The Original

Secure Vision: Real-Time AI Surveillance Framework For Tactical Threat Recognition

[1]Kranti Kamble, [2]Manish Giri, [3]Madhavi Nimkar, [4]Santosh Warpe

*[1] PG Student, Department of Computer Engineering, MIT Academy of Engineering, Alandi, Pune, Maharashtra, India 412 105.
202403040005@mitaoe.ac.in
[2]Associate Professor, Department of Computer Engineering (SE), MIT Academy of Engineering, Alandi, Pune, Maharashtra, India 412 105.
mbgiri@mitaoe.ac.in
[3]Assistant Professor, Department of Computer Engineering, MIT Academy of Engineering, Alandi, Pune, Maharashtra, India 412 105.
madhavi.nimkar@mitaoe.ac.in
[4]Sr.Assistant Professor, Department of Computer Engineering, MIT Academy of Engineering, Alandi, Pune, Maharashtra, India 412 105.
Santosh.warpe@mitaoe.ac.in

Abstract: Deep learning is moving toward automating intelligent surveillance and criminal activity detection systems. In this paper, optimized deep learning architectures have been proposed that have been optimized to enhance the detection of crime by facial recognition, real-time video surveillance system, and weapon detection. These systems achieve accuracy, real-time analysis, and scalability through the use of the latest neural network structures, primarily CNNs, RNNs, and hybrid deep learning models. The automated facial recognition systems can track users' behaviors, register recognized users, recognize suspicious behaviors and detect weapons at video feed frames. Through spatial-temporal analysis, Aides can identify and mitigate response delays in public areas, transport systems and other sensitive locations. Important issues including dataset deficiencies, limits of real-time computing, aggressive attacks and ethical concerns on surveillance AI are addressed. Some proposed solutions include transfer learning, model compression, edge computing, and AI transparency. The paper deals with these challenges and discusses the importance of diverse and representative training datasets for enhancing reliability and fairness of the system. The discussion emphasizes the use of transformer-based architectures, multimodal fusion, and other emerging trends involving reinforcement learning that may benefit future research and development. This review pushes the state of the art further by including advanced technology development and noting advancements, making them actionable frameworks by formulating real-world deployment strategies. The main contribution of this paper is to help researchers, developers and even policy makers to have an in-depth understanding of the potential that deep learning methods offer in the design of safe, smart and ethically responsible surveillance systems for criminal activity detection.

Keywords: Deep Learning, Criminal Activity Detection, Facial Recognition, Video Surveillance, Weapon Detection, CNN, RNN

