A
Major Project Report
On

"An AI-Based Student Tracking System for In-Depth Analysis of Student behavior"

Submitted in partial fulfillment of the

Requirements for the award of the degree of

Bachelor of Technology

In

Computer Science & Engineering –

Artificial Intelligence & Machine Learning

By

N. Shiva Kumar	– 20R21A6643
Vundela Vamsi	– 20R21A6657
Munigela Raviteja	– 20R21A6640
G. Abhinav Goud	– 20R21A6618

Under the guidance of
Mr. J Vijay Gopal
Associate Professor

Department of Computer Science & Engineering –
Artificial Intelligence & Machine Learning

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS
MLR INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
Affiliated to JNTUH, Approved by AICTE
Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

2024

INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
Affiliated to JNTUH,Approved by AICTE
Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

Department of Computer Science & Engineering-
Artificial Intelligence & Machine Learning

CERTIFICATE

This is to certify that the project entitled **"An AI-Based Student Tracking System for In-Depth Analysis of Student behavior"** has been submitted by **N Shiva Kumar(20R21A6643)**, **Vundela Vamsi (20R21A6657), Munigela Raviteja(20R21A6640), G Abhinav Goud(20R21A6618)** in partial fulfilment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad. The results embodied in this project have not been submitted to any other University or Institution for the award of any degree or diploma.

Internal Guide **Head of the Department**

Project coordinator **External Examiner**

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS

MLR INSTITUTE OF TECHNOLOGY

(UGC AUTONOMOUS)

Affiliated to JNTUH,Approved by AICTE

Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

NATIONAL BOARD of ACCREDITATION

NAAC A

Department of Computer Science & Engineering-
Artificial Intelligence & Machine Learning

DECLARATION

We hereby declare that the project entitled **"An AI-Based Student Tracking System for In-Depth Analysis of Student behavior"** is the work done during the period from **January 2024 to May 2024** and is submitted in partial fulfilment of the requirements for the award of degree of Bachelor of Technology in Computer Science & Engineering – Artificial Intelligence & Machine Learning from Jawaharlal Nehru Technology University, Hyderabad. The results embodied in this project have not been submitted to any other university or Institution for the award of any degree or diploma.

N. Shiva Kumar	–	**20R21A6643**
Vundela Vamsi	–	**20R21A6657**
Munigela Raviteja	–	**20R21A6640**
G. Abhinav Goud	–	**20R21A6618**

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS

MLR INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
Affiliated to JNTUH, Approved by AICTE
Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

NATIONAL BOARD OF ACCREDITATION

NAAC A

Department of Computer Science & Engineering-
Artificial Intelligence & Machine Learning

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that we now have the opportunity to express our guidance for all of them.

First of all, we would like to express our deep gratitude towards our internal guide **Mr. J VIJAY GOPAL,** Assistant Professor, Department of CSE -AIML for his support in the completion of our dissertation. We wish to express our sincere thanks to **Dr. K. SAI PRASAD,** HOD, Dept. of CSE-AIML and principal **Dr. K. SRINIVAS RAO** for providing the facilities to complete the dissertation.

We would like to thank all our faculty and friends for their help and constructive criticism during the project period. Finally, we are very much indebted to our parents for their moral support and encouragement to achieve goals.

N. Shiva Kumar – 20R21A6643
Vundela Vamsi – 20R21A6657
Munigela Raviteja – 20R21A6640
G. Abhinav Goud – 20R21A6618

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS

MLR INSTITUTE OF TECHNOLOGY

(UGC AUTONOMOUS)

Affiliated to JNTUH,Approved by AICTE

Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

Department of Computer Science & Engineering-
Artificial Intelligence & Machine Learning

ABSTRACT

The current trend of offline education has created difficulties in monitoring and supervising student behavior, resulting in disturbances and diversions during offline sessions. While several algorithms recognize these behaviors, their accuracy and efficiency are limited and mostly focused on single items. To solve this, we present a new AI-powered Student Tracking System that will revolutionize behavior analysis, attendance management, and incident detection in educational institutions with real-time monitoring. The system uses artificial intelligence (AI) methods, including Convolutional Neural Networks (CNNs), OpenCV, Face Recognition module, and YOLOv8 to record real-time insights regarding student behaviors such as napping in class, using mobile phones, and engaging in irregular activities which are also incorporated with real-time alerts using Twilio. The model is trained on the YOLOv8 algorithm, which is capable of capturing real-time video with faster Frames per Second(FPS). The system's key advantages include its capacity to accurately identify and record problems, particularly those involving property damage, allowing for timely actions and fostering responsibility. The user-friendly interface acts as a single center for instructors and administrators, providing real-time information on student conduct, attendance, and reported occurrences. The system evolves through continuous development, ensuring that it remains successful in addressing changing educational settings and dynamic student behaviors.

APPENDIX-1

LIST OF FIGURES

LIST OF FIGURES

APPENDIX-2

LIST OF TABLES

LIST OF TABLES

APPENDIX-3

LIST OF ABBREVIATIONS

ABBREVIATIONS

AI	**Artificial Intelligence**
YOLO	**You Only Look Once**
OpenCV	**Open Computer Vision**
FPS	**Frames per Second**
GPU	**Graphics Processing Unit**
CPU	**Central Processing Unit**
CCTV	**Closed Circuit Television**
CV	**Computer Vision**
iHMM	**Infinite Hidden Markov Model**
MCMC	**Markov Chain Monte Carlo**
LSTM	**Long Short Term Memory**
ISA	**Invariant Subspaces Analysis**
VB	**Variational Bayes**
EM	**Expectation Maximization**
SVM	**Support Vector Machines**

APPENDIX-4

REFERENCES

References

[1] O. Boiman and M. Irani, ”Detecting irregularities in images and in video,” Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1, Beijing, China, 2005, pp. 462-469 Vol. 1, doi: 10.1109/ICCV.2005.70.

[2] I. Pruteanu-Malinici and L. Carin, ”Infinite Hidden Markov Models and ISA Features for Unusual-Event Detection in Video,” 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 2007, pp. V - 137-V - 140, doi: 10.1109/ICIP.2007.4379784.

[3] F. Jiang, Y. Wu and A. K. Katsaggelos, ”Abnormal Event Detection from Surveillance Video by Dynamic Hierarchical Clustering,” 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 2007, pp. V - 145-V - 148, doi: 10.1109/ICIP.2007.4379786.

[4] I. Pruteanu-Malinici and L. Carin, ”Infinite Hidden Markov Models for Unusual-Event Detection in Video,” in IEEE Transactions on Image Processing, vol. 17, no. 5, pp. 811-822, May 2008, doi: 10.1109/TIP.2008.919359.

[5] M. Jager, C. Knoll and F. A. Hamprecht, ”Weakly Supervised Learning of a Classifier for Unusual Event Detection,” in IEEE Transactions on Image Processing, vol. 17, no. 9, pp. 1700-1708, Sept. 2008, doi: 10.1109/TIP.2008.2001043.

[6] T. Xiang and S. Gong, ”Video Behavior Profiling for Anomaly Detection,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 5, pp. 893-908, May 2008, doi: 10.1109/TPAMI.2007.70731.

[7] Y. -L. Hsueh, N. -H. Lin, C. -C. Chang, O. T. . -C. Chen and W. -N. Lie, ”Abnormal event detection using Bayesian networks at a smart home,” 2015 8th International Conference on Ubi-Media Computing (UMEDIA), Colombo, Sri Lanka, 2015, pp. 273-277, doi:

[8] K. Vignesh, G. Yadav and A. Sethi, ”Abnormal Event Detection on BMTT-PETS 2017 Surveillance Challenge,” 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, pp. 2161-2168, doi: 10.1109/CVPRW.2017.268.

[9] C. V. Amrutha, C. Jyotsna and J. Amudha, ”Deep Learning Approach for Suspicious Activity Detection from Surveillance Video,” 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 335-339, doi: 10.1109/ICIMIA48430.2020.9074920.

[10] Du, J. (2018, April). Understanding of object detection based on CNN family and YOLO. In Journal of Physics: Conference Series (Vol. 1004, p. 012029). IOP Publishing.

[11] X. Sun, S. Zhu, S. Wu and X. Jing, ”Weak Supervised Learning Based Abnormal Behavior Detection,” 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 2018, pp. 1580-1585, doi: 10.1109/ICPR.2018.8545345

[12] Z. Kain, A. Y. Ouness, I. E. Sayad, S. Abdul-Nabi and H. Kassem, ”Detecting Abnormal Events in University Areas,” 2018 International Conference on Computer and Applications (ICCA), Beirut, Lebanon, 2018, pp. 260-264, doi: 10.1109/COMAPP.2018.8460336.

[13] D. O. Esan, P. A. Owolawi and C. Tu, ”Detection of Anomalous Behavioural Patterns In University Environment Using CNN-LSTM,” 2020 IEEE 23rd International Conference on Information Fusion (FUSION), Rustenburg, South Africa, 2020, pp. 1-8, doi: 10.23919/FUSION45008.2020.9190406.

[14] R. Zheng, F. Jiang and R. Shen, ”Intelligent Student Behavior Analysis System for Real Classrooms,” ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 9244-9248, doi: 10.1109/ICASSP40776.2020.9053457.

[15] E. Elbasi, ”Reliable abnormal event detection from IoT surveillance systems,” 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), Paris, France, 2020, pp. 1-5, doi: 10.1109/IOTSMS52051.2020.9340162.

[16] H. S, J. D. Pushparaj and M. Malarvel, ”Computer Vision based Student Behavioral Tracking and Analysis using Deep Learning,” 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 772-777, doi: 10.1109/ICESC54411.2022.9885410.

[17] L. Li, M. Liu, L. Sun, Y. Li and N. Li, ”ET-YOLOv5s: Toward Deep Identification of Students’ in-Class Behaviors,” in IEEE Access, vol. 10, pp. 44200-44211, 2022, doi: 10.1109/ACCESS.2022.3169586.

[18] A. Tripathi, M. K. Gupta, C. Srivastava, P. Dixit and S. K. Pandey, ”Object Detection using YOLO: A Survey,” 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 747-752, doi: 10.1109/IC3I56241.2022.10073281.

[19] Y. Cui and H. Zou, ”Detection of Abnormal Behavioral States in Student Learning Based on Video Surveillance,” 2023 International Conference on Data Science and Network

Security (ICDSNS), Tiptur, India, 2023, pp. 01-06, doi: 10.1109/ICDSNS58469.2023.10245238.

[20] Trabelsi, Z.; Alnajjar, F.; Parambil, M.M.A.; Gochoo, M.; Ali, L. RealTime Attention Monitoring System for Classroom: A Deep Learning Approach for Student's Behavior Recognition. Big Data Cogn. Comput. 2023, 7, 48. https://doi.org/10.3390/bdcc7010048.

INDEX

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

The AI-Based Student Tracking System represents a ground breaking solution for educational institutions seeking to enhance behaviour analysis, attendance management, and incident detection. Leveraging advanced AI algorithms and hardware integration, the system provides real-time insights into student behaviours, streamlines attendance tracking processes, and ensures a safe and conducive learning environment. By combining cutting-edge technologies such as OpenCV and object detection with CNN, the system can accurately identify and categorize various student behaviours, including sleeping in class, using mobile phones, and engaging in disruptive activities. Additionally, the system promptly detects incidents such as property damage, enabling swift interventions to maintain campus safety. With its user-friendly interface and adaptability to changing educational needs, the AI-Based Student Tracking System offers educators and administrators a comprehensive solution to effectively monitor student behaviour and enhance overall educational experiences.

1.2 PURPOSE OF THE PROJECT

The primary purpose of the AI-Based Student Tracking System is to revolutionize behavior analysis, attendance management, and incident detection within educational institutions. By providing real-time insights into student behaviors and automating attendance tracking processes, the system aims to create a conducive learning environment that fosters student engagement and academic success. Additionally, the system seeks to enhance campus safety by promptly identifying and addressing incidents such as property damage. Overall, the project aims to leverage advanced AI technologies to optimize educational processes, improve student outcomes, and support educators in their efforts to create an enriching learning environment.

1.3 MOTIVATION

The motivation behind the AI-Based Student Tracking System stems from the growing need for effective tools to monitor student behavior, manage attendance, and ensure campus safety in educational institutions. With the increasing adoption of online learning and the challenges posed by shifting student behaviors, there is a pressing need for innovative solutions that can adapt to these changes and provide actionable insights for educators and administrators. The project is motivated by a desire to harness the potential of AI technologies to address these challenges comprehensively and empower educational institutions to create positive learning environments where students can thrive academically and personally. Additionally, the project is driven by a commitment to leveraging technology for social good and supporting the holistic development of students within educational settings.

CHAPTER 2

LITERATURE SURVEY

An extensive literature survey has been conducted by studying existing systems of Certificate verification and generation. A good number of research papers, journals, and publications have also been referred before formulating this survey.

2.1 EXISTING SYSTEM

In traditional educational environments, all the techniques of attendance tracking and behavior monitoring encounter substantial obstacles in maintaining an appropriate learning environment in physical classrooms. The current methods fail to effectively manage the changing environment of student behavior, which includes concerns such as students dozing off in class, using mobile phones during lectures, and engaging in disruptive activities. These issues need a more advanced and thorough approach to student monitoring to create an environment that promotes successful learning.

1		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/abstract/document/9074920	C.V Amrutha; C. Jyotsna; J. Amudha	Video Surveillance, Artificial Intelligence ,Machine Learning ,Deep Learning,Suspicious Behavior,Human Behavior,Academic Environment,Alert Message,Monitoring
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**

Deep Learning Approach for Suspicious Activity Detection from Surveillance Video	aims to detect suspicious or normal activity in an academic environment using deep learning techniques ,The problem that this solution aims to solve is the difficulty in differentiating between suspicious and normal human behavior in video surveillance footage	In the first part, features are computed from video frames. In the second part, based on the obtained features, a classifier predicts the class as suspicious or normal.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	The paper proposes a deep learning approach to detect suspicious or normal activity in an academic environment using video surveillance	The proposed system can detect suspicious activities in real-time and send an alert message to the corresponding authority.	It is important to note that deep learning models require a large amount of data for training and may not generalize well to new data.
2	The entire framework is divided into two parts. In the first part, the features are computed from video frames second part, based on the obtained features, the classifier predicts the class as suspicious or normal	It can differentiate various suspicious behaviors from live tracking of footages.	Additionally, the accuracy of such models may be affected by factors such as lighting conditions, camera angles, and occlusions

Major Impact Factors in this Work

the paper proposes a deep learning approach to detect suspicious or normal activity in an academic environment using video surveillance. The system sends an alert message to the

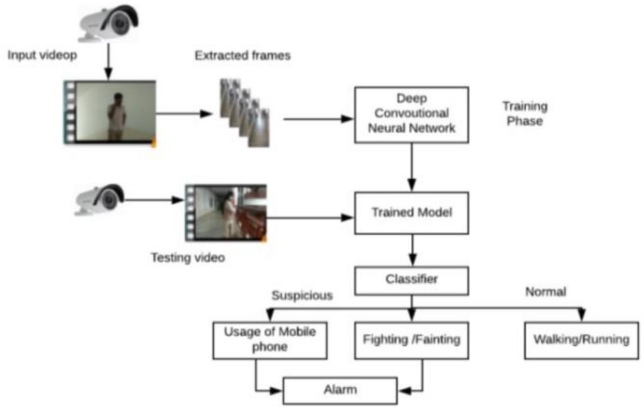
corresponding authority in case of predicting a suspicious activity

Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
Class of activity predicted (suspicious or normal) based on computed features	Features computed from video frames	Lighting Conditions Camera Quality	NA

Relationship Among the Above 4 Variables in This article			
Input and Output		**Feature of This Solution**	**Contribution & The Value of This Work**
Input	**Output**		
In the first part, features are computed from video frames.	based on the obtained features, a classifier predicts the class as suspicious or normal	To improve the secqurity for the community so that everyone can be safe and the rate of crimes would be decreased	To improve the secqurity for the community so that everyone can be safe and the rate of crimes would be decreased

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
The anamoly activities are increasing day by day and the crime ration is increasing day by day , This can help prevent untoward incidents and ensure safety in academic environments	There is a possibility that some activities may be misclassified as suspicious, leading to false alarms. This could cause unnecessary panic and confusion among students and faculty members

Analyse This Work by Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
There is a possibility that some activities may be misclassified as suspicious, leading to false alarms.	Anamoly detection	I. Introduction II. Literature Survey

This could cause unnecessary panic and confusion among students and faculty members		III. System Overview IV. Conclusion and future work V. References
Diagram/Flowchart		
		

---End of Paper 1---

2

Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/document/9885410	Hariharan S; J Daniel Pushparaj; Muthukumaran Malarvel	Covid-19,Global Pandemic,Virtual World,Video Conferencing Applications, Student Engagement ,Interaction,Face Identification,Face Spoofing ,CNN (Convolutional Neural Network),Gaze Tracking Head Positioning,Eye Tracking

The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	The Goal (Objective) of this Solution & What is the problem that needs to be solved	What are the components of it?
Computer Vision based Student Behavioral Tracking and Analysis using Deep Learning	The goal of this solution is to track student behavior in real-time virtual video conferencing applications using computer-vision and deep-learning. The problem that needs to be solved is the difficulty of monitoring many students at the same time by a single invigilator /faculty, thereby causing a drop in student engagement and interaction	Face identification for face spoofing using CNN Students' gaze tracking with head positioning & eye tracking Real-time virtual video conferencing applications Computer-vision & deep-learning Single-Shot-Multibox detector with ResNet-10 Architecture as the foundation Pre-trained Histogram Oriented Gradients (HOG) and linear SVM object detectors for facial landmarks

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

The proposed system uses a modular approach to track student behavior in real-time virtual video conferencing applications using computer-vision and deep-learning. The system uses a single-Shot-Multibox detector with ResNet-10 Architecture as the foundation to recognize and detect the student's face. The system then takes facial landmarks using pre-trained Histogram Oriented Gradients (HOG) and linear SVM object detectors.

	Process Steps	Advantage	Disadvantage (Limitation)
1	Face identification for face	The advantage of this	The disadvantage of this

	spoofing using CNN	system is that it can track student behavior in real-time virtual video conferencing applications using computer-vision and deep-learning.	system is that it requires a high-performance computer with a dedicated GPU to run efficiently
2	Students' gaze tracking with head positioning & eye tracking	This can help invigilators/faculty monitor many students at the same time, thereby increasing student engagement and interaction .	it may introduce bias or errors if not done carefully.

Major Impact Factors in this Work

Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
Facial Recognition Accuracy: How accurately the system identifies and verifies student faces. Gaze Tracking Precision: The system's ability to accurately track students' gaze and head positioning.	GPU Performance CPU and Memory Quality of Video Feed	Lighting Conditions Camera Quality Internet Connection Stabilit Eyewear or Accessories	NA

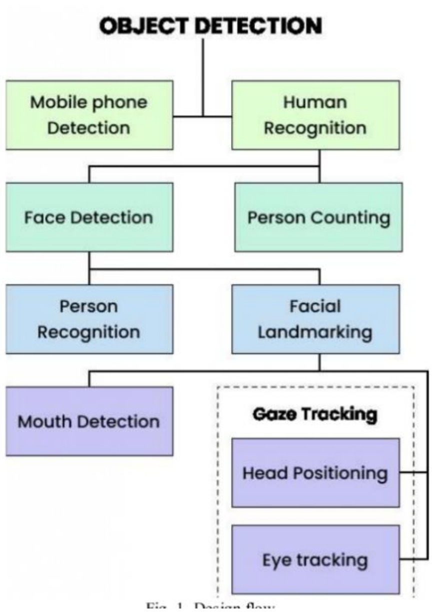
Relationship Among The Above 4 Variables in This article

the relationship among mediating (intervening) variables, moderating variables, dependent variables, and independent variables. The study focuses on optimizing the multi-modal image fusion architecture for medical image segmentation, with the segmentation accuracy as the dependent variable and the multi-modal image fusion architecture as the independent variable. The study does not examine the underlying mechanisms or processes that may mediate or moderate the relationship between the input images and the segmentation output.

<table>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>cameras and video footages</td><td>Track the students behavious accordingly</td></tr></table></td><td>The system uses a single-Shot-Multibox detector with ResNet-10 Architecture as the foundation to recognize and detect students' faces. Then, pre-trained Histogram Oriented Gradients (HOG) and linear SVM object detectors are used to take facial landmarks. With the captured features, the eyes are detected using dlib and the centers of eyeballs are extracted using OpenCV eye classifier</td><td>The proposed system provides a comprehensive solution for real-time student behavioral tracking in virtual video conferencing applications. It can be used by educators to monitor student engagement, attention, and participation during online classes</td></tr>
</table>

Positive Impact of this Solution in This Project Domain	Negative Impact of this Solution in This Project Domain
The proposed system provides a comprehensive solution for real-time student behavioral tracking in virtual video conferencing applications,"It can be used by educators to monitor student engagement, attention, and participation	The system may raise privacy concerns among students as it involves real time monitering of their behaviour.

during online classes".		
Analyse This Work By Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
Logically this is a good step that it analyses every student bahaviour an also tracks student in virtual video conferencing applications.	single-Shot-Multibox detector YOLOv2 dlib	1. Abstract 2. Introduction 3. Related Work 4. Experiment Results 5. Conclusion

Diagram/Flowchart


---End of Paper 2---

3

Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**

https://ieeexplore.ieee.org/abstract/document/9190406	Dorcas Oladayo Esan; Pius. A. Owolawi; Chuling Tu	Surveillance Systems,Anomalous Behavioral Patterns,False Detection (False Alarm) Errors,Performance,Semi-Supervised Techniques,Supervised Techniques,Accuracy Minimizing False Detection Errors,Crowded Environments,Convolutional Neural Network (CNN)
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
The preferred method for detecting abnormal behavior in a university context involves using a CNN-LSTM model. This model merges CNN for image feature extraction and LSTM for recognizing sequential patterns. CNN handles image features from frame sequences, while LSTM recalls vital information for detecting consecutive patterns.	The objective of the proposed solution is to detect anomalous behavioral patterns in a university environment using the Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) model. The solution aims to improve the accuracy of anomaly detection and minimize false detection errors in crowded environments. The main problem that	The paper addresses rising crime rates and the necessity for surveillance systems to identify unusual behaviors, emphasizing the challenge of false detections. It reviews prior works using various techniques like anomalies, biometrics, and motion patterns for anomaly detection. Introducing a CNN-LSTM system, it details its framework stages—image acquisition, preprocessing, and deep learning. The experimental setup outlines platforms, libraries, and model hyperparameters. Results showcase the CNN-LSTM's

With an accuracy of 86%, this model outperforms other detection methods, making it an effective tool for identifying unusual behavior in a university setting.	needs to be solved is the inaccurate detection of anomalous behavioral patterns in surveillance systems, which leads to false detection errors.	efficacy, comparing it to other methods, and emphasizing system accuracy. The conclusion underscores the CNN-LSTM's effectiveness while acknowledging the need for enhanced accuracy and reduced false positives in crowded scenarios.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

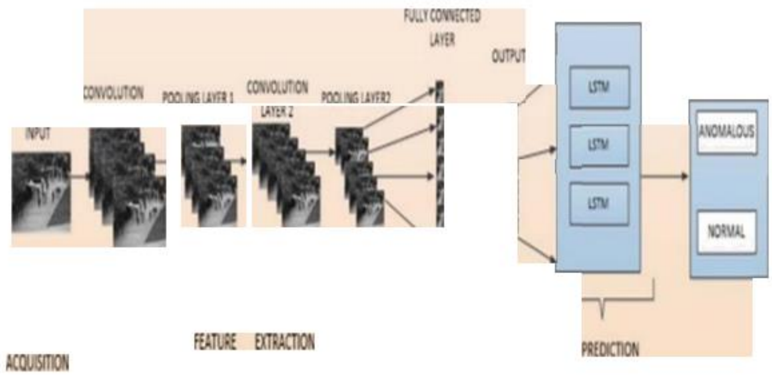
	Process Steps	**Advantage**	**Disadvantage (Limitation)**
1	Use a convolutional neural network (CNN) to extract image features from image frame sequences.	The proposed system can detect anomalous behavioral patterns in university environments.	The proposed system requires a large amount of labeled data for training.
2	Use a long short-term memory (LSTM) network to keep vital information for remembrance.	The system can help prevent crimes and ensure the safety of students and staff.	The high computational cost of training the model is a limitation of the proposed system.
3	Train the CNN-LSTM model on the University of California San Diego dataset, which contains various activities such as walking, running, and fighting.	The proposed system outperforms existing detection models with 86% accuracy.	disadvantage is that the results may be specific to the dataset and may not generalize well to other scenarios.

