Proceedings of the Third International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT-2025)
IEEE Xplore Part Number: CFP25CV1-ART; ISBN: 979-8-3315-2754-9

2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) | 979-8-3315-2754-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/IDCIOT64235.2025.10914712

Real Time Violence Detection System using YOLOv7 and Deep Learning Techniques

Sudha Senthilkumar
School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology,Vellore
Vellore, India
sudha.s@vit.ac.in

Gracy Agarwal
School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology,Vellore
Vellore, India
gracy.agarwal2021@vitstudent.ac.in

Arshia Shirish
School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology,Vellore
Vellore, India
arshia.shirish2021@vitstudent.ac.in

Shruti Kolte
School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology,Vellore
Vellore, India
shrutirajendra.kolte2021@vitstudent.ac.in

***Abstract* - In recent times, violence in public spaces such as schools, hospitals, and other community environments has increased a lot. Due to this alarming rise in violence CCTV cameras have been deployed everywhere to enhance safety and security. However, these systems depend on manual monitoring which is prone to human errors and cause delays in the identification of violent events and action which can lead to significant damage. To overcome these limitations, this study has proposed a violence detection and alarm system based on deep learning algorithms. The hybridization of YOLOv7 allows rapid recognition of possible threats by identifying human presence and movement, and the lightweight classifier provides a high accurate classification of the identified tasks while reducing the processing computational resources of the system. With this dual-model-based framework, it is guaranteed that both the speed and accuracy can be achieved, and that the system can be implemented on edge devices such as mobile devices. The function of the system is to identify violence while it is happening and to issue alerts to security or law enforcement authorities committed to rapid response. The system was tested in real-time scenarios and achieved 91% accuracy in detecting violent actions. The system incorporates a Telegram bot to send real-time alerts to the relevant authorities. By leveraging advanced deep learning techniques, this study seeks to close the gap between existing surveillance methods and the need for a more automated and accurate violence detection system.**

Keywords: Violence Detection, Real-time Surveillance, YOLOv7, MobileNet, LSTM, Deep Learning, Automated Alert System, Action Recognition, Computer Vision, Smart Surveillance

I. INTRODUCTION

Maintaining safety within public areas like schools, hospitals, or transit stations is of great importance in this modern world. Public violence cases are now an emerging threat anywhere around the world. Harassment on the streets, brutality without target, and public unrest have all become normal incidents that increase the individual's and society's exposure to risks. The public's safety should be ensured by the law enforcement agencies, but it is very challenging to react quickly when incidents of public violence occur.

While governments have installed surveillance systems in public places such as streets, transport centers and schools, the constant monitoring of feeds is arduous and tiring for human operators. Most security systems installed at these places rely on closed circuit video surveillance systems and depend on human observation, which is a prone to error. This old mode of observation leads to delayed responses towards violent acts, which could harm the people and properties. Recent studies on violence detection systems using deep learning models like YOLOv7, SSD, and Faster R-CNN have shown potential but are limited in real-time performance due to computational overhead and are unable to operate effectively on edge devices. Moreover, these models very often fail to capture the contextual complexity of violent actions which leads to reduced accuracy. To address these gaps, this research integrates YOLOv7 for rapid human detection, along with a lightweight classifier like MobileNet for efficient violent action recognition, and a real-time alerting mechanism using a Telegram bot. This combined approach ensures a scalable, efficient, and deployable solution that overcomes the limitations of prior works.

The opportunity to address these challenges emerges with the fast growth of machine learning and computer vision. Notifying all relevant authorities in real time as soon as an event occurs will prompt them to take action quickly to curb the event from getting out of hand. Deep learning-based violence detection systems are more accurate, faster, and more extensive in scope. Such systems will increase public

safety by providing significant reduction of workload in watching over surveillance tapes and promoting prevention of crime and order through the relevant authorities.

Recent deep learning breakthroughs open new avenues in the automation of surveillance systems. YOLOv7 is very effective for object detection as it has high speed and accuracy in identifying potential threats. On the other hand, light classifiers such as EfficientNet and MobileNet provide very optimized performance using very minimalistic computation which can be used on devices with low processing capacities. Still, current methods have not yet implemented these sophisticated models for the immediate detection and accurate action recognition in real-time environments. This work fills in the gap with the fusion of YOLOv7 and a lightweight classifier aimed at developing an automated system that enhances efficacy and accuracy in violence detection in surveillance scenarios.