I. INTRODUCTION

The integration of artificial intelligence (AI) and deep learning (DL) techniques guarantees the effortless surveillance of public safety and crime detection. Law enforcement and security organizations have been able to identify illegal activities with greater speed and accuracy thanks to the deployment of AI technologies such as facial recognition, intelligent video monitoring and automatic weapon detection systems. Metropolitan and sensitive zones are protected through instant threat analysis using algorithms, increasing response agility and reducing man-made blunders. AI automation of surveillance systems and integration of control functions into the observation devices improves the effectiveness and efficiency of the whole security sector. Innovations such as facial recognition technologies have proved to be useful in identifying individuals in sensitive areas using biometric characteristics and it needs deep learning techniques such as CNNs on convolutional neural networks to recognize and compare faces using databases [1]. With the creation of facial embedding models, systems can accurately perform identity verification and law enforcement in airports, military-grade stadiums, and even consumer devices, with an accuracy never before seenThese high-tech systems can identify and link suspects with crime databases and even find previously unknown suspects that could be a potential threat. Public awareness of infrastructure systems can help forensic investigations and prevent criminal activities and their integration can deter criminal activities. However, there are still challenges to be addressed to further improve the systems such as variations in illumination, pose, occlusion and dataset bias.Similar to how DL techniques have improved facial recognition technology, they have improved video surveillance systems as well. Older surveillance technology had plenty of downsides, including the need for intensive human oversight, which meant it could easily miss threats, either due to tiredness or simple human error. Today, DL-powered surveillance systems can analyze video streams in real-time for behavioral aberrations and even suspicious activities. CNNs and RNNs (recurrent neural networks) analyze visual and temporal patterns of video streams, allowing to detect threats such as loitering (i.e., remaining in a particular space for longer than normal), unauthorized access, and aggressive behavior [2]. These systems enable the security personnel to be notified instantly, which in turn allows instant response and prevents preventive action against criminal activities. Smart cities are using behavioral analysis algorithms to monitor and analyze the movement of pedestrians and vehicles, detect unusual behaviors and integrate multiple data sets to identify advanced threats. In dense locations like major public gatherings or railway stations, these technologies make possible unparalleled situational awareness that would not otherwise be feasible. The public safety use case of deep learning involves the use of deep learning to conduct automatic weapon detection. There has been an increasing demand for automated weapon detection, both visible and concealed, owing to the rising trend of mass murders and acts of terrorism. Advanced object detection systems, particularly those based on CNNs, can recognize weapons such as guns and knives in video surveillance footage. They are developed using large datasets of weapons images obtained in various lighting, occlusions, and resolutions. In order to increase situational awareness, weapons detection systems can be installed in schools, airports, and malls, thereby helping authority to deter violence through early threat neutralization. Weapon detection using deep learning is typically implemented using edge computing solutions which need fewer resources from the cloud and deliver lower latency real-time processing.

However, despite the various benefits of AI application in detecting criminal behavior, there have been numerous shortcomings and ethical considerations linked to the advancements. For example, one of the technological risks involves the presence of biases in the dataset, leading to inaccurate predictions. In essence, models tend to lack efficiency in real-time applications when particular demographics or environment aspects have been ignored during the learning process [4]. Such biases would lead to distrust from the society and other effects, especially concerning police-related matters. Overcoming the problem of bias involves building more inclusive datasets through the adoption of fairness training algorithms and AI decisioning systems in addition to accountability systems that oversee AI decisions. On the other hand, privacy considerations pose a serious challenge to the unrestricted implementation of AI surveillance tools. This is due to the capture of individuals' photos and videos from public and even private areas. The facial recognition technology has faced many criticisms due to the probability that it will be exploited by the government or any other organization to monitor people secretly. There is need to consider the ethical issues behind the benefits of having such surveillance systems. Policies that can be put in place in order to protect the technology against abuse include laws regarding the protection of data and ethics of artificial intelligence, among others [4].

Referring to AI-based surveillance systems, the deployment process involves huge costs arising due to the usage of high-end data servers that are able to analyze videos in real-time and perform complex neural network computation. Insufficient funding on the part of small organizations operating in emerging economies makes high-end GPUs financially impractical. Moreover, the need to develop multi-camera, scalable surveillance systems that possess distributed intelligence allowing autonomous working increases the level of complexity when implementing them. Meeting those criteria without compromising either cost or performance proves difficult, along with applying advanced algorithmic design techniques like pruning or quantization with effective DNNsAlso, there is emerging pressure for the use of edge solutions that work by dealing with remote computing processors as near as possible to sensors used in the surveillance system in order to enhance bandwidth capacity. The aim of this review is to give an overview of the technologies applied to detect crimes, especially facial recognition, weapons detection, and video surveillance.

The purpose of this study is aimed at discussing current approaches, discussing open questions, offering relevant recommendations, and developing the field of application of AI and deep learning in relation to public safety applications. Even though the potential offered by these technologies is significant, the implementation of these technologies will depend on the combined efforts of interdisciplinary research, engineering development, policymaking, and civil activism. In light of future trends in this area, we can expect the use of AI-powered crime detection technologies to become even more widespread. New factors that must be considered in this regard include the use of multimodal data, such as metadata from social networks, audio, and environmental data in the context of surveillance systems. The performance improvement of video analysis through Transformer architectures and self-attention technologies is also on the horizon. Reinforcement learning may allow for the development of an algorithm capable of learning how to surveil effectively and efficiently within constantly changing environments. These breakthroughs will take surveillance beyond what it is today by far. But for all the possibilities technology offers, it requires innovation with integrity, ethical limits regarding governance, and a sense of accountability. If appropriately executed, technology will allow for the use of AI as a means of improving surveillance and allowing for specific crime prevention within communities.