Major Impact Factors in this Work			
Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
Behavior prediction (normal or anomalous) based on the extracted image features using CNN-LSTM	Image frames input for CNN-LSTM model feature extraction	NA	NA

Relationship Among The Above 4 Variables in This article		
Input and Output	**Feature of This Solution**	**Contribution & The Value of This Work**
Input: The input for the proposed CNN-LSTM model is acquired image frames. These image frames are passed into a convolutional layer.. **Output:** The output of the CNN-LSTM model is the prediction of the behavior from the image frames. It classifies the behavior as either normal or anomalous.	CNN is used to extract features from image frames, while LSTM utilizes a gating mechanism to selectively retain important information for a longer period. The proposed solution combines the Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM) to improve the accuracy of surveillance systems in detecting anomalous behavioral patterns.	Contribution of the Work: The paper presents a performance analysis of the Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) in surveillance systems. It focuses on the detection of anomalous behavioral patterns in a university environment to assist security personnel in preventing hazardous activities such as crime. The proposed system outperforms existing detection models with an accuracy of 86%. Value of the Work:

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**

Analyse This Work By Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
The given document discusses a research paper on the detection of anomalous behavioral patterns in a university environment using the CNN-LSTM technique. The paper aims to address the challenge of inaccurate detection of anomalous behavior in surveillance systems, which can lead to false detection errors. The proposed system utilizes deep learning techniques to improve accuracy and minimize false detection errors in crowded environments.	Histogram-based methods Convolutional neural network with long short-term memory (CNN-LSTM) Mixture of probabilistic principal analysis	I. Abstract II. Introduction III. Related Work IV. Training Tasks V. Assumptions VI. Experiments VII. Conclusion and Future work
Diagram/Flowchart		



--End of Paper 3—

4		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://iopscience.iop.org/article/10.1088/1742-6596/1004/1/012029/pdf	Zhe Guo, Xiang Li	Object Detection, CNN, YOLO, mAP, FPS, YOLOV2
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that needs to be solved**	**What are the components of it?**
YOLOv22 is the current solution that is covered in Juan Du12's work, "Understanding of Object Detection Based on CNN Family and YOLO." The YOLO method, which is a	As stated in Juan Du's paper "Understanding of Object Detection Based on CNN Family and YOLO," the objective of the YOLOv2 solution is to increase the effectiveness and precision of object	1.item localization: To do this, bounding boxes must be drawn around each item in a picture or video. 2.Object Classification: In this step, the localized items are categorized into distinct groups.

representation of Convolutional Neural Networks (CNN) for object identification, has been improved upon by YOLOv2. It has a high capacity to represent the entire image and strikes an outstanding balance between speed and accuracy2. Because of this, YOLOv2 is a very effective and efficient method for object identification tasks.	identification jobs. There were two issues that needed to be resolved: 1. Speed: Although precise, traditional object identification techniques like Faster R-CNN were too slow for real-time applications. Their Frame Per Second (FPS) rate was slower than the real-time impact, ranging between 5 and 18. 2. Accuracy: While two-stage detectors like quicker R-CNN were more accurate than single-stage detectors like YOLO, the latter was quicker.	3.Grid Cells: YOLO creates a grid out of the input image and forecasts the class probabilities and bounding boxes for each grid cell. 4.Bounding Boxes: The system predicts many bounding boxes, each with a confidence score indicating the possibility of an item, for each grid cell. 5.Class Probabilities: To determine the object category, class probabilities are assigned to each bounding box. 6.Anchor Boxes: These are pre-made, grid-cell-associated bounding box forms with varying aspect ratios and sizes that are used to forecast the bounding box coordinates in relation to the anchor box shapes.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	backdrop and Core Solution- CNN: An overview of the backdrop and CNN's core solution is given at the outset of the study. It describes how	1. CNN: CNNs process images effectively. They are noise-resistant and attain good accuracy rates. Transfer learning is supported by CNNs,	CNN: CNNs have a substantial computational overhead even though they are effective at analyzing images. Large datasets are also

	image processing techniques might be sufficiently precise and quick, how computers could operate automobiles without the need for specialized sensors, and how assistive gadgets may provide users with real-time scene information.	meaning that after training for one activity, they may be utilized to do another task with little to no additional training. By automating the process of feature extraction, they can identify patterns in photos without requiring manual feature engineering.	necessary for them to attain high accuracy rates. The CNN may overfit, which means it gets overly specialized to the training dataset and performs badly on fresh data, if the dataset is too small. Additionally susceptible to hostile assaults are CNNs.
2	Faster R-CNN: The creation of Faster R-CNN, which attained a 76.41 Mean Average Precision (mAP), is covered in the publication. The Faster R-CNN's Frame Per Second (FPS) is still between 5 and 18, which is slower than the real-time effect. Therefore, speeding up object detection is the most important change that has to be done.	2. Faster R-CNN: By processing the entire picture and all ROIs using a shared convolutional layer as opposed to processing each ROI separately, Faster R-CNN outperforms R-CNN. As a result, object detection requires a lot less computing time. Faster R-CNN generates region suggestions by utilizing attention and CNN architecture based on deep learning, which not only saves a great deal of time but also enhances the accuracy of object recognition and localization in pictures.	Faster R-CNN: By processing the entire picture and all ROIs using a shared convolutional layer as opposed to processing each ROI separately, Faster R-CNN outperforms R-CNN. It still uses the laborious Selective Search technique, nevertheless, to generate region proposals. Additionally, it has trouble identifying objects in unique situations like occlusion, distortion, or small size.

Major Impact Factors in this Work			
Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
CNN for image feature extraction LSTM for sequential pattern recognition	Accuracy of anomaly detection (86%). Reduction of false detection errors	Lighting condition Camera quality Distance from camera Head movements	System latency

Relationship Among the Above 4 Variables in This article

Input and Output	Feature of This Solution	Contribution & The Value of This Work
Input: a picture including one or more items. This can be a picture or a still image from a live feed. The things in the picture might be anywhere in the picture **Output**: One or more bounding boxes, with a point, width, and height assigned to each. Every bounding box matches an object found in the picture. For every	1. Speed: Yolo doesn't deal with complicated pipelines, hence it is really quick. 45 frames per second (FPS) may be processed by it for photos. 2. Detection Accuracy: With relatively few background mistakes, Yolo outperforms other cutting-edge algorithms in terms of accuracy. 3. Generalization: Compared to previous object detection methods, YOLO performs generalized object	1. Real-Time Object identification: Real-time object identification is essential in many applications, including robotics, autonomous driving, and surveillance. It has been made possible by the use of CNN and YOLO models. 2. Increased Accuracy: The accuracy of object detection has increased dramatically with the usage of CNN and YOLO models. YOLO in particular has managed to get a high degree of accuracy with little losses.

and they could be of different sizes.	recognized object2, a class label is further supplied in addition to the bounding box. This label—which may say "car," "person," "dog," etc.—indicates the kind or class of the object.	representation more successfully and with no loss of precision. 4. Open-source: YoLO is an open-source project that allows developers and academics to freely utilize and contribute to its development. 5. Real-time Object Detection: To detect objects in real-time, the YOLO method uses convolutional neural networks (CNN). 6. One-Pass Detection: Yolo is a one-shot detector that processes an image using a fully convolutional neural network (CNN).	3. Versatility: The models can detect items in any given image and remove highlights. Because of this, they are adaptable and useful in a variety of contexts. 4.Overcoming Difficulties: CNN models, when used properly, may tackle problems such as diagnosing deformities, developing instructional or educational applications, etc. 5. Industry help: In the financial and other industries, the research offers industry help for feature extraction and targeted visual information.

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
1. Real-Time Object identification: This is a crucial feature for many applications, such as autonomous vehicles, robotics, and surveillance. CNN and YOLO models have been used to make it feasible. 2. Improved Accuracy: Using CNN and YOLO models has resulted in a significant improvement in object detection accuracy. Specifically, YOLO has achieved a high level of accuracy with little losses.	1.computer Resources: For training and inference, CNN and YOLO models demand a large amount of computer power. This might provide a problem for initiatives with constrained funding. 2. Localization Errors: When compared to other detection methods, Yolo has a tendency to generate more localization errors. This could affect certain apps' object detecting precision. 3. Real-World Challenges: The performance of

3. Versatility: The models are able to eliminate highlights and identify objects in any given picture. They are flexible and helpful in a range of situations as a result.	object identification can be impacted by images in real-world contexts that involve problems like noise, blurring, and rotational jitter..

Analyse This Work By Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
A thorough examination of the developments in object recognition algorithms is given in Juan Du's paper "Understanding of Object Detection Based on CNN Family and YOLO," which focuses in particular on Convolutional Neural Networks (CNN) and the You Only Look Once (YOLO) method. The author skillfully illustrates the speed constraints of Faster R-CNN and how YOLO gets around them. This is supported by statistical evidence, which lends further authority to the study. However, by incorporating more current developments or other well-liked methods in the field, the work may have been more thorough. The results add to continuing research aimed at increasing	1. Deep Learning Frameworks: Convolutional Neural Networks (CNNs) and its variations are frequently implemented and trained using frameworks such as TensorFlow, PyTorch, or Keras. 2. Object identification Algorithms: The article talks about a number of object identification techniques, including YOLO and Faster R-CNN. Deep learning frameworks are commonly utilized for the implementation of these algorithms. 3.Datasets: Training and assessing the effectiveness of object identification models depend on datasets. The COCO (Common Objects in Context), ImageNet, and PASCAL VOC datasets are	I. Introduction: This section of the study starts with outlining the significance of object detection in image processing as well as the difficulties it confronts. It talks about the progress of Convolutional Neural Networks (CNNs) since 2012 and its importance in object detection. II. Context and Primary Resolution CNN: An overview of the history of object detection and CNN's primary solution is given in this section. III.CNN Family and Faster R-CNN: The evolution of the CNN family, including Faster R-CNN, is covered in this work along with its Mean Average Precision (mAP) and Frame Per Second (FPS) capabilities.

the effectiveness and precision of these algorithms and have important ramifications for real-time object identification applications.	frequently utilized. 4.Evaluation Metrics: Two frequently used metrics for assessing the effectiveness of object detection models are Mean Average Precision (mAP) and Frame Per Second (FPS).	IV.YOLO: In this part, YOLO—a member of the CNN family—is introduced. It offers a fresh approach to object identification problems that is both easy to use and effective. It talks about the mAP, speed, and its comparison with Faster R-CNN.

Diagram/Flowchart

$$\mathbf{Pr}(\mathbf{Class_i}|\mathbf{Object}) \times \mathbf{Pr}(\mathbf{Object}) \times \mathbf{IOU}_{\mathbf{pred}}^{\mathbf{truth}} = \mathbf{Pr}\ (\mathbf{Class_i}) \times \mathbf{IOU}_{\mathbf{pred}}^{\mathbf{truth}}$$

---End of Paper 4---

5

Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/abstract/document/4470543	Lawrence Carin (SM'96–F'01) Iulian Pruteanu-Malinici	Background removal, Moving entities, Feature extraction, Infinite hidden Markov model (iHMM), Invariant subspaces, Complex scenes, Background subtraction or removal, Training data, Unusual events

The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	The Goal (Objective) of this Solution & What is the problem that need to be solved	What are the components of it?
1. Shift-Invariant Wavelets (SIWs): According to the publication, one method being evaluated for feature extraction is SIWs. By omitting subsampling, which is frequently carried out in the wavelet transform, SIWs are generated.	Objective of the Solution: The objective of the solution is to develop a detection algorithm for video analysis that can identify unusual events in complex scenes. The proposed algorithm aims to model normal behavior based on a large quantity of available data and detect events that have a low likelihood of being associated with the learned model.	Three feature extraction techniques are discussed in the document: Shift-Invariant Wavelets (SIWs), Independent Component Analysis (ICA), and Invariant Subspace Analysis (ISA). The document states that ISA yields superior performance for the purposes discussed, but the other two techniques are also widely used and warrant discussion and comparison. The document does not provide specific details on the relative merits of these techniques.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	State decomposition: An iHMM, which contains an unlimited number of hidden states, is used in this study to accomplish state decomposition. The	infinite hidden states: Using an iHMM with an infinite number of hidden states gives you greater freedom when modeling intricate scenes and	Computational complexity: Using Bayesian inference methods and an infinitely many hidden state iHMM can be computationally

	transition dynamics are exclusively controlled by the hyperparameters when the infinite parameters are integrated out using Bayesian analysis.	encapsulating the video data's sequential qualities.	costly, particularly when working with big datasets.
2	Bayesian inference: This method produces a posterior distribution on the appropriate number of HMM states, as opposed to choosing a predetermined number of states. This is accomplished by utilizing variational Bayesian inference, which is used in this study for the first time.	Bayesian inference: This method yields a posterior distribution on the appropriate number of HMM states and offers a systematic way to choose a model. This eliminates the need to choose the model structure by hand through trial and error..	Convergence problems: Because of nonconjugacy in the graphical model, the variational Bayesian inference technique employed in this study could need sampling steps. There aren't any convergence guarantees for this technique yet, even if the algorithm converged and performed well in the samples taken into consideration.

Major Impact Factors in this Work

Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
Detection of Unusual Events in Video Data	IHMM (Infinite Hidden Markov Model Bayesian Inference Feature Extraction Methods (SIWs, ICA, ISA	Lighting condition Camera quality	NA

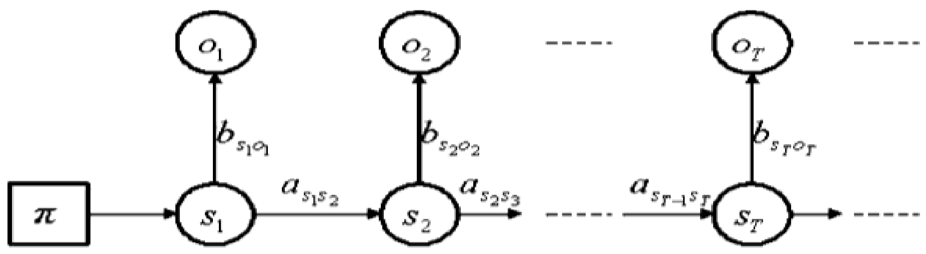
Relationship Among The Above 4 Variables in This article		
Input and Output	**Feature of This Solution**	**Contribution in This Work**
<table><tr><th>Input</th><th>Output</th></tr><tr><td>Details on applying the iHMM (Infinite Hidden Markov Model) to recognize unusual events in video footage are given in the accompanying material.</td><td>It provides information on the feature extraction techniques used, the convergence analysis of the MCMC algorithm, and the protocols for training and testing.</td></tr></table>	There are three feature extraction methods discussed in the document: Shift-Invariant Wavelets (SIWs), Independent Component Analysis (ICA), and Invariant Subspace Analysis (ISA). The document states that ISA yields the most accurate results for the purposes discussed. SIWs can be combined with the à trous algorithm to gain the benefits of both approaches. ICA is a statistical model that expresses observed data as a linear transformation of latent variables. ISA uses linear transformations and represents features as invariant subspaces.	1. The idea of an infinite Hidden Markov Model (iHMM) for video analysis is presented in this study. Because the iHMM can represent infinitely many hidden states, it is more flexible for modeling complicated video data. An actual application of the iHMM is given via the use of variational Bayesian inference and Bayesian analysis in this study. The performance of the iHMM is also contrasted with that of other traditional models by the authors, including DP-based Gaussian mixture models and maximum likelihood-based HMMs

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
Adaptability to broad contexts: The method may be used in a variety of video analysis scenarios since it is made to be flexible in varied environments.	Challenges with backdrop Removal and Extraction: The method attempts to steer clear of issues that arise while extracting moving entities from a backdrop environment and with

	background removal..

Analyse This Work by Critical Thinking	The Tools That Assessed this Work	What is the Structure of this Paper
The motivation behind this approach is to handle complex scenes where feature extraction on specific objects within the scene is difficult. Instead, feature extraction is performed on the entire scene. The authors also highlight that their proposed detection algorithm models normal behavior rather than unusual events, making it practical to obtain a sufficient set of training data. They discuss the complexity of time-series data and the challenges of model selection.	Infinite Hidden Markov Models (iHMM), Bayesian inference, and various feature extraction methods like Shift-Invariant Wavelets (SIWs), Independent Component Analysis (ICA), and Invariant Subspace Analysis (ISA).	I. Introduction II. Feature Extraction from Video III. Dirichlet Process and Hierarchical DP IV. Experimental Results V. Conclusions and Future Work

Diagram/Flowchart



--End of Paper 5—

6		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/document/10073281	Abhinandan Tripathi, Manish Kumar Gupta, Chaynika Srivastava, Pallavi Dixi, Shrawan Kumar Pandey	Object Detection, Computer Vision, YOLO, Regression
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
Object detection techniques based on You Only Look Once (YOLO) algorithm and aims to improve the accuracy.	The paper discusses various object detection techniques based on You Only Look Once (YOLO) algorithm. It also presents various modifications done on basic YOLO method and shows their analysis.	The paper presents a survey of various object detection techniques based on You Only Look Once (YOLO) algorithm. It aims to improve the accuracy of existing systems by presenting various modifications done on basic YOLO method and shows their analysis.
The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process		
YOLO (You Only Look Once) object detection method employs grid partitioning, bounding box prediction, and classification within each grid to detect objects, utilizing anchor boxes for multiple object identification. Techniques such as Intersection over Union (IOU) refine box selection, while multi-scale feature fusion improves performance, and constant optimization of loss functions and YOLO architecture modifications aim to enhance		

accuracy, speed, and adaptability across diverse applications, navigating the trade-offs between complexity and performance

	Process Steps	Advantage	Disadvantage (Limitation)
1	YOLO-based object detection partitions images into grids, predicts bounding boxes, and performs classification within each grid for object detection, utilizing anchor boxes for multiple object identification.	YOLO is notably fast as it performs detection in a single pass through the neural network, enabling real-time object detection applications, making it efficient for time-sensitive tasks.	YOLO might struggle with accurately localizing smaller objects within an image due to the grid-based approach, impacting precision for tiny or densely packed objects.
2	These models utilize techniques like Intersection over Union (IOU) for refining box selection and employ multi-scale feature fusion for improved performance, with specific model and application adaptations for domain-specific challenges.	It predicts multiple bounding boxes and class probabilities simultaneously, allowing for the identification of multiple objects within an image in one go.	When objects are heavily occluded or overlapping, YOLO may face challenges in accurately identifying and distinguishing them, affecting detection reliability.

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
In scenarios where high precision is crucial, such as in medical imaging or	Image or video frame	NA	The YOLO-based object detection process, including grid partitioning, box

scientific research, YOLO might not match the accuracy level achieved by certain slower but more precise object detection systems.			prediction, and more.

Relationship Among the Above 4 Variables in This article
In the context of YOLO-based object detection, the dependent variable is the system's output, signifying efficient real-time object detection with simultaneous classification and localization. The independent variable is the input, represented by image or video frames. While the provided information doesn't explicitly mention a moderating variable, the mediating variable is the intervening process of YOLO-based object detection. This process involves grid partitioning, bounding box prediction, and various techniques contributing to the final output. Together, these variables form a coherent system where image or video frames undergo YOLO-based detection, resulting in the efficient identification and localization of objects in real-time.

<table>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>Image or Video Frame</td><td>Bounding Boxes, Class Predictions, Confidence Scores</td></tr></table></td><td>Efficient real-time object detection with simultaneous classification and localization in diverse applications.</td><td>The contribution lies in providing a swift, unified approach for accurate object detection, streamlining real-time applications' decision-making processes.</td></tr>
</table>

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**

<table>
<tr><td colspan="2">The positive impact lies in enabling swift, accurate, and real-time object detection, vital for applications in diverse domains like surveillance, autonomous vehicles, and medical imaging, enhancing decision-making processes and automation.</td><td>One potential negative impact could be the challenge in precisely detecting small or occluded objects, reducing the model's accuracy in complex scenarios, such as tiny object identification or heavily obscured environments.</td></tr>
<tr><th>Analyse This Work By Critical Thinking</th><th>The Tools That Assessed this Work</th><th>What is the Structure of this Paper</th></tr>
<tr><td>This work presents YOLO as a versatile and rapid object detection framework, excelling in speed but facing challenges in accurate detection of small or occluded objects, necessitating continual trade-offs between speed and precision in diverse scenarios. Its applicability across various domains depends on context and the need for nuanced, domain-specific adjustments.</td><td>Assessment tools primarily involved Python-based frameworks (such as TensorFlow, PyTorch) for model development and training. Evaluation encompassed metrics like mAP (mean Average Precision) and loss functions, alongside domain-specific analyses tailored to the particular applications (e.g., object detection accuracy in surveillance, medical imaging, etc.).</td><td>I. Introduction
II. Feature Extraction from Video
III. Dirichlet Process and Hierarchical DP
IV. Experimental Results
V. Conclusions and Future Work</td></tr>
<tr><th colspan="3">Diagram/Flowchart</th></tr>
</table>

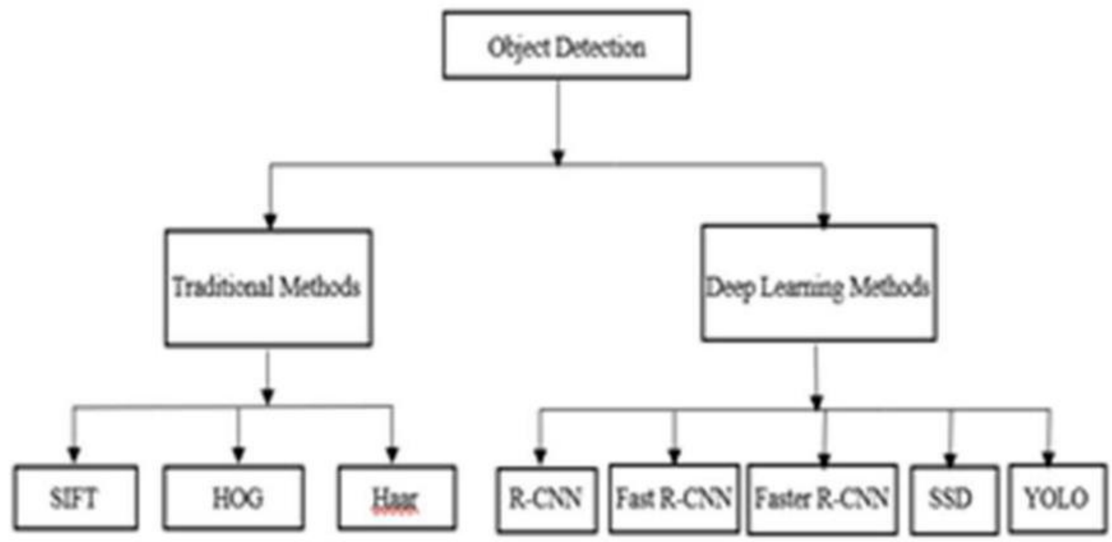


Figure 1: Classification of Object Detection Techniques

--End of Paper 6--

7		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/document/4379784	Iulian Pruteanu-Malinici and Lawrence Carin	Hidden Markov models, Dirichlet process, Variational Bayes
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
The paper proposes a solution for detecting unusual events in a	The goal of this solution is to develop a method for video feature extraction	The paper presents a method for detecting unusual events in video sequences using Invariant Subspace

video sequence. It uses invariant subspace analysis (ISA) to extract features from the video, and the time-evolving properties of these features are modeled via an infinite hidden Markov model (iHMM), which is trained using "normal"/"typical" video data.	that is effective in complex environments involving multiple and overlapping moving entities. The problem that needs to be solved is the extraction of meaningful and invariant features from video data in complex environments.	Analysis (ISA) and an infinite hidden Markov model (iHMM). ISA extracts features from the video, and iHMM models the time-evolving properties of these features. The iHMM is trained on "normal" video data, automatically determining the appropriate number of states and retaining a full posterior density function on all model parameters. Unusual events are identified when submitted sequential features yield a low likelihood in the trained iHMM.
The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process		
The work involves solving a problem of video-based feature extraction using a technique called Invariant Subspace Analysis (ISA). ISA is a traditional approach that uses linear transformations to compute features based on the inner product of input data with a particular basis. However, linear features lack invariance with respect to spatial shifts or phase changes. To address this, the work introduces a semi-parametric Hidden Markov Model (HMM) formulation, known as the Infinite Hidden Markov Model (iHMM). The iHMM allows for non-parametric learning of the underlying states in the video data.		

