Enhanced Classroom Management: A Smart AI-Based Attendance System Utilizing YOLOv11 for Face Recognition

P.Rishik[1], K.Megha sri[2], M. Srimann Reddy[3], Mr. Vaggela Rama Chandra Murthy Raju[4]

[1,2,3,4]Dept. of CSE, Vardhaman College of Engineering, Hyderabad, India

***Abstract*—In today's educational landscape, effective classroom management is crucial for promoting student engagement. Traditional attendance methods, such as roll calls and biometric systems, often lead to inefficiencies and errors. To address these issues, this study presents an AI-driven solution, "Enhanced Classroom Management: A Smart AI-Based Attendance System Utilizing YOLOv11 for Face Recognition." By leveraging advanced object detection, this system automates attendance tracking, reducing manual effort and improving accuracy.** [1]

The proposed system uses YOLOv11, a deep learning model known for its speed and precision in real-time object detection. Unlike biometric systems that require physical interaction, this technology enables automatic facial recognition for attendance recording. The AI-based framework enhances efficiency, ensuring real-time processing while minimizing disruptions during lectures. [2]

Evaluations show that the YOLOv11-powered system improves accuracy and reduces administrative workl oad. The findings highlight its potential to enhance security and reliability in attendance management. This research contributes to the growing field of AI in education, demonstrating how intelligent systems can create a more accountable learning environment.[3]

Index Terms--AI-based attendance system, face recognition, YOLOv11, real-time object detection, deep learning.

I.INTRODUCTION

Traditional attendance methods can be problematic in large classrooms, leading to inefficient use of class time. Manual processes, such as roll calls or sign-in sheets, are prone to errors and take up a lot of time, especially when there are many students. These methods can also interrupt the flow of the lesson, reducing the time available for teaching.

As classrooms grow larger, the drawbacks of manual attendance become even more noticeable. With a high number of students, these methods take longer and are more likely to lead to mistakes. The disruption caused by roll calls or signing sheets means less time is spent on actual teaching and engaging with students.

Automated systems provide a better solution by removing the need for manual attendance tracking. This allows instructors to focus more on their teaching and student interactions. The use of artificial intelligence (AI) and machine learning (ML) makes these systems even more powerful by offering real-time data and helping to identify patterns in student absence.

II.RELATED WORK

ID-Based Attendance Systems:

ID-based systems like AMS and TouchIn have streamlined attendance by automating the check-in process, making it faster and less prone to human error. They also allow for easy data storage and integration with school records. However, their inability to verify the actual identity of the user poses a serious issue. In cases where students share their ID cards or devices, the system has no way of detecting fraud, which can lead to inaccurate records.

This problem is even more noticeable in large classes where it's difficult for teachers to manually verify each student. As a result, while these systems offer convenience, they fall short in ensuring true accountability. This gap has led to a growing need for smarter solutions—ones that can combine automation with reliable identity verification, such as face recognition technologies.

Location-Based Attendance Systems:

Location-based attendance systems use Bluetooth or Wi-Fi to detect if a student is nearby, making the process hands-free and convenient. Students don't need to tap or scan anything, which keeps it simple.

However, these systems can be affected by signal issues. Factors like walls, interference, or signal overlap might mark students present even if they're outside the classroom. Some improvements, like Wi-Fi fingerprinting, help track location more accurately by analyzing signal patterns. Still, these systems often struggle to provide fully reliable data, especially in busy or complex environments.

This project focuses on enhancing classroom management through AI-powered automation to create a more efficient and accurate learning environment. The main goal is to develop a face recognition-based attendance system using the YOLOv11 model, known for its speed and precision. By recognizing students' faces as they enter the classroom, the system records attendance automatically in real time, reducing manual work for teachers.

It also aims to offer real-time updates so instructors can track attendance instantly and respond quickly if needed. Another important aspect is ensuring smooth integration with existing tools like student information and learning management systems, helping maintain accurate records and streamline administrative tasks.