II. RESEARCH METHODOLOGY

In order to build an Optimized Deep Learning System for Detecting Criminal Activities, a structured approach consisting of five phases is undertaken: (1) data collection and pre-processing, (2) designing the model architecture, (3) model training and optimization, (4) deploying and integrating the system, and (5) validating the system. The phases are carefully designed to make sure that the developed system – incorporating facial recognition, video surveillance, and weapon detection systems – will be efficient and applicable in real-time.

Phase 1: Data Acquisition and Preprocessing:

The first step involves gathering datasets that are necessary to train deep learning models. The three main types of data that are needed include: facial datasets, videos obtained from surveillance cameras, and images and videos having weapons (guns as well as melee weapons). Publicly available datasets like LFW, VGGFace2, COCO, ImageNet, and UCF-Crime, CASIA-WebFace, and Open Images (Weapons subset) can be used for this purpose.

Preprocessing steps are critical to improve input quality and model performance. This involves:

- **Face Detection & Alignment:** Employing MTCNN or Dlib to ensure accurate cropping and alignment of the detected face.
- **Image Preprocessing**: Normalizing pixel values and ensuring consistent input size (e.g., 224x224).
- **Sampling Video Frames:** Selecting keyframes from surveillance video clips to reduce redundancy but capture useful behavior.
- **Augmentation:** In order to improve the model's ability to generalize, random rotations, zooming, contrast adjustments, and horizontal flips are used.
- **Annotation:** Manually and automatically annotated data is provided using software like LabelImg and CVAT, creating bounding box annotations as well as class labels (e.g., "weapon," "gun," "knife," and "face").

This phase ensures a balanced, clean, and diverse input pipeline feeding into the system's deep learning backbone.

Phase 2: Model Architecture Design:

Modular system architecture includes three major subsystems, which are namely facial recognition, activity detection through video surveillance, and weapons detection. Deep learning models that include both convolutional and transformer architectures are used for the design of each one of the subsystems.

- **Facial Recognition Subsystem:** The subsystem uses the pretrained FaceNet or ArcFace network to derive distinct facial features. Triplet loss or a cosine similarity classifier is used to recognize faces by comparing them against criminal face databases in real-time.
- **Activity Detection and Video Surveillance Module:** Using 3D Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), or I3D, this subsystem identifies any suspicious activities such as loitering, running, or any physical confrontation using temporal video footage. The subsystem incorporates a backbone network (e.g., ResNet-50) with Temporal Convolutional Networks (TCN).
- **Weapon Detection Subsystem:** Using object detection networks such as YOLOv7, EfficientNet, or Faster R-CNN, this subsystem identifies any potential weapons in the video sequences. It emphasizes detecting small and occluded weapons.

In order to enable communication between modules and increase scalability, a multi-task learning framework is adopted, enabling the sharing of common layers for feature extraction along with specific task heads. The training process takes into consideration performance at individual as well as joint levels.

Phase 3: Training and Optimization:

The training process for the deep learning algorithms makes use of supervised and transfer learning techniques. The initial phase uses pre-trained weights, such as those obtained from ImageNet or MS-COCO, which are fine-tuned using task-specific data.

