Advanced Proctoring Framework for Exam Integrity

MITHUN S[1], MELVIN M SHAJAN[2], RAGAVENDIRAN S[3], SINDUJA K[4]

[1, 2, 3] *Student, Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore, India*

[4] *Assistant Professor, Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore, India*

Abstract- This research offers an intelligent invigilation system to maintain examination integrity by identifying unusual student behaviors through the use of deep learning. The model involves three phases: 1) verification of the student identity based on a face recognition method; 2) behavioral sampling to train the model, employing gesture analysis and convolutional 3D networks to analyze emotions; and 3) live video analysis of anomalous behavior, combining gesture and emotion analysis and student identification using face recognition. The model, trained on 4,000 training and 1,000 test images, classifies non-cheating activities with 99% accuracy and cheating activities with 97.6% accuracy. The suggested model performs better than other approaches, with accuracies of 98.4% for the detection of cheating behavior and 99.2% for non-cheating behavior, giving an overall accuracy of 98.8% and a low misclassification rate of 1.2%. Though the system exhibits strong accuracy, issues lie in scalability to larger classes with higher computational demands and requirements for more hardware for complete monitoring

Indexed Terms- Suspicious Activity Detection, Exam Integrity, Deep Learning, Face and Gesture Recognition, Emotion Analysis

I. INTRODUCTION

For any educational institution, examinations and evaluations 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. Students often break the rules of fair and impartial examinations supervised by invigilators by observing their neighbours' answers with head movements, turning back and sideways, whispering answers, extending their hands forward and backwards to exchange answer sheets, or copying answers from other materials [4]. Due to the prevalence of cheating and academic dishonesty, maintaining exam integrity presents significant challenges, even though it appears to be a simple responsibility for exam supervisors [5].

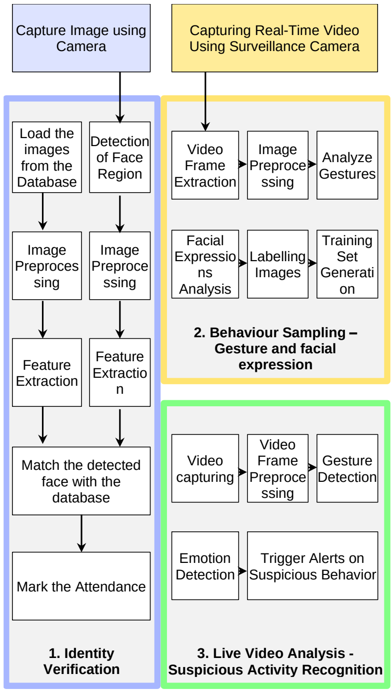

Capture Image using Camera
Capturing Real-Time Video Using Surveillance Camera
Load the images from the Database
Detection of Face Region
Image Preprocessing
Image Preprocessing
Feature Extraction
Feature Extraction
Match the detected face with the database
Mark the Attendance
1. Identity Verification
Video Frame Extraction
Image Preprocessing
Analyze Gestures
Facial Expressions Analysis
Labelling Images
Training Set Generation
2. Behaviour Sampling – Gesture and facial expression
Video capturing
Video Frame Preprocessing
Gesture Detection
Emotion Detection
Trigger Alerts on Suspicious Behavior
3. Live Video Analysis - Suspicious Activity Recognition

Figure 1 The Proposed Smart Surveillance System Framework

To reduce mistakes and human errors during exam invigilation, a few research 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, and fingerprint sensors, in addition to surveillance cameras, which 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: 1) verifying students' identities through facial recognition with a single-shot multi-box detector (SSD); 2) generating behavioral sampling through gesture analysis using You Only Look Once (YOLOv5) and emotional analysis using convolutional 3D networks (C3DN); and 3) analyzing real-time video by integrating gesture and emotional analysis along with pre-defined decision rules to classify malpractices from normal activities.

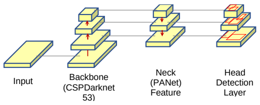


Figure 2 YOLO Architecture

II. 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, highlighting the need for automated systems. Several researchers have proposed various ML and DL models to recognize suspicious activities in surveillance videos.

