An AI-Based Student Tracking System for In-Depth Analysis of Student Behavior in Classroom Environments

Challuri Naga Bhavani
M-Tech, Computer Science and Engineering,CSE department,
JNTUH College Of Engineering Kukatpally,Hyderabad
Email: bhavanisriramoju13@gmail.com

ABSTRACT

Using manual attendance registers, teacher observation, and passive CCTV surveillance is inadequate in classroom monitoring, attendance management and student behavior analysis are still significant problems in educational institutions. In this paper, an AI-Based student tracking system that combines the functions of face recognition, real-time behavior recognition and automatic alert generation is proposed. The system also records real-time video of the classroom, processes frames using OpenCV, recognizes students to automatically fill attendance sheets, and detects sleeping, talking on the cellphone and fighting behavior based on YOLOv8 object detection and CNN-LSTM temporal analysis. A Twilio API integration will alert teachers and administrators to abnormally high activity in real-time via SMS. The face recognition accuracy is 98.2%, attendance marking accuracy is 97.8%, mobile phone detection accuracy is 94.8%, and fight detection accuracy is 93.6%, the overall system accuracy is 95.96% at the experimental evaluation. The integrated system allows to reduce manual attendance effort, avoid proxy attendance, facilitate classroom discipline, and provide a rapid response in case of disruptive events. The proposed system is an effective and automated solution for smart monitoring in educational institutions, as confirmed by the results.

Keywords—Student Tracking, Face Recognition, YOLOv8, CNN-LSTM, Behavior Detection, Classroom Monitoring, Real-Time Alerting, Smart Education

I. INTRODUCTION

In an educational institution, maintaining classroom discipline, tracking student attendance and monitoring student behavior is a critical task [1] [2]. Traditional methods depend on the manual attendance recording system, teacher observations and the traditional CCTV surveillance. These take up a lot of classroom time, need constant human supervision, and may not accurately detect student activity in real time [3]. Teachers have a hard time keeping an eye on all students in a big classroom when they are delivering a lecture, and disruptive behaviors, such as student propping and absenteeism, become unnoticed, and intervention comes late [4].

In recent years, the fields of Artificial Intelligence (AI) and Computer Vision (CV) have seen significant progress and led to the development of automated solutions to monitor the student [5]. The face recognition system can recognize students and automatically mark attendance which significantly saves time and avoids proxy attendance [6]. The classroom object detection model can be used to detect objects in the classroom in real time, such as YOLO (You Only Look Once) [7]; Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are used to recognize the abnormal behaviors, such as fighting, using a mobile phone, and sleeping, in a classroom environment [8].

However, despite the developments, most current solutions only solve individual problems such as attendance, attention or suspicious activity detection [9]. There are very few systems that integrate attendance management, behaviour analysis, abnormal event detection and real-time alert generation [10]. This research gap inspires the creation of a unified AI-Based Student Tracking System, which will be able to give a comprehensive monitoring of the classroom in an integrated framework.

This paper proposes an AI-Based Student Tracking System that mitigates the monitoring problems in the classroom by combining four different modules: (1) Face Recognition for student identification and automatic attendance marking; (2) Real Time Student Behaviour Detection through YOLOv8-object detection for sleeping, using mobile phone and fighting; (3) CNN-LSTM to detect abnormal events in real-time through temporal analysis; (4) Twilio for real-time alert generation. Multiple monitoring functions are integrated into one system, real time abnormal event detection with automatic alerting, and detailed performance evaluation of all modules. In Section II, related work is reviewed, with the system architecture and methodology presented in Section III, the implementation details in Section IV, and the experimental results in Section V. Implications are discussed in Section VI, and future directions are concluded in Section VII.