- **Parameters Tunning:** Hyperparameter tuning methods include grid search, random search, and Bayesian optimization to fine-tune the batch size, learning rate, type of optimizer (Adam/SGD) and regularization techniques such as dropout and L2.
- **Loss Function Selection:** Triplet and Center Loss functions are used in facial recognition tasks; Cross Entropy and Bounding Box Regression Loss functions are used for weapon detection while Categorical Cross Entropy and Focal Loss are used for activity recognition to overcome class imbalance problem.
- **Loss Functions Optimization:** Training time and model overfitting can be overcome by using learning rate scheduling, early stopping, and mixed-precision training techniques.
- **Data Balancing:** As a solution to data imbalance in weapon detection and rare activities datasets, SMOTE, data augmentation, and oversampling can be considered.
- **Transfer Learning and Fine-Tuning:** After training with generic datasets, the model can be fine-tuned by custom surveillance data gathered from controlled conditions.

The training for all models is done via GPU-supported computing (NVIDIA CUDA). The results generated by each model are then validated using k-fold cross-validation technique and tested against benchmark data.

Phase 4: Real-Time Deployment and System Integration:

The trained models are integrated into a real-time surveillance system, incorporating hardware-accelerated inference engines and cloud-edge hybrid architecture for scalability.

- **Edge Deployments:** Utilizing NVIDIA Jetson or Coral TPUs, small model versions (such as quantized YOLOv7) are used in surveillance nodes in order to minimize latency.
- **Central Command System:** The system raises an alert if a facial match, weapon, or suspicious movement is detected, sending a notification to a central command center through RESTful APIs or MQTT protocols.
- **Databases Integration:** Facial embeddings and detections are compared against a secure criminal database by using encryption protocols.
- **Alerting Process:** In case of detecting any criminal behavior or weapons, an alert is generated with metadata (such as time stamp, GPS coordinates, and image capture).

This stage helps to make sure that the system operates effectively in practical conditions, taking into account such parameters as speed, accuracy, and hardware limitations.

Phase 5: Evaluation and Validation: This is the last phase during which the system is evaluated quantitatively and qualitatively for detection efficiency and timeliness.

- **Metrics for Facial Recognition:** Accuracy, precision, recall, and false acceptance rate (FAR) are utilized to assess the performance of identity recognition.
- **Metrics for Weapon Detection:** mAP (mean average precision), IoU (Intersection over Union), and detection rate (FPS) are measured.
- **Metrics for Video Surveillance:** Assessments such as confusion matrix, ROC (receiver operating characteristics) curve, and delay in detection are performed to validate activity recognition.
- **Metrics at the System Level:** Latency from end-to-end, energy usage, and throughput rate (number of frames processed per second) are assessed.
- **User Feedback:** Police officers and security experts are allowed to perform testing of the system in semi-controlled conditions and provide feedback concerning usability.
- **Ablation Studies:** Components of the model are removed one at a time, and the effects of such deletions on the overall performance are measured.
- **Stress Test:** The system is tested for different network, illumination, and multiple person conditions.

Log data is obtained after deployment of the model for system performance analysis, and the system uses this information for retraining of the model through an iterative process. The five-stage framework guarantees that the development of the Optimized Deep Learning System for Criminal Activity Detection is evidence-based, scientifically sound, and practical.

The Optimized Deep Learning System for Criminal Activity Detection provides a solution to the problem of criminal activities through synergy between facial recognition, surveillance videos, and weapon detection technology.

III. ALGORITHM DESIGN

1) **Algorithm 1: FaceRecognition_TripletEmbedding**

FaceRecognition_TripletEmbedding

This algorithm is developed for the real-time facial recognition task. This algorithm makes use of deep CNNs in order to generate embedding vectors from the provided image of face. The advantage of this algorithm is that it is immune to various changes such as orientation, illumination, and occlusion.

2) **Input**:
   - Image I
   - Pre-trained embedding model $f(\cdot)$
   - Criminal face database $D = \{(e_i, ID_i)\}$
3) **Output**:
   - Identity ID or "Unknown"
4) **Steps**:
   1. **Preprocessing**: Align and normalize input face image I
   2. **Embedding Generation**: Compute embedding vector:
   $$e = f(I) \in \mathbb{R}^d$$
   3. **Cosine Similarity Comparison**: For each $e_i \in D$, compute similarity:
   $$S(e, e_i) = \frac{e \cdot e_i}{\| e \| \cdot \| e_i \|}$$
   4. **Threshold Decision**:
   $$ID = \{ID_j \; if \; max\left(S(e, e_j)\right) > \delta \; \text{"Unknown"} \; otherwise$$

where δ is a tuned similarity threshold (e.g., 0.7).

