Transfer Learning in Detecting E-Assessment Malpractice from a Proctored Video Recordings.

Uzoma Chidimma Anthonia, Dr. Bernadine Ifeoma Onah, Dr. Hippolyte Michael Tapamo, Afonughe Endurance.

1, 2, 4 Department of Computer Science, University of Nigeria, Nsukka (UNN), Nigeria

3 Department of Informatics, University of Yaoundé 1, Cameroon.

DOI: https://doi.org/10.51244/IJRSI.2025.120800090

Received: 04 Aug 2025; Accepted: 09 Aug 2025; Published: 08 September 2025

ABSTRACT

E-assessment Malpractice Image classification is very important in education applications to develop an effeactive studies. In this paper, we use imagenet benchmark dataset to classify five types of examination taking and malpractices involved (a) cheat from text books/notes/papers. (b) Using a phone to call. (c) asking a friend in the test room (d) operating a phone and (e) a normal exam taking without cheating. Due to the small number of training dataset, our classification systems evaluate deep transfer learning for feature extraction. . During these exams it is difficult to keep track of every student's screen at the same time to check if anyone is showing fraudulent behaviour. Even when recording all students' activities during exam and watching it afterwards to depend if they cheated is very labour intensive. This thesis uses a special type of Convolutional neural network called Inception which is are widely known for their effectiveness in image classification on students' proctored video recordings to determine if they show any malpractice behaviors, allowing us to build a framework to automate this labour intensive process system. The objective of this study is to increase the classification accuracy, speed the training time and avoid the overfitting. In this study, we trained our architecture to involve minimal pre-processing for 30-epoch number in order to study its impact on classification performance and consuming time. In addition, the paper benefits acceptable results with small number of epoch in limited time. Our interpretations confirm that transfer learning provides reliable results in the case of small dataset. The proposed system outperforms the state-of-the-art methods and achieve 96.8% classification accuracy.

Keywords: Convolutional Neural Network, Exam Malpractice, Classification, Deep learning, E-assessment, Transfer learning.

INTRODUCTION

Information, communication and technology (ICT) rapid evolution have caused many changes in all spectrums and especially in education (Adeyemi, Ogunlere & Akwaronwu, 2025). The innovations have transformed the traditional methods of learning and made introductions of new models like the distance learning and online resources, which have become the part of the contemporary learning process. Though these developments have led to improved accessibility and flexibility, they have emerged associated with new problems particularly in maintaining academic integrity amid online tests (Hussain, Qureshi & Malik, 2024). Studies have shown that there is increased use of academic cheating in e-assessments. As an example, 74 percent of schoolchildren confessed in 2013 that cheating during online tests would be quite easy, and 29 percent of students admitted to having committed such a malpractice. Such an increasing need demands automatic systems that can secure the validity of online exams. Some early efforts at automation used conventional machine learning to classify images (Jantos, 2021). These approaches however, although effective to some degree, are time consuming and demanding requiring preprocessing and design of feature extraction by experts due to their complicated nature. Their precision is proportional to the handcraft qualities of the features used; the scalability and robustness are minimal. Deep learning approaches specifically Convolutional Neural Networks (CNNs) have served as a more adequate solution to overcome these weaknesses. The CNNs have been proved outstanding in applications of detection, localization, segmentation, and classification in various fields including medical imaging (Rawat & Wang, 2017). Their effectiveness lies in the existence of large labeled datasets, high-potent GPUs, and the use of top-notch architectures that are capable of enhancing accuracy at each passing time. The paper presents a transfer learning-oriented methodology to recognize various cheating phenomena in online exams over the Online Exam Proctoring (OEP) data (Zhu et al., 2021). The cheating actions will involve the use of the notes, calling by using a phone, seeking help by another person, impersonation and normal non-cheating conduct. The amalgamation of the proposed system has three prominent elements, interface, and video processing and frame classification. The frames of exams of 24 students recorded by the pair of cameras are converted to high-quality input by de-duplicating them and converting the input to input of high quality. A fine-tuned CNN (Convolution Neural Network) model that is based on InceptionV3 is then used to classify these frames with transfer learning to achieve better efficiency and accuracy (Ghosh, Das & Nasipuri, 2019). Transfer learning can be used to apply the parts of the knowledge learned by the models to this task to save a considerable amount of time and computation efforts, without a significant decrease in the sensory detection accuracy (Ribani & Marengoni, 2019). The findings will thereafter be availed to the instructors, supplying concrete evidence of dubious acts. The proposed framework seeks to improve academic integrity by providing a scalable white label solution to malpractice detection in e-assessments where the challenge of malpractice is one of the most critical concerns in current digital education.