This research work aims to develop a real-time violence detection and alert system that combines YOLOv7 for rapid human detection with a lightweight classifier, such as MobileNet, for accurate violent action recognition. The system will be optimized for deployment on edge devices like surveillance cameras and mobiles, ensuring low computational resources and power are used. The project includes testing and integrating the appropriate deep learning models, developing an automated alert mechanism to notify security personnel upon detection of violent behavior, and optimizing the system for real-time performance. The scope also involves testing and validating the system across various real-world applications. This project aims to reduce dependence on human monitoring, improve response times to violent incidents, and increase public safety. The expected outcome is a scalable, efficient, and deployable model that provides a practical solution that improves public safety in various public spaces.

II. RELATED WORKS

The use of deep learning techniques for detecting violence has seen a lot of growth in last few years, especially with the progress made in data models like YOLOv7, SSD, Faster R-CNN and lightweight classifiers such as MobileNet. These models provide various benefits that increase the accuracy and speed of detection systems, in real-time surveillance applications. This review compiles insights from some research papers to give a complete understanding of the latest developments and highlight opportunities for integrating YOLOv7 with lightweight classifiers like MobileNet to enhance the accuracy violence detection.

Rahmawati et al. [1] compared the performance of Faster R-CNN, YOLOv3, and SSD in fog conditions. Faster R-CNN achieved the highest accuracy, but at the expense of speed. The results indicate the need for fast models like YOLOv3 with lightweight classifiers to benefit from speed gain without significant losses in accuracy under challenging conditions.

Heda and Sahare [2] gave a proper comparison of YOLOv3, YOLOv4 and YOLOv5. YOLOv3, though older, still remains the fastest and is best for real-time human detection. However, the study also demonstrated that newer models such asYOLOv5 achieve better accuracy, suggesting that combining YOLOv3 with a lightweight classifier could help ensure a more balanced approach, based on both speed and detection performance.

Wang, et al. [3] optimized YOLOv3 for face detection using network pruning techniques, drastically reducing model size and computation time with minimal loss in the precision. This showed all the advantages of model compression.
and exemplified the prospect of similar approaches - such as MobileNet, used to optimize models further for violence detection systems in resource-constrained
environments.

Kumar et al. (2024a) [4] utilized YOLOv7 in their framework for violent object detection where the potential issue was dealing with identification of weapons such as bat, knife, and gun in real time. The use of robust architectures, like deep feature extracting layers, break the limitations of earlier YOLO releases which did not accurately detect small objects. This optimization led to a mean average precision of 89.5%, proving the efficiency of this approach in improving violence detection systems, especially in resource-constrained smart city environments. However, this study did not the recognition of human actions and relies on computationally heavy architectures, making it unsuitable for low-resource environments.

Chen et al. [5] utilized YOLOv7's innovative architecture, especially its extended layer aggregation networks (E-ELAN), to improve feature learning and convergence. YOLOv7 effectively addressed obstacles in real-time object detection, such as recognizing small objects and handling high computational costs. This is done by optimizing gradient paths and including large kernel convolutions. These enhancements demonstrated the versatility of YOLOv7 in diverse applications. Such methodologies give the prospect of using other hybrid approaches, like integrating MobileNet, to increase real-time system performance. While this study is effective in optimizing gradient paths for small object detection, the study did not combine YOLOv7 with additional classifiers for contextual understanding and has high computational resource usage.

Yang et al. [6] integrated YOLOv7 with Bi-Level Routing Attention (BRA) to enhance object detection performance in densely populated classroom environments. This modification took into account challenges like occlusions and misclassification of visually similar behaviors by introducing

a dynamic sparse attention mechanism that improved query-aware sparsity. This optimized YOLOv7-BRA achieved a 2.2% improvement in mAP@0.5. This highlights the potential of incorporating advanced attention mechanisms like BRA in educational behavior detection systems. Although the attention mechanism improved performance in visually dense environments, it added computational overhead. The study was also limited to specific use cases like classrooms and did not generalize to public violence detection.