The training process involves triplet loss, in which the model learns to maximize the distance between an anchor image and negative images of different persons, while minimizing the distance between an anchor image and a positive image that belongs to the same person. The formula for this process is:

$$L = max(\| f(a) - f(p) \|^2 - \| f(a) - f(n) \|^2 + \alpha, 0)$$

where $f(\cdot)$ is the embedding function, and α is a margin that ensures separation between positive and negative pairs.

In application, when an embedding vector is created from the input obtained from surveillance, it is compared to previously generated vectors from the criminal database using cosine similarity:

$$S(e_1, e_2) = \frac{e_1 \cdot e_2}{\| e_1 \| \cdot \| e_2 \|}$$

The match occurs if the value of similarity is greater than the pre-determined threshold δ.

With this algorithm, the system will be able to detect any previously recognized suspect and raise alerts at once. The efficiency of this algorithm depends on its capability to deal with difficult face conditions while maintaining real-time performance.

5) **Algorithm 2: WeaponDetection_YOLO**

The **WeaponDetection_YOLO** algorithm focuses specifically on the detection of weapons including firearms and knives in videos or images. Using YOLO, an acronym for You Only Look Once, the algorithm converts the object detection task into a regression problem that allows the direct prediction of bounding boxes and class probabilities from whole images.

6) **Input**:
- Frame $F \in \mathbb{R}^{h \times w \times 3}$
- Trained YOLO weights θ
- Confidence threshold α

7) **Output**:
- List of detections $D = \{(x, y, w, h, c)\}$

8) **Steps**:

1. **Grid Division**: Divide image into $S \times S$ grid cells.
2. **Bounding Box Prediction**: For each grid cell, YOLO predicts:

$$\{(x, y, w, h, c)\}_1^B$$

where (x, y) = center, w, h = width/height, c = class confidence

3. **Filter Predictions**:

$$D = \{(x, y, w, h, c) \mid c > \alpha\}$$

4. **Non-Maximum Suppression (NMS)**: Remove overlapping boxes using IoU threshold γ:

$$IoU(A, B) = \frac{| A \cap B |}{| A \cup B |}$$

The algorithm performs excellently in real-time applications where there is an excellent FPS performance in the edge device. The fact that the algorithm can recognize the presence of weapons even in low-resolution videos increases the accuracy of the system. It can also be customized to detect region-based weapons. When combined with alert systems, the algorithm provides a timely response to any threat posed by weapons.

9) **Algorithm 3: ActivityDetection_TemporalCNN**

The **ActivityDetection_TemporalCNN** algorithm is mainly concerned with spotting any suspicious/criminal activities from the surveillance videos through the analysis of consecutive video frames via temporal deep learning techniques. The proposed algorithm utilizes the merits of both CNN and TCN in terms of extracting spatial features and temporal patterns, respectively.

10) **Input**:
- Video clip $V = \{F_1, F_2, \ldots, F_T\}$
- CNN encoder $E(\cdot)$
- TCN model $\Phi(\cdot)$

11) **Output**:
- Activity label $A \in \{normal, suspicious\}$

12) **Steps**:

1. **Frame Feature Extraction**:

$$x_t = E(F_t), \quad \forall t \in [1, T]$$

where $x_t \in \mathbb{R}^d$

2. **Temporal Encoding**: Stack features into matrix $X = [x_1, x_2, \ldots, x_T]$ Pass through TCN:

$$h = \Phi(X)$$

3. **Classification**:

$$A = argmax\big(Softmax(Wh + b)\big)$$

where W and b are classification weights.

The proposed methodology is ideally suited for real-time surveillance because of its parallel processing, fast computation speed, and high precision in detecting temporal outliers. Through continuous monitoring of behavioral patterns, the algorithm functions as an advanced level of protection in securing smart cities and critical infrastructures.