Problem Statement

The risk of examination malpractice has increased as e-learning and use of online assessment gains rapid momentum. The current and conventional systems in proctoring, including Safe Exam Browser and Online Exam Proctor, do not seem to suffice because they lack scalability, accuracy, and access to the internet during exams. Proctored video recordings can be manually monitored, but this is labor intensive and not applicable to large-scale assessment. Current solutions all bases on simple detection of faces with OpenCV, making it impossible to detect in full the possibility of suspect actions such as external communication or access of unauthorized external resources. In addition, there are no strong automated options available, which said, questions the credibility of online credentials, leaving institutions with problems of student integrity validation. Efforts to adopt advanced image classification technologies in fraud detection have also been limited by computational limitations. An efficient, scalable, and accurate framework built on deep learning and transfer learning is necessary in automating detection of malpractice on proctored video recordings.

LITERATURE REVIEW

Overview of e-assessment security methods.

E-assessment security focuses on ensuring the integrity, authenticity, and fairness of online examinations. Various methods have been developed to prevent and detect malpractice during remote assessments:

Authentication and Identity Verification: Techniques like password authentication, two-factor authentication (2FA), biometric verification (face recognition, fingerprint), and ID card checks ensure the candidate's identity before the exam begins (Al-Mutairi & Al-Sahli, 2024).

Secure Browsers and Lockdown Tools: Applications such as Safe Exam Browser restrict access to external websites, applications, and communication tools during the exam, preventing unauthorized information searches.

AI-Powered Remote Proctoring: Proctoring systems monitor candidates using webcams, microphones, and screen recordings. Advanced tools incorporate facial recognition, gaze tracking, and posture analysis to detect suspicious behavior.

Plagiarism Detection and Content Protection: Anti-cheating measures include randomizing questions, time limits, and plagiarism detection software for essay-based assessments.

Live and Automated Proctoring: Exams may be monitored by human proctors in real-time or through AI systems that analyze video streams and flag anomalies for review.

Traditional approaches to malpractice detection

Traditional methods for detecting malpractice in e-assessments have primarily focused on preventive measures and basic monitoring techniques rather than advanced automated systems. These approaches include:

Table 1: Traditional Approaches to Malpractice Detection

Approach	Description	Limitations
Human Proctoring	Invigilators monitor candidates via live video streams.	Labor-intensive, costly, and not scalable for large numbers of candidates.
Secure Browser Tools	Software (e.g., Safe Exam Browser) restricts access to other apps and websites during exams.	Ineffective for open-book exams and cannot prevent external communication.
Plagiarism Detection	Tools like Turnitin check for copied or reused content in written assessments.	Limited to text-based assessments; does not address behavioral cheating.
Face Detection	Facial recognition used for initial identity verification at the start of the exam.	No continuous monitoring; cannot detect behavioral malpractice.
Manual Video Review	Proctors review recorded sessions to detect suspicious activities.	Time-consuming, prone to human error, and lacks real-time intervention.