	Process Steps	Advantage	Disadvantage (Limitation)
1	The paper uses Invariant Subspace Analysis (ISA), a traditional method employing linear transformations, to extract features from video data. ISA features, designed to be	The advantage of Invariant Subspace Analysis (ISA) features is highlighted in the paper; specifically, their smaller dimension (40) is emphasized, as opposed to shift-invariant	According to the paper, there are implementation issues with shift-invariant wavelets because of their high dimension and insufficient training data. This restriction may have

	invariant to spatial shifts or phase changes, enhance their suitability for video sequence analysis, aiming to improve overall robustness and effectiveness in feature extraction.	wavelets' higher dimension (1,200). The larger dimension of shift-invariant wavelets (1,200) presents implementation challenges that are lessened by this reduced dimensionality (40).	an impact on the suggested algorithm's efficacy and accuracy in real-world situations where it may be hard or impossible to find training data.
2	The paper models the time-evolving properties of Invariant Subspace Analysis (ISA) features with an Infinite Hidden Markov Model (iHMM). The iHMM is trained on "normal" video data to grasp sequential characteristics of features. It autonomously determines the suitable number of Hidden Markov Model (HMM) states. The iHMM retains a comprehensive posterior density function on all model parameters.	According to the paper, shift-invariant wavelets and Independent Component Analysis (ICA) features provide less stable log-likelihoods than Invariant Subspace Analysis (ISA) features. Compared to the sensitivity seen in ICA features and shift-invariant wavelets, ISA features show less variability in the likelihood distribution in the presence of small changes (shifts) in the input image, yielding more stable and trustworthy results.	In comparing the MCMC and VB implementations for the iHMM, the paper points out that while the VB approach is efficient, it has the potential to identify abnormal scenarios during testing incorrectly. This restriction suggests that irregular events in video sequences may not always be accurately detected by the VB implementation.

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
Detection of unusual events	Video sequences collected in an	NA	Invariant Subspace Analysis (ISA) features, Infinite

	outdoor environment		Hidden Markov Model (iHMM)

Relationship Among the Above 4 Variables in This article
the type of video sequences (independent variable) affects the detection of unusual events (dependent variable) through the mediation of Invariant Subspace Analysis (ISA) features and the Infinite Hidden Markov Model (iHMM). The presence of these mediating variables explains how changes in the independent variable contribute to variations in the dependent variable. The role of any moderating variable is not explicitly discussed in the paper.

Input and Output		Feature of This Solution	Contribution & The Value of This Work
Input	**Output**		
video sequences collected in an outdoor environment	detection of unusual events in the video sequences.	The method compares the efficacy of invariant subspace analysis (ISA) with other approaches and highlights ISA's applicability for implementation because of its smaller dimension. The solution makes use of invariant feature subspaces, linear subspace representation, and an infinite hidden Markov model (iHMM) trained on "normal" events to detect unusual events in video data.	This paper presents a novel approach to anomaly detection by combining independent subspace analysis (ISA) for video data with the infinite hidden Markov model (iHMM) framework. This integration improves the algorithm's ability to detect anomalous events in video sequences by enabling non-parametric state learning and invariant feature extraction.

Positive Impact of this Solution in This Project Domain	Negative Impact of this Solution in This Project Domain
In this project, there are significant advantages to integrating the infinite hidden Markov model (iHMM) with invariant subspace analysis (ISA). In order to detect anomalies in video sequences	Large video datasets may not be as efficiently trained using the MCMC training method due to its high computational cost and requirement for a significant amount of CPU time.

more precisely, ISA features guarantee a stable log-likelihood.	

Analyse This Work By Critical Thinking	The Tools That Assessed this Work	What is the Structure of this Paper
When gathering sufficient unusual data is not feasible, the iHMM framework with independent subspace analysis performs exceptionally well in anomaly detection. We compare two inference tools: MCMC and VB. While VB provides effective training, it may not be completely accurate in detecting abnormal scenarios. Subsequent research aims to integrate data from neighboring blocks and investigate the joint application of MCMC and VB techniques (MCMC-VB) to improve efficiency.	The paper compares and contrasts the efficacy of three feature extraction methods for video analysis: Invariant Subspace Analysis (ISA), Shift-Invariant Wavelets (SIW), and Independent Component Analysis (ICA). Furthermore, to improve unusual-event detection in video sequences, two inference tools—Markov Chain Monte Carlo (MCMC) and Variational Bayes (VB)—are utilized for the infinite hidden Markov model (iHMM), with VB providing a viable substitute for MCMC in iHMM training.	I. Abstract II. Introduction III. Invariant Subspace Analysis (ISA) IV. Infinite Hidden Markov Model (iHMM) V. Variational Bayesian Inference

Diagram/Flowchart

--End of Paper 7—

8	
Reference in APA	

format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/abstract/document/4579745	Mark Jäger, Christian Knoll, and Fred A. Hamprecht	One-class learning, outlier detection, state-space models, time series classification, weak labels
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
Currently available methods for identifying anomalous events in industrial image sequences combine one-class and two-class classification frameworks into one. Using temporal dynamics, hidden Markov models (HMMs) are used to train the system on data with weak labels, showing a low false positive rate. Its effectiveness is confirmed by validation	The aim of the solution is to identify errors in image sequences from industrial processes by using temporal dynamics and spatial appearance to classify frames as regular or error frames. The difficulty in training statistical learning systems stems from the lack of positive examples for error frames. In order to accurately detect and classify error events, the solution uses state-space models and time series	PCA is a technique used in incremental learning for dimension reduction that splits the feature space into a principal subspace and residual error. The Regular Sequence Model (RSM), which learns to optimize states using the Bayesian Information Criterion, is learned for regular sequences, while Hidden Markov Models (HMMs) model sequences. By using distinct HMMs for each error class and a change detection algorithm to identify and categorize anomalous events in the data, the Error Sequence Model (ESM) expands on the Root Sequence Model (RSM).

using industrial data from laser welding processes.	classification techniques.	

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	Dimension Reduction: Principal Component Analysis (PCA) is used to compute a linear subspace from sequences without errors. The recorded images are then projected into this subspace to reduce the feature dimension.	The work demonstrates that applying Hidden Markov Models (HMMs) to process dynamics improves classification performance over techniques such as Gaussian Mixture Models (GMMs) and temporally-independent polynomial classifiers..	An extremely unbalanced data set was used for the experiments, according to the paper, with most sequences being regular and only a small percentage having error events. This can make it difficult to train automated systems using statistical learning.
2	Incremental Model Building: A Hidden Markov Model (HMM) is trained for the regular sequences (RSM) using the reduced features. The RSM represents the class of sequences without errors.	The efficacy of the HMM-based classification system in identifying sequences with anomalous events amidst typical process fluctuations is demonstrated by its low False Positive rate of 1.8%.	Absence of Positive Examples: The aberrations of interest are infrequent in many event detection applications, such as the laser welding process covered in the paper. This indicates that there aren't enough good examples, which makes it challenging to effectively train

			automated systems.

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
Improved classification performance	Unusual event detection	NA	Principal subspace features, DFFS signal

Relationship Among the Above 4 Variables in This article

The framework aims to enhance classification performance (dependent variable) by detecting unusual events (independent variable). The process involves mediating variables such as principal subspace features and the DFFS signal, which play a role in improving the system's ability to identify and classify anomalous events in industrial image sequences. The specific moderating variable is not specified in the provided information.

Input and Output		Feature of This Solution	Contribution & The Value of This Work
Input	**Output**	The combination of features from the principal subspace (PC1,2,3) and the residual subspace (DFFS) enables a satisfactory classification performance. The principal subspace features detect changes in overall brightness and translations of the melt pool, while the DFFS signal detects deformations not observed in the training data set of regular sequences.	This paper presents a new classification scheme for identifying anomalous events in industrial laser welding process image sequences. When compared to methods without temporal information, the framework's use of hidden Markov models (HMMs) to simulate temporal dependencies greatly improves classification performance.
Unusual event detection	Improved classification performance		

Positive Impact of this Solution in This	Negative Impact of this Solution in This

Project Domain	Project Domain
Significant benefits can be obtained from the suggested framework for event detection and incremental learning in industrial processes. Through the use of principal component analysis (PCA) and incremental learning to address imbalanced datasets, along with the incorporation of a hidden Markov model (HMM), the system improves the accuracy of unusual event detection.	Within the project domain, the suggested solution encounters difficulties, most notably because of an unbalanced dataset that contains few positive examples of infrequent error events in industrial processes. Because industrial scenes are complex, it can be challenging to discern between regular variations and real error events, which could result in false positives or negatives.

Analyse This Work By Critical Thinking	The Tools That Assessed this Work	What is the Structure of this Paper
The paper presents a framework that uses hidden Markov models (HMMs) as the basis for a classification system to identify anomalous events in industrial laser welding image sequences. The framework employs incremental learning through dimension reduction and model building to address imbalanced datasets.	The study models the dynamics of industrial laser welding using Hidden Markov Models (HMMs), demonstrating how well they perform in classification when compared to Gaussian Mixture Models (GMMs), which are deficient in temporal information. Furthermore, a polynomial classifier is used in the study to compare HMMs to a discriminative approach with temporal smoothing. Principal Component Analysis (PCA) is incorporated into the suggested framework to	I. Introduction II. Related Work III. Methodology a. Subspace Methods b. Continuous Hidden Markov Models c. Learning Procedure IV. Experimental Results V. Conclusion

	reduce dimensions in the detection of unusual events.	
Diagram/Flowchart		

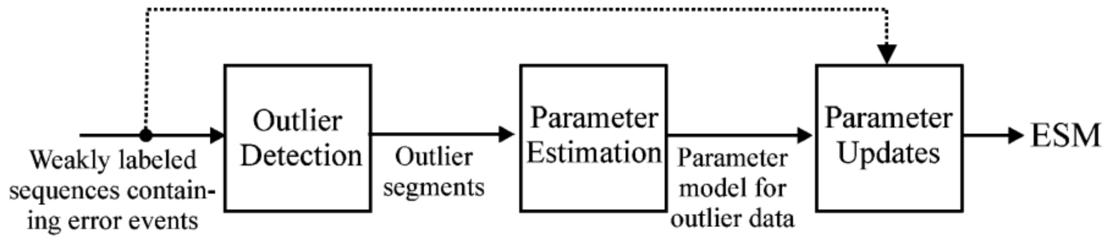


Fig. 3. Schematic overview to estimate the parameters of the error sequence model (ESM).

---End of Paper 8---

9		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://openaccess.thecvf.com/content_cvpr_2017_workshops/w34/papers/Vignesh_Abnormal_Event_Detection_CVPR_2017_paper.pdf	Kothapalli Vignesh, Gaurav Yadav, Amit Sethi	Surveillance system, Multiple cameras, Event detection, CNN, LSTM
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**

The document refers to the existing solution as the BMTT-PETS 2017 Surveillance Challenge. It's a technique for creating a trainable surveillance system that can identify possible threats on moving assets like cars and provide situational awareness.	The goal of the solution is to develop a trainable surveillance system that can detect anomalous activity in real-world scenarios with the least amount of video data possible for mobile assets, especially vehicles.	The suggested approach to abnormal event detection combines the use of a long short-term memory (LSTM) network for temporal feature extraction, a VGG-16-based CNN for spatial feature extraction, and background subtraction using a mixture of Gaussians. Whereas the LSTM concentrates on the temporal relationships between frames, the CNN records spatial features.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	Following background subtraction, a Convolutional Neural Network (CNN) extracts spatial features from each frame, distinguishing between normal and abnormal frames through training.	The study illustrates the value of training from scratch by showing better results when utilizing a scratch model as opposed to a pre-trained one.	The study illustrates the value of training from scratch by showing better results when utilizing a scratch model as opposed to a pre-trained one.
2	To capture long-term dependencies, a Long Short-Term Memory (LSTM) network processes feature vectors from the	As shown in Table 2, employing a Support Vector Machine (SVM) in conjunction with a Convolutional Neural	As shown in Table 2, employing a Support Vector Machine (SVM) in conjunction with a Convolutional Neural

	CNN, learning temporal patterns in the frames.	Network (CNN) for feature classification produces superior outcomes compared to softmax.	Network (CNN) for feature classification produces superior outcomes compared to softmax.

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
Classification of frames as normal or abnormal	Image sequences captured by surveillance cameras	NA	Features computed by the LSTM network

Relationship Among the Above 4 Variables in This article

Understanding these variables is essential to grasp how the proposed method for abnormal event detection works. Imagine the image sequences from surveillance cameras as the starting point, like the input. The LSTM network processes these images and calculates features. These features, in turn, play a key role in deciding whether a frame is normal or abnormal. This process contributes significantly to how well the surveillance system can identify unusual events. However, having more information about any additional factors that might influence this relationship would help us understand it even better.

Input and Output	Feature of This Solution	Contribution & The Value of This Work
Input: image sequence; **Output**: classification of frames	The suggested technique uses Gaussians mixtures for background subtraction and uses a 2.5 standard deviation threshold to mark	In order to improve accuracy, the paper presents a method for video activity recognition that combines deep learning, motion-based, and shape-

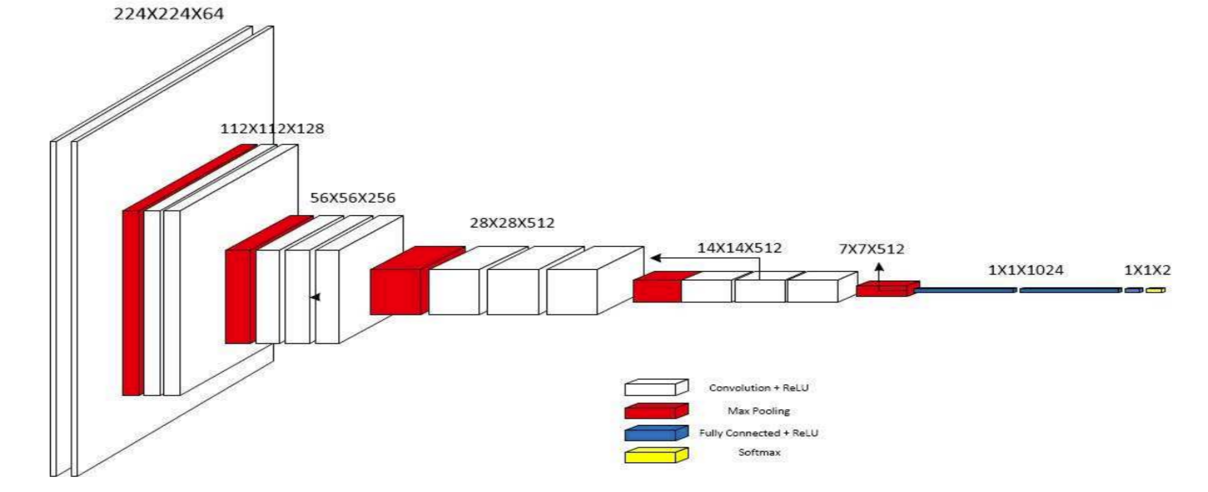
captured by surveillance cameras	as either normal or abnormal based on the features computed by the LSTM network.	foreground pixels. A VGG-16-based deep CNN is used to extract spatial features, highlighting the significance of depth for effectiveness. By feeding CNN output to an LSTM network, temporal relationships are taken into account to improve model accuracy and allow the network to learn long-range dependencies in sequential data.	based techniques. Incorporating trajectory analysis and interest points enhances the accuracy of feature extraction, resulting in more resilient recognition. Deep learning integration improves the system's capacity to recognize intricate activity, which makes it useful for tasks like threat and surveillance identification.

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
A scratch model was used to improve classification accuracy, an output layer of Support Vector Machines (SVM) was added to improve results, an LSTM network was added to improve recognition accuracy, and CNN, LSTM, and SVM were integrated to achieve superior performance. These positive impacts on the project domain are documented in the document.	The proposed solution may face challenges in the project domain, including limited training data impacting generalization, time-consuming training due to the complexity of CNN and LSTM, dependency on background subtraction effectiveness, and subjectivity in feature extraction methods. These factors may affect the accuracy and reliability of abnormal event detection in real-world scenarios.

Analyse This Work By Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
The document presents a method for a trainable surveillance system to detect abnormal events in human group activities with limited	Using Convolutional Neural Networks (CNN) for feature extraction, Support Vector Machines (SVM) for classification and	I. Introduction II. Related Work III. Proposed Methodology a. Background

video data. Challenges in activity recognition, including environmental variations and actor movements, are highlighted. The related work covers motion-based, shape-based, and deep learning-based approaches. The proposed method involves background subtraction, CNN-based feature extraction, and LSTM-based learning of temporal dependencies between frames.	background subtraction, Linear SVM for classification scores, Mixtures of Gaussians for background subtraction, Bag-of-Words Models for feature extraction, and LSTM Networks for sequence learning are just a few of the tools and techniques for video analysis and activity recognition that are covered in the document.	Subtraction b. Feature Extraction c. Long Short-Term Memory (LSTM) Network d. Classification IV. Experimental Results V. Conclusion

Diagram/Flowchart



---End of Paper 9--

10		
Reference in APA format		
URL of the Reference	**Authors Names and**	**Keywords in this Reference**

	Emails	
https://www.mdpi.com/2504-2289/7/1/48	Zouheir Trabelsi, Fady Alnajjar, Medha Mohan Ambali Parambil, Munkhjargal Gochoo, and Luqman Ali.	Education, Deep learning, Attention assessment, Student behavior dataset, Emotion recognition, Object detection, YOLOv5
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
The current monitoring system utilizes the DeepSORT algorithm to track students in a classroom. DeepSORT employs a two-part matching cascade, leveraging position and velocity information from current and previous images to link observed bounding boxes to tracks.	The proposed solution aims to create a real-time vision-based classroom system using YOLOv5 models for monitoring student emotions, attendance, and attention levels. It addresses the limitations of traditional methods in tracking these crucial metrics, offering instructors real-time insights to enhance classroom dynamics and improve academic outcomes.	The proposed system comprises a Data Acquisition Module that uses a camera to capture real-time classroom images, particularly students, connected to a desktop computer. The YOLOv5-based module employs the YOLOv5 network structure, encompassing the backbone, neck, and head parts, for Action Behavior Detection, Emotion Detection, and Facial Recognition.
The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process		

The proposed model uses a hybrid image fusion technique to effectively combine the MRI and CT images of brain and provide high quality fused images with minimal or no distortion.

	Process Steps	**Advantage**	**Disadvantage (Limitation)**
1	Classroom images are captured in real-time using a high-definition camera connected to a desktop computer. Multiple cameras may be used to cover a larger classroom.	The paper focuses on developing an automatic attention assessment system for classrooms. This system accurately measures students' attention levels using typical classroom behaviors and actions as a prediction metric..	Due to ethical constraints, the study's supporting data are not openly accessible. Interested researchers must request the data directly from the corresponding author, potentially impeding the replication and verification of the study's results.
2	Images and videos are collected from the web/Internet and from a classroom setting. These images are labeled using an open-source image annotation tool called LabelImg.	The proposed system utilizes deep learning algorithms to monitor students' behavioral and emotional patterns. This helps educators gain insight into how students are feeling during the lecture, allowing them to adapt their teaching approach accordingly.	Privacy concerns led to the decision not to record videos during the study, opting for numerical CSV files. While safeguarding student privacy, this choice limits the analysis of visual cues and non-verbal behaviors that could offer further insights into attention levels.

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
attention level of student	classroom behaviors and actions	NA	NA

Relationship Among the Above 4 Variables in This article
The dependent variable in the given document is the attention level of students, and the independent variable is the classroom behaviors and actions of the students. There is no information provided about moderating or mediating variables in the given document.

Input and Output		Feature of This Solution	Contribution in This Work
Input	**Output**	The three main components of the suggested automatic attention assessment system are an evaluation module, YOLOv5-based action behavior detection, and data collection. Real-time classroom images are captured by a high-definition camera, and student attendance is tracked through the use of a face recognition algorithm that uses Haar Cascade to match student faces with a database that has been stored.	In order to precisely measure students' attention levels during lectures, the research presents an automatic attention assessment system that uses deep learning. The benefit of the system is its real-time monitoring, which helps teachers modify their lesson plans in response to emotional and behavioral patterns.
images of the students	The output of the system includes various metrics and reports that provide insights into students' attention levels, behaviors, and emotions.		

Positive Impact of this Solution in This Project Domain	Negative Impact of this Solution in This Project Domain
By improving engagement through real-time insights into behavior and emotions,	One negative impact of this solution in the project domain is the potential for overfitting. If

the deep learning and computer vision utilized in the real-time vision-based classroom system has a positive impact on education. A student-centered learning environment can be created by instructors by personalizing their teaching strategies.	the hyperparameter settings are poorly chosen, it can lead to overfitting, which means that the model may perform well on the training data but not generalize well to new, unseen data. This can affect the accuracy and reliability of the system's predictions.

