Deep Learning-Based Smart Invigilation System for Enhanced Exam Integrity

Saravanan Arumugam*

Department of Computing, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

Received 08 August 2024; received in revised form 02 November 2024; accepted 06 November 2024

DOI: https://doi.org/10.46604/peti.2024.14105

Abstract

This study proposes a smart invigilation system to preserve exam integrity by detecting suspicious student behaviors using deep learning. The model consists of three phases, i.e., student identity verification using face recognition, behavioral sampling for model training utilizing gesture analysis and convolutional 3D networks for emotion analysis, and live video analysis of suspicious activities integrating gesture, emotional analysis, and face recognition. The model is evaluated using 4,000 training and 1,000 test images and identifies non-cheating activities with 99% accuracy and cheating activities with 97.6% accuracy. The proposed model outperforms other methods, achieving accuracies of 98.4% for identifying cheating behaviors and 99.2% for non-cheating behaviors, resulting in an overall accuracy of 98.8% and a low misclassification rate of 1.2%. While the system demonstrates high accuracy, challenges remain in scaling to larger classrooms due to increased computational demand and the need for additional hardware to ensure comprehensive monitoring.

Keywords: suspicious activity detection, exam integrity, deep learning, face and gesture recognition, emotion analysis

1. Introduction

For numerous educational institutions, examinations and evaluations often play a crucial role in assessing students' knowledge, capabilities, and proficiency across a wide range of subjects and courses related to their respective disciplines [1]. These examinations, which may be in the form of written tests, projects, assignments, presentations, or online tests, are not only mandatory but also fundamental for assessing the intellectual level and academic performance of students [2]. These forms of assessment help determine students' theoretical and practical knowledge, as well as their competence level.

Despite various assessment methods, written exams remain the most popular and conventional evaluation method. This method involves providing students with question papers and requiring them to write their answers within the allotted time, under the supervision of the invigilators [3]. Invigilators are responsible for maintaining the integrity and fairness of assessments by preventing dishonest activities from students. Specifically, actions including observing their neighbors' answers with head movements, turning back and sideways, whispering answers, extending their hands forward and backward to exchange answer sheets, or copying answers from other materials are signs of students breaking the rules of fair and impartial examinations supervised by invigilators [4]. Due to the prevalence of cheating and academic dishonesty, maintaining exam integrity presents significant challenges, even though it appears to be a literal sense of simple responsibility for exam supervisors [5].

* Corresponding author. E-mail address: saravanan.a@cit.edu.in

English language proofreader: Chih-Wei Chang

To reduce misunderstandings and human errors during exam invigilation, a few studies have suggested automated invigilation systems for monitoring students during their examinations [6]. These proposed systems have likely utilized a variety of hardware, such as microphones, speakers, fingerprint sensors, and surveillance cameras, which, however, can incur additional expenses [7-9]. Furthermore, existing studies have utilized deep learning (DL) methods like convolutional neural networks (CNN) or simple machine learning (ML) algorithms like support vector machines (SVM) or random forest (RF) to evaluate the captured images. However, the performance of these methods needs further improvement [10]. Moreover, these methods have been able to capture and assess a limited number of students at a given time frame during the examination.

Nevertheless, the time taken to process the images has been significantly high [11-12]. Thus, it is necessary to propose a smart monitoring system that operates at a lower cost with high accuracy for monitoring students during examinations. The problem addressed in this study is the inadequacy of current automated invigilation systems in effectively and efficiently monitoring student behavior during exams, which results in insufficient detection of academic dishonesty and compromised exam integrity.

This study presents a novel approach to addressing the challenges of academic dishonesty in examinations through an automated invigilation system that utilizes DL algorithms for facial, gesture, and emotion recognition. The primary objective is to develop a smart exam invigilation system that captures suspicious dishonest activities and malpractice in real-time examinations at higher education institutions, thereby preserving exam integrity. The specific objectives of the research are to maintain exam integrity, reduce human errors, alleviate invigilator workload, and assess student emotions to detect suspicious activities.

The proposed smart invigilation system employs closed-circuit television (CCTV) to capture student images during exams and operates in three phases using DL techniques. First, students' identities are verified through facial recognition with a single-shot multi-box detector (SSD); second, behavioral sampling is generated through gesture analysis using You Only Look Once version 5 (YOLOv5) and emotional analysis using convolutional 3D networks (C3DN); third, real-time video by integrating gesture and emotional analysis is analyzed along with pre-defined decision rules to classify malpractices from normal activities. Thus, the contributions of the research include:

(1) Development of student identity verification using facial recognition employing SSD.
(2) Development of a gesture analysis model using YOLOv5.
(3) Development of an emotion analysis model using C3DN.
(4) Creation of a dataset and generation of behavioral sampling records.
(5) Implementation of the smart invigilation system with facial, gesture, and emotion analysis.
(6) Assessment and evaluation of the performance of each phase, along with identification of limitations.

With improved performance and integrity, this automated invigilation system represents a significant advancement in examination monitoring by leveraging DL technologies for the comprehensive detection of academic dishonesty. By integrating facial, gesture, and emotion recognition, the proposed solution enhances the accuracy and efficiency of exam invigilation and fosters a fair assessment environment, ultimately contributing to the preservation of academic integrity in higher education.

This paper is structured as follows: Section 2 delves into a review of relevant literature in the field of smart invigilation. Section 3 presents the proposed methodology, providing a detailed explanation of the framework and the working procedure. Section 4 elaborates on the database used, the implementation of the model, and the performance metrics used for assessing the proposed model. Section 5 discusses the results obtained for the proposed model as well as the study's limitations. Finally, Section 6 concludes the work along with future recommendations.

2. Related Works

Owing to technological advancement and digitalization, surveillance cameras like CCTV play a significant role in humans' daily activities. Not only do shopping malls and stores use these surveillance cameras for security, but educational institutions also use them to detect and mitigate suspicious activities. However, monitoring these activities manually is a tedious and time-consuming process with a high potential for human error. Such an inefficiency highlights the need for automated systems. Several researchers have proposed various ML and DL models to recognize suspicious activities in surveillance videos.

Hernándeza et al. [13] developed a model to detect and prevent cheating in online assessments by analyzing student personalities, stress situations, and cheating practices using a model of data mining to detect cheats (DMDC) and Weka data mining. A model proposed by Atoum et al. [14] introduced a system that uses six components to detect user verification, text, voice, active window, gaze estimation, and phone, accurately identifying cheating during online exams using multimedia data from 24 subjects. Kamalov et al. [6] proposed a novel method for identifying potential cheating cases on final exams through a post-exam analysis of student grades. The method employs long-short-term memory (LSTM) and kernel density estimation (KDE)-based outlier detection to identify potential cheating cases, achieving high accuracy, and thereby enhancing academic integrity in course assessments.

Hoque et al. [7] proposed a framework for traditional examination systems, reducing invigilators, eliminating student malpractices, and requiring educational institutions to maintain a database using a parallax data acquisition tool. Examinants undergo biometric authentication before entering the hall, while invigilators use CCTV cameras and ultra-sensitive microphones to monitor physical and vocal malpractice during the exam. Tiong and Lee [15] developed an e-cheating intelligence agent using internet protocol (IP) and behavior detectors to monitor student behavior, prevent malicious practices, and integrate with online learning programs.