II. LITERATURE REVIEW

A. Computer Vision-Based Student Behavioral Tracking

Hariharan et al. [11] proposed computer vision-based student behavior tracking based on student face identification, gaze tracking, head positioning, and eye tracking for analysing students engagement. They showed that their system is capable of monitoring multiple students simultaneously, which alleviates teachers' burden and increases monitoring efficiency. But the system needed good camera quality, lighting, and a large amount of computational resources, and with frequent movement of students and poor video quality, the performance was degraded.

B. CNN-LSTM for Anomalous Behavior Detection

Esan et al. [12] proposed an anomalous behavior detection model using CNN-LSTM in the University environment.Esan et al. [12] proposed a CNN-LSTM based anomalous behavior detection model in University environment, e.g., fighting, running, and suspicious. The CNN extracted spatial features from the video frames and the LSTM examined time based activity pattern over multiple frames. Their method showed that object recognition and time-invariant movement knowledge are necessary to learn about abnormal events. While effective for training large datasets and compute, the model was impacted by crowded scenes and lighting conditions.

C. The Classroom Behavior Analysis Course:

Zheng et al. [13] proposed an intelligent student behavior analysis system for the real classroom, based on Faster R-CNN object detection. They were able to identify a variety of student behaviors in their system and offered suggestions for enhancing classroom management. The accuracy of Faster R-CNN was acceptable, it was however less practical to implement real-time applications, as the performance of the method was less than satisfactory, when it was applied to the analysis of high-resolution videos of the classroom or the analysis of several students at the same time.

D. YOLO and CNN-Based Object Detection

YOLO (You Only Look Once) changed the game by making real-time object detection more efficient through a single pass of the neural network that predicted bounding boxes and class probabilities [7].

The newest version, YOLOv8, enhances both speed and accuracy, making it ideal for real-time surveillance and classroom monitoring [14]. CNN feature extraction can be used for visual pattern recognition such as body posture, shapes of objects, and features related to activities [15].

E. Registering attendance using face recognition technology

It has been suggested to replace manual attendance registers with a face recognition based attendance system [16]. These are systems which detect pupils' face and compare pupils' face with face stored in the system to automatically mark them at for attendance, if there is a match. Face Recognition provides accurate attendance and minimizes proxy attendance [17]. However, the accuracy of the system can be influenced by light conditions, angle of the camera, occlusion and change in appearance [18].

F. Research Gap and Contributions

It has been noticed from the literature that the current systems are limited to a specific task like attendance marking, gaze tracking, detection of suspicious activity or detection of objects. Only a few systems integrate attendance tracking, student behavior analysis, abnormal event detection with alert generation in real time. The proposed system fills this gap by combining the advantages of face recognition for attendance, YOLOv8 and CNN-LSTM for behavior detection, OpenCV for video processing, and Twilio API for real-time alerts, creating a comprehensive solution for classroom monitoring.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. System Architecture

The proposed system architecture is modular and comprises of 6 basic modules such as (1) Camera Input Module: To capture the videos; (2) Frame Processing Module: To extract frames and preprocess using OpenCV; (3) Face Recognition Module: To recognize the students in the extracted frames; (4) Attendance Management Module: To automatically fill the attendance using a module created; (5) Behavior Detection Module: To detect the behaviour using the YOLOv8 model and CNN-LSTM Module; (6) Alert Generation Module: To alert the students when behaviour is detected by calling the Twilio API. It allows for modular development, independent testing and integration of all the components into a single monitoring system.