Hernándeza et al. (2006) developed a model to detect and prevent cheating in online assessments by analyzing student personalities, stress situations, and cheating practices using a DMDC model and Weka data mining [13]. A model proposed by Atoum et al. (2017) 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 [14]. The study by Kamalov et al. (2021) proposed a novel method for identifying potential cheating cases on final exams through a post-exam analysis of student grades [6]. 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. (2020) 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 [7]. 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 (2021) developed an e-cheating intelligence agent using IP and behavior detectors to monitor student behavior, prevent malicious practices, and integrate with online learning programs [15].

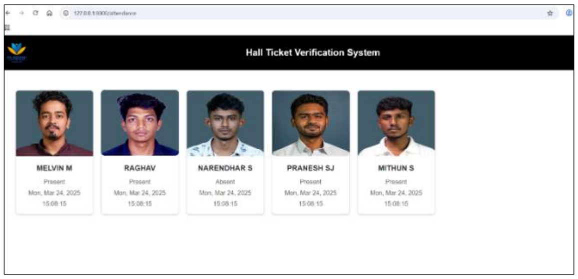
Kohli et al. (2022) [16] developed a real-time computer vision system using 3D CNN, object detector methods, OpenCV, and Google Tensor Flow to predict exam fraud with a 95% correlation. Mahmood et al. (2022) [17] developed a DL exam invigilation system using a Faster Regional Convolution Neural Network and face recognition, achieving 99.5% and 98.5% accuracy, respectively. Genemo (2022) [18] developed "L4-BranchedActionNet" using surveillance footage for identifying suspicious student behavior during exams, achieving 92.99% accuracy on CUI-EXAM and 89.79% accuracy on CIFAR-100. However, performance needs more improvement. Similar to this work, Asad et al. (2023) [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 Roa'a (2022) [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 (2023) [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 (2023) [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 (2023) [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 (2023) [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. (2023) [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. (2022) [25] developed a facial recognition system using computer vision and DL algorithms for online learning invigilation, detecting student behaviors and abnormalities. Cîrneanu et al. (2023) [26] studied the evolution of neural network architectures in FER, focusing on CNN-based ones and analyzing gestures and emotions for student cheating detection. Nishchal et al. (2020) [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 labelled instances. Verma et al. [11] employed a multi-modal DL approach to monitor students, detect emotions, estimate head pose, and track mouth movements, aiming to replace human proctors.

Thus, the literature reflects a considerable number of studies in this area. The overview of crucial studies, their performance, and shortcomings is illustrated in Table 1, wherein all these studies used synthesized data for testing. The shortcomings and gaps in research highlighted prove the need for developing a more vigorous, accurate, and extensible review monitoring system to improve exam integrity. Though past studies have improved on exam integrity, the suggested model fills various key gaps: it optimizes scalability by managing large areas efficiently and involves thorough analysis through gesture detection and emotion sensing. These improvements put our model in a better position to counteract the variability and complexity of human behavior during exams, ultimately offering a more viable solution for maintaining exam integrity.

The statement of the envisioned smart surveillance system based on behavioral sampling for maintaining exam integrity through DL methods is given in Fig. 1. The architecture consists of three stages: 1) identity authentication of the students during exams; 2) sampling of student behavior using gesture and face expression analysis; and 3) live video monitoring of suspicious action detection. The initial step comprises pre-processing the images based on a face recognition model and comparing faces to database faces. The behavioral sampling step includes recording video, pre-processing frames, gestural and emotive detection, image labelling, and making a training set. Live video analysis in step three entails recognizing gestures and emotion in real-time and initiating alarm for suspicious actions. Specifically, the SSD was utilized for face region detection, YOLOv5 was utilized for gesture analysis, C3DN were utilized for emotion analysis, and pre-defined decision rules were utilized to classify the images. The phases involved in the proposed model are explained below.