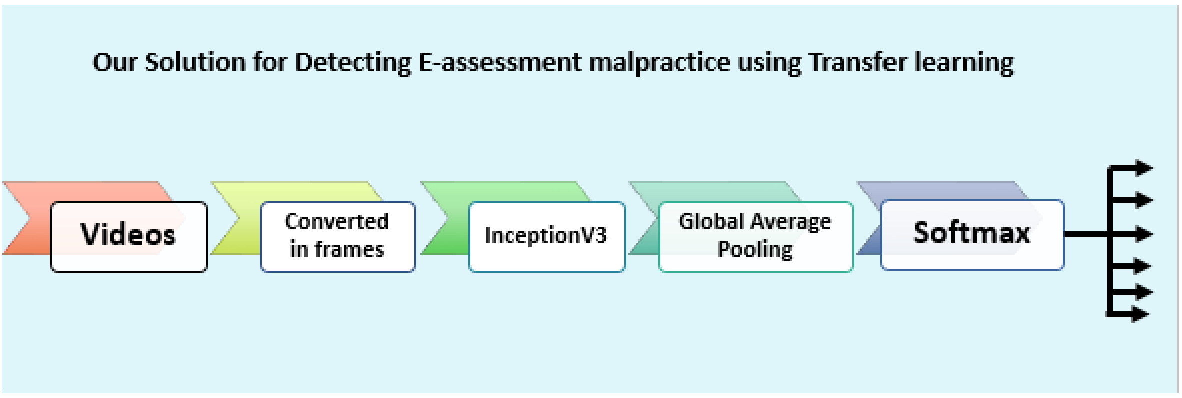
Role of computer vision and deep learning in video analysis

Computer vision and deep learning have ushered a new way of analyzing videos by proposing automated, precise, and scalable answers to understanding complex visual data (Manakitsa et al., 2024). These technologies play a very important role in establishing the integrity of online examination in the case of detecting e-assessment malpractice. Through computer vision, systems can trace the video streams of proctored sessions and detect possible suspicious activities including movements of hands, abnormal head movements, or references to unauthorized communications (Taherdoost, 2023). This is further improved using deep learning, specifically, Convolutional Neural Networks (CNNs), in which a system can learn what are convoluted patterns within a large dataset, and thus select minute cheating behaviors that human invigilators may overlook (Yulita et al., 2023). Facial recognition guarantees constant verification of identity during the exam, whereas pose estimation and gaze tracking provide a possible tool of evaluating attention and engagement levels. Other capabilities of the systems, such as temporal analysis of the behavior using 3D CNNs or Recurrent Neural Networks (RNNs) help to determine the behavioral patterns over the time and enhance accuracy in detecting an anomaly (Sukumaran & Manoharan, 2024). The methods greatly limit the shortcomings of manual monitoring which is prone to errors and very tedious. Furthermore, deep learning models can decode and segment features of a single frame and video resume, providing scale to analysis in real time (Mortezapour, Perumal & Mohamed, 2024). Transfer learning allows using pre-trained models to reduce the computational cost at the expense of a decreased detection rate (Karim & van Zyl, 2021). The pairing of computer-vision and deep learning given the combination provides a strong solution towards bringing automation to malpractice detection in e-assessments and promoting credibility and trust in online educational systems by bringing down the reliance of human proctors.

METHODOLOGY

The study uses a Research Developmental Methodology, which is a combination of structured video processing and deep learning methods to discover any form of malpractice in proctored e-assessment. This starts by recording videos of invigilated online tests. These frames are interpolated at a set interval to also limit the size of the recording on student activity, over time.