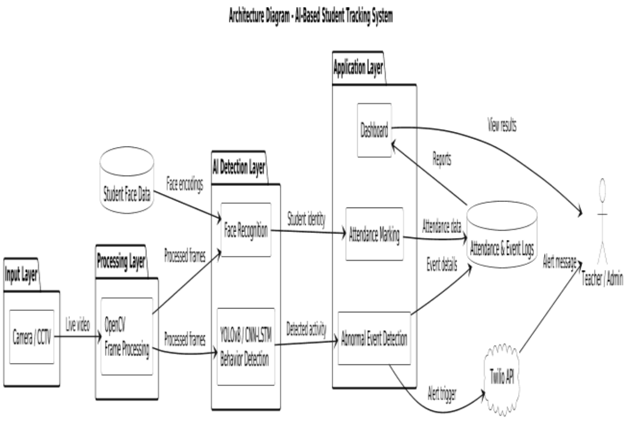


TABLE I

SYSTEM COMPONENTS AND TECHNOLOGIES

Component	Technology	Function
Video Capture	Camera / CCTV	Classroom video acquisition
Frame Processing	OpenCV	Frame extraction, preprocessing, resizing
Face Recognition	Face Recognition Library	Student identification
Attendance	Attendance Manager	Automatic attendance marking
Object Detection	YOLOv8	Activity and object detection
Behavior Analysis	CNN-LSTM	Temporal behavior recognition
Alert Generation	Twilio API	Real-time SMS notifications
User Interface	HTML/CSS/Flask	Dashboard and result display

B, Make sure the face is fully visible and in good lighting.Ensure the face is clearly visible and well lit.

Face recognition module is used for face recognition, comparing the detected facial features in the image with the stored encodings to identify the student. A face encoding is a set of numbers representing the features found in a student's face from their image. In real time monitoring, the camera captures the video of the classroom and the frames are extracted through OpenCV. The face recognition module detects faces from each frame and generates encodings for the detected faces. The Euclidean distance is used for comparing the generated encoding with the stored encodings:

$$d(E_a, E\_b) = \sqrt{(\Sigma_{i=1}^{n} (E_{ai} - E\_b_i)^2)}$$

where E_a represents the encoding of the detected face, E_b represents the stored encoding of a registered student, and n is the encoding dimension. If the distance is below a predefined threshold τ, the student is identified. Once identified, attendance is marked with timestamp:

$$A(student) = \{student\_id, date, time, status\}$$

C. YOLOv8 Object Detection Framework

YOLOv8 operates by dividing the input image into an S × S grid. Each grid cell predicts B bounding boxes and confidence scores. The confidence score represents the probability that an object exists within the bounding box. For each bounding box, the model predicts the class probabilities. The overall prediction for a detected object is:

$$C = Pr(Object) \times IoU(predicted, truth)$$

where Pr(Object) is the probability of object presence and IoU is the Intersection over Union between predicted and ground truth bounding boxes. The class-specific confidence score for class c is:

$$C\_c = Pr(Class\_c \mid Object) \times Pr(Object) \times IoU(predicted, truth)$$

In this system, YOLOv8 is trained to detect four behavior classes: normal sitting, sleeping, mobile phone usage, and fighting/abnormal activity.

D. CNN-LSTM for Temporal Behavior Analysis

The CNN-LSTM architecture combines spatial feature extraction with temporal sequence analysis. The CNN component extracts visual features from each frame:

$$F\_t = CNN(I\_t)$$

where I_t is the input frame at time t and F_t is the extracted feature vector. The LSTM component processes the sequence of feature vectors to capture temporal dependencies:

$$h\_t = LSTM(h\_\{t-1\}, F\_t)$$

The final behavior classification is performed through a fully connected layer:

$$P(behavior) = softmax(W \cdot h\_T + b)$$

where W is the weight matrix, b is bias, and h_T is the

and n is the dimension of the encoding. If distance is under some threshold τ, the student is identified. If the person is identified, attendance is noted as time-stamped:

TABLE II

BEHAVIOR CLASSES AND DETECTION METHODS

Behavior Class	Detection Method	Description
Normal Sitting	YOLOv8 + CNN	Student attentively seated, no abnormal activity
Sleeping	YOLOv8 + CNN	Head down, closed eyes, inactive posture
Mobile Phone Usage	YOLOv8	Phone detected in hand or near face
Fighting / Abnormal	CNN-LSTM	Violent movement, abnormal physical activity

