization of keystrokes as a form of continuous authentication. This approach can determine whether the user who initiated the initial authentication process at the beginning of the exam is the same as the one currently logged in. Overall, these proposed extensions hold promise for further improving the effectiveness and reliability of the system.

REFERENCES

[1] Abbas, M. A. E., & Hameed, S. (2022). A Systematic Review of Deep Learning Based Online Exam Proctoring Systems for Abnormal Student Behaviour Detection.

[2] Al Shbail, M. O., Esra'a, B., Alshurafat, H., Ananzeh, H., & Al Kurdi, B. H. (2021). Factors affecting online cheating by accounting students: the relevance of social factors and the fraud triangle model factors. *Academy of Strategic Management Journal, 20*, 1-16.

[3] Ali, Z. H. (2023). Cheating Detection in online exams using machine learning. *Journal Of AL-Turath University College, 2*(35).

[4] Alin, P., Arendt, A., & Gurell, S. (2023). Addressing cheating in virtual proctored examinations: toward a framework of relevant mitigation strategies. *Assessment & Evaluation in Higher Education, 48*(3), 262-275.

[5] Atoum, Y., Chen, L., Liu, A. X., Hsu, S. D., & Liu, X. (2017). Automated online exam proctoring. *IEEE transactions on multimedia, 19*(7), 1609-1624.

[6] Bawarith, R., Basuhail, A., Fattouh, A., & Gamalel-Din, S. (2017). E-exam cheating detection system. *International Journal of Advanced Computer Science and Applications, 8*(4).

[7] Bilen, E., & Matros, A. (2021). Online cheating amid COVID-19. *Journal of Economic Behavior & Organization, 182*, 196-211.

[8] Chuang, C. Y., Craig, S. D., & Femiani, J. (2017). Detecting probable cheating during online assessments based on time delay and head pose. *Higher Education Research & Development, 36*(6), 1123-1137.

[9] Cleophas, C., Hoennige, C., Meisel, F., & Meyer, P. (2023). Who's cheating? mining patterns of collusion from text and events in online exams. *INFORMS Transactions on Education, 23*(2), 84-94.

[10] D'Souza, K. A., & Siegfeldt, D. V. (2017). A conceptual framework for detecting cheating in online and take-home exams. *Decision Sciences Journal of Innovative Education, 15*(4), 370-391.

[11] EL Rhezzali, N., Hilal, I., & Hnida, M. (2023). Cheating Detection in Online Exams *Digital Technologies and Applications: Proceedings of ICDTA'23, Fez, Morocco, Volume 1* (pp. 431-440): Springer.

[12] Ferretti, S., & Roccetti, M. (2006). *AC/DC: an algorithm for cheating detection by cheating.* Paper presented at the Proceedings of the 2006 international workshop on Network and operating systems support for digital audio and video.

[13] Heriyati, D., Sari, R. L., Ekasari, W. F., & Kurnianto, S. (2023). Understanding contract cheating behavior among indonesian university students: An application of the theory of planned behavior. *Journal of Academic Ethics*, 1-24.

[14] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

[15] Hylton, K., Levy, Y., & Dringus, L. P. (2016). Utilizing webcam-based proctoring to deter misconduct in online exams. *Computers & Education, 92*, 53-63.

[16] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260.

[17] Kamalov, F., Sulieman, H., & Santandreu Calonge, D. (2021). Machine learning based approach to exam cheating detection. *Plos one, 16*(8), e0254340.

[18] Keresztury, B., & Cser, L. (2013). New cheating methods in the electronic teaching era. *Procedia-Social and Behavioral Sciences, 93*, 1516-1520.

[19] Ketab, S. S., Clarke, N. L., & Dowland, P. S. (2022). A robust e-invigilation system employing multimodal biometric authentication.

[20] Khan, A. R., Saba, T., Khan, M. Z., Fati, S. M., & Khan, M. U. G. (2022). Classification of human's activities from gesture recognition in live videos using deep learning. *Concurrency and Computation: Practice and Experience, 34*(10), e6825.

[21] Krambia Kapardis, M., & Spanoudis, G. (2022). Lessons learned during Covid-19 concerning cheating in e-examinations by university students. *Journal of Financial Crime, 29*(2), 506-518.

[22] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436-444.

[23] Liu, F., Chen, Z., & Wang, J. (2019). Video image target monitoring based on RNN-LSTM. *Multimedia Tools and Applications, 78*, 4527-4544.