Analyse This Work by Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
The document delves into an automatic attention assessment system for classrooms, using deep learning to measure students' attention based on behaviors. Emphasizing real-time engagement visualization, the system aids instructors in adapting teaching for better student outcomes. It discusses AI integration in education, especially in assessment and behavior analysis.	YOLOv5s: A pre-trained model used for training the system on masked and unmasked images. High-definition camera: Mounted in the classroom to capture students' actions and behaviors. CSV format: Used to save student attendance data.	I. Introduction II. Related Work III. Proposed Methodology IV. Experimental Results V. Conclusion
Diagram/Flowchart		

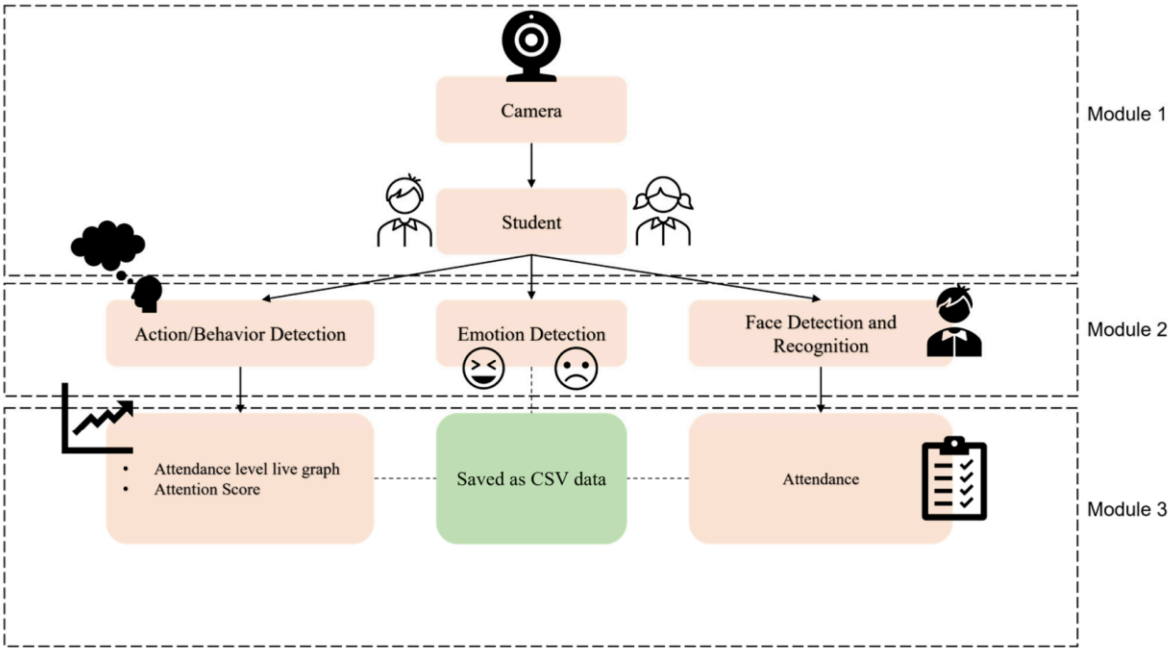


Figure 1. The general framework of the proposed system.

---End of Paper 10--

11		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/document/9053457	Zheng , Fei Jiang ,Ruimin Shen zhengr@sjtu.edu.cn,jiangf@sjtu.edu.cn, rmshen@sjtu.edu.cn	student behavior detection, Faster RCNN, scale-aware detection head, feature fusion
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**

The current solution for the intelligent student behavior analysis system utilizes an improved Faster R-CNN algorithm. This enhanced version addresses challenges in real classrooms by incorporating a scale-aware detection head for scale variations.	The solution aims to develop an intelligent system for automatically detecting hand-raising, standing, and sleeping behaviors in recorded classrooms. It seeks to overcome challenges like scale variations, low resolution, and imbalanced behavior samples..	The intelligent student behavior analysis system utilizes the Faster R-CNN algorithm with ResNet-101 for behavior detection from recorded classroom video streams. It features a scale-aware detection head to handle scale variations, a feature fusion strategy for capturing detailed and semantic information, and employs Online Hard Example Mining (OHEM) to address class imbalance. The overall architecture integrates these components to analyze student behaviors effectively.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	Building a large-scale student behavior corpus	Provides a dataset for behavior recognition	Time-consuming and requires manual labelling
2	Using Faster R-CNN for student behavior analysis	Classical object detection model	May not perform well on complex real classroom scenarios

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable

<table>
<tr><td>Mean Average Precision (mAP) The performance metric indicating the accuracy of the student behavior detection system.</td><td>Implementation of Enhanced Faster R-CNN Algorithm The manipulated variable, representing the technology used for student behavior analysis</td><td>Online Hard Example Mining (OHEM) A variable that influences the strength or direction of the relationship between the independent variable (algorithm) and the dependent variable (mAP).</td><td>Feature Fusion Strategy An intervening variable that explains the process through which the independent variable (algorithm) influences the dependent variable (mAP).</td></tr>
<tr><th colspan="4">Relationship Among The Above 4 Variables in This article</th></tr>
<tr><td colspan="4">In this study, the dependent variable is the Mean Average Precision (mAP), which measures the accuracy of the intelligent student behavior analysis system. The independent variable is the implementation of the enhanced Faster R-CNN algorithm, representing the technology used for student behavior analysis.</td></tr>
</table>

<table>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>Input refers to the video streams from recorded classrooms that are used as the model input for the intelligent student</td><td>The output of the intelligent student behavior analysis system is the detection results for hand-</td></tr></table></td><td>The solution introduces a feature fusion strategy to enhance accuracy in detecting small objects, like sleeping and hand-raising behaviors. This strategy efficiently captures both low-level and high-level semantic information, achieving impressive real classroom results without heavy computation costs. Additionally, a scale-aware detection head is proposed to</td><td>The study's contributions include a substantial dataset, an intelligent system for real classrooms using enhanced object detection, and advanced algorithms (Faster R-CNN) addressing scale variations and imbalanced samples. The value of the work lies in its practical applications for teaching quality, providing insights into teaching effectiveness and student development through</td></tr>
</table>

behavior analysis system.	raising, standing, and sleeping behaviors of students in recorded classrooms	handle scale variations in student behaviors, utilizing different dilation rates to detect objects of various sizes and improving detection capabilities for large-scale variations, including small objects.	automatic behavior detection. The improved algorithms and large-scale dataset contribute to advancing student behavior analysis, potentially enhancing educational practices.

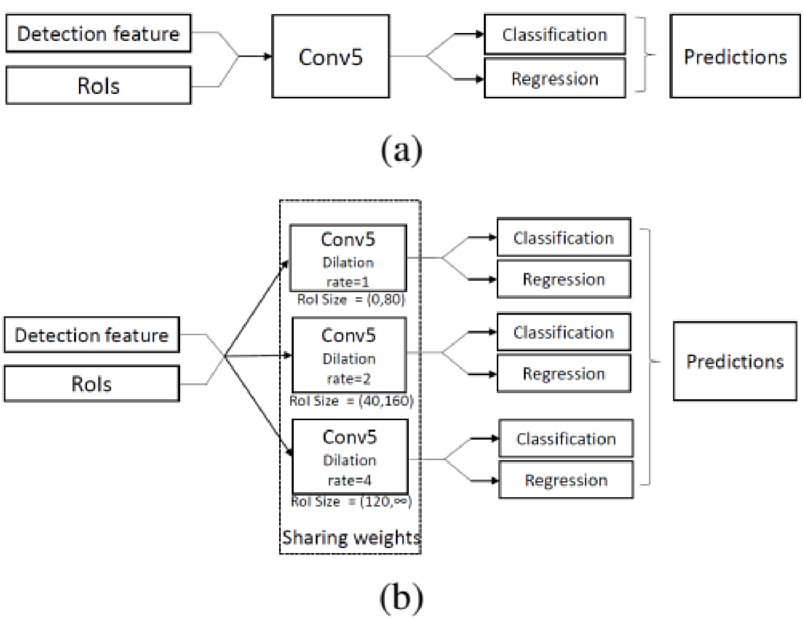
Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
The proposed intelligent student behavior analysis system, with an enhanced Faster R-CNN algorithm, improves detection accuracy by 3.4%. Tailored for real classrooms, it addresses scale variations and low-resolution behaviors efficiently. The system, featuring a scale-aware detection head and feature fusion, maintains practicality without heavy computational demands.	The proposed solution has potential negative impacts in the project domain, including challenges in accurately detecting small objects, persistent class imbalance issues, potential resource-intensive methods affecting performance, dependence on pre-recorded videos limiting real-time applicability, and concerns about generalizability to diverse classroom environments.

Analyse This Work By Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
The document presents an intelligent student behavior analysis system for real classrooms, addressing challenges like scale variations and class imbalances. It introduces an enhanced Faster R-CNN	the tools used for assessing the work involves existing behavior detection algorithms such as hand-crafted feature-based, pose-estimation-based, and object-detection-based algorithms. The proposed solution	I. Introduction II. Related Works III. Methods IV. Experiments V. Conclusion VI. References

model with a scale-aware detection head and Online Hard Example Mining (OHEM). The system achieves a 3.4% increase in mean Average Precision (mAP) on a real dataset, demonstrating its potential for practical applications in classrooms.	introduces an improved Faster R-CNN algorithm for student behavior analysis, suggesting the likely involvement of various tools and technologies for tasks such as data processing, feature extraction, model training, and evaluation.	

Diagram/Flowchart

(a) Original detection head in Faster R-CNN. (b) Our scale-aware detection head using dilated convolution for classification and regression.



---End of Paper 11---

12

Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this**

		Reference
https://ieeexplore.ieee.org/document/4359346	Tao Xiang, Shaogang Gong	Medical image fusion, G-CNNs, Gabor representation, convolutional neural network, fuzzy neural network.
The Name of the Current Solution (Technique/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
The current solution proposed in the document is called "Video Behavior Profiling for Anomaly Detection". designed to automatically profile behaviors and detect anomalies in surveillance videos..	The solution aims to create an automated system for analyzing surveillance videos, detecting abnormal behavior, and recognizing normal behavior. The key problem it addresses is the need for an automated method to analyze large volumes of surveillance video data, identifying abnormal behavior patterns without manual labelling, and learning from previously observed patterns in real-time.	The solution involves various components, including the distribution of elements, where a multimodal distribution indicates relevance. Eigenvectors are utilized for data clustering, assumed to follow a mixture of two Gaussians.
The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process		

	Process Steps	Advantage	Disadvantage (Limitation)

1	Behavior Representation: The paper addresses the problem of behavior representation, where each behavior pattern is represented as a feature vector of fixed length. This is achieved through a data clustering approach and dimension reduction using a spectral clustering algorithm.	The representation of behavior patterns as fixed-length feature vectors allows for easier data clustering and analysis.	Conventional clustering approaches may not be directly applicable due to the variable length of feature vectors.
2	Model Selection: The next step is to determine the number of clusters, which is done through automatic model selection. The largest eigenvectors of the normalized affinity matrix are used for data clustering, reducing the data dimensionality.	Automatic model selection helps determine the optimal number of clusters, reducing the need for manual intervention.	The selection of the number of clusters is still an unknown factor and may require further optimization.

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
Detection Results of Student Behaviors The outcome or response	Algorithm Modifications, Dataset Characteristics,	Domain Knowledge A potential variable that may moderate the relationship	Behavior Profiling and Anomaly Detection A variable that

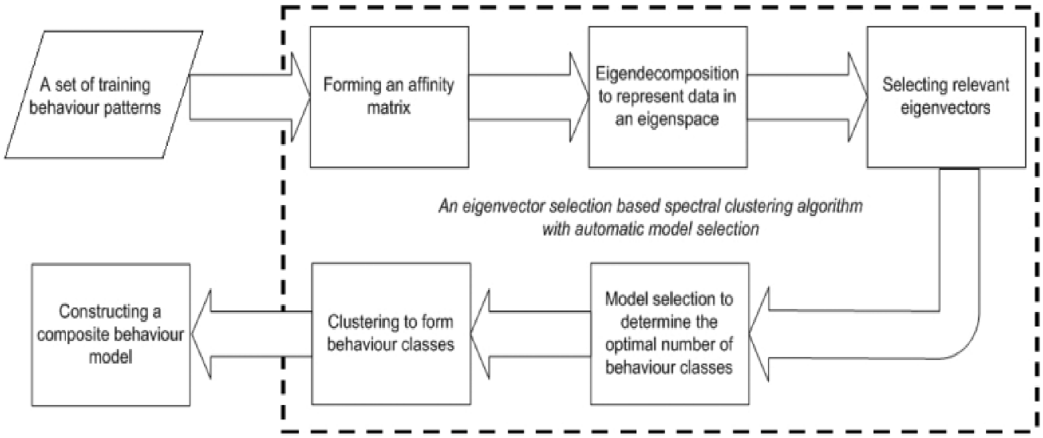
variable being measured	Model Architecture, Training Strategies Variables manipulated or controlled by the researcher.	between independent variables and the dependent variable. Domain knowledge could influence the system's performance.	explains the process or mechanism through which independent variables impact the dependent variable.

Relationship Among the Above 4 Variables in This article
In this research, the dependent variable is the detection results of anomalies in surveillance videos. The independent variables encompass various factors such as algorithm modifications, dataset characteristics, model architecture, and training strategies

Input and Output		**Feature of This Solution**	**Contribution & The Value of this Work**
Input	**Output**	The proposed solution for behavior profiling in surveillance videos has key features including a composite generative behavior model for handling complex behaviors, on-the-fly anomaly detection for real-time applications, automatic model selection to determine clusters, dimension reduction addressing the "curse	This work introduces a novel unsupervised framework for robust online behavior recognition and anomaly detection. Key components include discrete event-based behavior representation, a DBN-based affinity measure, spectral clustering with feature and model selection, a composite behavior model based on a mixture of DBNs, a runtime accumulative anomaly measure, and
surveillance video data generated continuously by closed-circuit television (CCTV) systems	output is detection results of student behaviors		

	of dimensionality" problem, and event-based behavior representation for accuracy.	an online LRT-based normal behavior recognition method..
Positive Impact of this Solution in This Project Domain		**Negative Impact of this Solution in This Project Domain**
The proposed approach improves anomaly detection through a robust composite generative behavior model in real-time surveillance. It emphasizes scalability to handle diverse behavior patterns, leveraging domain knowledge for informed decision-making in unfamiliar domains. The framework's flexibility allows customization with various segmentation, representation, and affinity techniques. Additionally, the approach contributes to ongoing research by addressing open questions in video anomaly detection, emphasizing knowledge generalization and accelerating the learning process with human supervision.		The solution presented in the document has limitations, including a lack of adaptability to new observations after training, a dependency on domain knowledge that humans naturally possess, challenges in generalizing knowledge.
Analyse This Work by Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
The proposed method effectively combines Gabor representation, multi-CNNs, and fuzzy neural network to The paper demonstrates the effectiveness and robustness of the proposed video behavior profiling framework for anomaly detection through experiments with noisy and sparse datasets from indoor and outdoor surveillance scenarios. Interestingly, the results indicate that a behavior model trained with unlabeled data outperforms labeled	The paper uses tools like Dynamic Bayesian Networks (DBNs), Spectral Clustering Algorithm, EM Algorithm, Bayesian Information Criterion (BIC), Online Segmentation Algorithm, Event Detection, Model	I. Introduction II. Related Work III. Behavior Pattern Representation IV. Behavior profiling V. Online anomaly detection

data in detecting anomalies from unseen videos. The online Likelihood Ratio Test (LRT)-based behavior recognition approach is highlighted as advantageous compared to commonly used methods.	Selection, and Affinity Matrix to develop a behavior profiling framework for anomaly detection in video sequences.	and normal VI. Behavior recognition VII. Experiments VIII. Discussions and conclusions

Diagram/Flowchart
 A block diagram illustrating our behavior profiling approach

---End of Paper 12---

13

Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/document/8545345	Xian Sun; Songhao Zhu; Songsong Wu; Xiaoyuan Jing	Abnormal Behavior Detection, Weak Supervised Detection, Temporal Consistency, Candidate Action Fragment,Corresponding Classifier

The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	The Goal (Objective) of this Solution & What is the problem that needs to be solved	What are the components of it?
The current solution proposed in the document is a weak supervised abnormal behavior detection method that utilizes temporal consistency. It constructs temporal gram matrices for a given video sequence and forms pairs of behavior units (candidate action fragments) based on the temporal consistency and smoothness of human behavior.	The solution aims to create a model for detecting abnormal behavior in videos by segmenting behaviors into actions and focusing on temporal consistency and smooth movement. It utilizes sparse reconstruction to improve the efficiency of behavior classification. The problem it addresses is the detection of anomalies in video sequences.	Temporal Gram Matrices: These capture temporal information in a video sequence. Candidate Action Fragments: These exploit smoothness in human behavior to identify start and end frames of abnormal behavior

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	Construction of Timing Sequence	Efficiently constructs sub-sequences with fixed length and stride	May result in loss of information due to fixed length and stride

2	Temporal Consistency	Explores Gram matrix to identify best candidates for behavior units	Requires a threshold for similarity score, which may not always be accurate

Major Impact Factors in this Work

Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
Abnormal Behavior Detection Results The outcome variable representing the effectiveness of the proposed weakly supervised abnormal behavior detection method in identifying abnormal behavior.	Algorithm Modifications, Dataset Features, Variables manipulated or controlled by the researchers to assess their impact on the dependent variable, including adjustments in algorithm parameters and dataset features.	Domain Knowledge A potential variable that may moderate the relationship between independent variables and the dependent variable, influencing the system's performance based on domain expertise	Temporal Consistency and Sparse Reconstruction A variable that explains the process or mechanism through which the independent variables impact the dependent variable..

Relationship Among The Above 4 Variables in This article

In this study, the dependent variable is the outcome of abnormal behavior detection, assessing the success of the proposed weakly supervised method. The independent variables consist of various factors such as algorithm modifications, dataset features, model architecture, and training techniques, all subject to manipulation to understand their influence.

Input and Output	**Feature of This Solution**	**Contribution & The Value of this Work**

Input	Output		
The method for abnormal behavior detection uses two matrices: Y (training samples) and D0 (initial dictionary).	The proposed abnormal behavior detection method utilizes temporal consistency and sparse reconstruction, representing behavior as a combination of normal training samples..	The key feature of this solution is its ability to detect abnormal behavior in video sequences by utilizing temporal consistency and sparse reconstruction. This unique approach allows behavior to be represented as a combination of normal training samples, resulting in more accurate detection of anomalies. The method's effectiveness is demonstrated through its superior performance compared to other methods, as evidenced by experimental results on datasets like CAVIAR and Crossing.	The work presents a weakly supervised method for detecting abnormal behavior based on temporal consistency, leveraging human behavior's smoothness to form behavior units. It uses sparse reconstruction, showing superior detection accuracy for single-person and two-person interaction anomalies compared to other methods. Its value lies in its verified effectiveness on common databases like CAVIAR and Crossing, demonstrating its superiority through experimental results..

Positive Impact of this Solution in This Project Domain	Negative Impact of this Solution in This Project Domain
Improved Anomaly Detection: The proposed method utilizes the network PCANet to extract more effective features from low-level features, such as image saliency information and multi-scale optical flow histogram. This leads to more accurate anomaly detection, as demonstrated by the experimental results on public databases.	Limited Training Data: The weak supervised learning approach used in this project relies on only one ordered behavior sequence for training. This limitation can result in a lack of diversity in the training data, which may affect the model's ability to accurately detect abnormal behavior in complex scenarios.

Analyse This Work By	The Tools That Assessed	What is the Structure of

Critical Thinking	this Work	this Paper
The work introduces a weakly supervised method for detecting abnormal behavior, emphasizing temporal consistency. It constructs behavior units in video sequences using temporal gram matrices and smoothness characteristics. Through sparse reconstruction, it detects anomalies..	The Tools That assessed this work are public databases of CAVIAR and Crossing, and the proposed method was compared with four other methods. The experimental results were evaluated using metrics such as receiver operating characteristic (ROC) curves and the area under the ROC (AUC).	I. Introduction II. Weak Supervised Learning III. Expermental and Analysis IV. Conclusion V. References

Diagram/Flowchart
 (a) Fighting  (b) Fainting  (c) Robbing the cell phone Figure 6. Examples of abnormal behavior on crossing dataset.

---End of Paper 13---

14	
Reference in APA	

format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/document/9761925	Lina Li, Minghan Liu, Liyan Sun, Yupeng Li Nianfeng Li	Deep learning, student behavior recognition, YOLOv5s, generative adversarial network, tiny object detection
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
The current solution proposed in the document is called ET-YOLOv5s. It is a deep learning model designed for the identification of students' in-class behaviors. The model is trained using a specific set of network parameters and achieves an ideal training effect	The proposed solution, ET-YOLOv5s, aims to enhance the identification of students' in-class behaviors by addressing the inefficiencies of existing methods that focus on single student behavior. It seeks to improve performance by simultaneously recognizing multiple students' behaviors in the classroom.	The YOLOv5s network comprises three main parts: Backbone, Neck, and Head. The Backbone includes Focus, CSPNet, and SPP for efficient training. The Neck incorporates FPN and PAN to aggregate parameters and address scale issues. The Head, the final part, has three modules for object detection at different scales, using BCELogits, BCEclsloss, and GIoU loss functions for optimizing bounding box, category, and intersection scale predictions.
The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process		
The proposed framework consists of several steps, each with its advantages and disadvantages:		

	Process Steps	Advantage	Disadvantage (Limitation)
1	Data Collection and Labeling:Classroom surveillance videos were converted into static images and labeled using LabelImg to mark each student's location and behavior with bounding boxes.	Real classroom surveillance videos provide authentic data for behavior recognition.	The process of converting videos into static images may result in the loss of temporal information.
2	Data Set Preparation: A dataset of 826 images was created from labeled classroom surveillance videos, representing eleven student behaviors with an unbalanced distribution across behaviors.	The labeled data set allows for supervised training of the model.	The imbalance in the number of images for each behavior may affect the model's ability to detect certain behaviors.

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
Abnormal Behavior Detection Results The outcome variable representing the success of ET-YOLOv5s in identifying various student behaviors in a classroom setting.	Algorithm Modifications, Dataset Features, Model Architecture, Training Techniques Variables manipulated or controlled by the researchers to assess their impact on the dependent variable.	NA	Tiny Object Detection Module, FPN, PAN, GIoU loss Function | A variable that explains the process or mechanism through which the independent variables impact the dependent variable.

Relationship Among the Above 4 Variables in This article
In this study, the primary focus is on the dependent variable, representing the results of abnormal behavior detection through ET-YOLOv5s. The independent variables encompass a range of factors, including algorithm modifications, dataset features, model architecture, and training techniques. These variables are carefully adjusted to assess their influence on the effectiveness of the abnormal behavior detection model.

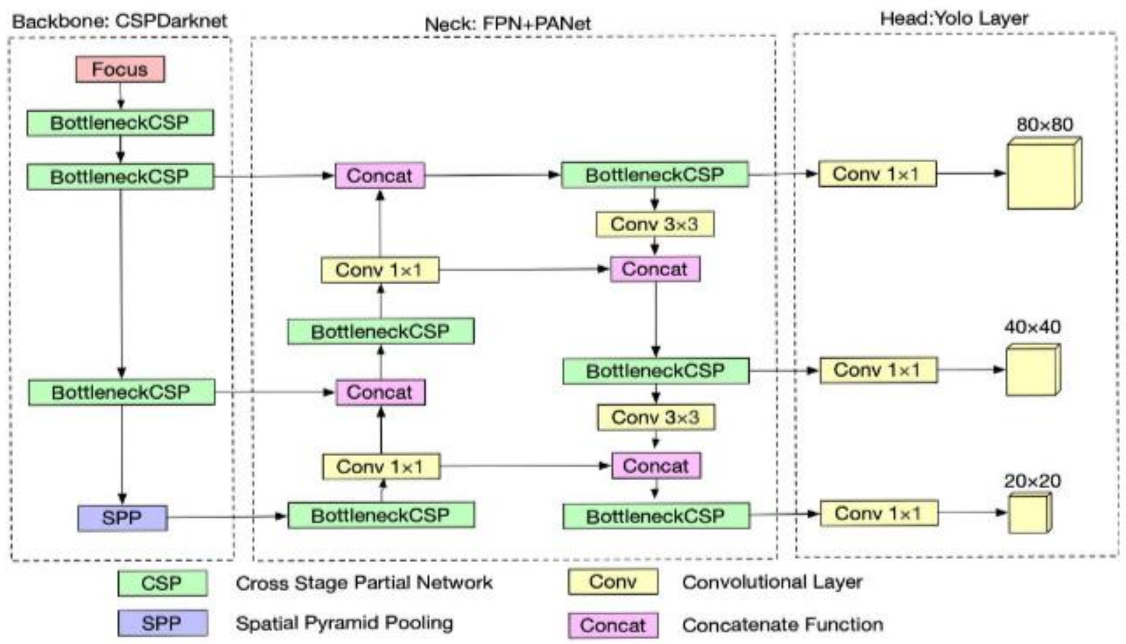
Input and Output		Feature of This Solution	Contribution & The Value of This Work
Input	**Output**	The ET-YOLOv5s solution integrates a tiny object detection module with features like FPN and PAN for enhanced detection of small objects. It optimizes network parameters using the BottleneckCSP and SPP	This work contributes by developing an efficient solution for detecting students' behaviors in classrooms, combining YOLOv5s with a tiny object detection module to enhance accuracy.
A dataset of real classroom surveillance videos	recognizing various student behaviors such as playing		

	mobile phones, sitting.	modules while employing the GIoU loss function to improve overall accuracy.	