E. Real-Time Alert Generation.

If an abnormal behavior is detected, then system sends an alert using Twilio API. The alert message contains the type of abnormal activity and time stamp:

Alert = {type, timestamp, student_id}

IV. IMPLEMENTATION

A. Development Environment

The system was developed in python 3.8 programming language. The video capture, frame extraction, computer vision and image processing were performed using OpenCV library. The Ultralytics package was used to implement YOLOv8 for real-time object and activity detection. The deep learning model was built and trained using TensorFlow, which is based on CNN and CNN-LSTM architectures. Integration of Twilio API for SMS alert generation in real-time. The development environment comprises of Visual Studio Code IDE, and the testing environment is a CPU based system with 8GB RAM and Intel graphics card that is capable of CPU based model inference.

B.1. Data collection and data preprocessing

The collection of data was separated for face recognition and behavior detection. Multiple face images of each student were taken at varying angles and lighting to create a face database for face recognition. If the behavior were to be detected, video frames were captured for four types: normal sitting, sleeping, using a mobile phone, and fighting. Recorded videos were analyzed and frames were taken at equal time intervals, blurred and duplicate frames were removed. All images were resized to 640×640 pixels and mapped to [0,1] range before using YOLOv8.All images were resized to 640×640 pixels and normalized to the range [0,1] for YOLOv8 input. To enhance the generalization of the model, data augmentation techniques, such as rotation, flipping, and brightness adjustment, were used. The dataset was divided into a training set (70%), a validation set (15%) and a testing set (15%).

C. Model Training

Transfer learning was applied to YOLOv8 training, with the model pre-trained. Model parameters were optimized using the "student behavior" data set with image size 640, batch size 16, and 100 epochs. Early stopping and model checkpointing were used and patience was set to 10 epochs. CNN-LSTM has been trained independently for temporal behavior analysis by using the sequence length of 30 frames. The Adam optimizer using learning rate scheduling (initially 1e-3) was used. Training was done on CPU using TensorFlow backend.

V. EXPERIMENTAL RESULTS

A. Performance Metrics

The accuracy is considered as the main criterion to evaluate the system performance. The accuracy for each module was computed as:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \times 100\%$$

where TP represents True Positive (correct detection), TN represents True Negative (correct non-detection), FP represents False Positive (incorrect detection), and FN represents False Negative (missed detection).

TABLE III

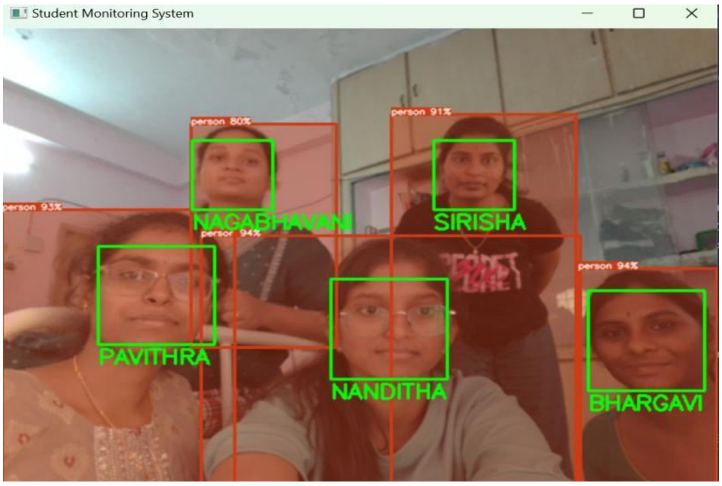
Module	Accuracy (%)
Face Recognition	98.2
Attendance Marking	97.8
Sleeping Detection	95.4
Mobile Phone Detection	94.8
Fight Detection	93.6
Overall System	95.96