Thomas et al. [7] proposed a real-time violence detection system that integrates MobileNetV2 with Cloud Firestore for storing and analyzing the incident data. Their system showed efficient detection with low latency due to MobileNetV2's lightweight architecture. However, the reliance on cloud-based infrastructure introduced challenges related to privacy and network dependency. This emphasizes the significance of edge-device-based solutions like the one proposed in this research to ensure enhanced security and real-time applicability.

Moh. W. Ahdi et al. [8] demonstrated the use of EfficientNet-B0 for classifying paddy diseases, showcasing its ability to achieve high accuracy with minimal computational requirements. This aligns with the goals of this research, where a similar lightweight classifier like MobileNet is leveraged for violence recognition for resource-constrained environments.

Veltmeijer et al. [9] introduced a subgroup analysis method for real-time violence detection and localization, which improves system reliability in complex environments. While their work focuses mainly on localization, it provides insights into improving detection accuracy under challenging scenarios which helped this research.

Kumar et al. (2024b) [10] explored MobileNetV2 for violence detection and achieved high performance on benchmark datasets. However, the study did not include an integrated alerting system or optimization for edge devices, which are critical aspects that are addressed in this research.

Siddique et al. [11] explored real-time violent activity detection using various deep learning models like ConvLSTM, CNN-BiLSTM, VGG16-BiLSTM, CNN-Transformer, and C3D. The study focused on extracting spatiotemporal features for efficient detection. CNN-BiLSTM achieved the highest accuracy of 83.33%, followed by ConvLSTM and C3D, both achieving 80% accuracy. CNN-Transformer demonstrated robust performance with an accuracy of 76.76%. VGG16-BiLSTM was the least accurate at 70%. The study highlights the potential of combining convolutional and recurrent architectures for violence detection however, the models tested require high computational resources.

Hsairi et al. [12] explored violence detection using deep learning techniques and focused on image-based datasets to enhance security. The study evaluated several models like sequential CNN, MobileNetV2, and VGG-16. Among these, VGG-16 achieved the highest accuracy of 71%, outperforming the other models when fine-tuned with transfer learning and data augmentation techniques. Sequential CNN and MobileNetV2 models demonstrated reasonable performance and were less effective compared to VGG-16. The study highlights the potential of VGG-16 for integration into surveillance systems but further optimizations are required.

Table 1 summarizes the gaps and limitations of the research papers discussed-

Research Paper	Gap	Limitation
Rahmawati et al. (2023)	Focused on foggy conditions and no focus on real-time violence detection.	Faster R-CNN had high accuracy but was computationally expensive and lacked precision.
Yang et al. (2024)	Limited to classroom behavior detection.	Bi-Level Routing Attention (BRA) added computational overhead and approach is not generalized to public domains.
Thomas and Balamurugan (2024)	Reliance on cloud infrastructure with no edge device optimization.	Privacy and network dependency issues due to Cloud Firestore-based architecture.
Kumar et al.(2024b)	No integrated alerting mechanism or optimization for edge devices.	Focused on MobileNetV2 but lacked real-time violence classification integration.

Table 1 – Summary of gaps and limitations

Despite advancements in violence detection systems, existing solutions fall short in delivering the required speed, accuracy, and efficiency for real-time surveillance. Models like YOLOv7 and Faster R-CNN have trade-offs between speed and precision and has not been effectively combined with lightweight classifiers for action recognition. Additionally, the integration of real-time alert mechanisms and optimization for edge devices remains underexplored. This research addresses these gaps by proposing a hybrid system

that combines YOLOv7 with MobileNet and incorporates a Telegram bot for immediate notifications.

The selection of YOLOv7 for human detection was due to its ability to achieve a good trade-off between speed and accuracy and thus suitable to be used in real-time tasks. The extended layer aggregation network in YOLOv7 provides effective detection despite being in a cluttered dynamic environment, while preserving computational efficiency. For violence recognition, MobileNet-BiLSTM was selected as, for one thing, it integrates the lightweight, robust structure of MobileNet and the temporal model learning ability of BiLSTM. This integration permits the system to accurately capture temporal sequences of violent behavior while achieving low computational complexity. MobileNet-BiLSTM provides a practical compromise between accuracy and speed, and, therefore, is applicable for real-time deployment on edge devices with limited resources. This will provide the best overall performance for a violence detection system, ensuring a high speed and accuracy on resource constrained devices.