Kohli et al. [16] developed a real-time computer vision system using 3D CNN, object detector methods, OpenCV, and Google Tensor Flow to predict exam fraud with a 95% correlation. Mahmood et al. [17] developed a DL exam invigilation system using a faster regional convolution neural network (Faster R-CNN) and face recognition, achieving 99.5% and 98.5% accuracy, respectively. Genemo [18] developed "L4-BranchedActionNet" using surveillance footage for identifying suspicious student behavior during exams, achieving 92.99% accuracy on CUI-EXAM, a benchmark dataset for exam monitoring, and 89.79% accuracy on CIFAR-100, a widely used image repository with 100 classes for classification tasks. However, performance needs more improvement. Similar to this work, Asad et al. [10] developed a DL-based CNN model using cameras to detect cheating patterns, generating reports for invigilators and aiding in effective exam cheating prevention strategies.

The technique proposed by Al_airaji et al. [19] detects cheating by analyzing students' head and iris movements, identifying shared abnormal behavior, and alerting authorities, reducing manual monitoring error rates. Kadthim and Ali [20] developed a model using multiple linear regression, SVM, RF, and k-nearest neighbour (KNN) classifiers for student score prediction, achieving a 96% accuracy rate. Alsabhan [21] developed an ML method using the 7WiseUp behavior dataset to identify exam-cheating incidents, improving student well-being and academic performance with a 90% accuracy rate. Zhou and Jiao [22] utilized the stacking ensemble ML algorithm to detect cheating behaviors in students' responses, revealing superior performance in item responses and summary statistics. Chang and Chang [23] utilized feature representation methods and ML algorithms to identify cheating in multiple-choice tests, using visual detection and small-sample examples. Ong et al. [24] proposed a model utilizing CCTV cameras to monitor students for cheating, achieving 83% accuracy with training on 50 behavior videos, thereby enhancing exam integrity.

Emotions revealed by the students also played a significant role in detecting cheating activities. However, only a few studies focused on emotion analysis. Ozdamli et al. [25] developed a facial recognition system using computer vision and DL algorithms for online learning invigilation, detecting student behaviors and abnormalities. Cîrneanu et al. [26] studied the evolution of neural network architectures in facial emotion recognition (FER), focusing on CNN-based ones and analyzing gestures and emotions for student cheating detection.

Nishchal et al. [4] utilized OpenPose for posture detection, AlexNet for cheating types, and sentiment analysis for emotion analysis, claiming that combining these methods improved cheating detection performance. Recently, Liu et al. [9] utilized multiple-instance learning to identify cheating behaviors in online exams, enabling precise annotations from labeled instances. Verma et al. [11] employed a multimodal DL approach to monitor students, detect emotions, estimate head pose, and track mouth movements, aiming to replace human proctors.

Therefore, the literature indicates a significant number of studies in this field. The summary of key studies, their performance, and limitations is presented in Table 1, where all these studies employed synthesized datasets for evaluation. The limitations and research gaps identified highlight the importance of developing a more robust, accurate, and scalable exam monitoring system to enhance exam integrity. While previous research has made strides in exam integrity, the proposed model addresses several critical gaps: it enhances scalability by efficiently managing large spaces and incorporates comprehensive analysis through gesture and emotion detection. These advancements position our model to better respond to the variability and complexities of human behavior during examinations, ultimately providing a more effective solution for ensuring exam integrity.

Table 1 Summary of key literature on exam integrity systems

Author	Year	Technique used	Achievements	Model performance	Limitations
El Kohli et al. [16]	2022	DL	Novel fraud detection approach	F1-score: 0.80 Precision: 0.78	High complexity; evaluation with small datasets
Mahmood et al. [17]	2022	DL (multiple algorithms)	Intelligent system with high accuracy	Accuracy: 94.7%	High computational complexity
Genemo [18]	2022	DL	Better cheating behavior detection	Accuracy: 85%	Limited scalability; potential false positives
Kadthim and Ali [20]	2023	ML	Improved online exam detection	Accuracy: 90.2%	Accuracy varies with cheating methods
Asad et al. [10]	2023	ML	Effective cheating detection	Accuracy: 88% F1-score: 0.79	Performance reliant on input data quality
Alsabhan [21]	2023	ML and DL	Accurate and effective model	Accuracy: 93.1%	Risk of overfitting; high computational demands
Zhou and Jiao [22]	2023	Stacking ensemble	Improved detection in large assessments	Precision: 0.551 Recall: 0.638	Training and deployment complexity
Verma et al. [11]	2024	DL	High accuracy and scalable	Accuracy: 92.3% Precision: 0.91	False positives; privacy concerns; implementation challenges
Liu et al. [9]	2024	ML	Improved accuracy and adaptability	AUC* score: 87.58%	Generalization issues; labeling challenges; high resource needs

*AUC: Area under curve

3. Proposed Methodology

The overview of the proposed smart surveillance system that uses behavioral sampling to ensure exam integrity using DL techniques is presented in Fig. 1. The framework involves three phases:

(1) identity verification of students during examinations;

(2) behavior sampling using gesture and facial expression analysis;

(3) live video analysis for suspicious activity recognition.

The first phase involves the pre-processing of images using a face recognition model and comparing faces with those in the database. Behavioral sampling, i.e., the second phase, involves capturing video, pre-processing the frames, detecting gestures and expressions, labeling images, and creating a training set. The third phase, namely live video analysis, involves detecting gestures and emotions in real time and triggering alerts for suspicious activities. More specifically, the SSD was employed for face region detection, YOLOv5 was used for gesture analysis, C3DN was used for emotion analysis, and predefined decision rules were applied to classify the images. The details of the phases involved in the proposed model are discussed below.

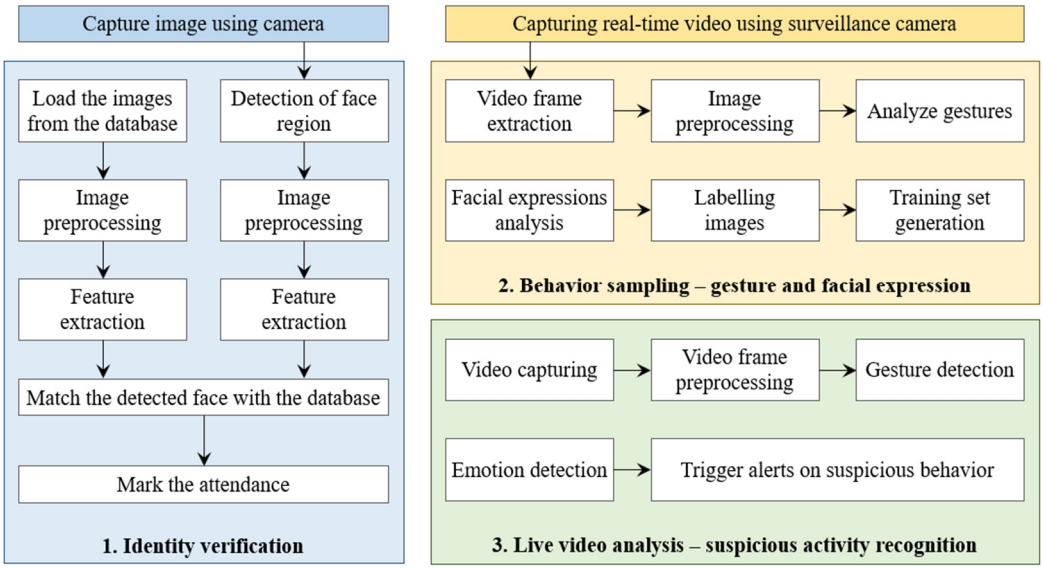

Fig. 1 Proposed smart surveillance system framework

3.1. Phase 1: *Identity verification*

Verifying the identity of individuals entering the examination hall is the initial step. Primarily, the students' facial images are captured and stored in an offline database to verify their identities when entering the hall. The camera records students as they enter the hall, and the video is then processed into frames for identity verification. The student database is organized in a structured folder, referred to as a directory, with each file assigned a unique identifier for proper referencing during further processing. This phase commences with loading and storing the student database in a local folder (the directory). These images are subsequently preprocessed, and features are extracted to compare with the live image to verify the identity of individuals.