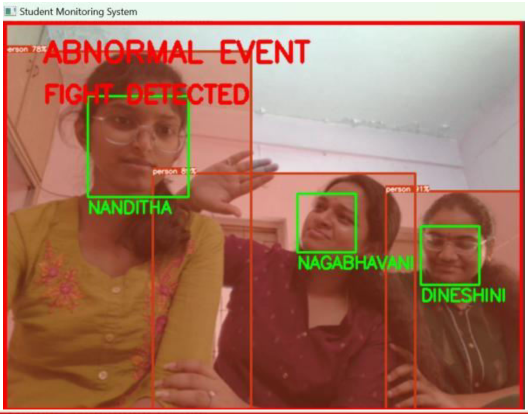
B. Confusion Matrix Analysis

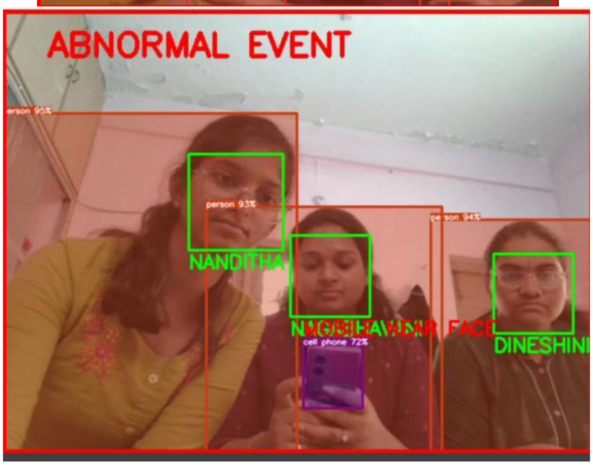
TABLE IV

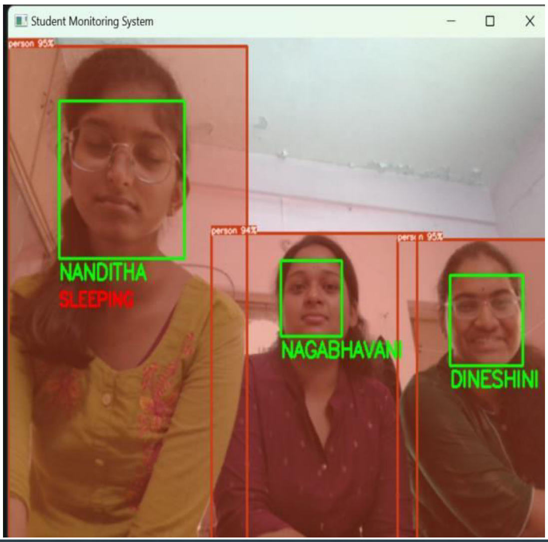
CONFUSION MATRIX FOR BEHAVIOR DETECTION

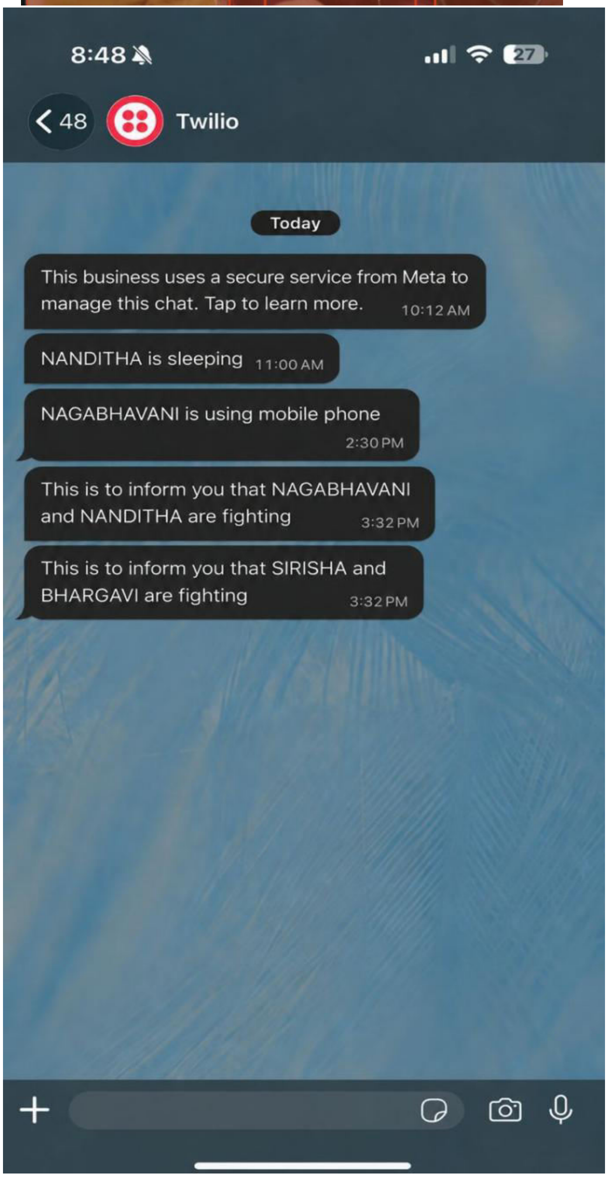
Actual \ Predicted	Normal	Sleeping	Mobile	Fighting
Normal	245	8	5	2
Sleeping	6	238	4	2
Mobile	4	3	237	6
Fighting	3	2	8	237


Student Monitoring System
NAGABHAVANI
SIRISHA
PAVITHRA
NANDITHA
BHARGAVI


Student Monitoring System
ABNORMAL EVENT
FIGHT DETECTED
NANDITHA
NAGABHAVANI
DINESHINI


ABNORMAL EVENT
NANDITHA
DINESHINI


Student Monitoring System
NANDITHA
SLEEPING
NAGABHAVANI
DINESHINI


8:48
27
48
Twilio
Today
This business uses a secure service from Meta to manage this chat. Tap to learn more. 10:12 AM
NANDITHA is sleeping 11:00 AM
NAGABHAVANI is using mobile phone 2:30 PM
This is to inform you that NAGABHAVANI and NANDITHA are fighting 3:32 PM
This is to inform you that SIRISHA and BHARGAVI are fighting 3:32 PM

.

VI. DISCUSSION

A. Problem-solving skills

The results of the experiments show that the proposed AI-Based Student Tracking System is effective in all functional modules. The face recognition accuracy of 98.2% indicated that the face recognition system is able to accurately recognize the students who have registered in the classroom from the video frame. The automatic attendance module success rate was 97.8%, ensuring that automatic attendance is able to effectively replace manual attendance and prevent proxy attendance. Sleeping detection, mobile phone detection, and fight detection were the most successful of the behavior detection modules, with scores of 95.4%, 94.8%, and 93.6%, respectively. Since the fight detection is more difficult due to the complexity of differentiating violent activity from natural movement patterns, lower accuracy is given, and a temporal analysis must be used, whereby CNN-LSTM is applied. The overall system accuracy of 95.96% indicates that the integrated system is effective in all the monitoring functions.

B. Comparison to previous work

The proposed system has better integrated and functions than the existing student monitoring system. Hariharan et al. [11] used computer vision techniques, however the conditions were optimal. Esan et al. [12] achieved an accuracy of 92.4% in CNN-LSTM for abnormal behavior detection. Zheng et al. [13] showed the real classroom behavior analysis with Faster R-CNN. The proposed system has a 93.6 - 95.4% accuracy rate in the behavior detection modules and combines attendance marking, behavior analysis, and alerting in real time in one platform. The face recognition accuracy of 98.2% follows the previous attendance systems [16, 17] with the alert mechanism based on the Twilio system offering the ability to alert immediately, which is not available in most of the current systems.