Preprocessing includes resizing (to 299x299 pixels), normalization of the pixel values, and denoising of the data by using a Gaussian filtering to increase clarity and homogeneity, conducted in each frame. The frames are then labeled as normal frames or suspicious frames according to the observed behavior patterns giving a labeled dataset to both train and test. InceptionV3 is the main convolutional neural network of which we use transfer learning with pre-trained weights used on ImageNet. This model has been chosen due to its balance between efficiency (computationally) and accuracy (as a classifier). Nonetheless, in future generations, the performance will be benchmarked with Resnet 50 and EfficientNet- B0 to support model selection. The latter parts of the network consist of Global Average Pooling (GAP) layer to prevent overfitting and, after it, a Softmax activation response to associate the behavior with preset classes. With a training-to-testing ratio of 80:20, the dataset was divided, and 5-fold cross-validation was used in order to guarantee the model generalizability and minimize the variance in the model performance indicators. Countermeasures like no overlapping of sessions between training and testing were put into place to address prevention of data leakage. We used accuracy, precision, recall, and F1-score as our key performance measures in order to assess the performance of the models. The model obtained an accuracy of 96%, although it is in future interest to determine the robustness in alternative video environments and deployment settings. Ethical compliance was highly followed. Informed consent was gained in collection of data and prior ethical clearance was obtained in the host institution. The data of all the videos was anonymized and secured in accordance with GDPR and institutional review board (IRB) regulations. This framework has a scalable and robust archive solution that supports automatic academic integrity monitoring of digital assessment.


Our Solution for Detecting E-assessment malpractice using Transfer learning
Videos
Converted in frames
InceptionV3
Global Average Pooling
Softmax

.

Figure 1: Procedure for E-assessment Malpractice using Transfer Learning

Data Collection

Our dataset was gotten from OEP (https://www.cse.msu.edu/computervision/OEP_database.tar.g),

Which consists of three parts: an interface, video processing and frame classification. (a) Cheating from text books/notes/papers, (b) using a phone to call a friend, (c) asking a friend in the test room, and (e) having another person take the exam other than the test taker. This tool which feature recordings of the entire monitor of 24 students during the time of exam uses two cameras in a way that the face of the student and the monitor were clearly visible in both cameras. The datasets would send videos of all the students' recordings to a pipeline that consists of a series of methods. This pipeline will be used to separate the videos into frames. The first part, video processing, would shorten the video from a two to three-hour long video down to a few thousand frames, leaving as few duplicates or similar looking frames as possible.

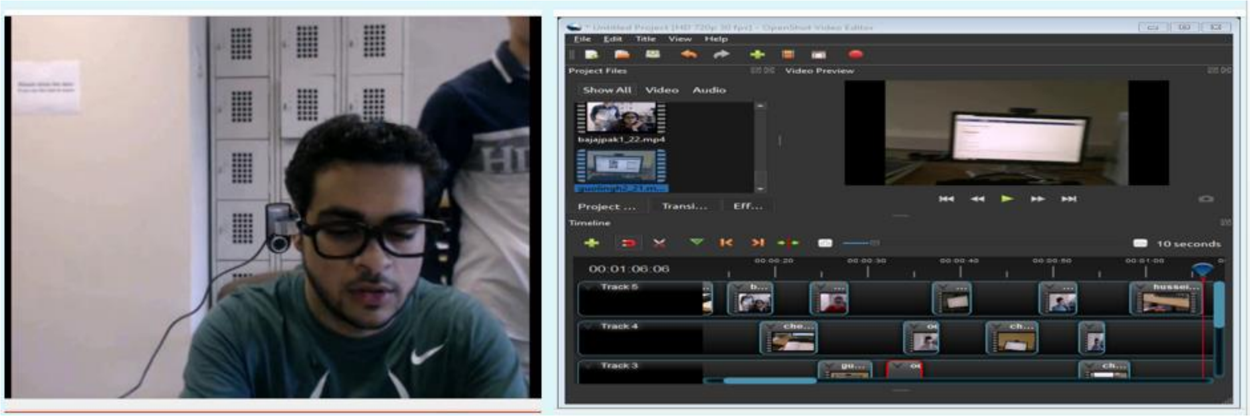


Figure 2: OEP dataset illustrating video to frames processing with Openshot

The combination of these two types of subjects enriches the database with various e-assessment malpractice techniques, as well as the sense of engagement in real exams. For each of 24 sessions, we collect the audio and two videos from both cameras as seen in Fig. 2. Each session varied in length with an average time of 17 minutes.