Image preprocessing: In this step, several techniques are employed to prepare student images for analysis. Initially, images are loaded from a directory using OpenCV, with each image uniquely identified for processing. The images are converted to grayscale and resized to a standardized dimension. Subsequently, pixel values are normalized to the range [0, 1] using a min-max approach and then scaled to [−1, 1] through mean normalization with predefined functions. This scaling enhances training stability and convergence for neural networks, as centering data around zero improves performance. These preprocessing steps optimize the images for feature extraction and model training, ensuring suitability for comprehensive analysis.

Feature extraction and database embedding: A pre-trained face recognition model, InceptionResnetV1, from the facenet_pytorch library, extracts embeddings from the database images. These embeddings represent essential facial features such as edges, corners, the overall structure of the face, and the spatial relationships between facial landmarks (e.g., eyes, nose, and mouth). Trained on the VGGFace2 dataset, which contains over 3.3 million images of more than 9,000 identities, the model generates feature vectors encapsulating these key characteristics in a high-dimensional space. The database stores these features along with corresponding person IDs, enabling efficient face comparison and recognition to function based on unique embeddings derived from diverse conditions [27].

Image acquisition and face detection: This step entails capturing the student's image using the OpenCV library to communicate with the camera and capture a single frame. Subsequently, SSD, a DL-based face detection model, is utilized to locate and extract the facial region from the captured image [28]. SSD was selected for its real-time performance and high accuracy in detecting faces at different scales and orientations. In contrast to multi-stage models like Faster R-CNN, SSD conducts detection in a single pass, enhancing efficiency for rapid processing applications, such as real-time exam invigilation. SSD processes an image by dividing it into a grid of cells, predicting multiple bounding boxes of varying sizes and aspect ratios. Each box includes parameters such as width, height, center coordinates, and probability scores that indicate the likelihood of face presence. Non-maximum suppression (NMS) eliminates overlapping boxes with lower scores, reducing false alarms. Additionally, SSD resizes, normalizes, and converts the image format, extracting high-level features that represent unique face characteristics for subsequent comparisons.

Identity verification: After feature extraction from the captured image, the next step involves comparing these features to the embeddings stored in the database of registered students. Cosine similarity is used for this comparison, focusing on direction rather than magnitude in high-dimensional data. If a match is found, the system verifies the student's identity and records attendance automatically. A threshold of 0.75 is employed to ensure reliable detection of genuine matches while minimizing false positives. If no match is found, the user is notified.

3.2. Phase 2: Behavior sampling – gesture and facial expression

This phase focuses on generating the training dataset, which further enhances the optimization of model performance. It generates training samples based on student behaviors, including gesture identification with head orientation and emotion analysis, classifying them as cheating, non-cheating, or suspicious activities. The system captures a video clip and converts it to frames, which are then preprocessed using DL techniques such as YOLOv5 for gesture detection and C3DN for emotion analysis.

Image acquisition and preparation: Initially, video is captured through real-time acquisition of frames from a live camera feed or a pre-recorded video source. The frames are continuously read, timestamped, displayed in real-time, and saved periodically until the process is manually stopped. Concerning analysis, frames are extracted at specified intervals (e.g., every second), saved as individual image files, and preprocessed by applying resizing, normalization, and scaling. This preprocessing step preserves color information, ensuring accurate object detection, especially for identifying head orientations during further analysis.

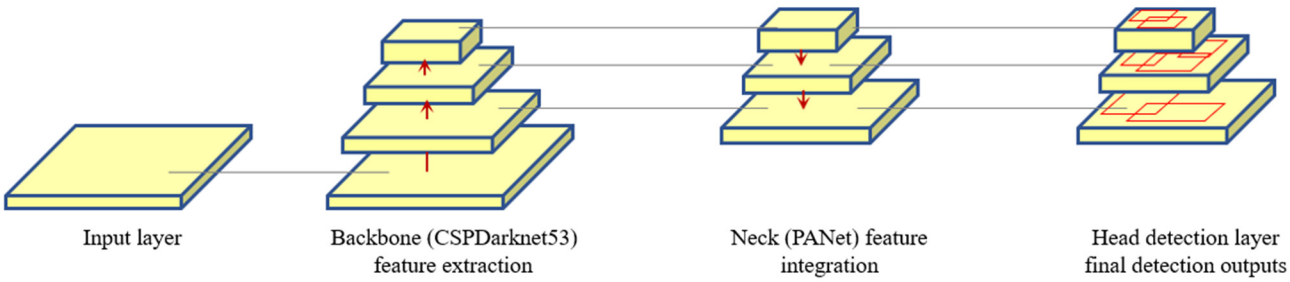

Fig. 2 YOLOv5 architecture

Gesture detection: The YOLOv5 model, a DL-based object detection algorithm, is used for gesture detection [29]. YOLOv5 is chosen for its optimal balance of speed and accuracy, making it suitable for real-time exam invigilation. Unlike slower and more computationally intensive algorithms like RetinaNet, its single-stage detection approach minimizes latency while maintaining high performance. The model partitions the input image into a grid and predicts bounding boxes and class probabilities within each grid cell. Fig. 2 illustrates the YOLOv5 architecture, which begins with a backbone (CSPDarknet53) that extracts intricate features from input frames via convolutional layers. The neck component, such as the path aggregation

network (PANet), integrates features across various scales to enhance detection capabilities. YOLOv5's head predicts bounding boxes, objectness scores, and class probabilities for each grid cell, optimizing detection accuracy for head orientations (left, right, up, down, front, and back) and activities like cheating (left, right, and back movements) or normal (front, up, and down movements).

Emotion detection: The proposed model utilizes the C3DN for emotion analysis from facial expressions in video frames, featuring 3D convolutional layers that operate across spatial and temporal dimensions, as shown in Fig. 3 [30]. The C3DN is chosen for its ability to capture spatial and temporal patterns in video data, enhancing sensitivity to subtle emotional cues and temporal dynamics compared to traditional 2D convolutional networks. Initially, faces are identified and isolated within bounding boxes for focused analysis. The C3DN model, trained on labeled datasets for emotion recognition, processes facial features, evaluating subtle cues like eyebrow movement and mouth shape. It predicts positive emotions (happy, neutral, and sad) and negative states (anxiety, fear, and stress) for identifying cheating behaviors. The model employs 3D convolutional and pooling layers to efficiently handle larger inputs, followed by fully connected layers that flatten feature maps, leading to a Softmax layer that outputs probabilities for each emotion class.