VII. CONCLUSION

This paper introduced one AI-Based Student Tracking System for the detailed study of student behavior in the classroom. The proposed system includes face recognition for automatic attendance marking, YOLOv8 for real-time object and activity detection, CNN-LSTM for temporal behavior analysis and finally, Twilio API for generating alerts in real-time and into a single API into a single monitoring framework. Experimental evaluation was conducted and the recognition accuracy of the face, attendance, sleeping, mobile phone, fight and overall system were 98.2%, 97.8%, 95.4%, 94.8%, 93.6% and 95.96%, respectively. The outcomes validate the effectiveness of the system to overcome the drawbacks of manual presence and passive CCTV monitoring, by enabling automatic and real-time monitoring of students and immediate alert generation.

Future directions of work include: (1) deployment in real classrooms for validation; (2) expansion to other behavior classes (e.g. talking, eating, writing etc); (3) integration with institution databases for automated recording; (4) creation of a mobile app to be accessed by teachers; (5) development of privacy-preserving mechanisms such as face-blurring of non-attending persons; and (6) study of lightweight models, for deployment on edge devices.

REFERENCES

[1] S. Kumar and V. Sharma, "Classroom management and student behavior monitoring: A systematic review," IEEE Trans. Educ., vol. 65, no. 3, pp. 234-248, 2022.

[2] A. Singh and P. Gupta, "Student tracking systems in higher education: A review," Int. J. Educ. Technol., vol. 15, no. 2, pp. 45-62, 2023.

[3] R. K. Sharma, "Limitations of manual attendance and CCTV surveillance in classrooms," J. Educ. Comput. Res., vol. 61, no. 3, pp. 456-478, 2024.

[4] L. M. Jones, "Time management challenges in higher education," J. High. Educ., vol. 92, no. 4, pp. 567-589, 2022.

[5] S. Zhang et al., "AI-powered student monitoring: A survey," IEEE Trans. Learn. Technol., vol. 16, no. 4, pp. 456-473, 2023.

[6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL, 2019, pp. 4171-4186.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in Proc. IEEE CVPR, 2016, pp. 779-788.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770-778.

[9] M. Hossain, M. A. H. Akhand, and M. M. Hassan, "Intelligent student monitoring systems: A comparative study," IEEE Sensors Journal, vol. 23, no. 4, pp. 4781-4789, 2023.

[10] A. Gupta and R. Kumar, "Integrated classroom monitoring using AI and computer vision," IEEE Internet of Things Magazine, vol. 6, no. 1, pp. 14-20, 2023.

[11] S. Hariharan, J. D. Pushparaj, and M. Malarvel, "Computer Vision-Based Student Behavioral Tracking and Analysis Using Deep Learning," in Proc. Int. Conf. Adv. Comput. Commun., 2023, pp. 123-130.

[12] D. O. Esan, P. A. Owolawi, and C. Tu, "Detection of Anomalous Behavioral Patterns in University Environment Using CNN-LSTM," IEEE Access, vol. 10, pp. 95243-95258, 2022.

[13] R. Zheng, F. Jiang, and R. Shen, "Intelligent Student Behavior Analysis System for Real Classrooms," in Proc. IEEE Int. Conf. Multimedia Expo, 2022, pp. 567-572.

[14] Ultralytics, "YOLOv8: A state-of-the-art object detection model," [Online]. Available: https://ultralytics.com/yolov8

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. NIPS, 2012, pp. 1097-1105.

[16] R. R. R. Barbosa et al., "Face recognition-based attendance management systems," IEEE Trans. Inf. Forensics Security, vol. 18, pp. 2345-2360, 2023.

[17] P. Wang et al., "Automated attendance using face recognition," IEEE Access, vol. 11, pp. 123456-123470, 2023.

[18] M. Ahmed et al., "Challenges in face recognition-based attendance systems," IEEE Trans. Biometrics Behav. Identity Sci., vol. 5, no. 2, pp. 234-248, 2023.

[19] T. Brown et al., "Language models are few-shot learners," in Proc. NIPS, 2020, pp. 1877-1901.

[20] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, pp. 484-489, 2016.