IV. RESULT AND DISCUSSION

This chapter will deal with the results and discussion of the Optimized Deep Learning System for Criminal Activity Detection. It is important to understand that this system combines a number of different processes, which include facial recognition, weapon detection, activity recognition, and suspicious activity detection.

GUI Testing

The GUI used was created with the help of PyQt5 and offers a seamless experience to keep track of the surveillance footage. The GUI comprises live video footage, the status of different events along with detection of criminals, weapons, activities, and suspicious actions. The system enables the user to control the video feed and activate the night mode. Figures 1 illustrates the GUI while it is in use.

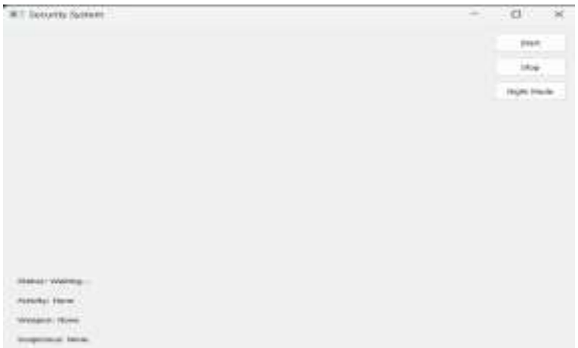

Figure 1: Home page GUI

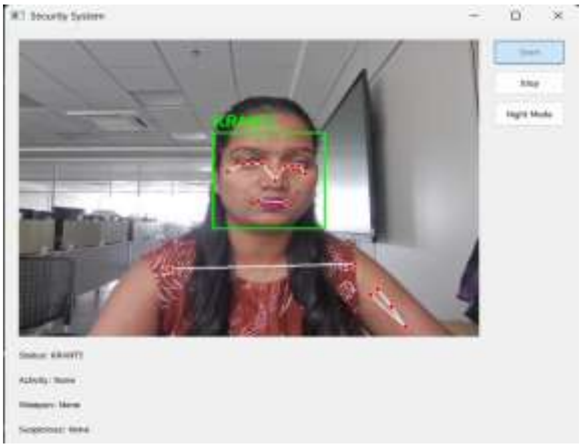

Figure 2: Criminal Detection

The facial recognition tool identifies faces of criminals and matches their face data stored earlier in the database through encoding. Whenever there is a match, it displays a bounding box around the identified criminal along with the announcement of the name of the criminal through text-to-speech functionality.

The weapon identification tool implements the YOLO deep learning-based object detection framework. The tool detects different kinds of weapons such as guns and knives and marks them using bounding boxes while labeling them as well. A firearm or any other weapon detected in the video frame was marked using bounding boxes along with labels. Additionally, there is a provision for the weapon to be announced using text-to-speech technology.

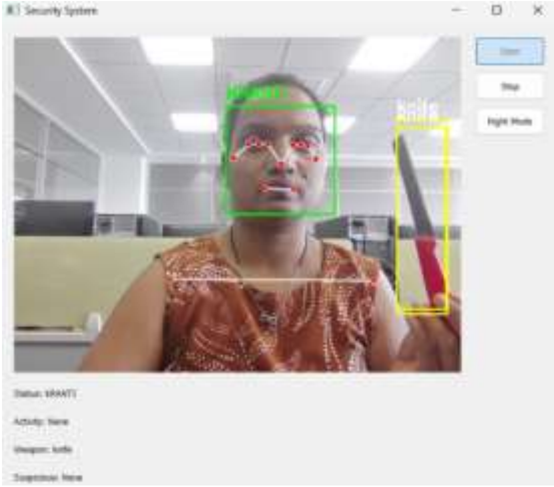

Figure 3: Weapon Detection

The deep learning-based model for the detection of criminal behavior that uses facial recognition technology, along with video surveillance and weapons detection is intended to improve public security by recognizing criminal behavior.