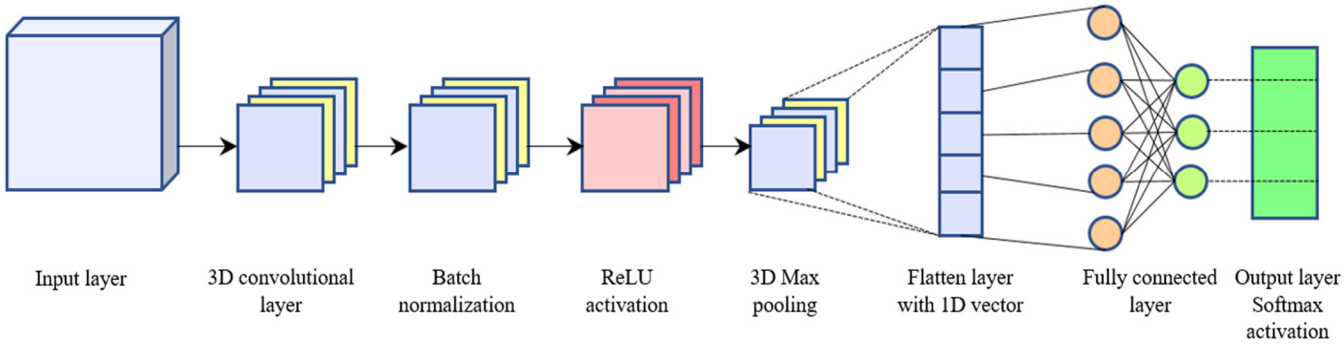

Fig. 3 C3DN architecture

The results are visualized by annotating frames with predicted emotion labels alongside the corresponding facial regions, providing insights into emotional responses captured in real-time video data. This detailed analysis assists in predicting suspicious activities based on emotional states, enhancing the accuracy and reliability of the behavioral assessment system.

Training set generation: Upon detecting gestures and emotions, images are categorized as 'cheating', 'normal', or 'suspicious' based on specific criteria, primarily involving detected head orientations and emotions. If the head orientation is frontal or upward with emotions classified as neutral, happy, or sad, it is labeled as 'no cheating'. On the other hand, head orientations to the left, right, or backward with emotions indicating fear or anxiety are categorized as 'cheating'. Alternatively, if the head is oriented frontally downward and emotions suggest fear or anxiety, it is labeled as 'suspicious'. These labeled training datasets are preserved for subsequent analysis or model training.

3.3. Phase 3: Live video analysis – suspicious activity recognition

Live video analysis involves continuous capture from a surveillance camera and recording frames in real-time. Frames are extracted at one-second intervals for preprocessing, including resizing, normalization, and scaling to enhance object detection. Next, YOLOv5 is used to identify student gestures through head movements, predicting bounding boxes and orientations—left, right, up, down, front, and back. The C3DN model analyzes facial expressions to determine emotions like happiness, sadness, fear, and anxiety. Detected orientations and emotions classify behaviors as 'cheating', 'normal', or 'suspicious'. An SSD model recognizes the student's face, comparing it with a database using cosine similarity to trigger alerts. Fig. 4 illustrates the workflow of the proposed model, while Algorithm 1 outlines the pseudocode for the overall implementation of the proposed model.

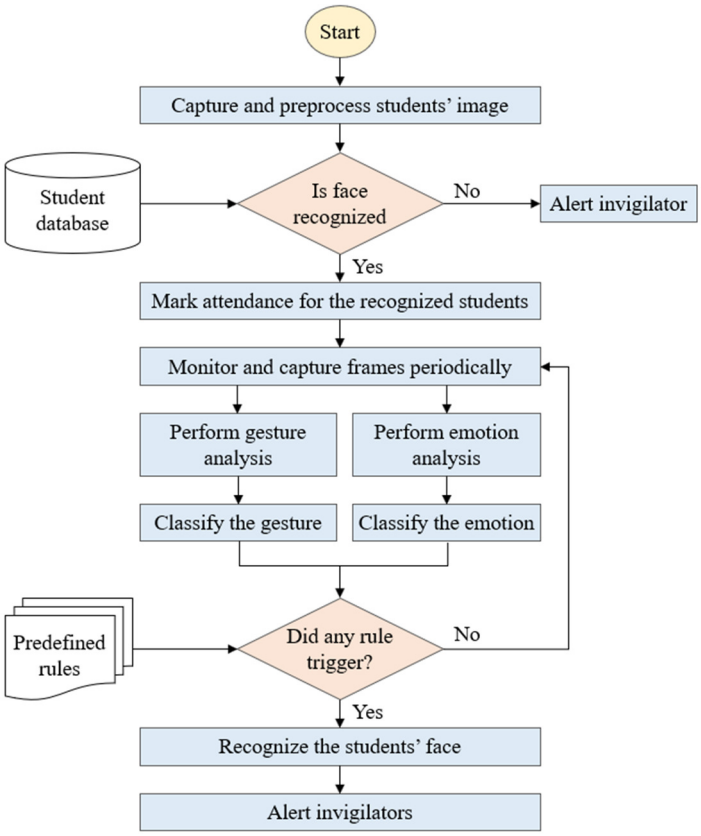

Fig. 4 Workflow of the proposed model

Algorithm 1: Pseudocode for proposed smart invigilation system

```
Procedure Smart_Invigilation()
Begin
   // Initialize Components
   Load Libraries (OpenCV, TensorFlow, Keras)
   Initialize Camera Feed from CCTV
   Load Pre-trained Models (Facial Recognition, Gesture Detection, Emotion Detection)

   // Start of Examination - Capture students entering the hall
   For Each Frame DO
      Preprocess Frame (Grayscale, Resize, Normalize)
      Detected_Faces = Detect_Faces(Facial_Recognition_Model, Frame)
      For Each Detected_Face DO
         Extract Facial_Embeddings
         Student_ID = Compare_Embeddings(Embeddings, Stored_Profiles)
         If Student_ID is not null then
            Mark Attendance to Student_ID
         End if
      End for
   End for

   // During Examination - Monitor students
   While Examination DO
      Capture Video Frame
      Preprocess Frame (Grayscale, Resize, Normalize)
      Detected_Faces = Detect_Faces(Facial_Recognition_Model, Frame)
      For Each Detected_Face DO
         Extract Facial_Embeddings
         Student_ID = Compare_Embeddings(Embeddings, Stored_Profiles)
         If Student_ID is not null then
            Record Student_ID
         End if
         ROI = Extract_ROI(Detected_Face, Frame)
         Record Gesture_Result = Detect_Gesture(Gesture_Detection_Model, ROI)
         Record Emotion_Result = Detect_Emotion(Emotion_Detection_Model, Detected_Face)
      End for
```

```
    // Data Integration and Comparison with Predefined Rules
    Behavior_Result = Integrate_Results(Student_ID, Gesture_Result, Emotion_Result)
    If Compare_with_Predefined_Rules(Behavior_Result) is Suspicious or Cheating then
       Trigger_Alert(Invigilator)
       Log Incident_Details
    End if
  End while

  // End of Examination
  Stop Recording
  Generate Report (Summary of Examination Conduct)
End Procedure
```

4. Experimental Analysis

This section details the experimental setup for system implementation, encompassing the dataset generation, the hardware and software configurations, the hyperparameters used for various learning models, and the performance metrics employed to evaluate the proposed study and compare it with existing systems.