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
The positive impact of the proposed solution, ET-YOLOv5s, in the project domain includes improved detection accuracy through high-resolution image generation and deep identification. It addresses data imbalance by emphasizing the annotation of underrepresented types, leading to enhanced model performance.	The negative impact of the proposed solution, ET-YOLOv5s, includes challenges related to unbalanced sample sizes, limited performance in distinguishing specific behaviors (like drinking and unknown actions), and the potential for misclassification or missed detection of student behaviors. Additionally, the solution's effectiveness is dependent on extended training time and sufficient diverse data collection, which could pose challenges if these resources are limited.

Analyse This Work by Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
The document introduces ET-YOLOv5s, a solution for recognizing students' behaviors in classrooms. It combines ESRGAN for high-definition image generation and a tiny object detection module with YOLOv5s for improved recognition accuracy, particularly for smaller objects.	The document discusses the use of tools such as the ET-YOLOv5s model (an improved version of YOLOv5s) and ESRGAN for enhancing recognition accuracy and detecting smaller objects in the classroom.	I. Introduction II. Proposed multiple behaviors recognition III. Solution IV. Experiments V. Discussion VI. Conclusion VII. References

Diagram/Flowchart



YOLOv5s network structure.

---End of Paper 14---

15		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/document/10245238	Yong Cui ,Hua Zou	video surveillance, student learning, abnormal behavior, status detection
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that needs to be solved**	**What are the components of it?**
The current solution, Video Surveillance Technology, utilizes cameras and	The solution's goal is to improve education quality by using video surveillance to	The system includes monitoring tools (cameras, sensors), real-time video

sensors to collect student behavior data, employing algorithms like facial and behavior recognition. This technology detects abnormal behavior states, offering insights into students' learning behavior for improved education quality.	monitor and analyze students' learning behavior, particularly in detecting abnormal behavior. It addresses the limitation of traditional methods by employing machine vision, computer vision, and machine learning to automate the identification of abnormal behavior states, facilitating proactive support for teachers and students.	processing for facial and behavior recognition, data analysis, video surveillance software selection, stable environmental settings, student privacy emphasis, system testing, security measures, and both quantitative and qualitative analysis.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

The proposed MS-DAYOLO framework improves the robustness and accuracy of object detection in cross-domain scenarios, making it a promising solution for real-world applications.

	Process Steps	Advantage	Disadvantage (Limitation)
1	Install Monitoring Equipment: Cameras and sensors are installed in classrooms or student dormitories to collect student behavior data.	Provides a means to collect student behavior data in real-time	Raises concerns about privacy and security issues related to the use of video surveillance technology.
2	Video Stream Processing: The collected video stream data is transmitted to the server for real-time processing.	Enables real-time analysis of student behavior, allowing for timely intervention and support	Requires significant computational resources and may result in false positives or negatives in behavior recognition.

Major Impact Factors in this Work			
Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
Detected Behavioral States of Students The outcome variable representing the states of students' behavior as detected by the video surveillance system, including normal and abnormal behavior	Install Monitoring Equipment, Video Stream Processing, Data Analysis | Variables manipulated or controlled by the researchers to assess their impact on the dependent variable.	NA	Video Surveillance Technology, Algorithm Technology, Privacy Measures Variables that explain the process or mechanism through which the independent variables impact the dependent variable..
Relationship Among The Above 4 Variables in This article			
the dependent variable is the "Detected Behavioral States of Students," representing the outcomes of the video surveillance system. The independent variables include the steps of installing monitoring equipment, video stream processing, and data analysis.			

Input and Output		Feature of This Solution	Contribution & The Value of This Work
Input	**Output**	This solution utilizes video surveillance to monitor and detect abnormal behavior states in students during learning. Cameras and sensors in classrooms collect behavior data processed through algorithms, enabling analysis of	This study contributes a video surveillance method using computer vision and machine learning to detect abnormal student behaviors, supporting timely intervention for
The input for the system includes video data collected from	The system collects and processes data through video surveillance		

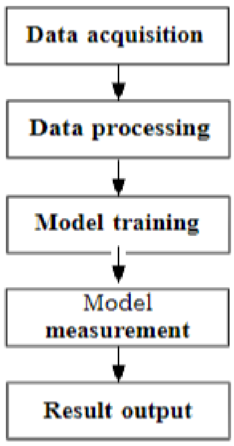
surveillance cameras installed in classrooms	and algorithm technology, and then outputs the detected behavioral states of students.	posture, facial features, and behavior. The system aids teachers in understanding students' learning situations, facilitating timely support.	teachers and schools. The practical application in education, emphasizing algorithm improvements and privacy protection, enhances teaching by enabling monitoring of abnormal behavior for better understanding and assistance.

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
The video surveillance method enhances student learning by accurately detecting abnormal behaviors through computer vision and machine learning. It improves education quality, ensures privacy, conducts thorough analyses for continuous improvement, and stays updated with the latest technologies. Overall, the solution facilitates timely support, maintaining privacy and relevance in education.	The use of video surveillance for monitoring students' learning behavior presents potential negative impacts. These include concerns about privacy violation, limited information capture affecting detection accuracy, challenges in algorithm adaptability to diverse learning scenarios, and ethical considerations related to anonymous data use.

Analyse This Work By Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
The document explores video surveillance for detecting abnormal student behavior, prioritizing privacy and security. It covers the paper's structure, methodology, application, system	The paper utilizes a comprehensive approach, integrating video surveillance, computer vision, machine learning, multi-face feature extraction, and deep learning to monitor and detect abnormal	I. Introduction II. Application of video surveillance in detecting abnormal behavior states of students III. Student

development, and testing, emphasizing clear problem statements, technology selection, and robust testing. The importance of protecting student privacy, utilizing anonymous data, and conducting both quantitative and qualitative analyses for evaluating algorithm effectiveness is highlighted.	student behaviors during learning. This involves real-time processing of data from cameras and sensors, with computer vision identifying facial, posture, and behavior patterns. Machine learning compares the data to normal behavior, and the study explores multi-face feature extraction algorithms.	behavior detection under video surveillance IV. System upsration test results V. Conclusions VI. References

Diagram/Flowchart



Behavioral State Detection Process

---End of Paper 15---

16		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/document/8460336/	Zahraa kain, Abir Y Ouness, Ismail EI Sayad,	Magnitude, normal, abnormal, optical flow, histogram of

	Samih Abdul-Nabi, Hussien Kassem.	magnitude, analyzes.
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
Awareness of normal and abnormal.	The aim is to findout the normal and abnormal events in the university by dividing the video frame into zones.	The system divides the video frame into zones, computes the magnitude of optical flow in each zone, and analyzes these data to classify it as normal or abnormal events based on a logical threshold.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	To findout the normal and abnormal events in the university by dividing the video frame into zones.	We can control the abnormal activities that are done by the students in the university premises.	Sometimes it may cause the disturbance for the privacy of the students.
2	The system divides the video frame into zones, computes the magnitude of optical flow in each zone, and analyzes these data to classify it as normal or abnormal events based on a	A pleasant environment will be maintained in the university without any abnormal activities of students.	May be it is not possible to moniter all the students in the university at a time.

	logical threshold.		

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
Normal and abnormal events detection. And event classification.	Video frame division into zones. Video frame zones.	Logical threshold.	NA

Relationship Among the Above 4 Variables in This article

Input and Output	Feature of This Solution	Contribution & The Value of This Work
Input: Video . Output: Normal and abnormal activities.	The system divides the video frame into zones, computes the magnitude of optical flow in each zone, and analyzes these data to classify it as normal or abnormal events based on a logical threshold.	Good to have this knowledge from this paper as we also want to restrict the abnormal activities of the students.

Positive Impact of this Solution in This Project Domain	Negative Impact of this Solution in This Project Domain
We can identify and solve unusual events that may pose a threat to the safety and security of students, faculty, and staff in university areas.	Sometimes it may cause the disturbance for the privacy of the students.Students may feel uncomfortable due to thhis

Analyse This Work by	The Tools That Assessed	What is the Structure of

Critical Thinking	this Work	this Paper
This work is good, as they are trying to restrict the abnormal activities of the students in the universities.	CCTV cameras.	I. Abstract II. Dividing the video frame into zones. III. Computing the magnitude of optical flow in each zone. IV. Analyzing these data and classifying it based on a logical threshold as normal or abnormal events. V. Conclusion

Diagram/Flowchart

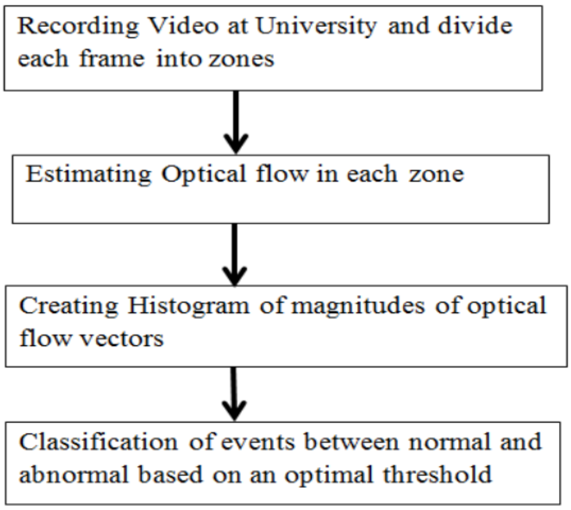


Figure 1: A detailed block diagram of proposed model

---End of Paper 16—

17		
Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**

https://ieeexplore.ieee.org/document/1541291	Oren Boiman, Michal Irani	Saliency, visual attention,chunks of data,query, spatiotemporal, overwhelmed.
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that needs to be solved**	**What are the components of it?**
The Ensemble Search Algorithm efficiently detects video patches through graph-based Bayesian inference, with constraints on arrangement and descriptors. It enables inference and generalization from a few examples to broader image patterns, providing a unified framework for various computer vision tasks, including attention, recognizing suspicious behaviors, and identifying unusual objects.	The solution aims to develop an approach that, from a few examples, can automatically detect irregularities and unusual patterns in images and videos without explicit definitions. It addresses limitations in existing methods tied to predefined object classes and centers, striving to overcome these constraints using a probabilistic graphical model and an efficient inference algorithm for detecting and generalizing from a small set of examples.	Observed variables,Hidden database variables,Bayesian dependencies,patch descriptors.
The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process		

	Process Steps	Advantage	Disadvantage (Limitation)

1	The paper introduces an approach leveraging a Graph-Based Bayesian Inference Algorithm to detect patches in video sequences and infer patterns from a few examples.	The paper introduces an approach for inferring and generalizing from a few examples to determine the validity of a broader context of image patterns and behaviors, even in the absence of specific, previously seen configurations.	The paper highlights a limitation in the current method, noting its reliance on predefined object classes and centers, making it unsuitable for detecting irregularities in scenarios without predefined object classes.
2	It redefines saliency and visual attention, presenting a unified framework for computer vision problems.	The paper introduces an efficient graph-based Bayesian inference algorithm for detecting large patches at multiple spatio-temporal scales. It imposes constraints on both the relative geometric arrangement and descriptors of the detected patches.	Previous methods used restrictive representations like trajectories or a single small descriptor vector for a frame, which were either too specific or too global, limiting effectiveness in detecting irregularities.

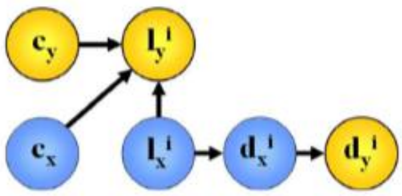
Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening) variable
Identification of irregularities.	Graph-Based Bayesian Inference Algorithm. Redefined saliency and visual attention.	NA	NA

<table>
<tr><th colspan="3">Relationship Among The Above 4 Variables in This article</th></tr>
<tr><td colspan="3"></td></tr>
<tr><th>Input and Output</th><th>Feature of This Solution</th><th>Contribution & The Value of This Work</th></tr>
<tr><td><table><tr><th>Input</th><th>Output</th></tr><tr><td>Video and input query.</td><td>unusual behaviors in the video sequence.</td></tr></table></td><td>The solution utilizes a probabilistic graphical model and inference algorithm to detect irregularities and infer new configurations efficiently from a database. It incorporates patch descriptors, geometric arrangement, multi-scale search, and patch composition for speed in handling large ensembles.</td><td>Good to have this knowledge from this paper as we also want to restrict the abnormal activities of the students.</td></tr>
</table>

<table>
<tr><th>Positive Impact of this Solution in This Project Domain</th><th>Negative Impact of this Solution in This Project Domain</th></tr>
<tr><td>The proposed solution positively impacts computer vision by enabling efficient inference, accurate patch detection, introducing a new interpretation of saliency and visual attention, and providing a unified framework for diverse problems.</td><td>The paper identifies limitations in current methods, such as constraints to predefined object classes, use of overly specific or global representations, complexity in the inference process, and reliance on simplifying statistical assumptions that may not always be applicable.</td></tr>
</table>

<table>
<tr><th>Analyse This Work By Critical Thinking</th><th>The Tools That Assessed this Work</th><th>What is the Structure of this Paper</th></tr>
<tr><td>This work is good, as they are trying to restrict the abnormal activities of the students in the universities.</td><td>Python, computer vision, machine learning.</td><td>I. Introduction
II. Inference by composition
III. Applications</td></tr>
</table>

		IV. Conclusion
Diagram/Flowchart		
 Figure 4. The probabilistic graphical model.		

---End of Paper 17—

18

Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/abstract/document/9340162	Ersin Elbasi	Integration for Intelligent Systems, Event recognition,Face, body, and hand detection, Fall detection,Convolutional neural network,Moving object activities
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**

Integrating the intelligent system is the current solution.	The solution uses IoT, machine learning, and probabilistic methods to automate the detection and recognition of moving objects in video for efficient surveillance in diverse environments, addressing limitations of manual observation.	Structure of Internet of Things for Surveillance Systems, Machine Learning Algorithms, Methodology, Literature Review.	
The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process			
	Process Steps	**Advantage**	**Disadvantage (Limitation)**
1	Environment modeling for each video taken from IoT, Motion segmentation using image difference method.	Provides a representation of the video environment.	May require significant computational resources.
2	Cluster-based object tracking.	Tracks the movement of objects over time.	May struggle with occlusions or objects that move out of frame.
Major Impact Factors in this Work			
Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
Detection and recognition of moving objects in video	Image difference method for motion segmentation	Presence of occlusions or objects moving out of frame in object tracking	NA
Relationship Among the Above 4 Variables in This article			

Input and Output		Feature of This Solution	Contribution in This Work
Input	**Output**		
The input data consists of video sequences captured from cameras, and the output is the recognition and	detection of human activities.	The solution combines low-level and high-level processing techniques, utilizes machine learning algorithms, integrates data from multiple cameras, and leverages IoT technology to accurately recognize and classify human activities in surveillance systems.	The work develops IoT surveillance systems using machine learning, addressing data and security challenges. It has valuable applications in security, healthcare, and wildlife monitoring, improving safety through accurate event recognition and real-time alerts.

Positive Impact of this Solution in This Project Domain	Negative Impact of this Solution in This Project Domain
The proposed solution in this project domain offers improved accuracy, real-time alerts, enhanced security, efficient monitoring, and high classification accuracy. These advantages contribute to the effectiveness and reliability of the surveillance system, making it a valuable tool for ensuring safety and security in various domains.	While IoT-based surveillance systems offer many advantages in terms of data collection and analysis, there are potential negative impacts that need to be considered and addressed to ensure the success and effectiveness of the system in this project domain.

Analyse This Work By Critical Thinking	The Tools That Assessed this Work	What is the Structure of this Paper
This work is good, as they are trying to findout the abnormal activities in the sequence of the video frames.	the paper utilizes edge computing, face/body/hand detection algorithms, fall detection algorithms, CNN, and various machine learning	I. Abstract II. Introduction III. Literature Review IV. Discussions

	algorithms for event detection and recognition in IoT surveillance systems.	V. Conclusion
Diagram/Flowchart		
 Fig 1. Structure of Internet of Things for surveillance systems		

--End of Paper 18--

19

Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/abstract/document/7297468	Yu-Ling Hsueh†, Nien-Hung Lin†, Chia-Che Chang†, Oscal T.-C. Chen§, and Wen-Nung Lie.	Abnormal events, Alert messages, Related work, Environmental settings and data collection methods, Model construction procedures, Features.
The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	**The Goal (Objective) of this Solution & What is the problem that need to be solved**	**What are the components of it?**
The current solution for smart home activity	The solution aims to detect abnormal events in a smart	The solution utilizes a Bayesian Network and graphical model to

recognition utilizes a hybrid user-assisted approach, incorporating a human preference model and active learning. This minimizes data annotation, improving efficiency by reducing manual labeling efforts for users.	home using non-wearable sensors like IP cameras. It addresses the challenge of accurately monitoring multiple users' activities by analyzing trajectories and voices, issuing alerts for potential abnormal events to enhance user safety and security.	detect abnormal events in a smart home. Constructed via search algorithms or expert configuration, the model undergoes parameter learning from historical data. Performance evaluation through experimental results ensures accurate detection within a reasonable response time.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	Related work , Environmental Settings and Data Collection .	Researchers have already focused on recognizing daily life activities using wearable sensors, providing a foundation for this work.	The related work may have limitations or biases that could affect the accuracy of the proposed model.
2	Model Construction Procedures.	Data collected from a smart home environment provides real-life scenarios for analysis.	The data collection process may have limitations or biases that could affect the generalizability of the model.

Major Impact Factors in this Work

Dependent Variable	Independent Variable	Moderating variable	Mediating (Intervening)

			variable
Detection of abnormal events in a smart home.	Active learning methods.	Quality of data collected from smart home sensors.	Features extracted from trajectories and voices.

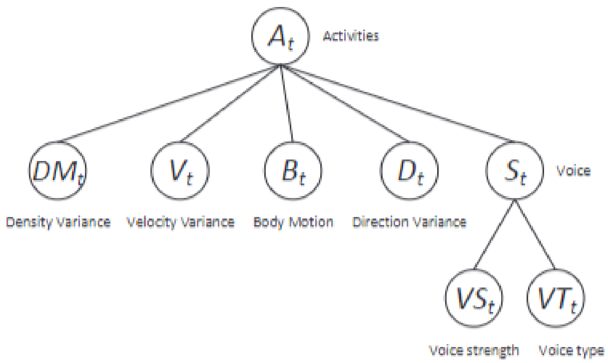
Relationship Among the Above 4 Variables in This article

Input and Output	**Feature of This Solution**	**Contribution & The Value of This Work**
<table><tr><th>Input</th><th>Output</th></tr><tr><td>Data collected from non-wearable sensors</td><td>Event detection model</td></tr></table>	This solution highlights the significance of feature selection before constructing a probabilistic model. The accuracy and performance of the detection model rely on the relevance and non-redundancy of the chosen features.	This work develops a Bayesian Network model using non-wearable IP camera data to detect abnormal events in smart homes, offering accurate alerts by analyzing ,user trajectories, voices, and videos.

Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**
The proposed solution offers several positive impacts in the project domain of activity recognition in smart homes. It improves accuracy, reduces data annotation efforts, enables real-time activity inference, ensures non-intrusive data collection, and achieves high precision and accuracy in event detection.	While the proposed solution offers the potential for accurate event detection in a smart home environment, it also presents challenges related to privacy, data coverage, system reliability, computational complexity, and user acceptance. These negative impacts need to be carefully considered and addressed to ensure the successful implementation and adoption of the system.

Analyse This Work By Critical Thinking	The Tools That Assessed this Work	What is the Structure of this Paper
This work is good ,as they are trying to develop a event detection system to analyse the abnormal activities.	Weka, Bayes network editor.	I. Abstract II. Introduction III. Literature review IV. The proposed model V. Results and discussions VI. Conclusion VII. References

Diagram/Flowchart



--End of Paper 19--

20

Reference in APA format		
URL of the Reference	**Authors Names and Emails**	**Keywords in this Reference**
https://ieeexplore.ieee.org/abstract/document/4379786	Fan jiang,ying Wu, Aggelos K.Katsaggelos.	Abnormal event detection, Surveillance video, Dynamic hierarchical clustering, HMM-based similarity, Spectral clustering, Overfitting problem.

The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc)	The Goal (Objective) of this Solution & What is the problem that need to be solved	What are the components of it?
The paper presents a novel dynamic hierarchical clustering (DHC) approach for improving abnormal event detection in surveillance video. It addresses overfitting issues from HMM-based methods.	The solution aims to detect abnormalities in surveillance videos by employing a clustering-based approach. It addresses the challenge of identifying potential threats in specific situations by automatically processing video streams and characterizing actions.	The paper proposes a clustering-based method with a multi-sample similarity measure for detecting abnormalities in surveillance video. Introducing a dynamic hierarchical clustering (DHC) method to address overfitting, the authors iteratively correct training errors. Experimental results show the method's improvement over a baseline, utilizing a single-sample similarity measure and spectral clustering.

The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process

	Process Steps	Advantage	Disadvantage (Limitation)
1	Initialization and Distances between two groups are calculated using specific equations.	Each trajectory is fitted with an HMM, providing a better representation of the data.	Initial overfitting may occur as each HMM is trained on just one trajectory.
2	Merging: The two groups with the smallest distance are merged into one if a	Distances between groups are calculated, allowing for comparison and merging.	The specific equations used may not capture all aspects of similarity.

	certain criterion is satisfied.		

Major Impact Factors in this Work			
Dependent Variable	**Independent Variable**	**Moderating variable**	**Mediating (Intervening) variable**
Detection of abnormal events in surveillance videos.	Use of Dynamic Hierarchical Clustering (DHC) method	Specific equations used in distances calculation	Features extracted from trajectories and voices

Relationship Among the Above 4 Variables in This article

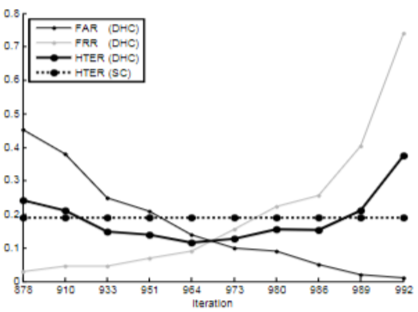
Input and Output		**Feature of This Solution**	**Contribution & The Value of This Work**
Input	**Output**		
Surveillance videos.	Abnormal event detection.	The proposed solution for abnormal event detection in surveillance video employs Dynamic Hierarchical Clustering (DHC), a robust algorithm that iteratively reclassifies and retrains data groups at different levels, correcting errors from overfitting. DHC's self-adjusting capability in both model training and data clustering enhances its reliability for abnormal event detection.	The DHC approach tackles overfitting in abnormal event detection using HMMs on multiple samples, correcting errors and showing superiority over a baseline in surveillance video experiments. Its self-adjusting capability in model training and data clustering adds substantial value to abnormal event detection.