III.LITERATURE REVIEW

X. Cao, Z. Wang, Y. Zhao, and F. Su. "Scale aggregation network for accurate and efficient crowd counting." ECCV, 2018.

This paper introduces a Scale Aggregation Network (SAN) to improve crowd counting, a task that's challenging due to varying crowd densities, scale, and occlusions. The authors address these challenges by designing a network that combines multi-scale features to capture both global and local patterns in crowd scenes. The SAN uses deep learning to extract features at different resolutions and merges them for a more accurate count of individuals, even in crowded settings. Through experiments on various benchmark datasets, the authors show that SAN outperforms existing methods in terms of both accuracy and efficiency. This approach proves robust across different crowd densities and scene complexities, making it a valuable tool for applications like public safety, event monitoring, and urban planning.

A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei. "What's the point: Semantic segmentation with point supervision." ECCV, 2016.

In this paper, the authors propose a novel approach to semantic segmentation using point-level supervision instead of costly and time-consuming pixel-level annotations. By using a sparse set of labeled points, the system achieves competitive results, especially in tasks like object segmentation. This method is ideal for large datasets or when fine-grained segmentation isn't essential. The paper demonstrates that point-level supervision reduces annotation costs while maintaining accuracy, making it a practical alternative to traditional pixel-based methods, particularly in resource-constrained environments.

M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. "The PASCAL Visual Object Classes Challenge: A Retrospective." IJCV, 111(1), 2015.

This paper offers a retrospective look at the PASCAL Visual Object Classes (VOC) Challenge, a key benchmark in computer vision. Established in 2005, the VOC Challenge focused on evaluating object detection, classification, and segmentation models using a shared set of annotated images. It played a major role in advancing object recognition by providing standardized datasets and evaluation metrics. The authors reflect on how the challenge led to key improvements in object detection and recognition, as well as its ability to adapt over time to address new problems like action recognition and segmentation. Despite some limitations in diversity and complexity, the VOC Challenge has left a lasting impact on the field, influencing current trends and the development of new benchmarks and algorithms.

M. Gao, A. Li, R. Yu, V. I. Morariu, and L. S. Davis. "C-WSL: Count-Guided Weakly Supervised Localization." ECCV, 2018.

In this paper, the authors introduce C-WSL, a method designed to improve object localization using weak supervision. Traditional object detection methods require precise annotations, which are often costly and time-consuming. C-WSL simplifies this by using count annotations instead of detailed pixel-level or bounding box labels. The method uses the total count of objects in an image to guide localization, making it more efficient and less reliant on detailed annotations. Through experiments on various benchmark datasets, C-WSL proves to be effective at both localizing objects and estimating their counts, outperforming other weakly supervised approaches. This technique is especially valuable for large-scale datasets where detailed annotations are hard to obtain, marking a significant advancement in weakly supervised learning.

R. Girshick. "Fast R-CNN." ICCV, 2015.
Fast R-CNN is an important improvement in object detection, building on the earlier R-CNN model. Girshick introduces key changes to make the model faster and more efficient. Unlike R-CNN, which requires extracting regions of interest (ROIs) and running them through separate classifiers, Fast R-CNN combines these steps into a single network. The model performs feature extraction on the whole image, applies ROI pooling to get fixed-size feature maps from the ROIs, and uses a softmax layer for classification. Bounding box regression then refines the object locations. This results in a significant speedup and better memory efficiency compared to R-CNN. Fast R-CNN has become one of the most widely used techniques in object detection, influencing later models like Faster R-CNN and Mask R-CNN, and has had a major impact on both research and practical applications in computer vision.

P. Chattopadhyay, R. Vedantam, R. R. Selvaraju, D. Batra, and D. Parikh. "Counting everyday objects in everyday scenes." CVPR, 2017.

This paper introduces a new way to count objects in everyday scenes, without the need for detecting each object individually. The authors propose a method that directly estimates the number of objects by using deep learning techniques to analyze the entire scene. Their approach handles challenges like occlusion, varying object sizes, and cluttered backgrounds by focusing on overall scene context. The method uses a model to predict counts based on features from the whole scene, improving accuracy and robustness. Tested on several real-world datasets, the model proves effective for counting a range of objects in various environments, making it a practical solution for applications like inventory systems and surveillance. This approach reduces the need for costly object detection annotations, offering a scalable solution for real-world object counting and scene understanding.