4.1. Dataset used

The proposed model initially creates a database by live-capturing individual students from different angles during classes, storing 1,000 images for recognition. These images, with a resolution of 1920 × 1080, are preprocessed for efficient processing and memory management. This database is deployed to train the SSD model for detecting and recognizing student faces. Furthermore, the system captures and converts student video during examinations into frames to detect suspicious activities. By processing frames sequentially, instead of storing entire video streams, the system efficiently manages memory, even with a large number of students in the hall.

A training dataset for suspicious activity identification was collected, containing 2,000 images from multifarious classroom settings: 1,000 depicting students engaging in cheating behaviors and 1,000 showing genuine behaviors. This dataset represents both traditional and modern learning environments, encompassing a spectrum of typical student behaviors during exams. It captures a range of responses throughout the testing period, ensuring an accurate reflection of real-life situations. This diversity is critical for enhancing the model's ability to detect suspicious activity and generalize across different contexts, thereby supporting a more effective automated invigilation system. To further enrich the dataset's representation of real-world exam scenarios, data augmentation techniques were applied, increasing the dataset size to 5,000 images, with 80% allocated for training and 20% for testing.

Students' head movements were monitored using a training set that includes orientations: left, right, up, down, frontward, and backward, to train YOLOv5. Moreover, a training set for the emotions of students was created with six classes: happy, neutral, sad, anxiety, fear, and stress, used to train C3DN. These classifications label images as 'cheating', 'no cheating', and 'suspicious', either manually or semi-automatically based on a predetermined rule. Following training, the model is evaluated using test data from the exam dataset. The video is divided into frames, with each frame examined for head movement and emotion before being categorized as 'cheating', 'no cheating', or 'suspicious'.

Furthermore, this research acknowledges the importance of student privacy in video monitoring. It emphasizes the need for secure storage and management of recorded footage, ensuring that access is restricted to authorized individuals only. To mitigate potential biases in facial, gesture, and emotion analysis, extensive training with diverse datasets will be employed to enhance detection accuracy. The performance of the system will be audited regularly to identify and resolve any inconsistencies. By highlighting these measures, the research aims to reinforce the ethical framework surrounding automated invigilation systems in the educational environment.

4.2. Experimental setup

The hardware used for the analysis includes two HIKVISION EZVIZ CS-BW3824B0 cameras, along with an NVR 8-CHANNEL and 2TB AV HDD, strategically placed to monitor all students. Additionally, the system utilizes a Logitech Brio Ultra HD Pro USB camera for high-definition image capture.

Regarding processing, an Acer WS laptop equipped with an Intel i5 (12th Gen) processor, 16GB DDR4 RAM, RTX 3050 6GB GPU, and 512GB SSD storage is employed for real-time processing, accelerating DL computations for facial recognition and gesture analysis while ensuring effective data processing. The code is developed using Python within the Jupyter IDE framework, utilizing OpenCV and other necessary libraries.

Moreover, memory is managed by employing batch processing techniques, where the number of frames captured is reduced to optimize processing. Libraries like NumPy are used to allocate memory efficiently, and both the batch size and resolution are controlled to ensure smooth processing, especially in the context of the massive number of students presenting in the examination hall.

Apropos of feature extraction, the SSD model utilizes a modified VGG16 architecture with transfer learning. The VGG16 model, pre-trained on a large dataset like ImageNet, consists of 13 convolutional layers and 3 fully connected layers designed to extract detailed features from input images. The initial convolutional layers of VGG16 use 64 filters in the first two layers, followed by 128 filters in the next two layers, and 256 filters in the following three layers.

The SSD architecture extends VGG16 by adding additional convolutional layers with 512, 1024, 256, and 128 filters, enabling it to detect objects at various scales. The model integrates multi-box loss as its primary loss function and utilizes stochastic gradient descent (SGD) with momentum as the optimizer for efficient training. Key hyperparameters include a batch size of 1, momentum set to 0.9, weight decay of 0.0005, a localization loss weight of 1.0, and a confidence threshold of 0.01.

The YOLOv5 model detects gestures by identifying human heads and their orientations in each video frame. It employs transfer learning for feature extraction, pre-trained on large datasets like common objects in context (COCO), with CSPDarknet53 as its backbone. CSPDarknet53 includes 29 convolutional layers that capture rich features, starting with 32 filters and followed by layers with 64, 128, 256, and 512 filters for high-level feature extraction. The neck component, PANet, integrates features across scales to enhance detection, with filtering ranging from 64 to 256 filters. The final output layer has 255 filters (3 anchor boxes per grid cell × 4 bounding box coordinates + 1 objectness score + 80 class probabilities). Key training hyperparameters include a learning rate of 0.01, a batch size of 1, momentum at 0.937, weight decay of 0.0005, and a confidence threshold of 0.01, with training over 120,000 steps across 300 epochs.

Concerning emotion detection, the C3DN model captures spatial and temporal features from facial expressions in video frames. The architecture includes four convolutional layers: the first with 32 filters for initial feature extraction, the second with 64 filters, the third with 128 filters, and the fourth with 256 filters for intricate facial patterns. 3D pooling layers follow each convolutional layer to reduce spatial and temporal dimensions while retaining information. Fully connected layers flatten the feature maps for classification. The final layer is a Softmax layer that outputs probabilities for each emotion class (e.g., happy, sad, or neutral). Key hyperparameters include a learning rate of 0.01, a batch size of 1, momentum at 0.9, weight decay of 0.0005, and training over 120,000 steps across 100 epochs using SGD with momentum.

Thus, the outputs from the identity verification phase initiate the gesture analysis phase. Once the system confirms a student's identity, it employs YOLOv5 to monitor specific gestures. Concurrently, emotional analysis via C3DN evaluates the student's emotional state. The results from these analyses are integrated using predefined decision rules, which classify activities as either normal or indicative of potential malpractice.

4.3. Performance measure

The proposed model is evaluated using an annotated dataset with 5-fold cross-validation, allocating 80% of the data as the training set and 20% as the test set. The model's performance is assessed individually across three phases using a confusion matrix, which includes four values as follows: true positive (TP) for correctly predicted positives (e.g., detecting cheating), true negative (TN) for correctly predicted negatives (e.g., detecting no cheating), false positive (FP) for incorrectly flagged positives, and false negative (FN) for missed detections of cheating.

Evaluation metrics include accuracy, precision, recall, F1-score, specificity, false discovery rate, error rate, and Cohen's Kappa statistics. Accuracy measures the proportion of correctly classified instances, while precision assesses the accuracy of positive predictions. Specificity and recall measure the ability to identify actual negatives and positives, respectively; the F1-score balances precision and recall. The false discovery rate measures the percentage of false positives, and the error rate calculates incorrect predictions relative to the total predictions. Finally, Cohen's Kappa measures the agreement between two raters, accounting for chance agreement. Specifically, to sum up, the formulas for these metrics are discussed below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{5}$$

$$\text{False discovery rate} = \frac{\text{FP}}{\text{FP} + \text{TP}} \tag{6}$$

$$\text{Error rate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \tag{7}$$

$$\text{Kappa statistics} = \frac{P_o - P_e}{1 - P_e} \tag{8}$$

Here, P_o and P_e are observed and expected agreements by chance.