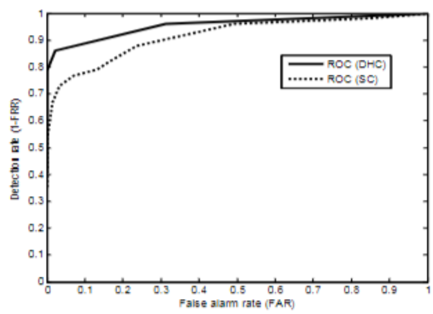
Positive Impact of this Solution in This Project Domain	**Negative Impact of this Solution in This Project Domain**

The DHC approach positively impacts abnormal event detection in surveillance by addressing overfitting, correcting errors, and outperforming baselines. Its self-adjusting capability enhances adaptability and system robustness, marking a valuable advancement in the project domain.	Negative impacts of the proposed solution include potential issues with model overfitting due to dataset limitations, data shortage leading to suboptimal performance, computational complexity, and sensitivity to similarity measures. These factors may affect the accuracy and efficiency of abnormal event detection in surveillance.

Analyse This Work By Critical Thinking	**The Tools That Assessed this Work**	**What is the Structure of this Paper**
The work is good, as they are trying to improve the detection of the abnormal activities in the surveillance video.	The study compared two abnormal event detection tools for surveillance videos: one with single-sample-based similarity using HMMs and spectral clustering, and the other utilizing multi-sample-based similarity with dynamic hierarchical clustering (DHC). The assessment aimed to gauge their performance in real surveillance video scenarios.	I. Abstract II. Introduction III. Related Works IV. Materials and Methods V. Proposed Fusion Method VI. Experimental Results and Discussion VII. Conclusions VIII. 8) References

Diagram/Flowchart





--End of Paper 20--

2.2COMPARISION TABLE:

Author	Year	Approach	Description
Oren Boiman, Michal Irani	2005	Graph-based Bayesian inference algorithm.	This work detects visual irregularities by assembling observed segments like a puzzle from database chunks, assessing likelihood based on ease of assembly. Utilizing probabilistic graphical inference, it is applied to saliency and suspicious behavior recognition.
Iulian Pruteanu-Malinici and Lawrence Carin	2007	Invariant Subspace Analysis(ISA), Infinite Hidden Markov Model (iHMM)	To overcome the linear transformation drawback which is not robust to change spatial position or phase Invariant Subspace Analysis(ISA). iHMM it can adopt complexity the model doesn’t require the same parameters to be passed each time. It explores and adopts to each set of observations.
Fan jiang, ying Wu, Aggelos K.Katsaggelos.	2007	Dynamic Hierarchical Clustering.	Dynamic Hierarchical Clustering is a clustering-based approach for detecting abnormalities in surveillance video. It addresses the overfitting problem by using a multi-sample-based similarity measure. The algorithm starts with initializing each trajectory as a group and fitting it with a Hidden Markov Model.

Lawrence Carin (SM’96–F’01) Iulian Pruteanu-Malinic	2008	Hidden Markov Models(HMM)	They developed a detection algorithm for video analysis that can identify unusual events in complex scenes by using IHMM (Infinite Hidden Markov Model), Bayesian Inference, Feature Extraction Methods (SIWs, ICA, ISA)
Mark Jäger, Christian Knoll, and Fred A. Hamprecht	2008	HMMs with Gaussian mixture models	The proposed model takes input high dimensional data and converts it into a simple form. It uses HMM with a Gaussian Mixture
Tao Xiang; Shaogang Gong	2008	Dynamic Bayesian Networks (DBN)	The approach in the article uses Dynamic Bayesian Networks (DBNs) to model behavior patterns. Each behavior pattern is represented using a DBN, which is trained on the training data using the expectation-maximization (EM) algorithm.
Yu-Ling Hsueh†, Nien-Hung Lin†, Chia-Che Chang†, Oscal T.-C. Chen§, and Wen-Nung Lie.	2015	Bayesian Network.	A Bayesian Network is a probabilistic graphical model that represents the probabilistic relationships among a set of random variables. It is a directed acyclic graph (DAG) that consists of nodes representing random variables and edges representing the dependencies between them.

Kothapalli Vignesh, Gaurav Yadav, Amit Sethi	2017	CNN, LSTM	CNN is used to extract features from each frame and it distinguishes between normal and abnormal frames. These features fed into LSTM which further processes.
C.V Amrutha; C. Jyotsna; J. Amudha	2018	CNN model (Deep Learning)	They proposed the system uses a pre-trained model called VGG-16(Visual Geometry Group), which is trained on the ImageNet dataset
Juan Du	2018	Object Detection(YOLO and CNN family)	This paper briefly discusses the current algorithms that are in use for object detection such as CNN, Faster RCNN, YOLO, YOLOv2
Xian Sun; Songhao Zhu; Songsong Wu; Xiaoyuan Jing	2018	Weak Supervised Learning	The article proposes a weakly supervised detection method based on temporal consistency for abnormal behavior. The method constructs temporal gram matrices for a given video sequence and forms behavior units based on human behavior's temporal consistency and smoothness.
Zahraa kain, Abir Y Ouness, Ismail EI Sayad, Samih Abdul-Nabi, Hussien Kassem.	2018	Optical flow, C++ using Open CV, Histogram of magnitudes.	Recording the video and dividing each frame into zones then estimating the optical flow in each zone and creating a histogram of the magnitude of optical flow vectors finally classifying the normal and abnormal activities based on an optimal threshold.
Dorcas Oladayo	2020	Histogram-based	The proposed solution is to detect

Esan; Pius. A. Owolawi; Chuling Tu		methods Convolutional neural network with long short-term memory (CNN-LSTM)	anomalous behavioral patterns in a university environment using the Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) model.
Rui Zheng; Fei Jiang; Ruimin Shen	2020	Faster R-CNN (Region-based Convolutional Neural Network) and Online Hard Example Mining (OHEM)	The article proposes an improved version of the Faster R-CNN algorithm for student behavior analysis. The algorithm uses a scale-aware detection head with different dilation rates to handle the large-scale variations in student behaviors. The article utilizes Online Hard Example Mining (OHEM) during training to address the class imbalances in the dataset
Ersin Elbasi.	2020	Naïve Bayes Classifier, Iterative Classifier Optimizer, Decision Tree , Random Forest, Multilayer Perception.	The paper introduces a machine learning-based surveillance system achieving over 96% accuracy in recognizing human activities from diverse cameras through the Internet of Things.
Hariharan S; J Daniel Pushparaj; Muthukumaran Malarvel	2022	DL and CV algorithms	This is done by DL and CV algorithms which would perform modules such as augmented expression reading, Gaze tracking with head positioning, facial recognition, and Eye tracking. The proposed system will give the analytical output in a graphical representation.

Lina Li; Minghan Liu; Liyan Sun; Yupeng Li; Nianfeng Li	2022	ET-YOLOv5 Along with ESRGAN (Enhanced Super-Resolution Generative Adversarial Network)	The proposed model uses the ET-YOLOv5s approach is a fast and effective solution proposed in the article for identifying multiple students' behaviors in a classroom environment. It is an improved version of the YOLOv5s (You Only Look Once) algorithm, which is implemented in Pytorch for fast and accurate detection.ESRGAN is a component used in the ET-YOLOv5s approach to generate high-resolution images from low-resolution ones.
Abhinandan Tripathi, Manish Kumar Gupta, Chaynika Srivastava, Pallavi Dixi, Shrawan Kumar Pandey	2022	Object Detection using YOLO	YOLO is used to detect objects which divide each frame into an NxN grid. It provides a confidence score for each frame. If an object is present in more than one grid then the center is selected by using Intersection over Union.
Yong Cui; Hua Zou	2023	Computer Vision and TensorFlow Framework	Computer vision is a key technology used in the proposed approach to detect abnormal behavior states during students' learning and The TensorFlow framework is utilized in the proposed approach for machine learning tasks..
Zouheir Trabelsi, Fady Alnajjar,	2023	YOLOv5	YOLO is a fast and accurate object detection algorithm which

Medha Mohan Ambali Parambil, Munkhjargal Gochoo, and Luqman Ali.			detects various student behaviors. It uses DeepSort Algorithm to continuously monitor individual students.

2.3 WORK EVALUATION TABLE:

	Work Goal	System's Components	System's Mechanism	Features /Characteristics	Performance	Advantages	Limitations /Disadvantages	Results
C.V Amrutha; C. Jyotsna; J. Amudha	Detect suspicious or normal activity in an academic environment using deep learning techniques	Features computed from video frames, Classifier for prediction	Deep learning approach for video surveillance	Real-time detection, alert message generation	Can differentiae suspicious behaviors, Alerts authorities	Requires large data for training, Affected by environmental factors	Not specified	Valuble contribution to video surveillance, Alert generation

Hariharan S; J Daniel Pushparaj; Muthukumaran Malarvel	Real-time student behavioral tracking in virtual video conferencing applications	Face identification (CNN), Gaze tracking, Head positioning, Eye tracking	Modular approach using computer vision & deep learning	Face identification using CNN, Gaze tracking with head positioning & eye tracking, Real-time virtual video conferencing applications.	Efficient tracking and monitoring	Increases student engagement & interaction	Requires high-performance computer with dedicated GPU	Comprehensive real-time behavioral tracking in virtual video conferencing applications
Dorcas Oladayo Esan; Pius. A. Owolawi; Chuling Tu	Detection of anomalous behavioral patterns using CNN-LSTM in surveillance	CNN for image feature extraction, LSTM for recognizing sequential	CNN extracts image features, LSTM recalls sequential patterns	CNN for feature extraction, LSTM for recognizing sequences, High computational cost, Accurac	High	Improved accuracy, Minimized false detections	Specific to dataset, High computational cost	Outperforms other detection methods with 86% accuracy

	system	patterns		y of 86%				
Juan Du	YOLO v2 for object detection using CNN	Object Localization, Object Classification, Grid Cells, Bounding Boxes.	Utilizes a grid-based approach for object identification	Speed of 45 frames per second (FPS), High detection accuracy, Generalization ability	High	Real-time object identification, High accuracy, Versatility	High computational overhead for CNN, Localization errors in YOLO	Achieves FPS of 155 and an mAP of 78.62, balancing speed and accuracy efficiently
Lawrence Carin (SM'96–F'01)	Unusual Event Detection Algorithm for Video Analysis	Shift-Invariant Wavelets (SIWs), Independent Component Analysis (ICA).	Bayesian inference, iHMM for state decomposition, ISA for feature extraction	Adaptability, Complexity handling, Temporal information capture	High	Flexibility, Scene complexity handling, No need for background removal	Computational complexity, Convergence issues, Background-related limitations	Model focusing on normal behavior with potential for unusual event detection
Abhinandan Tripathi, Manish	Object detection techniques based	YOLO-based object detection process	YOLO employs grid partitioning, boundin	Real-time object detection with simultan	Continuous optimization of loss functions and	Efficient real-time object detection.	Challenges in accurately localizing	Making the bounding box from the identified images.

Kumar Gupta, Chaynika Srivastava, Pallavi Dixi, Shrawan Kumar Pandey (2022)	on You Only Look Once (YOLO) algorithm and aims to improve accuracy.	Grid partitioning Bounding box prediction Classification within each grid	g box prediction, and classification within each grid for object detection.	eous classification and localization	YOLO architecture modifications aim to enhance accuracy, speed, and adaptability.		smaller objects.	
Iulian Pruteanu-Malinici and Lawrence Carin (2008)	Develop a method for video feature extraction effective in complex environment	Invariant Subspace Analysis (ISA), Infinite Hidden Markov Model (iHMM), Anomaly	ISA features for invariant feature extraction, iHMM trained on "normal" video data, detectio	Small dimension ISA features, stability in likelihood distribution, non-parametric state learning, efficient	NA	Improved anomaly detection, stable log-likelihood, non-parametric state learning	Implementation issues with shift-invariant wavelets, potential inaccuracies in VB	-.

	s involving multiple moving entities	Detection	n of unusual events through sequential features	training and testing with iHMM and ISA			implementation, limitations in gathering specific anomalous data	
Mark Jäger, Christian Knoll, and Fred A. Hamprecht(2008)	Identify errors in image sequences from industrial processes.	Principal Component Analysis (PCA)	Dimension Reduction (PCA) Model Selection	Unusual event detection	NA	Improved classification performance	Extremely unbalanced dataset	Improved industrial quality control for identifying abnormalities in image sequences.
Kothapalli Vignesh, Gaurav Yadav	Develop a trainable surveillance system for	Background subtraction, CNN, LSTM, Linear	Background subtraction using Gaussian mixture	Effective use of Gaussian mixtures for background subtracti	Recognition accuracy improvement demonstrated through	- Improved accuracy in video activity recognit	- Limited training data impacting generalization -	Recognition accuracy improvement demonstrated through experiment

, Amit Sethi (2017)	detecting abnormal events, especially in mobile assets like vehicles, with minimal video data.	SVM	s, CNN for spatial features, LSTM for temporal dependencies, Linear SVM for final classification	on - Importance of depth (using VGG-16) for spatial feature extraction - Integration of CNN, LSTM, and SVM for superior performance	experimentation	ion - Effective training, particularly for small datasets	Time-consuming training due to CNN and LSTM complexity - Dependency on background subtraction effectiveness - Subjectivity in feature extraction methods	ation
Zouheir Trabelsi, Fady Alnajjar, Medh	Develop a robust project management system	Database, User Interface, Server	Client-Server Architecture	Real-time collaboration, Task tracking, Reportin	Maximum performance	Enhanced collaboration	Initial implementation time	Increased productivity

a Mohan Ambali Parambil, Munkhjargal Gochoo, and Luqman Ali (2020)				g				
Zheng, Fei Jiang ,Ruimin Shen	Develop an intelligent system for automatically detecting student behaviors in	Faster R-CNN algorithm with ResNet-101	Scale-aware detection head Feature fusion strategy Online Hard Example Mining (OHEM)	Building a large-scale student behavior corpus - Using Faster R-CNN for student behavior analysis - Scale-aware detection	Improved accuracy in detecting small objects - Efficient handling of scale variations - Mitigation of class imbalances through	3.4% increase in mean Average Precision (mAP) - Practical applicability in real classroo	Time-consuming manual labeling for large-scale dataset - Potential challenges in comple	Impressive real classroom results with improved algorithms and large-scale dataset.

	recorded classrooms.			head - Feature fusion strategy	OHEM	ms	x scenarios - Resource-intensive methods	
Tao Xiang, Shaogang Gong	Automated surveillance video analysis, anomaly detection	Discrete-scene event detection - Spectral clustering - Generative behavior model - Online Likelihood Ratio Test	1. Behavior Representation: Data clustering and dimension reduction using spectral clustering. 2. Model Selection: Automatic model selection	Fixed-length feature vector representation - Automatic model selection for optimal cluster determination - Detection of abnormal behavior patterns - Real-time	Effective and robust anomaly detection with indoor and outdoor surveillance datasets	- Easier data clustering and analysis with fixed-length feature vectors - Automatic model selection reduces manual intervention - Discovering	Variable length of feature vectors may pose challenges in clustering - Unknown factor in the selection of the number of clusters - Clusteri	Positive impact on anomaly detection through robust behavior model.

			n using eigenvectors.	anomaly detection - Composite generative behavior model - Event-based behavior		natural grouping aids in building a model for normal behavior	ng problem with an unknown number of clusters	
Xian Sun, Songhao Zhu, Songsong Wu, Xiaoyuan Jing	Weakly supervised abnormal behavior detection utilizing temporal consistency	Temporal Gram Matrices - Candidate Action Fragments - Corresponding Classifier - Sparse Reconstruction	1. Construction of Timing Sequence: Efficiently constructs sub-sequences with fixed length and stride. 2. Temporal Consistency:	Temporal Gram Matrices for capturing temporal information. - Candidate Action Fragments for identifying start and end frames of abnormal behavior	Efficiently constructs sub-sequences. Temporal consistency explores Gram matrix. Weakly supervised learning of behavior sequence with a	Improved anomaly detection through effective feature extraction. - Unsupervised detection adaptable to different scenario	- Limited training data reliance on one ordered behavior sequence. - Dependency on artificial features. - Threshold setting challen	Improved detection accuracy in various scenarios.

			Explores Gram matrix to identify the best candidates for behavior units. 3. Weak Supervised Learning of Behavior.	. - Corresponding Classifier for training and classification. - Sparse Reconstruction for detecting abnormal behavior.	classifier.	s. - Higher detection accuracy for various abnormal behaviors. - Efficient and effective methods using temporal consistency and smoothness	ges for sparse reconstruction. - Limited comparison with supervised methods.	
Lina Li, Minghan Liu, Liyan Sun, Yupeng Li Nianf	Enhance identification of students' in-class behavi	Backbone, Neck, Head (Focus, CSPNet, SPP, FPN, PAN, BCELo	ET-YOLOv5s model trained with labeled dataset, incorporating	Real classroom surveillance videos, 826 image dataset, Modified	Improved detection accuracy of multiple student behaviors	Authentic data collection using real classroom videos - High-	Loss of temporal information during video-to-image convers	Success in recognizing multiple student behaviors simultaneously, promising for enhancing

eng Li	ors	gits, BCEclsloss, GIoU loss)	tiny object detection module	YOLOv5s network structure		resolution image generation - Improved model performance with tiny object detection	ion - Imbalance in labeled dataset distribution - Computational resource and time-intensive training	teaching quality and monitoring classroom behaviors
Yong Cui, Hua Zou	Improve education quality through Video Surveillance Technology	Monitoring tools (cameras, sensors), real-time video processing (facial and behavior recognition), data	Video Surveillance Technology uses cameras and sensors to collect student behavior data, employing facial and	Real-time data processing, continuous algorithm enhancement, comparison with similar studies	Improved detection accuracy of abnormal behavior states, timely intervention for teachers and students	Real-time analysis for timely intervention - Privacy and security concerns - Computational resource-intensive video	Loss of temporal information during video-to-image conversion - Imbalance in labeled dataset distribution - Comput	Success in recognizing multiple student behaviors simultaneously, showing promise for enhancing teaching quality and monitoring classroom behaviors.

		analysis, video surveillance software, stable environmental settings, privacy emphasis, system testing, security measures	behavior recognition algorithms. Focus on algorithm enhancement and standard research paper structure			stream processing - Potential false positives or negatives in behavior recognition	ational resource and time-intensive training	
Zahraa Kain, Abir Y Ounes s, Ismail El Sayad, Samih Abdul -Nabi,	Awareness of normal and abnormal events in university areas.	Video input, Video Frame Division into Zones, Magnitude of Optical Flow, Logical Thresho	Divides the video frame into zones, computes optical flow, classifies events as normal	Control of abnormal activities, Privacy concerns.	Not specified.	Safety and security enhancement, Restricts abnormal activities.	Privacy concerns for students, Potential discomfort, No specified mediating variable	Identification of normal and abnormal activities

Hussien Kassem (Year: Not specified).		ld.	or abnormal.				.	
Oren Boiman, Michal Irani (Year: Not specified).	Development of Ensemble Search Algorithm for Irregularity Detection in Images and Videos.	Graph-Based Bayesian Inference Algorithm, Redefined Saliency and Visual Attention, Ensemble Search Algorithm.	Detects video patches, infers patterns through graph-based Bayesian inference, handles large ensembles efficiently.	Unified framework for computer vision tasks, Recognition of suspicious behaviors, Identification of unusual objects.	Not specified.	Efficient inference, Accurate patch detection, Unified framework for diverse computer vision problems.	Challenges in explaining valid ensembles, Reliance on predefined object classes, Overly specific or global representation.	Identification of irregularities in images and videos.
Ersin Elbasi (Year: Not	Integration for Intelli	Internet of Things (IoT),	Automated detection and	Improved safety through real-time	Not specified.	Efficient monitoring,	Potential need for signific	Improved accuracy in event recognition

specified).	gent Systems - Detection and Recognition of Moving Objects in Video.	Machine Learning Algorithms, Methodology, Literature Review.	recognition of moving objects, Efficient surveillance, IoT, Machine Learning, Probabilistic Methods.	detection, High classification accuracy.		High classification accuracy, Real-time detection and recognition.	ant computational resources, Challenges in occlusion handling, Limitations in detecting complex activities.	, Enhanced security and safety in surveillance systems.
Yu-Ling Hsueh et al. (Year: Not specified).	Smart Home Activity Recognition - Abnormal Event Detection.	Bayesian Network, Graphical Model, Search Algorithms, Active Learning, Human	Hybrid user-assisted approach, Non-wearable sensors (IP cameras), Efficient data annotati	Improved efficiency through reduced manual labeling, Accurate detection within a reasonable response	Not specified.	Reduced manual labeling efforts, Accurate and real-time abnormal event detection, Minimi	Potential biases in related work, Limitations in data collection process, Expert knowledge	Bayesian Network model development, Accurate alerts through trajectory and voice analysis, Improved accuracy in smart

		Preference Model.	on, Real-time abnormal event detection.	time.		zed data annotation.	required for graphical model construction.	home event detection.
Fan Jiang, Ying Wu, Aggelos K. Katsaggelos (Year: Not specified).	Abnormal Event Detection in Surveillance Video.	Dynamic Hierarchical Clustering (DHC), HMM-Based Similarity, Spectral Clustering.	Iterative correction of errors from overfitting, Automatic calculation of the number of clusters, Multi-sample similarity measure	Superior effectiveness demonstrated in experimental results, Robustness against overfitting.	Not specified.	Robust abnormal event detection, Improved performance over baseline methods.	Potential issues with model overfitting, Sensitivity to similarity measures, Computational complexity, Data shortage.	Improvement over baseline in surveillance video experiments, Self-adjusting capability enhances reliability for abnormal event detection.

CHAPTER 3

PROPOSED SYSTEM

3.1 PROPOSED SYSTEM

The proposed system of the student tracking system integrates cutting-edge AI technologies to effectively monitor student behavior and enhance classroom management. At its core, the system utilizes YOLOv8, a state-of-the-art object detection algorithm, to accurately identify and track various activities such as sleeping in class, using mobile devices, or engaging in disruptive behavior. This algorithm is complemented by the utilization of Convolutional Neural Networks (CNNs), which are trained on a dataset of behavioral patterns to enable robust and real-time analysis. Furthermore, the system incorporates computer vision tools like OpenCV for image processing, enabling the extraction of meaningful features from visual data captured by surveillance cameras installed in classrooms and college premises. These features are then fed into the YOLOv8 model for object detection and behavior classification. The architecture also includes a centralized dashboard interface accessible to educators and administrators, providing real-time insights into student behavior, attendance records, and incident detection.

3.2 ADVANTAGES OF PROPOSED SYSTEM

The proposed AI-powered student tracking system offers several advantages over the existing manual methods and basic surveillance systems:

1. Advanced Behavior Analysis: The AI-powered system utilizes sophisticated algorithms to accurately analyze student behaviors in real-time. It can detect various behaviors such as sleeping in class, using mobile phones, and engaging in disruptive activities, providing educators with valuable insights into student conduct.

2. Automated Attendance Management: Unlike manual attendance tracking methods, the proposed system automates the process of attendance management using smart technology. Specialized cameras and AI algorithms enable automatic identification of students present in the classroom, reducing the burden on educators and ensuring accurate attendance records.

3. Enhanced Incident Detection: The system's AI-driven analysis enables rapid detection of incidents, including cases of property damage or conflicts among students. Real-time alerts and

notifications ensure that educators and administrators can respond promptly to potential threats or disruptions, thereby enhancing campus safety.

4. Comprehensive Dashboard: The user-friendly interface of the system provides educators and administrators with a centralized dashboard for monitoring student behavior, attendance records, and documented incidents. This comprehensive view enables informed decision-making and proactive intervention strategies.