Annotating and preprocessing for training

Human annotation and labeling is performed offline after collecting the data by viewing the two videos and audio simultaneously and using Openshot video editor and DVDvideosoft Studio. The labeling of one cheat instance consists of three pieces of information: the start time, end time and type of cheating. We label 5 different types of cheating behaviors: (1) cheating from a book, notes or any text found on paper. (2) talking to a person in the room. (3) using the Internet. (4) asking a friend a question over the phone. (5) using a phone. The labeling process for every session is done carefully and required nearly 30∼35 minutes per session. Section 3.4 illustrates examples of different types of cheating from various subjects. The total duration of all types of cheating is reported to be 7, 235 seconds. The total number of cheat behaviors performed by all subjects is equal to 569 instances, varying in the type and duration of cheating. The five cheat types defined in our system cover all kinds of cheating behaviors we could manually identify in the collected OEP dataset. It is reasonable to assume that they are also the most common cheating techniques in the real world. Note that the techniques used within a specific type can vary from one subject to another, increasing the level of difficulty in detecting some of the instances. For example, some student may open a book in front of them to cheat from, while others hide the book behind the computer screen or below the desk introducing partial occlusion. Moreover, some students talk in a room with another person asking for help where both are visible in one of the cameras, while others might talk with another student who is not visible in any of the two videos. Some speak with a low voice (i.e., whispering), while others speak normally. Many other variations are also present in this dataset, since we did not constrain the subjects in how to cheat.

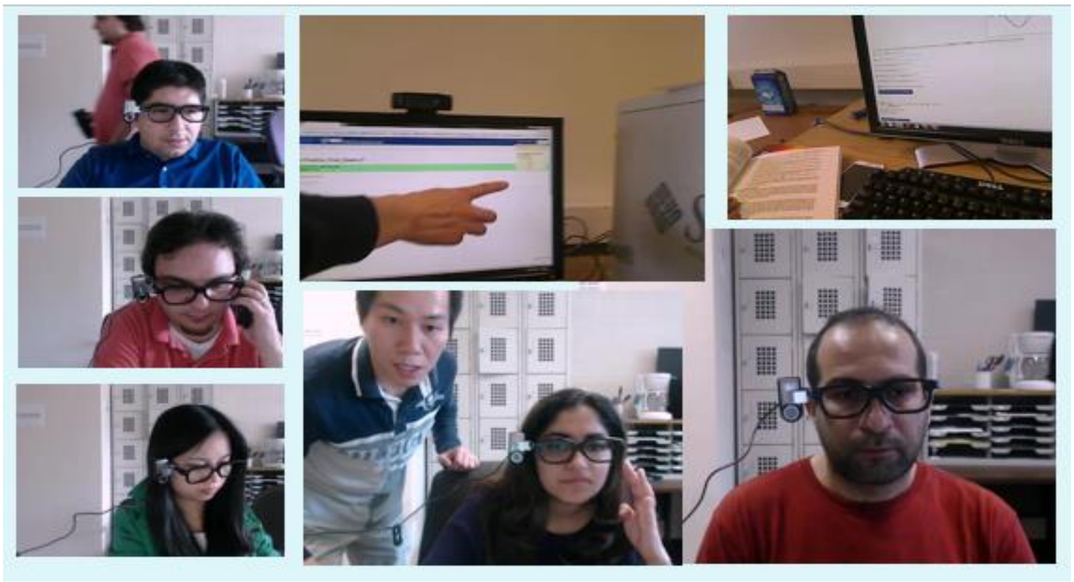


Figure 3: the InceptionV3 malpractice classifier combines the 4 classes of malpractice and one class of normal exam taking instances as shown.