127.0.0.1:8000/attendance

Hall Ticket Verification System

MELVIN M
Present
Mon, Mar 24, 2025
15:08:15

RAGHAV
Present
Mon, Mar 24, 2025
15:08:15

NARENDHAR S
Absent
Mon, Mar 24, 2025
15:08:15

PRANESH SJ
Present
Mon, Mar 24, 2025
15:08:15

MITHUN S
Present
Mon, Mar 24, 2025
15:08:15

Phase 1: Identity verification

The first step is to authenticate the identity of people entering the examination hall. The facial images of the students are taken and stored in an offline database to authenticate their identities when they enter the hall. Students are photographed by the camera as they enter the hall, and the video is processed into frames for authentication of identity. The student database is kept in an organized folder, also known as a directory, with every file being given a specific identifier for easier referencing in further processing. This step starts with loading and saving the student database in a local folder (the directory). The images are preprocessed, features are extracted, and matched against the live image to authenticate the identity of the people.

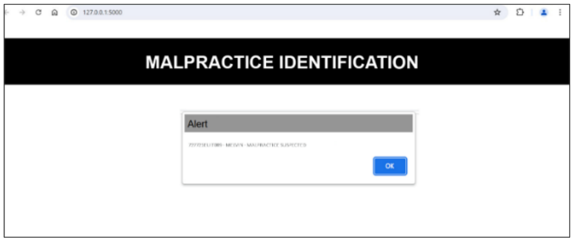
Image preprocessing: Here, a number of methods are used to pre-process student images for analysis. First, images are read from a directory with OpenCV, and each image is given a unique identifier for processing. Images are converted to grayscale and resized to a standard dimension. Second, pixel values are normalized to [0, 1] using a min-max method and then scaled to [−1, 1] by mean normalization with given functions. This scaling improves convergence and stability during training for neural networks, as data centring around zero facilitates better performance. All these preprocessing operations prepare images for model training and feature extraction so that the images will be well-suited for overall analysis.

Feature extraction and database embedding: A pre-trained face recognition network, InceptionResnetV1, from the facenet_pytorch library extracts embeddings from the database images. These embeddings capture important facial features like edges, corners, the general structure of the face, and the spatial relationships between facial landmarks (e.g., eyes, nose, mouth). Trained on the VGGFace2 dataset, comprising over 3.3 million images of over 9,000 identities, the model produces feature vectors that capture these distinguishing features in a high-dimensional space. The database stores the features along with corresponding person IDs, allowing for efficient face comparison and recognition based on distinctive embeddings produced under varied conditions [27]. Acquiring the image and detecting the face: This part involves taking a picture of the student using OpenCV library in order to talk to the webcam and take just one frame.

SSD was chosen for its real-time capability and high accuracy at detecting faces of varying scales and orientations. Unlike multi-stage models such as Faster R-CNN, SSD performs detection in one pass, improving efficiency for applications requiring quick processing, including real-time invigilation of exams. SSD treats an image by splitting it into a cell grid, predicting several bounding boxes with different sizes and aspect ratios. Each box has dimensions like width, height, centre coordinates, and probability scores representing the probability of face existence. Non-Maximum Suppression (NMS) removes overlapping boxes with lower scores, minimizing false alarms. Moreover, SSD resizes, normalizes, and compresses the image form, extracting high-level face features characterizing distinctive face traits to be used in comparison. Identity verification: Following feature extraction from the input image, subsequent steps involve matching those features with the embeddings saved in the database of enrolled students. Comparison is carried out using cosine similarity with its emphasis on direction, not magnitude, in high-dimensional space.

If it finds a match, the system automatically verifies the identity of the student and marks attendance.

A 0.75 threshold is used to avoid false positives while ensuring correct identification of real matches. If there is no match, the user is alerted



MALPRACTICE IDENTIFICATION

Alert

727721EUIT089 - MEGVIN - MALPRACTICE SUSPECTED

Phase 2: Behavior sampling – gesture and facial expression

This step emphasizes creating the training dataset that is used to optimize the model's performance. It creates samples for training on the basis of student behaviors such as head orientation and gesture identification with emotion analysis, classifying them as non-cheating, cheating, or suspicious activity. The system takes a video clip and processes it into frames and preprocessed the frames using DL methods like YOLOv5 for identifying gestures and C3DN for emotion analysis.