5. Adaptability and Continuous Improvement: The system's adaptability allows it to evolve over time through data analysis and user feedback. By continuously refining its AI algorithms and processes, the system can effectively respond to changing educational needs and evolving student behaviors, ensuring its relevance and effectiveness in the long term.

Overall, the proposed AI-powered student tracking system offers a comprehensive solution to transform behavior analysis, attendance management, and incident detection in educational institutions, fostering a positive learning environment and supporting student success.

3.3 SYSTEM REQUIREMENTS

3.3.1 SOFTWARE REQUIREMENTS

Below are the software requirements:

1. Python (version 3.x recommended)
2. Twilio module
3. Label Img / Roboflow
4. OpenCV
5. Face Recognition Module
6. YOLOv8
7. dlib library
8. ONNX runtime library
9. pillow library

10. Ultralytics

3.3.2 HARDWARE REQUIREMENTS

Hardware requirements for successful development and implementation are as follows:

1. Raspberry Pi
2. IP enabled Camera

3.3.3 IMPLEMENTATION TECHNOLOGIES

You Only Look Once (YOLO)

Deep learning models, like YOLOv8, are now fundamental in several fields, including robotics, video surveillance, and autonomous driving. Using computer vision techniques and machine learning algorithms, the Yolov8 architecture can detect and localize objects in pictures and videos with remarkable speed and precision. These are the YOLO architecture's step-by-step characteristics. The first input in this process is a photograph, which travels through a convolutional neural network's "backbone." The relevant features are extracted by the backbone network

The neck, a component of the model, receives features gleaned from the backbone. The features are combined and examined on several scale dimensions to provide a sufficient comprehension or depiction of the image's content. This v8 model employs an anchor-free prediction, in contrast to the previous iterations of YOLO that use anchor-based detection. This involves predicting the object's center point directly rather than using pre-established anchor boxes, which is good because it increases efficiency and reduces the need for bounding box prediction forked Class and bounding boxes.

YOLOv8 divides the input image into cells, creating a grid-like structure. The bounding boxes and class probabilities for each bounding box are predicted by each grid cell. These odds indicate the likelihood that a certain item class—such as a car or person—will be found in that box. Several bounding boxes can overlap in an image since YOLOv8 predicts numerous bounding boxes for the same object. YOLOv8 use the NMS technique to choose the most likely bounding box for each object and discard many boxes with a high boundary overlap to prevent this and get rid of a lot of redundant bounding boxes. The result of YOLOv8 is an image with bounding boxes formed around observed items, class names, and confidence scores.

CNN:

Convolutional Neural Networks (CNNs) are pivotal in object detection, serving to extract meaningful features from input images. These networks, often deployed as backbone architectures like ResNet or MobileNet, learn hierarchical representations of visual patterns through convolutional layers. In two-stage object detection frameworks such as Faster R-CNN, CNNs also facilitate region proposal generation, identifying potential object locations via region proposal networks (RPNs). Subsequently, CNNs contribute to both object classification and localization tasks, predicting object categories and refining bounding box coordinates. With pre-training on large datasets followed by fine-tuning on detection-specific data, CNNs adapt to diverse object detection tasks. Their ability to automatically learn discriminative features enables the accurate detection and localization of objects within images, making CNNs indispensable in modern object detection systems.

CHAPTER 4

SYSTEM DESIGN

4.1 PROPOSED SYSTEM ARCHITECTURE

The presented system architecture aims to assess irregular activities, such as students sleeping in classrooms or using mobile phones during lessons, to uphold the quality of education. Initially, data is gathered from cameras or other sensors. For model evaluation, datasets like "CIVAR" can be employed. Key processes like image processing and video capture are facilitated by the "OpenCV" module.

Subsequently, data pre-processing tasks, including cleaning, normalization, and other transformations, are performed. The prepared data is then fed into the YOLO algorithm. This algorithm divides images into frames to locate target objects. Upon finding an object, the system tracks it. If the item is not located immediately due to potential concerns such as image overlaps, a solution known as "Intersection Over Union (IOU)" is used, followed by a new search for the target object.

After successful object tracking, the system evaluates whether the individual being tracked is participating in any inappropriate behavior. If an abnormality is discovered, the system notifies higher-ups. If no misbehavior is detected, the system will continue to track the individual. These phases are repeated repeatedly throughout the procedure.

4.2 MODULES

On an overall involves four main modules, which cater to the four main functions of this implementation, i.e., to identify children and provide interactive learning for childern and also to provide real time feedback and awareness to the society of major issues taking place now a days.

4.2.1 Tensorflow

This module involves TensorFlow which is an open-source machine learning framework developed by Google Brain, designed to facilitate the creation, training, and deployment of deep learning models. At its core, TensorFlow operates on a computational graph paradigm, where mathematical operations are represented as nodes and data flow through the edges. This graph-based approach enables efficient execution across multiple devices, including CPUs, GPUs, and TPUs, making TensorFlow suitable for a wide range of applications, from research prototypes to large-scale production systems. With its flexible architecture and extensive library of pre-built components, TensorFlow supports various machine learning tasks, including image recognition, natural language processing, and reinforcement learning.

One of the key features of TensorFlow is its scalability and distributed computing capabilities, allowing users to train complex models on large datasets distributed across multiple machines. TensorFlow also provides high-level APIs like Keras, which simplifies the process of building and training neural networks, making it accessible to both beginners and experienced developers. Additionally, TensorFlow's ecosystem includes tools for model debugging, visualization, and deployment, such as TensorFlow Serving and TensorFlow Lite, enabling seamless integration into production environments and edge devices.

Furthermore, TensorFlow fosters a vibrant community of researchers, developers, and practitioners, who contribute to its continuous improvement and expansion. Through initiatives like TensorFlow Research Cloud and TensorFlow Hub, users can access pre-trained models, datasets, and computational resources, accelerating the development and deployment of AI solutions. Overall, TensorFlow has emerged as a leading framework in the field of machine learning, empowering individuals and organizations worldwide to harness the power of deep learning for diverse applications, ranging from healthcare and finance to robotics and autonomous vehicles.

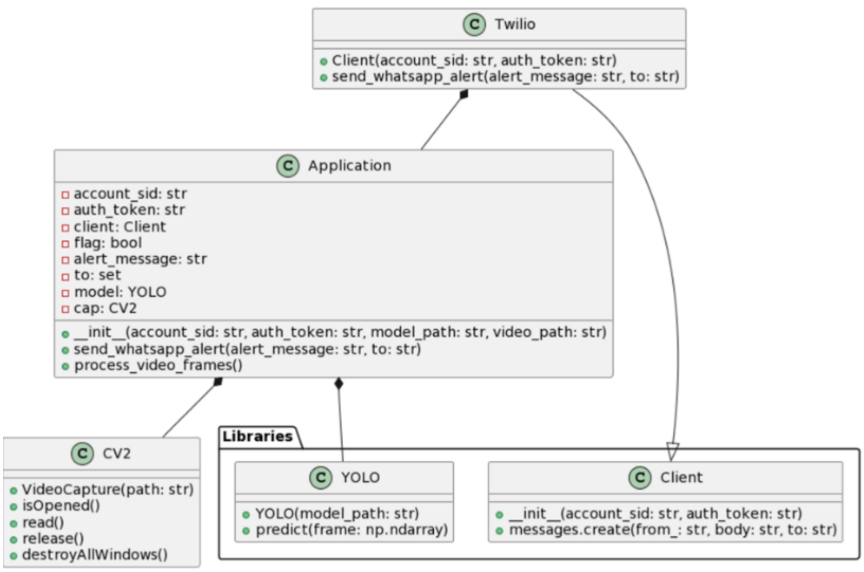


Figure 1: Use case Diagram of Proposed System

4.2.2 Ultralytics:

Ultralytics is a prominent company specializing in computer vision and deep learning solutions, particularly renowned for its development and open-source contributions to state-of-the-art object detection algorithms. Led by a team of expert researchers and engineers, Ultralytics has gained recognition for its cutting-edge work in the field of artificial intelligence (AI), focusing on empowering organizations across various industries with advanced visual recognition capabilities. The company's flagship product, the PyTorch-based YOLOv5 framework, stands out for its exceptional speed, accuracy, and ease of use, making it a popular choice for real-time object detection tasks. Furthermore, Ultralytics is committed to fostering collaboration and knowledge sharing within the AI community, evidenced by its extensive documentation, tutorials, and active participation in academic conferences and workshops. With a dedication to innovation and accessibility, Ultralytics continues to push the boundaries of computer vision technology, driving advancements that have profound implications for diverse applications, including autonomous vehicles, surveillance systems, and wildlife conservation efforts.

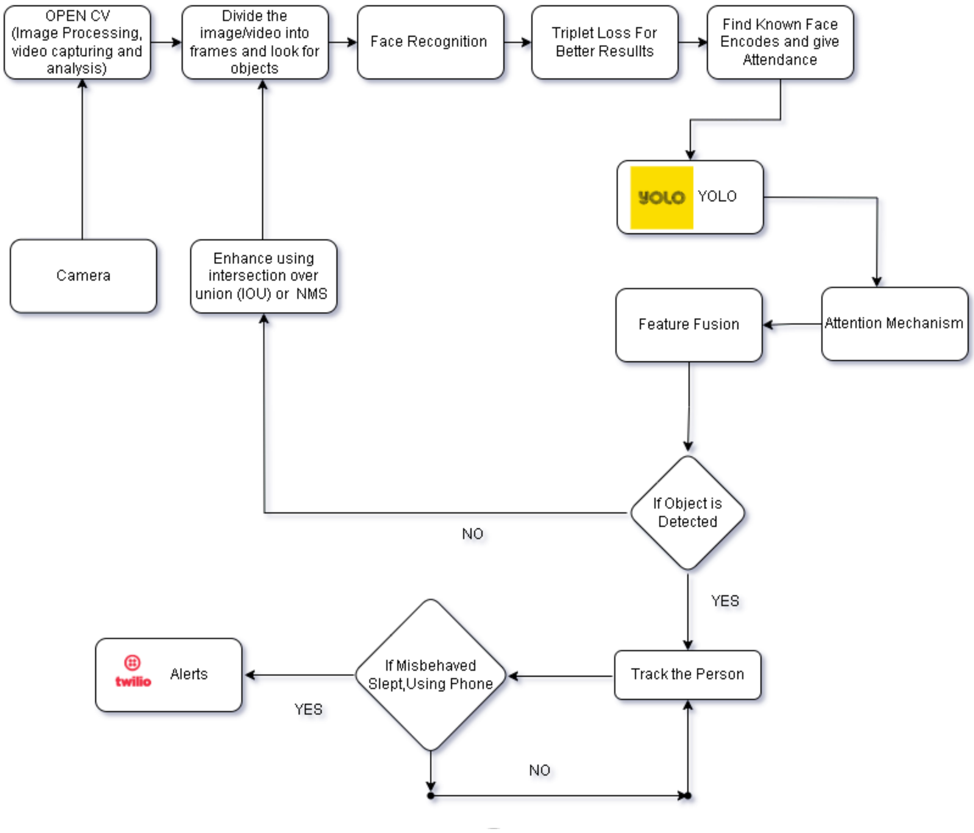


Fig 2: Activity diagram of the proposed system

4.2.3 Importing YOLO from Ultralytics:

Importing YOLO (You Only Look Once) from Ultralytics involves leveraging cutting-edge object detection capabilities for various applications. Ultralytics offers an open-source implementation of YOLO, a state-of-the-art deep-learning algorithm renowned for its speed and accuracy in detecting objects within images and videos. By integrating YOLO into projects, users can efficiently identify and track objects of interest in real time, making it invaluable for tasks such as surveillance, autonomous vehicles, and wildlife monitoring. Importing YOLO from Ultralytics provides access to pre-trained models trained on large-scale datasets, enabling rapid deployment without the need for extensive training data or computational resources. This accessibility democratizes the use of advanced computer vision technologies, empowering developers and researchers to tackle diverse challenges with ease.

Moreover, importing YOLO from Ultralytics offers seamless integration with popular deep-learning frameworks like PyTorch, facilitating flexible customization and extension for specific use cases. With its user-friendly interface and comprehensive documentation, Ultralytics' implementation of YOLO simplifies the development process, allowing users to focus on application-specific tasks rather than grappling with complex technical details. Additionally, Ultralytics provides ongoing support and updates, ensuring that users have access to the latest advancements and improvements in object detection technology. This commitment to continuous innovation enhances the reliability and performance of YOLO-based solutions, fostering confidence in its capabilities across diverse industries and domains.

Furthermore, importing YOLO from Ultralytics opens up possibilities for collaborative research and development initiatives, as the open-source nature of the project encourages knowledge sharing and community engagement. Developers and researchers can leverage the extensive ecosystem surrounding YOLO to exchange ideas, share best practices, and collaborate on addressing common challenges. By fostering a vibrant and inclusive community, Ultralytics promotes innovation and progress in the field of computer vision, ultimately driving advancements in object detection and related applications. In summary, importing YOLO from Ultralytics offers a powerful and accessible solution for harnessing the capabilities of deep learning in object detection, paving the way for transformative applications across diverse industries and domains.

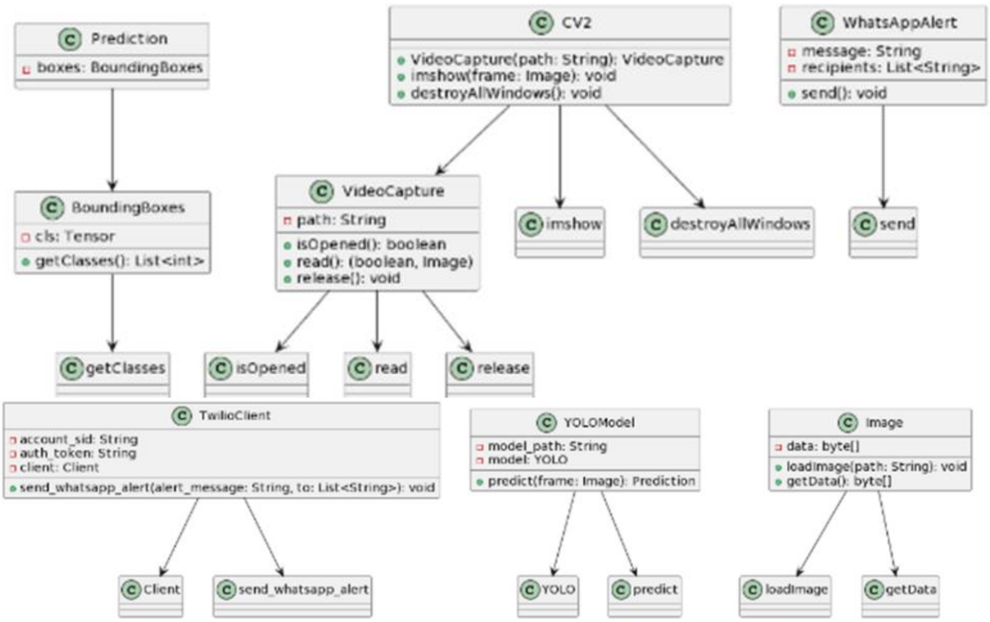


Fig 3: Class diagram of the proposed system

4.2.4 Twilio:

Twilio is a cloud communications platform that enables developers to integrate various communication channels, such as voice, messaging, and video, into their applications through simple APIs (Application Programming Interfaces). Founded in 2008, Twilio has become a leading provider of communication tools for businesses worldwide. Its platform offers scalable and reliable solutions that empower developers to build custom communication experiences tailored to their specific needs.

One of Twilio's core offerings is its Programmable Communications Cloud, which provides a suite of APIs for voice calls, SMS messaging, video conferencing, and more. Developers can easily embed these communication capabilities into their applications, allowing for seamless interactions with users across multiple channels. Twilio's APIs abstract away the complexities of telecom infrastructure, enabling developers to focus on building innovative communication features without having to manage the underlying infrastructure themselves.

In addition to its communication APIs, Twilio offers a range of other products and services to enhance the customer experience and streamline business operations. These include Twilio Flex, a fully programmable contact center platform; Twilio SendGrid, a cloud-based email delivery service; and Twilio Segment, a customer data platform for gathering, analyzing, and acting on user data.

Twilio's platform is highly flexible and customizable, making it suitable for businesses of all sizes and industries. Whether it's enabling real-time customer support, automating marketing campaigns, or enhancing collaboration among team members, Twilio provides the tools and infrastructure needed to create engaging and personalized communication experiences. With its commitment to innovation and customer success, Twilio continues to shape the future of communication technology and drive digital transformation across industries.

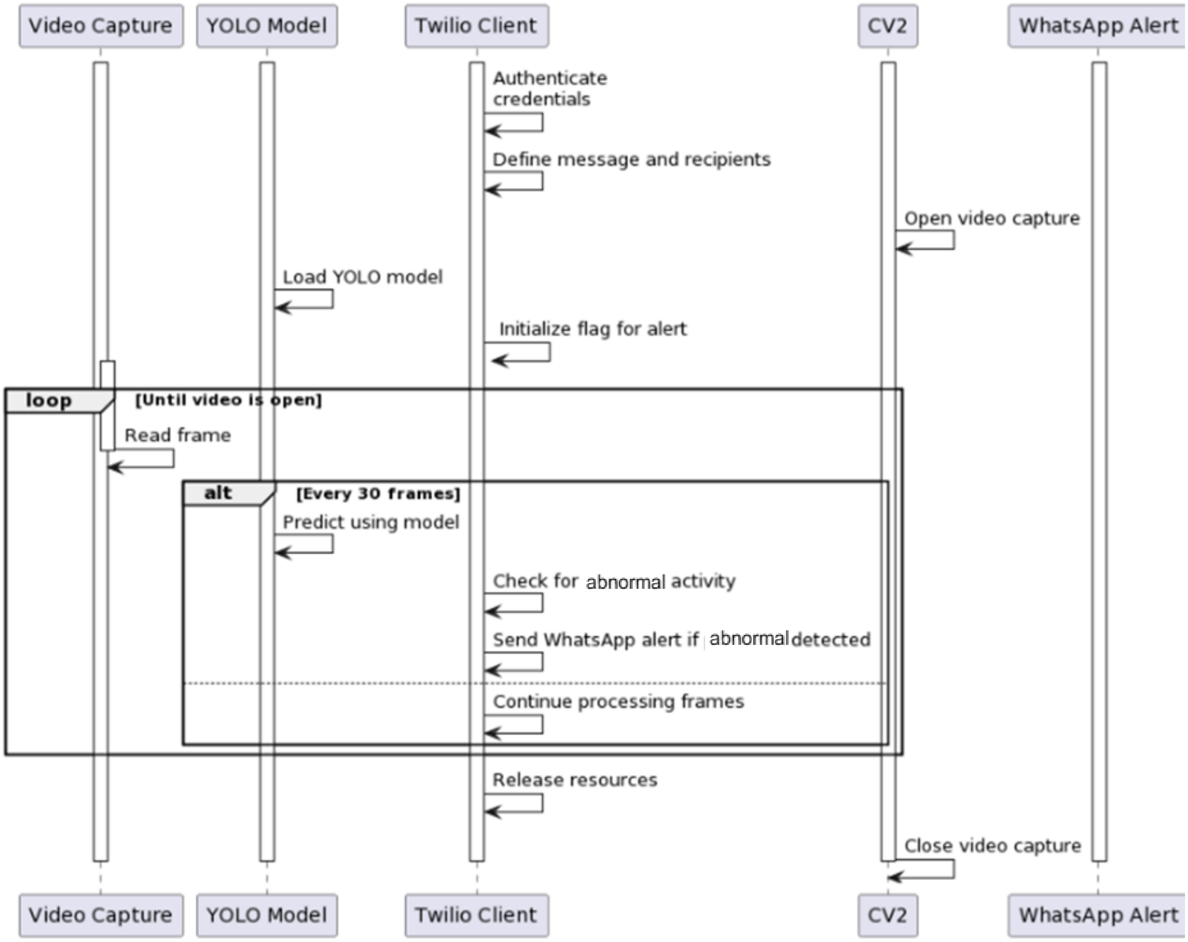


Fig 4: Sequence diagram of the proposed system

4.2.5 OpenCV:

OpenCV, short for Open Source Computer Vision Library, is a widely used open-source computer vision and machine learning software library. Originally developed by Intel, it now boasts a large community of contributors and users. OpenCV provides a comprehensive suite of functionalities for tasks such as image and video analysis, object detection and recognition, facial recognition, motion tracking, and augmented reality. It supports multiple programming languages including C++, Python, Java, and MATLAB, making it accessible to a broad range of developers and researchers.

One of the key strengths of OpenCV lies in its extensive collection of pre-trained models and algorithms, which enable rapid development and deployment of computer vision applications. These models cover a wide range of tasks, from basic image processing operations like filtering and thresholding to more advanced tasks like feature extraction and deep learning-based object detection. Additionally, OpenCV integrates seamlessly with other popular libraries and frameworks such as TensorFlow and PyTorch, allowing for interoperability and flexibility in building complex machine-learning pipelines.

Another notable feature of OpenCV is its platform independence, with support for various operating systems including Windows, Linux, macOS, Android, and iOS. This versatility enables developers to deploy their applications across different devices and environments, from desktop computers to mobile devices and embedded systems. Furthermore, OpenCV provides optimized implementations of its algorithms for different hardware architectures, leveraging parallel processing capabilities to achieve high performance and efficiency.

In recent years, OpenCV has become an indispensable tool in numerous domains including robotics, autonomous vehicles, medical imaging, security and surveillance, and augmented reality. Its ease of use, extensive documentation, and active community support have contributed to its widespread adoption and continued evolution. As computer vision continues to play an increasingly important role in various industries and applications, OpenCV remains at the forefront, driving innovation and empowering developers to turn their ideas into reality.

4.3 UML Diagrams

UML stands for Unified Modelling Language. UML is a standardized general-purpose modelling language in the field of object-oriented software engineering. In its current form, UML comprises of two major components: a Meta-model and a notation. The Unified Modelling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modelling and other non-software systems. The UML uses mostly graphical notations to express the design of software projects.

4.3.1 Use Case Diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

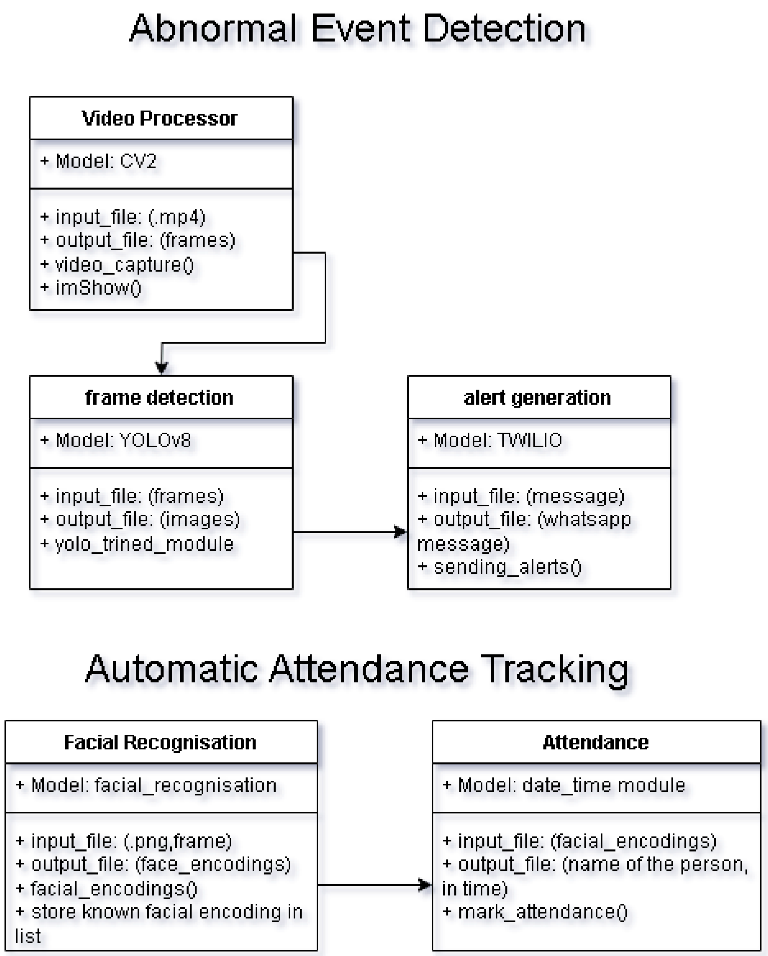


Figure 5: Use Case Diagram

CHAPTER 5

IMPLEMENTATION

5.1 BRIEF EXPLANATION OF IMPLEMENTATION

The existing problem can be solved by designing an artificial intelligence-based Student trackig system to analyse their behaviour i.e. an artificial intelligence model that detects sleeping, violence activity and using phone activity and immediately alerts the corresponding action-taking department.