III. METHODOLOGY

The proposed methodology for our real time alerting violence detection system is one that integrates advanced deep learning techniques in 3 main phases — rapid human detection in any footage or video, precise action recognition and an automated real time alert system. The aim of this project is to enhance the efficiency and accuracy of surveillance videos. This section comprehensively explains the key components of the method we have proposed. The following section will include Architectural Design, Model Selection, Detection Strategies and the Alerting Mechanism we have used to ensure a highly responsive framework.

The proposed system architecture, as illustrated in Figure 1, provides a modular and streamlined framework designed for efficient real-time violence detection and alerting on resource-constrained devices. This architecture outlines the flow of data through key components. There are multiple modular components, each optimized for a specific function. We have used a layered approach to ensure the handling of large amounts of video feeds while simultaneously making sure that the computational resources are used prudently. The main components are:

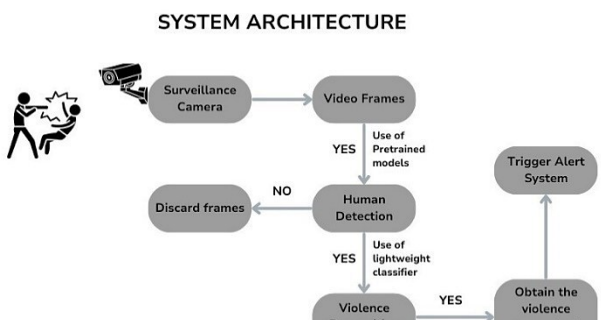

Fig. 1. System Architecture

A. Video Input and Frame Processing

CCTV surveillance footage from well-placed cameras serve as our systems primary data source. These video frames will be provided in real-time, followed by a preprocessing module and then resized as needed to match the input requirements of our deep learning modules.

B. Human Detection Module

We have employed the algorithm YOLOv7 for human detection. In the first step, we discard video frames in which no human can be identified, and then concentrate on video frames in which humans are identified.

C. Violence Detection Module

Violence detection module is carried out by MobileNet-BiLSTM, which classifies the action as violent or non-violent, based on the action categorization. In MobileNet spatial feature is extracted from frames, and in BiLSTM temporal patterns, which are extracted from frames, are processed to distinguish violent behavior accurately.

In order to obtain high accuracy, the system uses pre-trained models which have been fine-tuned on a balanced dataset including a variety of scenarios. Data augmentation procedures are used to generate variations in term of illumination, viewpoint, and crowd density, to improve model generalization. These strategies ensure robust performance in complex environments.

D. Alert System

Post violence detection, the alert module activates automatically and sends in a real time notification to the relevant authorities. This increases the overall response time to such incidents and helps the authorities be notified on time.

For detection of humans, we tried 3 main models which were namely MobileNet-SSD, RCNN, and YOLOv7, the main motivation behind choosing of this model was its balance between accuracy of detecting humans and its speed in doing so. This model incorporates an Extended Local Attention Network (ELAN) and advanced feature aggregation which makes it the top priority in detecting humans fast even in cluttered and dynamic surveillance environments.

For the violence detection, we used a combination of two models namely MobileNet-BiLSTM for their accuracy and computational efficiency. Both these models excel in balancing of speed along with precision during the classification even on devices where resources may be limited.

The detection process we have proposed in our system is focused on optimization of accuracy and efficiency. By combining YOLOv7 for human detection with action classification. This entire process together ensures relevant frame are the only ones undergoing analysis and enabling the real-time violence detection.