IV.PROPOSED METHODOLOGY

This study adopted a modular and systematic methodology to develop a smart classroom attendance system that utilizes computer vision and deep learning. The project involved two major stages: face detection using YOLOv11 and identity classification for accurate attendance logging. To accomplish this, we constructed a custom dataset, trained specialized models for both detection and recognition, and deployed an integrated pipeline that can operate in real-time using webcam input. Each phase of the system—data preparation, training, testing, and deployment—was designed to ensure high precision and practical applicability in classroom environments

A. DATASETS

1) FACE DETECTION DATASET

To enable effective face localization, a customized dataset was curated by capturing classroom images in diverse conditions—variations in student pose, lighting, and facial orientation. These images were annotated manually using the Roboflow platform, where bounding boxes were drawn around every visible student face. The labeled dataset was saved in YOLO format, consisting of image files (.jpg) and their corresponding label files (.txt) along with a .yaml configuration defining the dataset structure. This dataset was fundamental in training the YOLOv11 model to detect faces with high confidence and accuracy in a classroom setting.

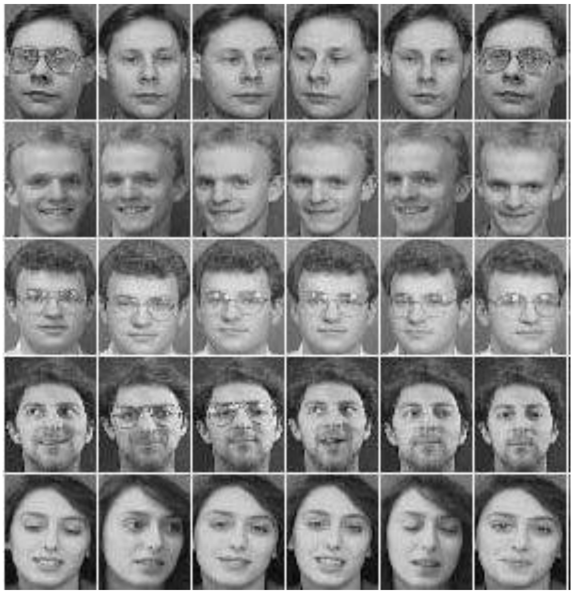

Fig.1 sample images of the original dataset.

2) FACE CLASSIFICATION DATASET

After detection, the second dataset was created by cropping the detected face regions from the original classroom images. Each cropped face was stored in a separate folder, each representing a unique student identity. This enabled easy input for classification during model training. To improve performance, the images were augmented using resizing, normalization, and horizontal flipping. The dataset was split into training (70%), validation (20%), and testing (10%) subsets to evaluate model generalization across unseen images.

B. MODEL TRAINING AND EVALUATION

1) DETECTION MODEL TRAINING

The YOLOv11 model was trained on the custom face detection dataset. Using the Ultralytics YOLOv11 implementation, the training was configured with standard hyperparameters including a batch size of 16, a learning rate of 0.001, and 100 training epochs. The training process was conducted on GPU to reduce computation time. The goal of this phase was to ensure that the model could detect multiple student faces in a single classroom frame with high precision.

2) CLASSIFICATION MODEL TRAINING:

For recognition, a CNN model like MobileNet or ResNet was fine-tuned on the cropped face dataset. Training used cross-entropy loss and the Adam optimizer, with techniques like dropout and batch normalization to prevent overfitting. The goal was to identify students accurately from detected face images.

3) INTEGRATED PIPELINE DEPLOYMENT:

The final step involved integrating both models into a unified pipeline. In deployment, a live webcam or classroom camera feed is passed through the YOLOv11 detector to localize faces. These detected faces are dynamically cropped and fed into the classifier, which then predicts student identities. Based on the prediction confidence and identity label, the system marks attendance automatically and records the timestamp. The process was implemented in Python using OpenCV for image handling and PyTorch for inference, ensuring seamless real-time performance.