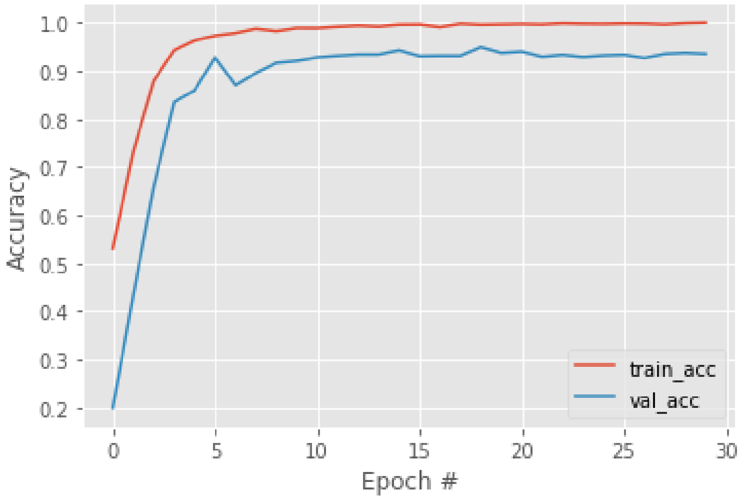
OEP dataset examples illustrating various cheat types. The examples are grouped in pairs showing both webcam and wearcam at a specific time of the exam. The subjects are cheating from books, notes, papers, smartphones, the Internet, or asking someone in the room and normal exam taking.

RESULTS

Table 2: Performance metrics

Classes	Precision	Recall	F1-score	Support	Accuracy
Checking from Book	0.97	0.93	0.95	150	
Calling someone on the phone	1.00	0.95	0.97	150	
Another person helping	0.97	0.97	0.97	150	
Normal example	0.97	0.99	0.98	150	
Operating a phone	0.91	0.97	0.94	150	
				750	0.96

Table 2 shows that the model performed well in all classes scored at 96% accuracy. Moreover, the F1-score peak of 0.98 is noted in the category of Normal example, which demonstrates an excellent capability to classify non-malpractice behavior correctly. The process of calling someone using the telephone achieves a perfect precision (1.00) with recall as 0.95, which is indicating minimal false positives. Although in each of the remaining classes the F1-scores are higher than 0.94, the precision of operating a phone (0.91) indicates that there are some misclassifications of this behavior in comparison to the rest. No doubt, overall, the model presents strong reliability and equal-performance in all situations of cheating and non-cheating.



Training and validation accuracy curve on their different data

Figure 4: Model Accuracy

The graph of Figure 4, shows that training and validation accuracy are rapidly rising over the first few epochs, and remain stable after approximately 10 epochs. The training accuracy has a close value of 1.0, whereas the validation accuracy reached a flat line of 0.94, which means almost no overfitting. The convergence in the curves indicates good extrapolation on unseen data.