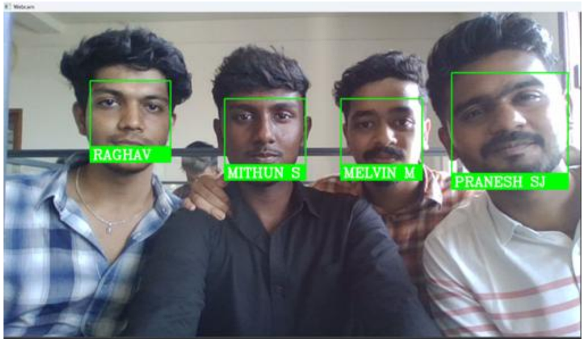
Image acquisition and preparation: First, video is recorded through real-time capture of frames from a live camera or pre-recorded video input. The frames are repeatedly read, stamped, shown in real-time, and stored intermittently until manually interrupted. For examination, frames are pulled at periodic intervals (e.g., per second), stored as individual image files, and preprocessed through resizing, normalization, and scaling. This pre-processing operation retains colour information to provide correct object detection, particularly in the determination of head orientations in subsequent analysis.

Gesture detection: Gesture detection is performed using the YOLOv5 model, which is a DL-based object detection algorithm [29]. YOLOv5 is utilized due to its best trade-off between speed and accuracy, and it is well-suited for real-time exam invigilation. In contrast to slower and more computationally expensive algorithms such as RetinaNet, its single-stage detection method reduces latency while ensuring high performance. The model divides the input image into a grid and predicts class probabilities and bounding boxes in each cell of the grid. Fig. 2 shows the YOLOv5 architecture, which starts with a backbone (CSPDarknet53) that extracts detailed features from input frames using convolutional layers. The neck module, e.g., the Path Aggregation Network (PANet), combines features from multiple scales to improve detection ability. YOLOv5's head outputs bounding boxes, objectness scores, and class probabilities per grid cell to maximize detection precision for head directions (left, right, up, down, front, and back) and activities such as cheating (left, right, and back) or normal (front, up, and down) movements.

Emotion recognition: The suggested model applies the C3DN (Convolutional 3D Network) to emotion recognition from facial expressions in video frames, with 3D convolutional layers that process both spatial and temporal dimensions, as indicated in Fig. 3 [30]. The C3DN is selected for its capability to extract spatial and temporal patterns in video data, with greater sensitivity to subtle emotional signals and temporal dynamics than with conventional 2D convolutional networks. First, faces are detected and separated in bounding boxes for detailed examination. The C3DN model, which has been trained on emotion recognition labeled datasets, analyzes facial features, considering subtle expressions such as eyebrow movement and mouth shape. It makes predictions of positive emotions (happy, neutral, sad) and negative emotions (anxiety, fear, stress) for detecting cheating behaviors. The model utilizes 3D convolutional and pooling layers to effectively process larger inputs, and then fully connected layers that flatten feature maps, culminating in a softmax layer that provides probabilities for every emotion class.

The findings are visualized through frame annotation of predicted emotion labels and their corresponding facial areas, gaining insight into emotional reactions extracted from real-time video streams. This detailed insight is useful for predicting suspicious activity from emotional states, increasing the reliability and accuracy of the behavioural assessment system.

Generation of training set: When gestures and emotions are detected, images are labeled as ‘cheating,’ ‘normal,’ or ‘suspicious’ depending on given criteria mainly including detected head orientations and emotions. When the head orientation is frontal or upward and emotions are neutral, happy, or sad, it is marked as ‘no cheating.’ Head orientations in the left, right, or backward directions and emotions that show fear or anxiety are marked as ‘cheating.’ Or if the head is frontally pointing downwards and emotions show fear or anxiety, it is marked as ‘suspicious.’ These labelled training data are stored for later analysis or model training.