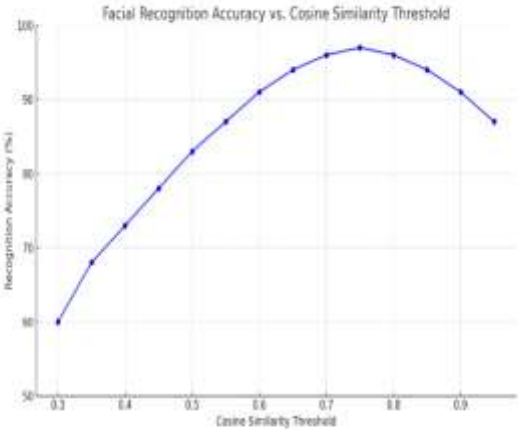

Figure 4: Recognition accuracy varies with cosine similarity thresholds.

The diagram below is used to show the correlation between the performance of the facial recognition system and the cosine similarity threshold. The accuracy of the system significantly increases with the increase in the cosine threshold from 0.3 to 0.85 and reaches its maximum point at the value of 0.85 which can be considered around 97%.

From the above point on, the accuracy starts decreasing indicating that strict values of threshold might lead to the exclusion of positive results. This proves that in order to ensure efficient automation in surveillance, the value of cosine similarity needs to be strictly specified and should be within 0.75 to 0.85.

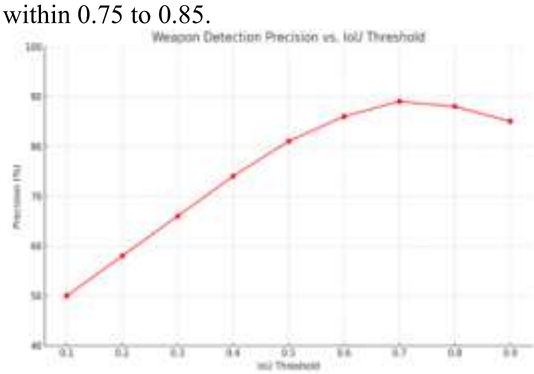

Figure 5: Displays precision changes with different IoU thresholds for bounding box filtering.

This figure 5 illustrates the impact of weapon detection accuracy on the IoU standards that have been set. Bounding box predictions are noted to increase proportionally with the increasing IoU standards from 0.1 to 0.7 with the highest rate recorded at 89% at 0.7. Too much bounding box predictions above 0.7 lead to decreases in precision implying that setting excessively high threshold levels on bounding box intersection may overlook the actual detections. Therefore, the fact remains that the level of IoU must be set appropriately in this scenario, and the recommended range will fall within the values of 0.6 to 0.7 in order to obtain a sufficient level of weapon detection precision.

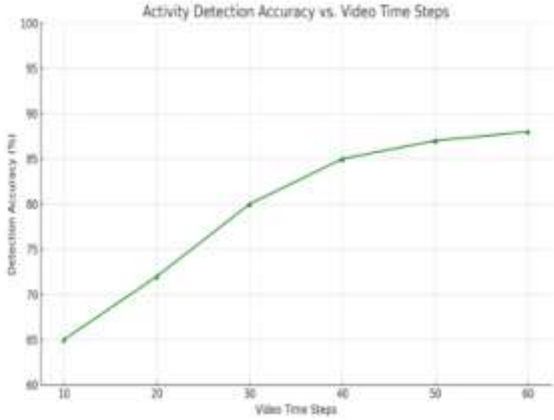

Figure 6: accuracy improvements as more time steps (video frames) are analyzed.

In figure 6 above, it is evident that the more the video time steps evaluated, the greater the accuracy of detection becomes. Alongside the increase in the number of frames, the accuracy of the model in detection also increases, starting with 10 frames, and the improvement is experienced up to 60 frames, where the accuracy rises from 65% to 88%. This demonstrates a beneficial feature in TCNs, whereby by using temporal dependencies over time, the detection becomes more reliable. Beyond 50 time steps, any further improvement becomes insignificant since the optimum point is achieved. This implies that 40 - 50 frames are normally sufficient for detection purposes.