Training and validation error curve on their different data

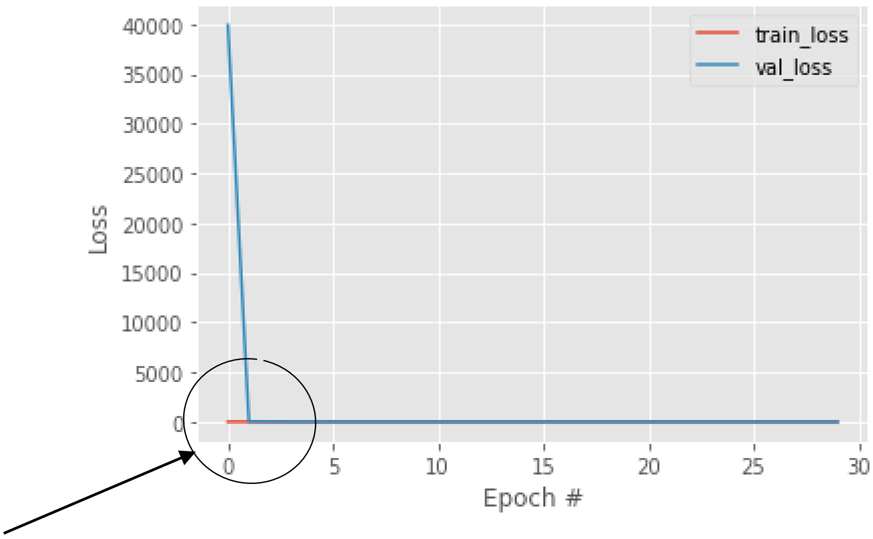


Figure 5: Model Loss

In Figure 5, the training and validation loss demonstrated in the graph reflects on the model obtained after 30 epochs. Both losses decline rapidly in a couple of epochs and converge close to zero, which demonstrates that convergence is achieved fast. The small distance between the training and the validation loss indicator low level of overfitting combined with high level of generalization. It means that this performance proves the efficiency and accuracy of model to minimize classification errors.

DISCUSSION

The assessment of the model based on InceptionV3 indicated an excellent assessment on all of the established categories of behaviors. Table 2 indicates that the model had an overall accuracy of 96 percent and F1-score of more than 0.94 on all classes. One of the most notable examples was in the category of normal behavior with the highest F1-score of 0.98 proving that the model can knowingly classify non-malpractice factors.

Comparatively, the least precision corresponded with the category of operating a phone with 0.91 indicating a minimal percentage of false positives- this could be related to similar gesture of using hands in a familiar behavior. However, the model had a balanced classification in all conditions, such as calling someone (precision: 1.00, recall: 0.95) reevaluating its good performance. The training dynamics within the 30 epochs are shown in figures 4 and 5. The validation and training accuracy increased quickly in the first 10 epochs with only a slight difference that converged at 0.94 showing that overfitting is minimal. The training loss decreased significantly and the little difference between training and validation loss represent high generalization ability. These findings indicate that the model is efficient and well regularized, which can extrapolate performance on new and unseen data. Although these performance metrics are encouraging, the wider value of this work is in its low resource application of video-based behavioral classification with a highly computationally motivated transfer learning paradigm. However, in contrast to the state of business before this research, which requires multi-camera arrangements or high-cost sensor integration, a single video stream as input to the system and frame-wise processing can generate results comparable or even better when combined with optimized CNN structures. Nevertheless, the novelty of the research is not progressive, and in the future, it is necessary to consider adding temporal models (e.g., LSTMs or transformers) to represent motion sequence better. Furthermore, an enriched reading of captured behaviors beyond face-emotion cues or eye-gaze-tracking might go beyond the binary classification system into finer-grained academic integrity analytics.

CONCLUSION

This research discussed the limitations in the current traditional methods of e-assessment malpractice identification, which are mainly dependant on face detection that cannot be able to capture the complex acts of malpractice. In closing this gap, a multi-media analytics system has been suggested to provide checks and balances to academic integrity in online assessment testing. The framework takes long proctored video, subdivides them into manageable frames, and identifies them according to fine-adjusted convolutional neural networks. Malpractice behaviors are divided into four classes of cheating and one of normal, and summarized by instructors using an easy interface. The implementation used state-of-art CNN network, InceptionV3, to extract low-level and significant high-level features in captured video frames such as gaze estimation, phone detection, and window activity. Such characteristics allow detecting suspicious activity in temporal sequences correctly, and detection becomes reliable. An experimental proof with the dataset of 24 test takers proved the ability of the framework to recognize 96 percent of cases of cheating across various scenarios. Outcomes affirm that deep learning and transfer learning have potential to detect malpractice in e-assessment in a scalable and automated manner. It is hoped that future study directions are based on the integration of multimodal conversations, real-time processing and even more responsive architecture to improve further system accuracy and dependability. These results indicate a major milestone in terms of maintaining credibility in online learning and building confidence when it comes to digital learning services.