Phase 3: Live video analysis - suspicious activity recognition

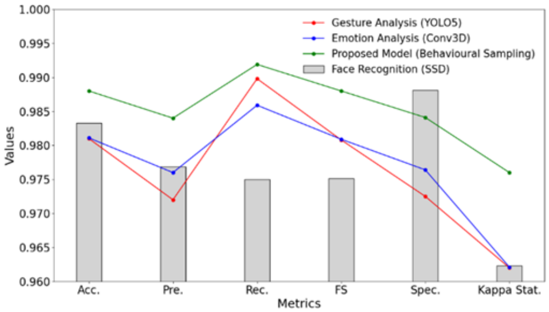
Live video analysis involves real-time continuous capture from a surveillance camera and recording frames in real-time. Frames are preprocessed by being extracted at one-second intervals, including resizing, normalization, and scaling to boost object detection. The second stage employs YOLOv5 to detect student gestures by tracking head movements, predicting bounding boxes and orientations—left, right, up, down, front, and back. The C3DN model identifies facial expressions to identify emotions such as happiness, sadness, fear, and anxiety. Identified orientations and emotions label behaviors as ‘cheating’, ‘normal’, or ‘suspicious’. An SSD model identifies the face of the student and compares it with a database based on cosine similarity to initiate alerts. Fig. 4 shows the workflow of the proposed model, and Algorithm 1 provides the pseudocode for the overall implementation of the proposed model.

III. EXPERIMENTAL ANALYSIS

Dataset Used

This section explains the experimental design for system deployment, including the generation of the dataset, hardware and software setups, hyperparameters for different learning models, and performance metrics utilized to measure the proposed study and compare it with current systems.

The model in question first develops a database through live-capturing individual students in multiple directions in the course of classes and saving 1,000 images for recognition. The images, which have a size of 1920 x 1080, are preprocessed to ensure efficient processing and management of memory. The database is employed in training the SSD model for face detection and recognition of students. The system also captures and transforms student video under examination into frames to identify suspicious behavior. By processing frames in sequence, rather than holding entire video streams in memory, the system conserves memory despite a high number of students in the hallway. A dataset for suspicious activity detection was gathered, with 2,000 images taken from different classroom environments: 1,000 of students performing cheating activities and 1,000 of authentic activities.



Performance comparison for various Suspicion Models

This data set accounts for both conventional and contemporary learning environments, capturing a range of typical student activities in an examination. It records a variety of responses throughout the test-taking period, providing a realistic illustration of real-world scenarios. This variety is invaluable in improving the model's capacity to identify unusual behavior and generalize across settings, thus informing an improved automated invigilation system. In order to additionally enrich the dataset's representation of actual examination settings, data augmentation methods were used, enlarging the dataset to 5,000 images, 80% of which were used for training and 20% for testing

Head movements of students were tracked with a training set containing orientations: left, right, up, down, front, and back to train YOLOv5. A training set for students' emotions was also developed with six classes: happy, neutral, sad, anxiety, fear, and stress, utilized to train C3DN. These categories mark images as 'cheating,' 'no cheating,' and 'suspicious,' either manually or semi-automatically according to a pre-defined rule. After training, the model is tested based on test data from the exam dataset. The video is broken down into frames and each frame is analyzed for head motion and emotion before being labeled as 'cheating,' 'no cheating,' or 'suspicious.'

In addition, this study recognizes the need for student privacy in video surveillance. It upholds the requirement for secure storage and handling of recorded content, with access limited only to authorized users. To counter possible biases in facial, gesture, and emotion detection, rigorous training across heterogeneous datasets will be utilized to boost detection accuracy. Regular auditing of system performance will be conducted to detect and address any discrepancies. Through the exposure of these measures, the study seeks to add strength to the ethical basis of automated invigilation systems in learning environment.

Experimental setup

The equipment utilized for analysis comprises two HIKVISION EZVIZ CS-BW3824B0 cameras, in addition to an NVR 8-CHANNEL and 2TB AV HDD, which is positioned to view all students. The system also employs a Logitech Brio Ultra HD Pro USB camera to capture high-definition images. For processing, an Acer WS laptop with an Intel i5 (12th Gen) processor, 16GB DDR4 RAM, RTX 3050 6GB GPU, and 512GB SSD storage is utilized for real-time processing, enhancing DL calculations for face recognition and gesture detection while maintaining efficient data processing. The code is coded on Python through the Jupyter IDE framework using OpenCV and other required libraries. Additionally, memory is handled by using batch processing methods, in which the number of frames taken is minimized to process optimally. Libraries such as NumPy are utilized to distribute memory efficiently, and both batch size and resolution are regulated to allow for efficient processing, particularly when the number of students in the examination room is considerable.