IV. EVALUATION AND RESULTS

A. Human Detection Model Evaluation

To further illustrate the performance differences among the models, the following screenshots showcase examples of human detection from each: YOLOv7, SSD MobileNet and Faster RCNN (TensorFlow Model). As seen in Figure 4, YOLOv7 consistently captures more individuals within each frame, demonstrating both precision and accuracy in detection. In comparison, Figure 2 illustrates that SSD MobileNet shows fewer detections with occasional missed subjects, while the TensorFlow model in Figure 3, though capable, often struggles with frame processing speed, resulting in fewer and sometimes
delayed detections. These visual results reinforce YOLOv7's effectiveness in providing comprehensive and real-time detections, supporting our quantitative findings.

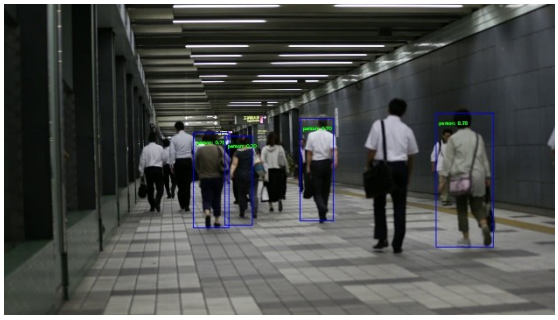

Fig. 2. MobileNet SSD Algorithm Output

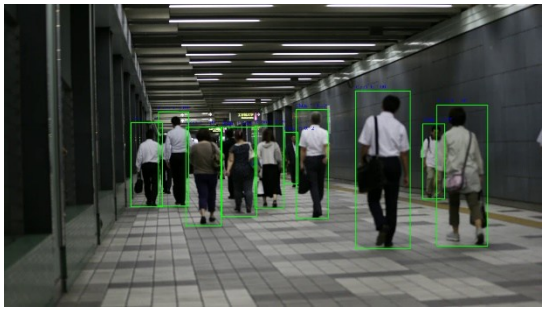

Fig. 3. Faster RCNN Algorithm Output

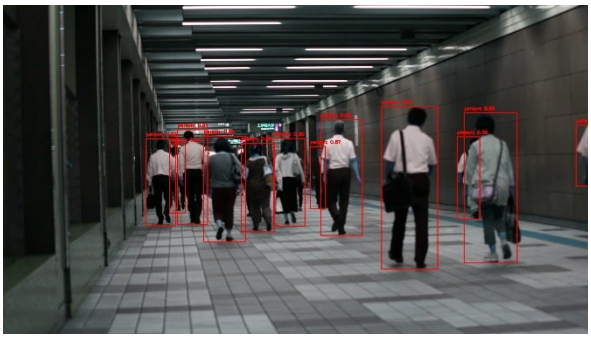

Fig. 4. YOLOv7 Algorithm Output

For evaluating Human Detection performance, we have tried 3 different models and compared them on the basis of time taken in detection for each frame and the total detections in the footage. The models we chose for our paper are — YOLOv7, SSD MobileNet Model and TensorFlow. These parameters were used to assess their speed and accuracy to create a robust real time detection system.

Table 2: Comparative Study of all models for Human Detection-

MODEL USED	TIME TAKEN TO DETECT OBJECT	TOTAL OBJECTS DETECTED
YOLOv7 model	0.0308 seconds	6262
SSD MobileNet model	0.1015 seconds	1092
TensorFlow (Faster RCNN) model	1.4498 seconds	3204

Key Metrics used:

- Average Time per Frame: As shown in Table 1, YOLOv7 demonstrated a significant advantage in speed, with an average processing time of 0.0308 seconds per frame. This was substantially faster than SSD MobileNet (0.1015 seconds) and the TensorFlow model (1.4498 seconds). Figure 5 compares the average time taken by different models. This shows that the YOLOv7 model is efficient in the aspect of time and can analyze the frames of video faster than the other two models, which makes it ideal for any real-time applications such as ours.

- Total Detections in Video: In terms of detection capability mentioned in Table 2, YOLOv7 detected 6,262 instances across the video, which is substantially higher than the detections by SSD MobileNet (1,092 detections) and the TensorFlow model (3,204 detections). Figure 6 compares the total detections achieved. The higher detection rate of YOLOv7 compared to TensorFlow and SSD

MobileNet make it even more of an ideal choice for our human detection process.