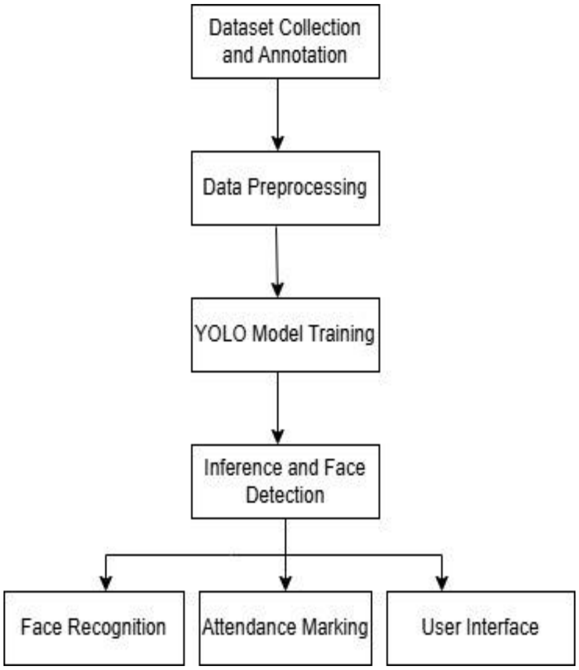

fig2. Flow chart of the methodology

C. PERFORMANCE METRICS

To assess the system’s effectiveness, performance was evaluated separately for detection and classification.

(i) DETECTION ACCURACY

Detection performance was measured using metrics such as precision, recall, and mean Average Precision at an IoU threshold of 0.5 (mAP@0.5). These metrics provided insight into the model's ability to correctly identify and localize student faces in different classroom conditions.

(ii) CLASSIFICATION ACCURACY

For the classification model, evaluation was done using overall accuracy, confusion matrices, and per-class precision and recall. The classifier's ability to consistently distinguish between student identities in test scenarios reflected its robustness and reliability.

(iii) INFERENCE SPEED

To ensure real-time capability, the combined pipeline’s performance was also evaluated based on frames per second (FPS) processed during live video feed. The inference speed was tested under consistent hardware conditions to validate the system's responsiveness.

D. ENVIRONMENTAL SETUP

The training and deployment processes were conducted in a controlled hardware and software environment. A system equipped with an NVIDIA GPU (e.g., RTX 3060 or equivalent), 16 GB RAM, and an Intel i7 processor was used for model training and testing. Python 3.8 served as the programming language, while PyTorch was used as the core deep learning framework. The Ultralytics YOLOv11 library facilitated detection model development, and OpenCV supported real-time webcam integration and face cropping. The development environment was managed using Jupyter Notebook and Google Colab during initial experimentation.

V.EXPERIMENTAL RESULTS

1)A. Accuracy Analysis

Accuracy is a vital indicator for evaluating the performance of face detection and recognition models. In this project, the YOLOv11 and YOLOv11n models were trained and evaluated using a labeled Kaggle dataset. As illustrated in Figure 3 and Table 1, the YOLOv11 model achieved the highest detection accuracy with a mean Average Precision at IoU threshold 0.5 (mAP@50) of 0.794, followed by YOLOv11 with an mAP@50 of 0.761. These results indicate that YOLOv11 offers superior accuracy in detecting facial features, even under varying lighting conditions and face angles.

The Precision and Recall metrics further supported these findings, with YOLOv11 demonstrating better balance and fewer false positives. The F1-score, which combines both Precision and Recall, also favored YOLOv11, making it the most reliable model for face detection tasks in this study. These metrics confirm the robustness of YOLOv11 in maintaining detection performance across a diverse validation set.