For extracting features, the SSD model uses a transfer learning-based modified VGG16 architecture. The VGG16 model, trained beforehand on a big dataset such as ImageNet, has 13 convolutional layers and 3 fully connected layers that are trained to extract high-level features from input images. The early convolutional layers of VGG16 utilize 64 filters in the first two layers, then 128 filters in the subsequent two layers, and 256 filters in the subsequent three layers. The SSD architecture takes VGG16 further by incorporating more convolutional layers with 512, 1024, 256, and 128 filters to enable it to recognize objects at different scales. The model incorporates multibox loss as the main loss function and stochastic gradient descent (SGD) with momentum as the optimizer for effective training. Some of the important hyperparameters used are a batch size of 1, momentum as 0.9, weight decay of 0.0005, localization loss weight of 1.0, and confidence threshold as 0.01.

In the end of all these the outcomes clearly demonstrate that the proposed system is able to offer a perfect and confidential solution for ITS, thus, resulting in a significant boost of vehicle network security.

The YOLOv5 model identifies gestures by detecting human heads and their directions in every video frame. Transfer learning is employed for extracting features, pre-trained using massive datasets such as COCO, and having CSPDarknet53 as a backbone. CSPDarknet53 consists of 29 convolution layers that extract high-level features, beginning with 32 filters followed by layers with 64, 128, 256, and 512 filters for extracting high-level features. The Neck module, Path Aggregation Network (PANet), combines features at different scales to improve detection, with filtering from 64 to 256 filters. The last output layer consists of 255 filters (3 anchor boxes per grid cell × 4 bounding box coordinates + 1 objectness score + 80 class probabilities). Major training hyperparameters are a learning rate of 0.01, batch size of 1, momentum of 0.937, weight decay of 0.0005, and a confidence threshold of 0.01, training over 120,000 steps in 300 epochs.

For detecting emotions, the C3DN model extracts spatial and temporal information from facial expression in frames of a video. It has four convolution layers: the first one with 32 filters for primary feature extraction, the second with 64 filters, the third one with 128 filters, and the fourth one with 256 filters for detailed facial patterns. Following each convolution layer are 3D pooling layers to compress spatial and temporal dimensions without losing information. Fully connected layers map the feature maps to classification. The last layer is a softmax layer that returns probabilities for each class of emotion (e.g., happy, sad, or neutral). Some of the important hyperparameters are a learning rate of 0.01, batch size of 1, momentum of 0.9, weight decay of 0.0005, and training for 120,000 steps in 100 epochs using SGD with momentum.

Therefore, the results from the identity verification stage trigger the gesture analysis stage. After the system has verified a student's identity, it employs YOLOv5 to track particular gestures. Simultaneously, emotional analysis through C3DN assesses the emotional state of the student. The outputs of these analyses are combined through predetermined decision rules, which categorize activities as normal or suggestive of possible malpractice

Performance measure

The model is tested with an annotated dataset through 5-fold cross-validation, where 80% of the data is used as the training set and 20% as the test set. The performance of the model is independently evaluated through three phases employing a confusion matrix of four measures: true positive (TP) in correctly identifying the positives, like identifying cheating (e.g., detecting cheatings), true negative (TN) in rightly rejecting the negatives (e.g., detecting not cheating), false positive (FP) in mistakenly asserting positives, and false negative (FN) in failing to identify cheating. Evaluation measures are accuracy, precision, recall, F1-score, specificity, false discovery rate, error rate, and Cohen's Kappa statistics.Accuracy calculates the ratio of correctly classified instances, whereas precision calculates the accuracy of positive predictions. Specificity and recall calculate the capacity to detect actual negatives and positives, respectively; the F1-score balances recall and precision. The false discovery rate calculates the proportion of false positives, and the error rate calculates incorrect predictions as a proportion of total predictions. Lastly, Cohen's Kappa indicates the extent to which two raters agree after adjusting for chance agreement. Formulas for the above metrics are explained below.