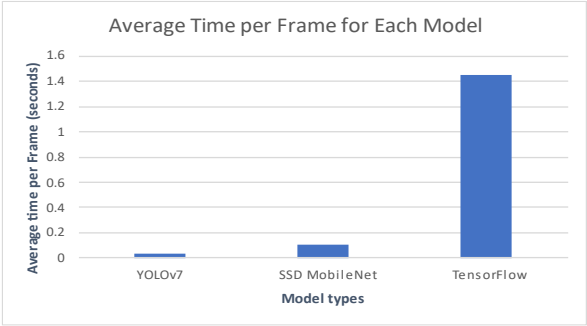

Fig. 5. Average time taken by each frame for all models

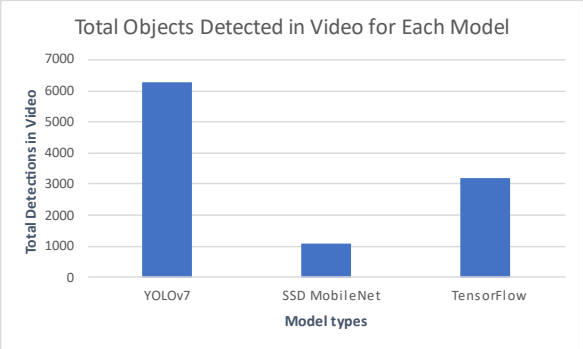

Fig. 6. Total detections of each model

The above evaluations of key metrics in the Human Detection models validate the YOLOv7 model to be the most suitable for our Violence detection system, where we favor real-time applicators and precision. Our systems requirement is fulfilled for both aspects making it our choice.

B. Violence Detection Model Evaluation

For the violence detection in surveillance footage, we have used MobileNet-BiLSTM system on a dataset with 1000 non-violent and 1000 violent videos. After execution, our method has achieved an Accuracy Score of 0.91.

As seen in Figure 7, the confusion matrix shown below exhibits a balanced detection of both behavior types, with 92 and 90 true positives for non-violence and violence respectively, while also maintaining low false positive and false negative rates.

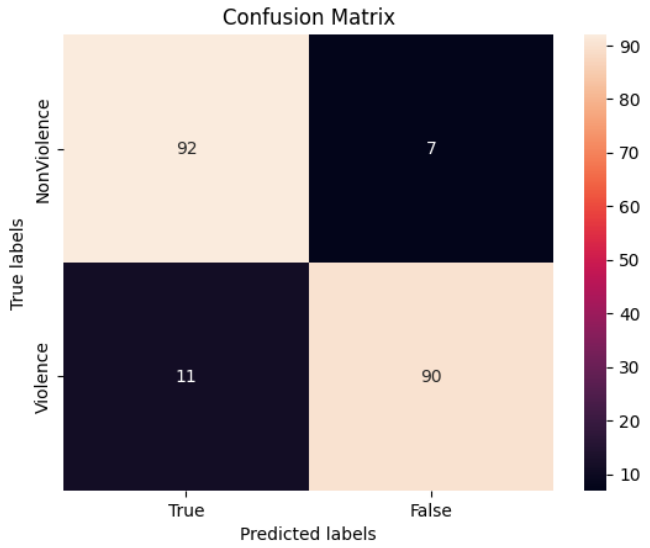

Fig. 7. Confusion Matrix

The classification report proves the robustness of our model, it also maintains consistent performance across metrics, showing its effectivity for real-time violence detection.
Figure 8 presents the evaluation metrics, demonstrating the model's robust performance:

- Precision: ratio of correctly predicted positive observations to the total number of predicted positive observations. For reference, class 1 (violence), a precision of 0.93 means that 93% of the instances predicted as violence are correctly assessed.
- Recall: ratio of correctly predicted positive observations to all observations in the actual class. For reference, class 0 (non-violence), recall of 0.93 means 93% of actual non-violence instances are detected.
- F1-Score: harmonic mean of precision and recall (balancing both). The F1 score in our model is 0.91 which indicates a good balance.

```
Classification Report is :
              precision    recall  f1-score   support

           0       0.89      0.93      0.91        99
           1       0.93      0.89      0.91       101

    accuracy                           0.91       200
   macro avg       0.91      0.91      0.91       200
weighted avg       0.91      0.91      0.91       200
```

Fig. 8. Evaluation metrics result

Figures 9 and 10 show the model's learning trends. The Accuracy and Loss trends after examining reflect consistency in the learning process. The final Loss value around 0.15-

0.25. This indicates a well-trained model capable of generalizing across the test set and our model also achieved around 94-97% Accuracy on the Testing Set.