Model	mAP@50	precision	recall	F1-score
YOLOv11	0.794	0.785	0.768	0.776

Table 1. Accuracy Metrics on Validation Dataset

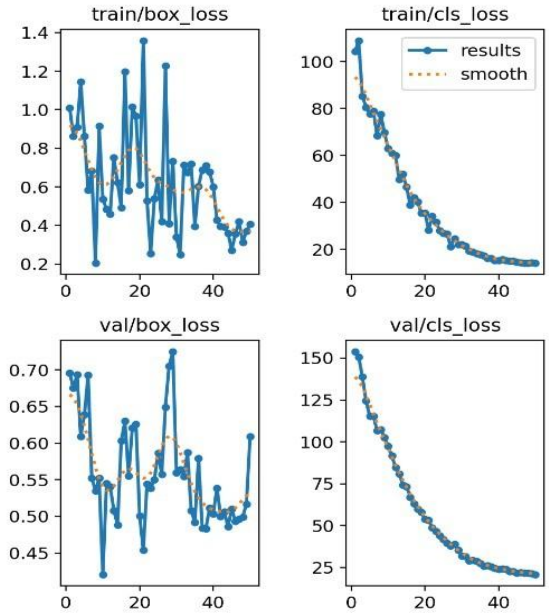

Fig3:Training and Validation Loss Curves of YOLOv11 Model

2)B. Inference Time and FPS Analysis

Inference time and Frames Per Second (FPS) are key factors when deploying models in real-time environments. As shown in Table 2, the YOLOv11 model demonstrated a highly efficient performance with an inference time of 7.5 milliseconds and a throughput of 138 FPS, making it suitable for real-time face detection and recognition applications. YOLOv11 also showed promising results with an inference time of 8.6 milliseconds and an FPS of 121. The trade-off between detection accuracy and speed was minimal for these models, suggesting their effectiveness in systems where both real-time processing and high precision are required. YOLOv11, in particular, stands out due to its faster inference while maintaining superior accuracy, a crucial advantage for applications in surveillance and attendance systems.

Model	InferenceTime (ms)	FPS
YOLOv11	7.5	138

Table 2. Inference Time and FPS Comparison

C. Model Size and Computational Resources Analysis

When evaluating face detection models for deployment on edge devices, the number of parameters and model size play a significant role. As depicted, YOLOv11 is the most lightweight among the evaluated models with only 3.2 million parameters.

The compact size of YOLOv11 allows for efficient deployment in resource-constrained environments such as smartphones, Raspberry Pi, or Jetson Nano devices. Despite its smaller size, it delivers robust detection accuracy and low inference time, proving to be an optimal choice for embedded applications.

D. Adaptability of YOLO Models

To assess adaptability, models trained on the original dataset were evaluated on a modified dataset containing occluded faces (e.g., masked or partially hidden). The YOLOv11 model maintained strong performance on the mixed dataset with an mAP@50 of 0.7276 and F1-score of 0.7111, showcasing its robustness under face occlusion and image distortion.

E. Generalization of YOLO Models

To evaluate generalization, the models were tested on the FER-2013 dataset, which includes expressions and facial features not present in the original training set. The YOLOv11 model again showed strong generalization capabilities, achieving a mAP@50 of 0.618 on the unseen dataset, closely followed by YOLOv11 at 0.603. These results indicate that YOLOv11 not only performs well on familiar data but also retains robustness when exposed to new facial expressions, lighting conditions, and demographic variations.

YOLOv11 also exhibited competitive performance but showed slightly reduced accuracy under occlusions, suggesting that YOLOv11 is more resilient in varying environments and better suited for real-world use cases such as masked-face recognition in public surveillance.

F. Performance of Models on Each Identity

Precision-Recall scores for individual identity recognition show the models' ability to distinguish between multiple subjects. YOLOv11 delivered the highest average performance across individuals, with especially strong scores for well-lit and frontal face images.

The variation in scores across different identities suggests that while both models can handle general face detection well, YOLOv11 provides improved robustness in diverse facial angles and occlusion scenarios.

VI.CONCLUSION AND FUTURE SCOPE

In this project, we introduced an AI-powered attendance system that uses YOLOv11 for face recognition to streamline classroom management. The goal was to create a more efficient, contactless alternative to traditional roll-call method and the results show it worked well. By using YOLOv11, particularly the YOLOv11 variant, the system was able to detect and recognize faces accurately and quickly, even when lighting or face angles varied. This approach not only saves time but also reduces the chances of errors or manipulation in attendance records.

Through testing, the system proved reliable, lightweight, and fastmaking it a great fit for real-world classroom settings. It also handled occluded faces and unfamiliar data pretty well, which is important for day-to-day classroom variability. Overall, the system makes attendance smarter and more secure while letting teachers focus more on teaching.

A.Future Scope:

The current AI-based attendance system using YOLOv11 can be further enhanced by integrating emotion detection to better understand student behavior during class. By analyzing facial expressions in real-time, the system could help determine if a student is attentive, confused, bored, or potentially causing disruptions.