Methods	Accuracy
Gesture Analysis	97.2
Emotion Analysis	97.6
Behavioural Sampling	98.4

IV. RESULTS AND DISCUSSIONS

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 included 1000 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

Thus, a frontal head orientation with neutral, happy, or sad emotions was classified as 'no cheating'; head orientations to the left, right, or back with emotions like anxiety, fear, and stress were classified as 'cheating'; and 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.

The analysis indicated that the SSD face recognition method identified 146 students correctly, incorrectly classifying 4 images, and had higher accuracy and precision of 0.9833 and 0.9769, respectively, with low false discovery rates and error rates of 0.0256 and 0.0167. Additionally, the comparison of the results of several student activity detection methods, including gesture analysis, emotion analysis, and the suggested model, demonstrated better performance based on several measures. The suggested model regularly recorded the highest values in precision (0.9840), accuracy (0.9880), recall (0.9919), F1-score (0.9880), specificity (0.9841), and Cohen’s Kappa (0.9760), showing better overall performance than gesture and emotion analysis.

Although gesture analysis revealed competitive performance with improved accuracy (0.9810), recall (0.9898), and error rate (0.0190), indicating negligible false negatives, it was slightly behind on other measures compared to emotion analysis and the suggested model. Likewise, emotion analysis revealed high precision (0.9760) and F1-score (0.9809), reflecting negligible false positives for effective detection of suspicious behavior. But for student activity detection in exams, the proposed model outperformed single gesture analysis and emotion detection approaches on all metrics. Class-wise accuracy for these models is shown in Table 3. The research demonstrated that gesture and emotion analysis were successful in predicting non-cheating activities and detecting cheating activities respectively, and their combination into the proposed model improved accuracy. The values were graphed as a bar in Fig. 9, where the bars are the face recognition method performance and the lines are the suspicious activity detection methods.

The suggested model was tested by taking live photographs of students, changing the number of students present in the exam hall. Fig. 10 shows the outcome. It can be seen that when the number of students is small, the model has 100% accuracy. But when the number of students increases, the accuracy drops because there is less visibility of images inside the classroom. Thus, the model worked best in small classrooms with a maximum of 30 students to provide complete coverage of the students. In large exam rooms with a capacity of up to 100 students, extra cameras were needed to capture all student information completely.

Although the model proposed shows better performance in identifying cheating, performing better than most models in the research community, there are some limitations that come with these improvements. These limitations are addressed below.

Scalability Problems: The model can identify and recognize faces and gestures in the exam hall with a seating capacity of at least 30 students. But as the number of students in the hall increases or the hall gets larger, the face and gesture detection accuracy may be compromised, leading to an increased error rate. This scalability issue points toward the necessity of additional research into optimizing the model for greater numbers, which could require more cameras.
High-Quality Image Dependence: Another significant aspect to consider is the effectiveness of the model, which largely depends on the input image quality. Lighting conditions and image resolution variability may negatively impact detection accuracy. Hence, robust image preprocessing methods and the feasibility of training the model using a variety of datasets with different image qualities need to be investigated.

Gesture Analysis Limitations: Another significant factor to consider is the effectiveness of the model, which largely depends on the quality of input images. Changes in lighting conditions and image resolution may negatively impact detection accuracy. Hence, it is crucial to investigate strong image preprocessing methods and the possibility of training the model using varied datasets that contain different image qualities.

Impact of Human Factors: Human factors, including stress and anxiety manifested by students when they are being tested, could also affect the efficacy of the model. These can cause a rise in false negatives in emotional analysis, which are not picked up and can enable cheating behavior to continue. For this, future models could include a larger number of training images that incorporate different emotional expressions and states.

Detection of Cheating Actions: Human behavior is multifaceted and dynamic, and the model may not detect or identify all cheating actions. For example, though head-down poses are regarded as non-cheating activities, students trying to cheat using self-help strategies—e.g., writing on hands, calculators, or mobile phones—may not be detected. To overcome this, future work should aim to expand the dataset size and use sophisticated DL approaches such as multi-task learning and attention mechanisms.

Comparison with Other Studies: Furthermore, this study does not make direct comparisons of results with other existing studies because the nature of each study's synthesized datasets is different, and therefore, such comparisons are not practical. Future work would be better off using standardized benchmark datasets or testing the model performance using commonly used metrics in related studies to allow for more significant comparisons.