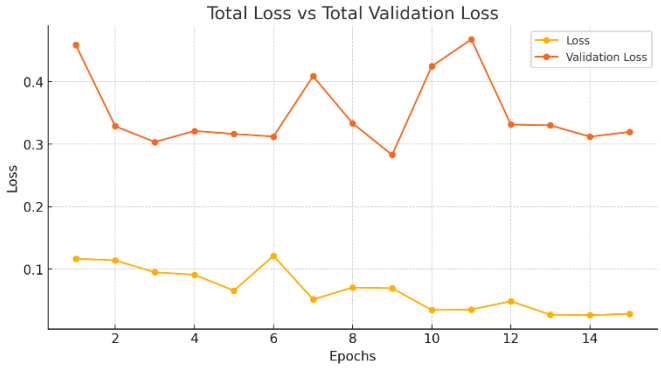

Fig. 9. Total loss Vs Validation loss

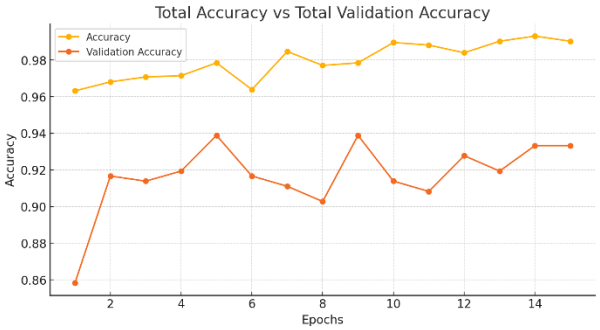

Fig. 10. Total Accuracy Vs Validation Accuracy

Precision can be enhanced by reducing false positives by using methods including raising the confidence threshold while detecting, standardizing the training data to include examples of edge cases and misclassifications, and incorporating contextual data such as object relationships. These improvements enhance the system's reliability in real-world scenarios.

To evaluate the performance and adaptability of our model we implemented two distinct prediction functions:

- Frame-by-Frame: Figure 11 and Figure 12 depict the frame-by-frame prediction results. In this function, the model processes each frame and classifies it into two categories (Violent and Non-Violent). This is useful for real time surveillance with the only drawback that it may lead to some noise in our prediction due to lack of context. This method prioritizes speed and rapid decisions making it ideal for our system

.

Fig. 11. Frame by Frame prediction of Violence

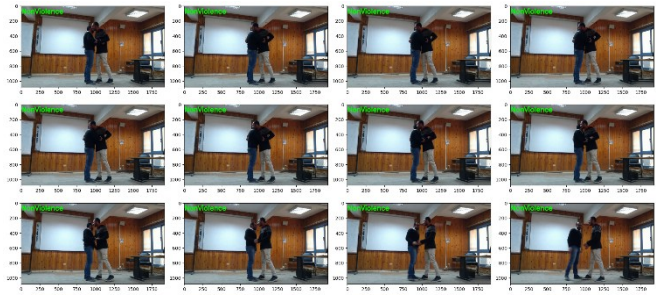

Fig. 12. Frame by Frame Prediction of Non-Violence

- Full Video Prediction: Figure 13 showcases the full video prediction approach with high confidence in detection. This is a holistic approach to view the entire footage all together, aggregating the predictions over all frames for better and more accurate results however it may be computationally heavy as compared to the first function and may make it un-ideal in some situations for real time violence detection. However, it is the best model for a detailed analysis of any surveillance footage.

Fig. 13. Full video prediction (Violence detected with confidence of 99%)

Table 3 gives a comparison of the accuracy values of different models.