5. Results and Discussion

The face recognition module that applies the SSD method was evaluated with images of 150 students. Feature extraction for these images was carried out, and the images were compared with the student database. The exam dataset comprised 1,000 test images, with the remaining images serving as the training set, while YOLOv5 evaluated gesture analysis based on head orientations.

YOLOv5, a head-orientation-based system, evaluated students' specific behaviors, classifying cheating and non-cheating activities based on left, right, and back head movements. Furthermore, the C3DN model was evaluated individually by analyzing facial features and subtle cues, identifying positive emotions as cheating and negative states as cheating. Finally, the proposed model classified test set images using head orientations, emotions, and pre-defined decision rules for classification. Figs. 5-8 display the obtained results.

Fig. 5 Face detection and recognition

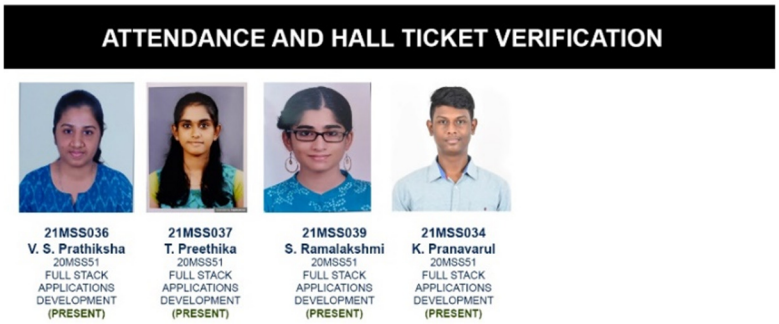

Fig. 6 Student identity verification

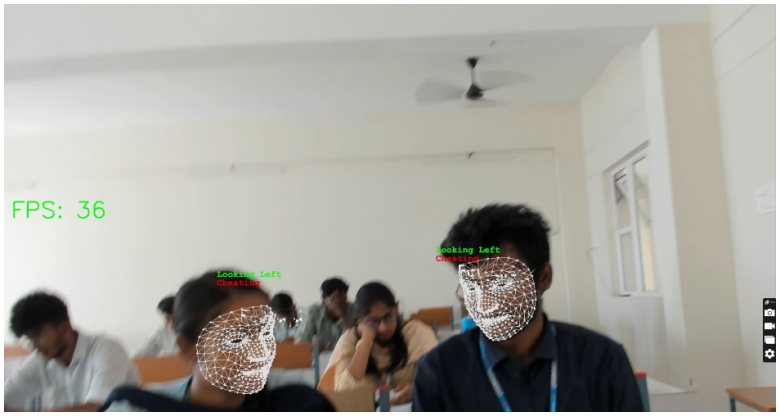

Fig. 7 Gesture detection

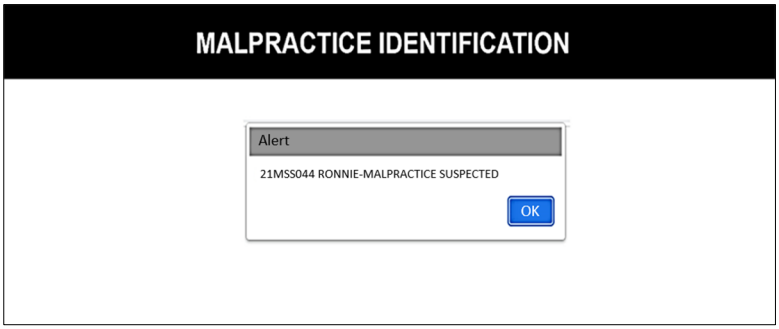

Fig. 8 Suspicious activity detection

Thus, a frontal head orientation with neutral, happy, or sad emotions was classified as ‘no cheating’; conversely, head orientations to the left, right, or back with emotions like anxiety, fear, and stress were classified as ‘cheating’; and, intermediately, a front, up, and downward head orientation with fear or anxiety emotions was classified as ‘suspicious’. The results of the various analyses are presented in Table 2.

Table 2 Performance comparison

Metrics	Face recognition (SSD)	Student activity detection		
		Gesture analysis (YOLOv5)	Emotion analysis (C3DN)	Proposed model (behavioral sampling)
Accuracy	0.9833	0.9810	0.9811	0.9880
Precision	0.9769	0.9720	0.9760	0.9840
Recall	0.9750	0.9898	0.9859	0.9919
F1-score	0.9751	0.9808	0.9809	0.9880
Specificity	0.9881	0.9725	0.9764	0.9841
False discovery rate	0.0256	0.0280	0.0240	0.0160
Error rate	0.0167	0.0190	0.0189	0.0120
Cohen's Kappa	0.9623	0.9620	0.9621	0.9760

The analysis showed that the SSD face recognition approach correctly identified 146 students, misclassifying 4 images, and achieved superior accuracy and precision of 0.9833 and 0.9769, respectively, with minimal false discovery rates and error rates of 0.0256 and 0.0167. Moreover, a comparison of results from manifold student activity detection methods, including gesture analysis, emotion analysis, and the proposed model, revealed improved performance across multiple metrics. The proposed model consistently achieved the highest values in accuracy (0.9880), precision (0.9840), recall (0.9919), F1-score (0.9880), specificity (0.9841), and Cohen's Kappa (0.9760), indicating superior overall performance compared to gesture and emotion analysis.

While gesture analysis showed competitive results with better accuracy (0.9810), recall (0.9898), and error rate (0.0190), signifying minimal false negatives, it slightly lagged across other metrics compared to emotion analysis and the proposed model. Similarly, emotion analysis demonstrated strong precision (0.9760) and F1-score (0.9809), indicating minimal false positives for effective detection of suspicious activities. However, concerning student activity detection during examinations, the proposed model consistently outperformed individual gesture analysis and emotion detection methods across various metrics. The class-wise accuracy for these models is presented in Table 3.

Table 3 Class-wise comparison of the student activity detection

Methods	Predicted values	Actual values		Accuracy (%)
		Cheating	No cheating	
Gesture Analysis (YOLOv5)	Cheating	TP: 486	FP: 5	97.2
	No cheating	FN: 14	TN: 495	99.0
Emotion analysis (C3DN)	Cheating	TP: 488	FP: 7	97.6
	No cheating	FN: 12	TN: 493	98.6
Proposed model (behavioral sampling)	Cheating	TP: 492	FP: 4	98.4
	No cheating	FN: 8	TN: 496	99.2

The study showed that gesture and emotion analysis were effective in predicting non-cheating activities and identifying cheating activities respectively, and their integration into the proposed model enhanced accuracy. The values were plotted as a graph, as shown in Fig. 9, in which the bars represent the performance of the face recognition method and the lines represent the suspicious activity detection methods.