REFERENCE

1. Adeyemi, J., Ogunlere, S., & Akwaronwu, B. (2025). Real-Time Detection of Examination Malpractices Using Convolutional Neural Networks and Video Surveillance: A Systematic Review with Meta-Analysis. British Journal of Computer, Networking and Information Technology, 8, 15–50. Retrieved from https://doi.org/10.52589/BJCNIT-QC5EELJE.
2. Al-Mutairi, A., & Al-Sahli, R. (2024). Secure Authentication System Based on Multi-Factor Authentication. Social Science Research Network (SSRN). Retrieved from https://doi.org/10.13140/RG.2.2.24880.74247.
3. Ghosh, S., Das, N., & Nasipuri, M. (2019). Reshaping Inputs for Convolutional Neural Networks—Some Common and Uncommon Methods. Pattern Recognition, 93, 332–348. Retrieved from https://doi.org/10.1016/j.patcog.2019.04.009.
4. Hussain, M., Qureshi, Z., & Malik, S. (2024). The Impact of Educational Technologies on Modern Education: Navigating Opportunities and Challenges. Global Educational Studies Review, IX(3), 21–30. Retrieved from https://doi.org/10.31703/gesr.2024(IX-III).03.
5. Jantos, A. (2021). Motives for Cheating in Summative E-Assessment in Higher Education - A Quantitative Analysis. In J. Zaharia & M. Deac (Eds.), Proceedings of the 13th International

Conference on Education and New Learning Technologies (pp. 1764-1770). IATED. Retrieved from https://doi.org/10.21125/edulearn.2021.1764.

6. Karim, Z., & van Zyl, T. L. (2021). Deep/Transfer Learning with Feature Space Ensemble Networks (FeatSpaceEnsNets) and Average Ensemble Networks (AvgEnsNets) for Change Detection Using DInSAR Sentinel-1 and Optical Sentinel-2 Satellite Data Fusion. Remote Sensing, 13(21), 4394. Retrieved from https://doi.org/10.3390/rs13214394.
7. Mortezapour Shiri, F., Perumal, T., & Mohamed, R. (2024). A Comprehensive Overview and Comparative Analysis on Deep Learning Model. Journal on Artificial Intelligence, 6(5), 301–360. Retrieved from https://doi.org/10.32604/jai.2024.054314.
8. Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation, 29, 2352–2449. Retrieved from https://doi.org/10.1162/NECO_a_00990.
9. Ribani, R., & Marengoni, M. (2019). A Survey of Transfer Learning for Convolutional Neural Networks. In C. C. Mariotti, M. F. Ortega, L. C. Quaresma, & L. A. Raimundo (Eds.), Proceedings - SIBGRAPI 2019 - XXIV Brazilian Symposium on Graphics and Image Processing (pp. 47-57). São Paulo, Brazil: IEEE. Retrieved from https://doi.org/10.1109/SIBGRAPI-T.2019.00010.
10. Sukumaran, A., & Manoharan, A. (2024). Multimodal Engagement Recognition From Image Traits Using Deep Learning Techniques. IEEE Access, PP(1), 1–1. Retrieved from https://doi.org/10.1109/ACCESS.2024.3353053.
11. Taherdoost, H. (2023). Deep Learning and Neural Networks: Decision-Making Implications. Symmetry, 15(9), 1723. Retrieved from https://doi.org/10.3390/sym15091723.
12. Yulita, I. N., Hariz, F. A., Suryana, I., & Prabuwono, A. S. (2023). Educational Innovation Faced with COVID-19: Deep Learning for Online Exam Cheating Detection. Education Sciences, 13(2), 194. Retrieved from https://doi.org/10.3390/educsci13020194.
13. Zhu, W., Braun, B., Chiang, L., & Romagnoli, J. (2021). Investigation of Transfer Learning for Image Classification and Impact on Training Sample Size. Chemometrics and Intelligent Laboratory Systems, 211, 104269. Retrieved from https://doi.org/10.1016/j.chemolab.2021.104269.