Model	Accuracy (%)
YOLOv7 + MobileNet (ours)	91
VCG	71

ConvLSTM	80
CNN-BiLSTM	83.33
VGG16-BiLstm	70
CNN-Transformer	76.76
C3D	80

Table 3

V. CONCLUSION

This research paper proposed a novel integrated framework for real-time fight detection and alerting in a surveillance system that can combine YOLOv7 with a lightweight classifier such as MobileNet to improve speed and accuracy of violence detection on edge devices. The framework developed will close major research gaps through suggesting approaches that incorporate both spatial and temporal detection along with real-time alert mechanisms which are optimized for resource-limited areas. Despite its performance, the existing system suffers from various limitations, including false alarms in highly complex environments, the inability to handle low light or occlusion states, and the requirement for more diverse datasets to mitigate the effects of wide-ranging situational differences. Furthermore, privacy and ethics issues still pose critical problems when realizing these systems in public spaces.

These results further establish the efficacy of the YOLOv7 and MobileNet framework for real-time violence detection with high accuracy and low computational overhead. Future improvements can be done by incorporating multi-modal learning methods where video data will be combined with audio input for better detection performance in cases where relying on only visual cues may not work. This can be further improved by research in lightweight classifiers like EfficientNet-B0, for further optimization of computational efficiency leading to higher usability on ultra-low-power and edge devices. Advanced techniques such as graph neural networks (GNNs), can be explored to model relationships between individuals in a crowded space. These developments will ensure that the system is adaptive and scalable to various surveillance environments.

REFERENCES

[1] Rahmawati, L., Rustad, S., Marjuni, A., Soeleman, M. A., & Andono, P. N. (2023, September). Foggy-Based Object Detection In Video Using Faster R-CNN, YOLOv3, and SSD. In *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 412-416). IEEE.

[2] Heda, L., & Sahare, P. (2023, April). Performance evaluation of YOLOv3, YOLOv4 and YOLOv5 for real-time human detection. In *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)* (pp. 1-6). IEEE.

[3] Wang, Y., Huang, L., Li, J., & Sun, T. (2023, December). Research on YOLOv3 Face Detection Network Model Based on Pruning. In *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)* (Vol. 11, pp. 349-352). IEEE.

[4] Kumar, P., Shih, G.-L., Guo, B.-L., Nagi, S. K., Manie, Y. C., Yao, C.-K., Arockiyadoss, M. A., & Peng, P.-C. (2024a). Enhancing smart city safety and utilizing AI expert systems for violence detection. *Future Internet, 16*(2), 50.

[5] Chen, Y., Yuan, X., Wu, R., Wang, J., Hou, Q., & Cheng, M.-M. (2023). YOLO-MS: Rethinking multi-scale representation learning for real-time object detection. *arXiv*.

[6] Yang, F., Wang, T., & Wang, X. (2024). Student classroom behavior detection based on YOLOv7-BRA and multi-model fusion. *arXiv*.

[7] Thomas, M., & Balamurugan, P. (2024, April). Real-Time Violence Detection and Alert System using MobileNetV2 and Cloud Firestore. In *2024 2nd International Conference on Networking and Communications (ICNWC)* (pp. 1-9). IEEE.

[8] Moh. W. Ahdi, K. Sjamsuri, A. Kunaefi, B. A. Nugroho, & A. Yusuf. (2023). *Convolutional Neural Network (CNN) EfficientNet-B0 Model Architecture for Paddy Diseases Classification*. 2023 14th International Conference on Information & Communication Technology and System (ICTS). IEEE.

[9] Veltmeijer, E., Franken, M., & Gerritsen, C. (2024). Real-time violence detection and localization through subgroup analysis. *Multimedia Tools and Applications*

[10] Kumar, R., Gupta, A., & Rajeswari, D. (2024b, June). Violence Detection System using MobileNetV2. In 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 1555-1560). IEEE.

[11] L. A. Siddique, R. Junhai, T. Reza, S. S. Khan, and T. Rahman, "Analysis of Real-Time Hostile Activity Detection from Spatiotemporal Features Using Time Distributed Deep CNNs, RNNs and Attention-Based Mechanisms," *arXiv*, vol. 2302.11027, 2023.

[12] L. Hsairi, S. M. Alosaimi, and G. A. Alharaz, "Violence Detection Using Deep Learning," *Arabian Journal for Science and Engineering*, vol. 49, no. 4, pp. 3603–3615, 2024.