1. Data collection and Data preprocessing
2. Implementing the YOLOv8 Model
3. Developing and training the model
4. Testing the model

5.1.1 DATA PREPARATION AND PREPROCESSING

Data is gathered for the initial data pipeline from a variety of internet sources, including youtube. We separate the video into the necessary sources that we find useful to our model and gather its frames, and we separate the necessary frames for identification. We proceed to analyze them before storing them in a database. Next, we obtain data sources from kaggle or roboflow for the second data pipeline. The added benefit of using kaggle or roboflow is that you have labeled data that does not require manual annotations, making our work faster and more effective. In the third data pipeline, to use the facial recognition feature of this system, which requires student faces for training, we scrape the college or school website to obtain the student faces. This allows us to use the system's automated attendance.

Combining all of the data pipelines results in a corpus of student face images for attendance tracking and violence data images that can be used for training purposes to detect abnormal network activity and training images to detect student use of mobile phones and dozing off in class.

5.1.2 IMPLEMENTING THE YOLOV8 MODEL

Implementing the yolov8 model involves several key steps to ensure the effective detection of guns and poachers. Firstly, the training data is organized and prepared by creating a .yaml file. This file includes the addresses of the training, testing, and validation data, along with information about the classes present in the dataset. Once the data is set up, the yolo model is imported from the ultralytics package, leveraging its powerful capabilities for object detection tasks. With the model imported, training begins using the collected dataset. During training, the yolov8 model learns to identify and localize guns and poachers within the images, optimizing its detection performance through iterative adjustments. By meticulously following these steps, the yolov8 model can be effectively trained to detect abnormal activity, using phone , sleeping activity contributing to improve the quality of education.

5.1.3 TRAINING AND DEVELOPING THE MODEL

Training and developing the model is a critical phase that significantly influences its performance. Several key considerations are vital throughout this process. Firstly, selecting an appropriate number of epochs is crucial to strike a balance between effective training and avoiding overfitting. Epoch selection involves iteratively training the model on the dataset, with each epoch representing a complete pass through the entire dataset. Additionally, model configuration plays a pivotal role in optimizing performance. Parameters such as batch size and image size are adjusted to fine-tune the model's behavior and improve its ability to detect guns and poachers accurately. Lastly, model evaluation is essential to assess its performance objectively. By evaluating the trained model on separate validation and test datasets, its ability to generalize and accurately detect poaching activity can be thoroughly assessed. These considerations collectively ensure that the trained model is robust, effective, and reliable in addressing the conservation challenges posed by poaching

5.2 SOURCE CODE

```
from ultralytics import YOLO
import twilio
from twilio.rest import Client
import cv2
import time # Import the time module
camera_label='CSM-A Camera --->'

account_sid = 'AC8836dfeb51a6f5ea7f0b97cf4e7b2696'
auth_token = 'ade6965347c16992392b4af33c216e2d'
twilio_phone_number = 'whatsapp:+14155238886'
```

```
cooldown_period = 60
recipient_phone_numbers = ['whatsapp:+919502152068' , 'whatsapp:+919000608068' , 'whatsapp:+919441841865']

client = Client(account_sid, auth_token)


def send_whatsapp_alert(alert_message, to):
	message = client.messages.create(
		from_=twilio_phone_number,
		body=alert_message,
		to=to
	)
	print(f" Message SID: { message.sid}")

def send_alerts(recipients, alert_message):
	print("sending alerts", alert_message)
	for recipient in recipients:
		send_whatsapp_alert(alert_message, recipient)

messages = {
	"sleep": "{} Student is sleeping in class 😴, Check the cameras immediately!".format(camera_label),
	"phone" : "{} Student is using phone 📲 in class, Check the cameras immediately !".format(camera_label),
	"violence" : "{} Violence activity Detected ❌, Check the cameras immediately !".format(camera_label)
}
send_status = { "sleep": False, "phone": False, "violence": False}

to = 'whatsapp:+918179533097'
```

```
# model_path = "/Users/apple/Desktop/final_year_project/yolov8_custoum/Yolo_runs_custom-2/Ravi_best.pt"
model_path="/Users/apple/Desktop/final_year_project/yolov8_custoum/Yolo_runs_custom-2/Major.pt"
model = YOLO(model_path)

# video_path = "/Users/apple/Desktop/final_year_project/yolov8_custoum/Yolo_runs_custom-2/fight2.mp4"
# cap = cv2.VideoCapture(video_path)
cap=cv2.VideoCapture(0)

while cap.isOpened():
	ret, frame = cap.read()
	if not ret:
		print("Error: Failed to grab a frame.")
		break
	# Reset message sent flags for each frame
	activity_messages_sent = {activity: False for activity in activity_messages_sent}
	res = model.predict(frame)
	res = res[0]
	res_tensor = res.boxes.cls
	res_numpy = res_tensor.numpy()

	for i in res_numpy:
		if(i==1 and i==2 and i==4):
			send_alerts(recipient_phone_numbers, messages["violence"])
			send_alerts(recipient_phone_numbers, messages["phone"])
			send_alerts(recipient_phone_numbers, messages["sleep"])
			send_status["sleep"] = True
			send_status["phone"] = True
			send_status["violence"] = True
			break
		if i == 4 and not send_status["sleep"]:
```

```
            send_alerts(recipient_phone_numbers, messages["sleep"])
            send_status["sleep"] = True
        if i == 1 and not send_status['phone']:
            send_alerts(recipient_phone_numbers, messages["phone"])
            send_status["phone"] = True
        if i == 2 and not send_status['violence']:
            send_alerts(recipient_phone_numbers, messages["violence"])
            send_status["violence"] = True
        if i not in [1, 2, 3]:
            print("No alert needed")
        # Display the frame with bounding boxes
        cv2.imshow('Object Detection', frame)
        if cv2.waitKey(1) & 0xFF == ord('q'):
            break
cap.release()
cv2.destroyAllWindows()
```

CHAPTER 6

RESULTS

The figures below illustrate the metrics provided by the model. Figure 8 depicts the confusion matrix resulting from model training, as well as the number of classes trained. Metrics show anticipated values on the left (y-axis) and true values on the x-axis. The confusion matrix is a table with rows representing real classes and columns representing anticipated classes. The confusion matrix's diagonal members reflect the number of correctly classified cases for each class, whereas the off-diagonal elements represent misclassification. A multiclass classification task will contain many rows. columns corresponding to each class.

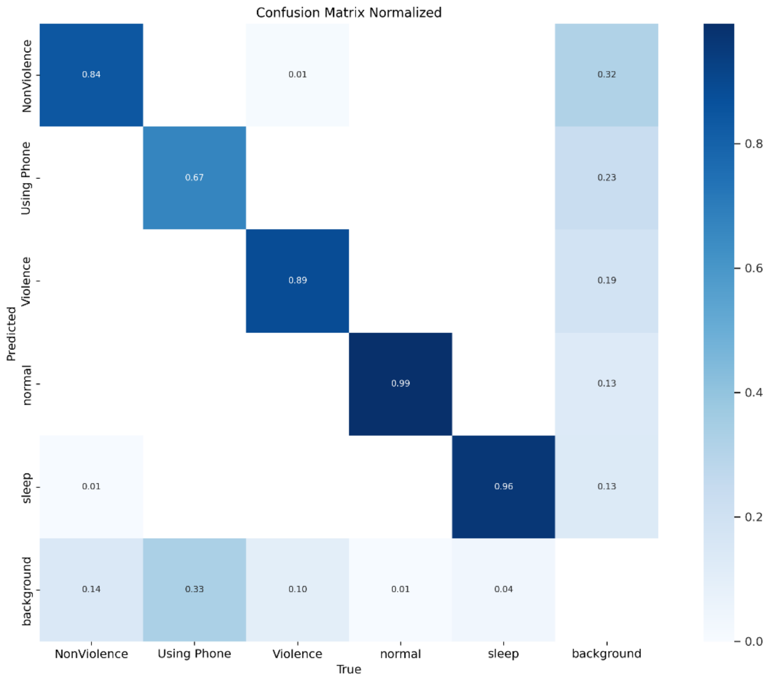


Figure 6: Confusion Matrix Normalized

Another measure in Figure 9 depicts the losses that are represented in various losses, such as trainbox loss and valbox loss, which frequently refer to the loss generated by bounding box regression. This loss represents an error in predicting the bounding box's coordinates. The terms train/cls loss and val/cls loss refer to the loss associated with object classification. This loss represents inaccuracies in predicting the items' class labels. They are steadily shrinking, demonstrating that the model is learning from its mistakes. On the other hand, the precision and recall curves are expanding exponentially, demonstrating that accuracy is always improving.

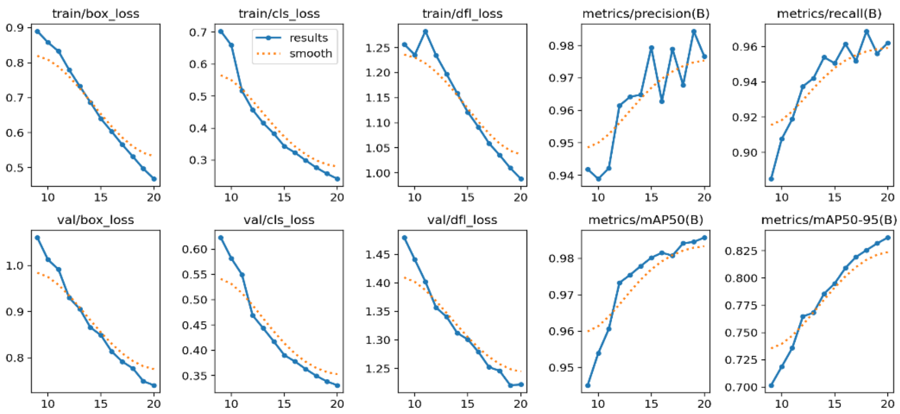


Figure 7: Different losses in the phases

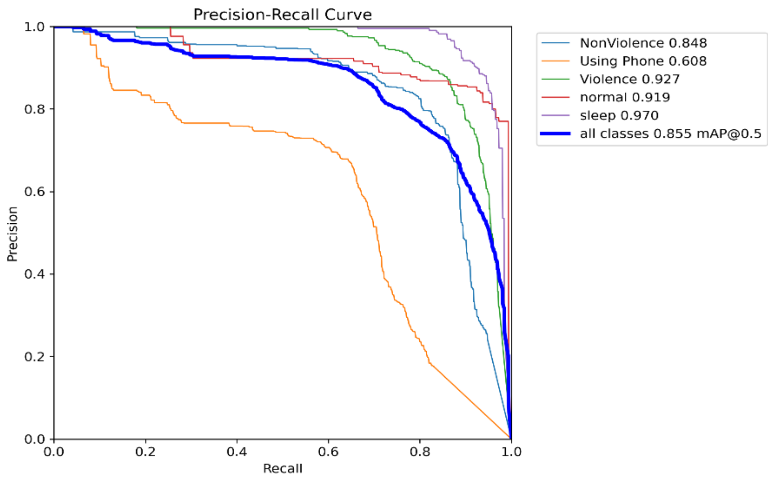


Figure 8: Precision Recall Curve

The next measure is the precision/recall curve which is in figure 11. It is created by adjusting the confidence level for classifying a detection as a true positive. Precision shows the fraction of correct detections (true positives) among all detections made by the model. It evaluates how accurate the model's detections are among those it claims to have made. Recall refers to the fraction of genuine positive detections acquired by the model out of all ground truth objects in the dataset. It assesses how well the model can locate all essential objects. The PR curve aids in evaluating the overall performance of the YOLO detector and selecting the best operating point based on the application's specific requirements.

Results from the model

The model uses a face recognition module to capture the face encodings of people. Figure 12 depicts the successful detection of a person by a facial recognition module, and based on the recognition, it writes the attendance and time of entry into the class in a CSV file. Using CSV files for data storage enables simple interaction with a variety of software programs and analytical tools. Figures 11 and 14 and 15 illustrate the model's detections for the classes of violence, sleeping, nonviolence, normal, and phone use. The bounding boxes also display the matching percentage and class names.

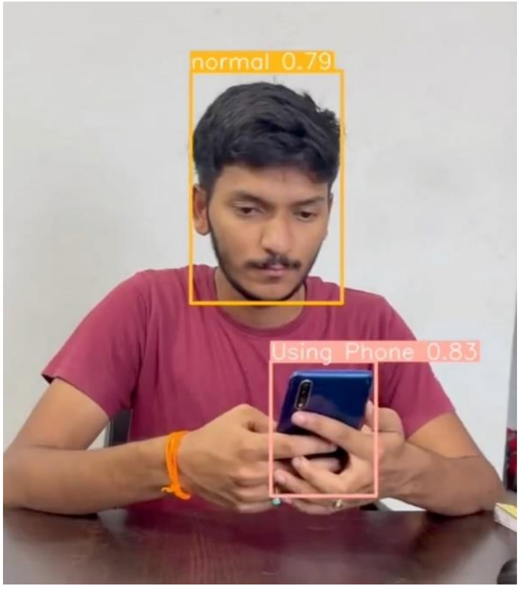


Figure 9: Realtime results from the model(Mobile Detection)

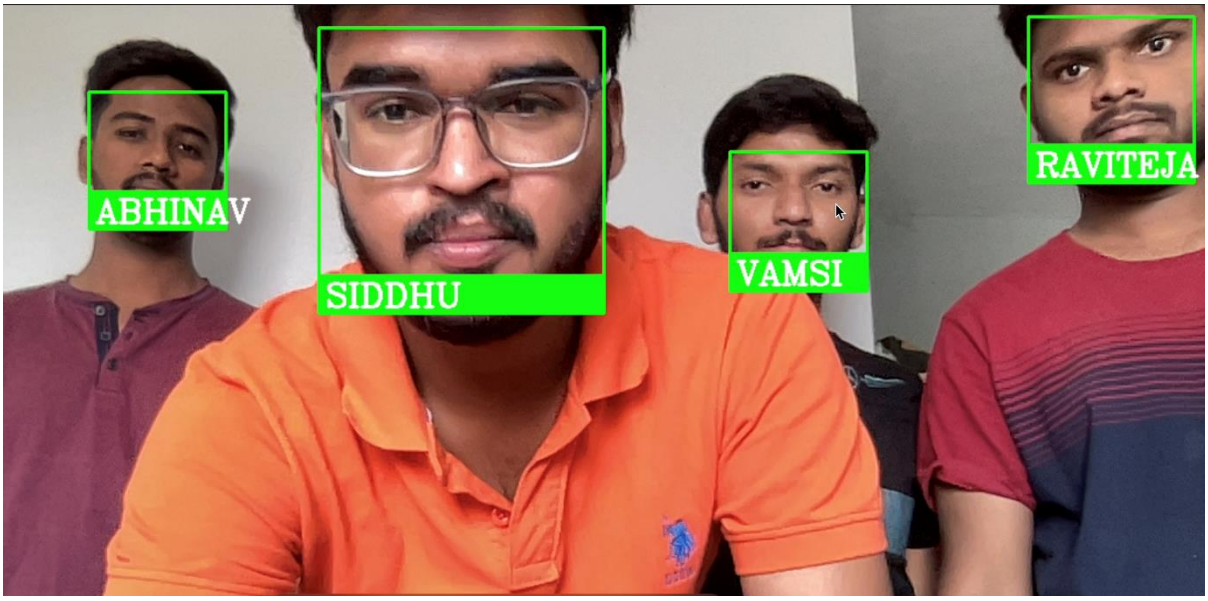


Figure 10: Realtime results from the Face recognition model

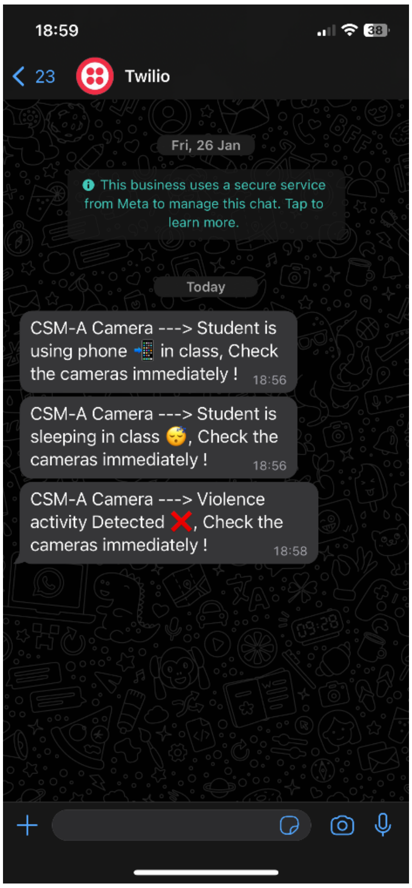


Figure 11: Realtime Messages In WhatsApp

Once the prediction is made, we match the classes by converting the tensor array into a numpy array. If the class we want is predicted, we use the Twilio module, which is depicted in Figure 13, to send the message.

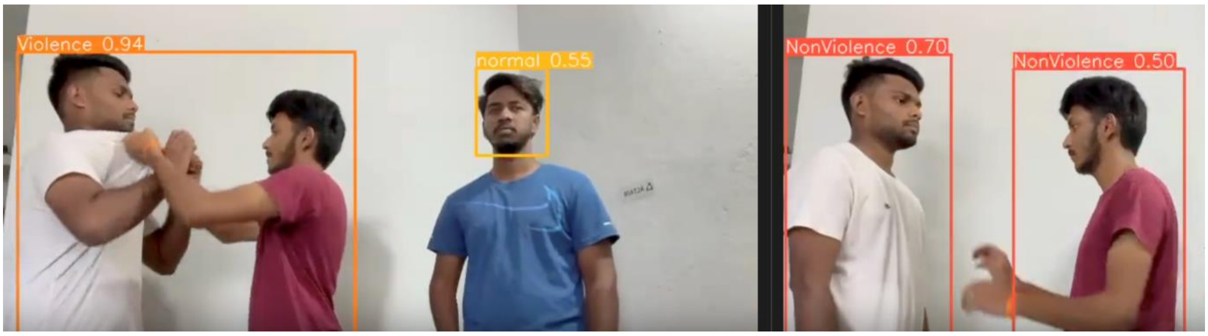


Figure 12: Realtime detection of violence

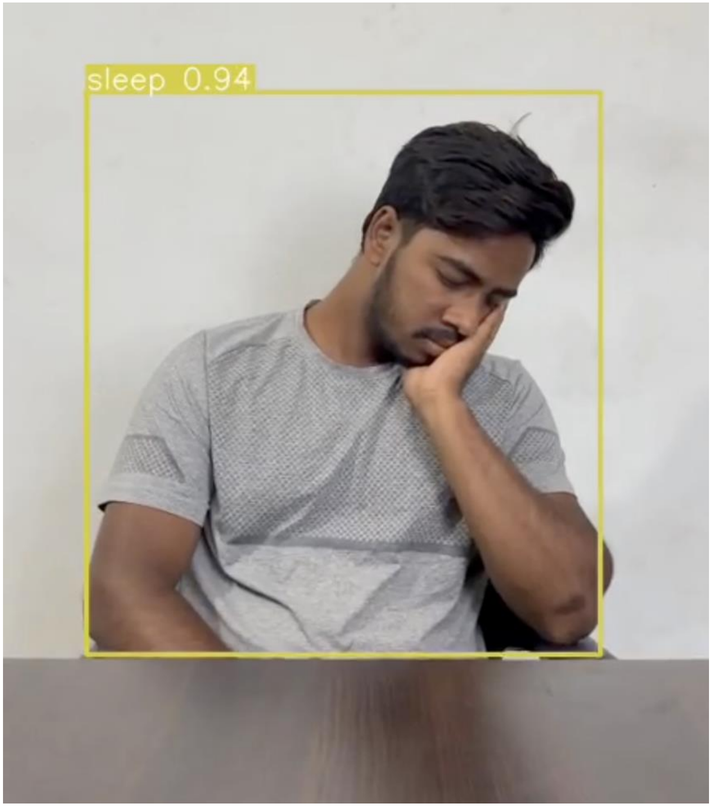


Figure 13: Realtime detection in sleep

CHAPTER 7

CONCLUSION

In summary, YOLOv8 is a significant development in computer vision's real-time object identification field. With its improved architecture and state-of-the-art features, its deep learning model allows for very accurate object recognition in a variety of applications.this study report offered a complete approach to tackling the issues of monitoring student behavior and creating a conducive learning atmosphere in offline educational environments. By integrating cutting-edge AI technologies, particularly the YOLOv8 algorithm, the proposed Student Tracking System provides a game-changing solution for behavior analysis, attendance management, and incident detection in educational settings. This article used the Student Tracking System to show how AI technologies like YOLOv8, CNNs, and face recognition modules can revolutionize behavior analysis and incident detection. The technology can precisely recognize and categorize a wide range of student behaviors, such as dozing in class, using mobile phones, and engaging in abnormal activities, thanks to strategically placed cameras and real-time monitoring. Furthermore, the system's integration with OpenCV for preprocessing, data collecting from various sources, and machine learning models improves accuracy and efficiency. The addition of features such as triplet loss for fine-grained categorization and non-maximum suppression for removing redundant detections enhances the system's performance. The suggested system architecture, as explained, offers educational institutions a scalable and customizable solution for successfully monitoring student behavior and ensuring a positive learning environment. The user-friendly interface and real-time warnings allow instructors and administrators to handle concerns quickly and instill accountability in students. The Student Tracking System, which combines novel technologies with realistic implementation tactics, represents a viable route for improving educational quality and student safety in offline educational settings.

FUTURE ENHANCEMENTS AND DISCUSSIONS

For future enhancements and discussions, the project can explore several avenues to improve its effectiveness, efficiency, and applicability. For future enhancements and discussions, several avenues can be explored to further improve the project's efficacy and relevance. Firstly, integrating advanced deep learning architectures such as transformer-based models or attention mechanisms could enhance the system's ability to capture complex temporal and spatial patterns in surveillance footage, potentially leading to more accurate anomaly detection and behavior recognition. Additionally, leveraging reinforcement learning techniques could enable the system to adapt and learn in real-time from its environment, improving its adaptability to dynamic surveillance scenarios. Furthermore, incorporating multimodal data fusion techniques, such as combining visual data with audio or text inputs, could provide a richer understanding of the observed environment and enhance the system's overall performance. Moreover, addressing privacy concerns and ethical considerations through the development of robust data anonymization and encryption methods is crucial for ensuring the responsible deployment of surveillance systems in educational settings. Finally, conducting extensive user studies and engaging with stakeholders, including educators, students, and privacy advocates, can provide valuable insights into the system's usability, acceptance, and potential societal impacts, guiding future development efforts toward creating more ethical and socially responsible surveillance solutions.