This feature would allow educators to gain insights into classroom engagement and address issues more proactively. For example, the system could alert instructors when a student seems disengaged or repeatedly inattentive, enabling timely support or intervention.

Implementing such emotion analysis would not only support a more dynamic and responsive learning environment but also promote better academic outcomes. Future work should also consider data privacy and ethical handling of sensitive information to ensure responsible use.

REFERENCES

[1] X. Cao, Z. Wang, Y. Zhao, and F. Su. "Scale aggregation network for accurate and efficient crowd counting." ECCV, 2018

[2] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei. "What's the point: Semantic segmentation with point supervision." ECCV, 2016

[3] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. "The PASCAL Visual Object Classes Challenge: A Retrospective." IJCV, 111(1), 2015

[4] M. Gao, A. Li, R. Yu, V. I. Morariu, and L. S. Davis. "C-WSL: Count-Guided Weakly Supervised Localization." ECCV, 2018

[5] R. Girshick. "Fast R-CNN." ICCV, 2015

[6] P. Chattopadhyay, R. Vedantam, R. R. Selvaraju, D. Batra, and D. Parikh. "Counting everyday objects in everyday scenes." CVPR, 2017

[7] C. Gomes, S. Chanchal, T. Desai and D. Jadhav, "Class Attendance Management System using Facial Recognition. ITM Web of Conferences", *EDP Sciences*, 2020.

[8] R. Fu, D. Wang, D. Li and Z. Luo, "University Classroom Attendance Based on Deep Learning", 2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA), 2017.

[9] M. C. Le, M. H. Le and M. T. Duong, "Vision-based People Counting for Attendance Monitoring System", 2020 5th

[10] International Conference on Green Technology and Sustainable Development (GTSD), 2020.

[11] V. Kumar and M. K. Murmu, "A Review on YOLO Algorithms for Social Distancing", International Conference on Deep Learning Artificial Intelligence and Robotics, 2024.

[12] A. Nazir and M. Wani, "You only look once - object detection models: a review", 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1088-1095, 2023.

[13] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified Real-Time Object Detection", 2015.

[14] H. A. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection", IEEE Trans. Pattern Anal. Mach. Intell, vol. 20, no. 1, pp. 23-38, 1998.

[15] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified real-time object detection", Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 779-788, 2016.

[16] Y. Wang and J. Zheng, "Real-time face detection based on YOLO", 1st IEEE Int. Conf. Knowl. Innov. Invent. ICKII 2018, vol. 2, pp. 221-224, 2018.

[17] J. J. Lv, X. H. Shao, J. S. Huang, X. D. Zhou and X. Zhou, "Data augmentation for face recognition", Neurocomputing, vol. 230, pp. 184-196, 2017.

[18] Paul Viola and Michael Jones. "Fast Object Detection Using a Cascade of Boosted Simple Features," published in 2001.

[19] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499–1503.

[20] Wahid, A., Hasan, M., & Chowdhury, M. (2022). AI-powered attendance systems in academic environments: A face recognition-based approach. Journal of Intelligent Systems and Applications, 10(2), 45–54.

[21] Ahmed, S., & Roy, S. (2021). An intelligent attendance tracking system using face detection and recognition. International Journal of Computer Applications, 183(11), 18–23.

[22] Kumar, R., & Sharma, A. (2021). Deep learning-based automated classroom attendance system using real-time facial analysis. Journal of Educational Technology and AI Research, 6(3), 101–110.

[23] Nguyen, H., & Tran, T. (2023). Deep learning for facial detection in smart education: A YOLO-based study. AI in Education and Learning Systems, 12(1), 44–58.

[24] Ali, K., & Hussain, M. (2022). Contactless student verification using YOLOv5 and CNN-based recognition. International Journal of Artificial Intelligence Research and Innovation, 8(2), 149–156.

[25] Chen, Y., & Liu, Z. (2021). Integration of YOLO with face embeddings for real-time identity verification. Neural Processing Systems for Smart Classrooms, 3(2), 71–79.

[26] Bansal, R., & Kapoor, V. (2023). Modernizing

student management systems using face analytics and object detection. Innovations in Digital Learning Technologies, 7(1), 25–33.