Lack of Cost Analysis: One of the main limitations of the suggested model is the lack of a cost analysis. Estimating the operating costs involved with the algorithm, including processing time, memory space, and energy consumption, is essential in understanding its practicability and feasibility for real-world usage. It is important for this analysis in determining opportunities for optimization and system efficiency, and future work must incorporate a cost analysis to realize resource demands and algorithm scalability.

CONCLUSION

This research presents a DL-driven smart invigilation system to maintain exam integrity by detecting fraudulent activity during examinations. The system involves three primary phases: (1) verification of student identity via SSD-based face recognition, (2) sampling of behavior via gesture analysis using YOLOv5 and emotion detection via C3DN, and (3) real-time live monitoring that combines gesture and emotion information to identify suspicious activities. With a 98.8% accuracy, the model overcomes major limitations of current solutions, improving academic dishonesty detection and strengthening exam integrity. The system provides automated invigilation, facilitates efficient resource allocation, and provides an equitable testing environment, simplifying exam processes, enhancing security, and minimizing human error for higher learning institutions.

Though the model possesses high accuracy, there are some limitations that guide the direction for future research. The model may perform better with bigger exam contexts by using enhanced hardware configurations or distributed camera setups. Another avenue of improvement is broadening the training dataset to incorporate a greater diversity of cheating motions. With the computational intensive nature of DL models, there should be cost assessment on the time of computation and memory used. Moreover, the inclusion of eye contact and head orientation in gesture analysis and investigation of sophisticated DL methods, including multi-task learning and attention mechanisms, may enhance the model's use in various testing environments.

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

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. Gamage, R. G. Pradeep and E. K. de Silva, "Rethinking Assessment: The Future of Examinations in Higher Education," Sustainability, vol. 14, no. 6, pp.1-15, 2022.

[3] M. A. Mulongo, "Effectiveness of University Examinations Management Strategies in Mitigating Examination Malpractices in Kenya", (Doctoral dissertation, Karatina University).

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

[5] O. L. Holden, M. E. Norris and V. A. Kuhlmeier, "Academic Integrity in Online Assessment: A Research Review," In Frontiers in Education (vol. 6, pp. 1-13), Frontiers Media SA, July 2021.

[6] F. Kamalov, H. Sulieman and D. Santandreu Calonge, "Machine Learning based Approach to Exam Cheating Detection," Plos One, vol. 16, no. 8, pp. 1-15, 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] W. Alsabhan, "Student Cheating Detection in Higher Education by Implementing Machine Learning and LSTM Techniques," Sensors, vol. 23, no. 8, pp. 1-21, 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, 2024.

[10] M. Asad, M. Abbas, A. Asim, A. Hafeez, M. M. Sadaf, A. U. Haq and M. Asif, "Suspicious Activity Detection During Physical Exams,". Available at SSRN 4676389, pp. 1-15, 2023.

[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," In Conference on Data Mining| DMIN (vol. 6, pp. 205), 2006.

[14] Y. Atoum, L. Chen, A. X. Liu, S. D. 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," arXiv preprint arXiv:2101.09841, pp. 1-9, 2021.

[16] S. El Kohli, Y. Jannaj, M. Maanan and Rhinane, "Deep learning: New Approach for Detecting Scholar Exams Fraud,". The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 46, pp. 103-107, 2022.

[17] F. Mahmood, J. Arshad, M. T. Ben Othman, M. F. Hayat, N. Bhatti, M. H. Jaffery, A. U. Rehman and H. Hamam, "Implementation of an Intelligent Exam Supervision System using Deep Learning Algorithms," Sensors, vol. 22, no. 17, pp. 1-21, 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] M. Roa'a, I. A. Aljazaery 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, pp. 1-21, 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, pp. 1-19, 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, pp. 1-32, 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, 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, pp. 1-18, 2020.

[29] L. Ling, J. Tao, G. Wu, “Research on Gesture Recognition based on YOLOv5,” In Chinese Control and Decision Conference (pp. 801-806). IEEE, May 2021.