The proposed model was evaluated by capturing live images of students, varying the number of individuals present in the exam hall. Fig. 10 displays the results. It is evident that when the number of students is minimal, the model achieves 100% accuracy. However, as the student count increases, the accuracy declines due to limited image visibility within the classroom. Therefore, the model performed best in smaller classrooms with up to 30 students, ensuring comprehensive student coverage. In contrast, in larger exam halls accommodating up to 100 students, additional cameras were required to fully capture all student details.

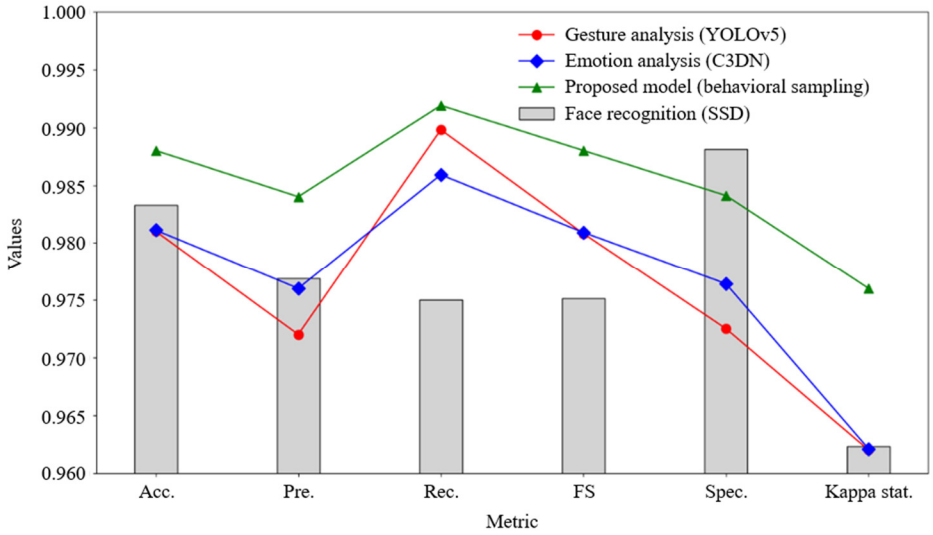

Fig. 9 Performance comparison for various suspicious detection models

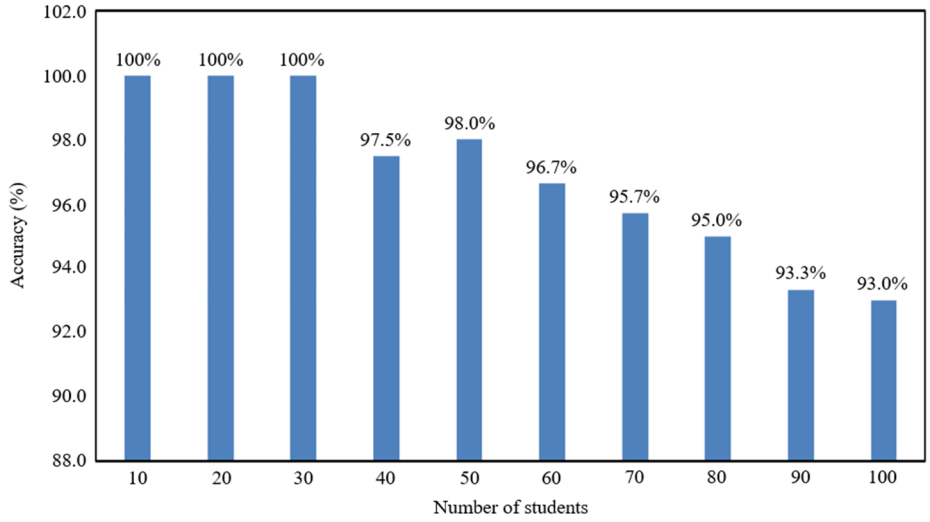

Fig. 10 Evaluation of the proposed model

While the proposed model demonstrates enhanced performance in detecting cheating, outperforming many existing models in the research field, it is important to recognize a few limitations that accompany these improvements. These limitations are discussed below.

Scalability issues: The model is capable of detecting and recognizing faces and gestures in the exam hall with a minimum seating capacity of 30 students. However, as the number of students in the hall increases, or the size of the hall expands, the accuracy of face and gesture detection may decrease, resulting in a higher error rate. This scalability concern highlights the need for further research into optimizing the model for larger groups, which may inevitably necessitate additional cameras and computational resources to maintain performance.

Dependence on high-quality images: Another important consideration is the effectiveness of this model, which is significantly influenced by the quality of input images. For instance, variations in lighting conditions and image resolution could adversely affect detection accuracy. Therefore, it is essential to explore robust image preprocessing techniques and the potential for training the model with diverse datasets that include varying image qualities.

Gesture analysis limitations: The effectiveness of the model in gesture analysis may be influenced by the types of gestures it recognizes. While the model currently relies on head orientation, incorporating additional gestures, such as hand or leg movements, could enhance prediction accuracy. Furthermore, including students' eye contact in the analysis could further improve its overall effectiveness.

Impact of human factors: The effectiveness of the model might also be impacted by human factors, such as stress and anxiety exhibited by students during exams. These factors can incur increased false negatives in emotion analysis, which may go undetected and potentially perpetuate cheating behaviors. To mitigate this, future iterations of the model could benefit from including a broader range of training images that capture various emotional expressions and states.

Recognition of cheating behaviors: Human behavior is complex and variable, and the model might not capture, cover, or recognize all possible cheating behaviors. For instance, while head-down positions are considered non-cheating activities, students attempting to cheat with self-help methods—such as writing on their hands, using calculators, or accessing mobile phones—may evade detection. To address this, future research should focus on increasing the dataset size and employing advanced DL techniques like multi-task learning and attention mechanisms.

Comparison with existing studies: Additionally, this study does not directly compare results with other existing studies due to the unique nature of each study's synthesized datasets, hindering the practicability of such comparisons. Future research should employ standardized benchmark datasets or evaluate the model's performance using commonly applied metrics in related studies to enable more meaningful comparisons.

Lack of cost analysis: A significant limitation of the proposed model is the absence of a comprehensive cost analysis. Evaluating the operational expenses associated with the algorithm, such as processing time, memory usage, and energy consumption, is crucial for understanding its feasibility and practicality in real-world applications. This analysis is essential for identifying optimization opportunities and assessing system efficiency, and future research should include a detailed cost analysis to understand resource requirements and algorithm scalability.

6. Conclusions

This study developed a DL-based smart invigilation system aimed at ensuring exam integrity by detecting dishonest behaviors during examinations. The system comprises three key phases: (1) student identity verification using SSD-based face recognition, (2) behavioral sampling through gesture analysis with YOLOv5 and emotion assessment using C3DN, and (3) live video monitoring that integrates gesture and emotion data to identify suspicious activities.

The model was evaluated using a dataset of 5,000 images, demonstrating high accuracy and robust performance in detecting dishonest activities, thereby addressing gaps in existing solutions. The key findings of the study include:

(1) High accuracy: The model achieves an overall accuracy of 98.8%, rendering promising solutions for monitoring exams.
(2) Enhanced detection of Academic Dishonesty: The system effectively improves the detection of dishonest behaviors, reinforcing exam integrity.
(3) Automation of invigilation: By automating the invigilation process, the system reduces reliance on manual monitoring, streamlining exam processes.
(4) Resource optimization: The system enables effective allocation of resources, creating a fair and efficient testing environment.
(5) Improved security: The model strengthens exam security and minimizes the risk of human error in higher education institutions.

Despite high accuracy, the model has limitations suggesting future research directions. Its effectiveness in larger exam settings could be improved through advanced hardware setups or distributed camera systems. Further enhancement may involve expanding the training dataset to cover a broader range of cheating behaviors. In light of the computational demands of DL models, a cost analysis of processing time and memory usage is advisable. Additionally, incorporating eye contact and head orientation in gesture analysis and exploring advanced DL techniques, such as multi-task learning and attention mechanisms, could strengthen the model's application in diverse testing environments.

Funding

This work was funded by the Research and Development Cell, Coimbatore Institute of Technology, Coimbatore under the Seed Money Project.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] S. Erduran, Y. El Masri, A. Cullinane, and Y. P. D. Ng, "Assessment of Practical Science in High Stakes Examinations: A Qualitative Analysis of High Performing English-Speaking Countries," International Journal of Science Education, vol. 42, no. 9, pp. 1544-1567, 2020.

[2] K. A. A. Gamage, R. G. G. R. Pradeep, and E. K. de Silva, "Rethinking Assessment: The Future of Examinations in Higher Education," Sustainability, vol. 14, no. 6, article no. 3552, 2022.

[3] M. A. Mulongo, "Effectiveness of University Examinations Management Strategies in Mitigating Examination Malpractices in Kenya," Ph.D. dissertation, Karatina University, Kenya, 2020.

[4] J. Nishchal, S. Reddy, and P. N. Navya, "Automated Cheating Detection in Exams Using Posture and Emotion Analysis," IEEE International Conference on Electronics, Computing and Communication Technologies, pp. 1-6, 2020.

[5] O. L. Holden, M. E. Norris, and V. A. Kuhlmeier, "Academic Integrity in Online Assessment: A Research Review," Frontiers in Education, vol. 6, article no. 639814, 2021.

[6] F. Kamalov, H. Sulieman, and D. Santandreu Calonge, "Machine Learning Based Approach to Exam Cheating Detection," PLoS ONE, vol. 16, no. 8, article no. e0254340, 2021.

[7] M. J. Hoque, M. R. Ahmed, M. J. Uddin, and M. M. A. Faisal, "Automation of Traditional Exam Invigilation Using CCTV and Bio-Metric," International Journal of Advanced Computer Science and Applications, vol. 11, no. 6, pp. 392-399, 2020.

[8] P. Alin, A. Arendt, and S. Gurell, "Addressing Cheating in Virtual Proctored Examinations: Toward a Framework of Relevant Mitigation Strategies," Assessment & Evaluation in Higher Education, vol. 48, no. 3, pp. 262-275, 2023.

[9] Y. Liu, J. Ren, J. Xu, X. Bai, R. Kaur, and F. Xia, "Multiple Instance Learning for Cheating Detection and Localization in Online Examinations," IEEE Transactions on Cognitive and Developmental Systems, vol. 16, no. 4, pp. 1315-1326, 2024.

[10] M. Asad, M. Abbas, A. Asim, A. Hafeez, M. M. Sadaf, A. U. Haq, et al., "Suspicious Activity Detection During Physical Exams," unpublished.

[11] P. Verma, N. Malhotra, R. Suri, and R. Kumar, "Automated Smart Artificial Intelligence-Based Proctoring System Using Deep Learning," Soft Computing, vol. 28, no. 4, pp. 3479-3489, 2024.

[12] J. Xue, W. Wu, and Q. Cheng, "Intelligent Invigilator System Based on Target Detection," Multimedia Tools and Applications, vol. 82, no. 29, pp. 44673-44695, 2023.

[13] J. A. Hernándeza, A. Ochoab, J. Muñozd, and G. Burlaka, "Detecting Cheats in Online Student Assessments Using Data Mining," International Conference on Data Mining, pp. 204-210, 2006.

[14] 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.

[15] 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," https://doi.org/10.48550/arXiv.2101.09841, 2021.

[16] S. El 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.

[17] F. Mahmood, J. Arshad, M. T. Ben Othman, M. F. Hayat, N. Bhatti, M. H. Jaffery, et al., "Implementation of an Intelligent Exam Supervision System Using Deep Learning Algorithms," Sensors, vol. 22, no. 17, article no. 6389, 2022.

[18] M. D. Genemo, "Suspicious Activity Recognition for Monitoring Cheating in Exams," Proceedings of the Indian National Science Academy, vol. 88, no. 1, pp. 1-10, 2022.

[19] 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, vol. 16, no. 08, pp. 124-137, 2022.

[20] R. K. Kadthim and Z. H. Ali, "Cheating Detection in Online Exams Using Machine Learning," Journal Of AL-Turath University College, vol. 2, no. 35, pp. 35-41, 2023.
[21] W. Alsabhan, "Student Cheating Detection in Higher Education by Implementing Machine Learning and LSTM Techniques," Sensors, vol. 23, no. 8, article no. 4149, 2023.
[22] T. Zhou and H. Jiao, "Exploration of the Stacking Ensemble Machine Learning Algorithm for Cheating Detection in Large-Scale Assessment," Educational and Psychological Measurement, vol. 83, no. 4, pp. 831-854, 2023.
[23] S. C. Chang and K. L. Chang, "Cheating Detection of Test Collusion: A Study on Machine Learning Techniques and Feature Representation," Educational Measurement: Issues and Practice, vol. 42, no. 2, pp. 62-73, 2023.
[24] S. Z. Ong, T. Connie, and M. K. O. Goh, "Cheating Detection for Online Examination Using Clustering Based Approach," JOIV: International Journal on Informatics Visualization, vol. 7, no. 3-2, pp. 2075-2085, 2023.
[25] F. Ozdamli, A. Aljarrah, D. Karagozlu, and M. Ababneh, "Facial Recognition System to Detect Student Emotions and Cheating in Distance Learning," Sustainability, vol. 14, no. 20, article no. 13230, 2022.
[26] A. L. Cîrneanu, D. Popescu, and D. Iordache, "New Trends in Emotion Recognition Using Image Analysis by Neural Networks, A Systematic Review," Sensors, vol. 23, no. 16, article no. 7092, 2023.
[27] S. Peng, H. Huang, W. Chen, L. Zhang, and W. Fang, "More Trainable Inception-ResNet for Face Recognition," Neurocomputing, vol. 411, pp. 9-19, 2020.
[28] A. Kumar, Z. J. Zhang, and H. Lyu, "Object Detection in Real Time Based on Improved Single Shot Multi-Box Detector Algorithm," EURASIP Journal on Wireless Communications and Networking, vol. 2020, no. 1, article no. 204, 2020.
[29] L. Ling, J. Tao, and G. Wu, "Research on Gesture Recognition Based on YOLOv5," 33rd Chinese Control and Decision Conference, pp. 801-806, 2021.
[30] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. W. Shalaby, "A 3D-Convolutional Neural Network Framework with Ensemble Learning Techniques for Multi-Modal Emotion Recognition," Egyptian Informatics Journal, vol. 22, no. 2, pp. 167-176, 2021.