The system will increase threat detection efficiency by lowering both false positive and negative results that can help the law enforcement agents deal with the situation better. Besides, it will increase the level of safety for the people due to timely notifications of any suspicious activities, gun violence, and criminals being tracked down by the security agents, who, thus, can interfere and stop any crime from taking place. It is going to be resource-efficient because it won't require human monitoring all the time. It will also have scalability and adaptability because it will grow with time, taking into account new types of situations and data. The system will reduce the number of false alarms and false detections, offering full-fledged monitoring. At last, it will be reliable, responsible for people's privacy and ethical standards, as well as resistant to any kind of evasion and attack.

V. CONCLUSION AND FUTURE WORK

This is an efficient socio-technical system for the identification of crimes based on facial recognition, video recognition, and detection of weapons as a single system. This is an impressive technology that contributes immensely towards ensuring public safety since the technology is able to offer real-time identification of threat situations hence the prevention of the criminal acts through intervention at the early stages. With automated monitoring of threats, there is no requirement for continued human surveillance. Since the technology makes use of multi-modality of systems, then the accuracy of the information being processed is highly accurate, thus reducing false negatives and positives. It is readily adaptable and applicable to high traffic zones and high-security areas in urban environments, thus making sure that evasion of such system is minimized. Privacy concerns are addressed using privacy technologies in accordance with regulations. Moreover, the capability of the system to learn from new input data allows evolution and development of its capabilities to detect objects with growing efficiency. In terms of long-term results, the possibility to achieve lower crime levels in conjunction with higher safety levels in public spaces is hoped for. What is more, the integration with current systems and structures proves effective, allowing high versatility and minimal disruptions in the process of implementation. Hence, it may be claimed that this project offers efficient help to law enforcement institutions in their work. However, the possibility of large scale usage of this tool can lead to reformation of current methodologies of crime prevention. In terms of improvement, one should aim at higher quality detection, better generalizability with multiple types of datasets, and faster operations. IoT integration and expansion into smart cities can increase efficiency and applicability. Testing of the system constantly with adjustments made to it and testing in critical situations will allow proving its reliability. Finally, implementation of the system in tandem with other methods of policing will help develop an ethics framework of surveillance.

VI. REFERENCES

1. Nazanin sedaghati et al., "Application of human activity/action recognition: a review" Multimedia Tools and Applications, Volume 84, pages 33475–33504, 2025.
2. KASHAF U. DUJA et al., "Video Surveillance Anomaly Detection: A Review on Deep Learning Benchmarks" IEEE Access, Volume 12, 2024.

3. H. L. GURURAJ et al., “A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges” IEEE Access, Volume 12, 2024.
4. S. Arul et al., “Deep Learning Crime Monitoring and Alerting System using Hybrid Feature Engineering technique with IoT Technology” IEEE Xplore: 2025
https://ieeexplore.ieee.org/document/11042792
5. Rohan Kuppili et al., “Real-Time Weapon Detection System for Enhanced Security Monitoring” IEEE Xplore: 2025
6. R. Roopa et al., “An Optimized Hybrid CNN-AlexNet Framework for Real-Time Crime Activity and Suspicious Object Detection” Springer, 2025
https://link.springer.com/chapter/10.1007/978-981-96-5238-9_33
7. Aditi Ganji et al., “Systematic Review on Models Used for Crime Detection Using PRISMA Approach” Springer, 2025
https://link.springer.com/chapter/10.1007/978-981-96-4148-2_28
8. Sagar Rajebhosale et al., “PublicVision: A Secure Smart Surveillance System for Crowd Behavior Recognition Zones” IRJET, Volume: 11 Issue:2 ,2024.
9. MARWA QARAQE et al., “PublicVision: A Secure Smart Surveillance System for Crowd Behavior Recognition” Volume 12, 2024.
10. Abdulraqeb Alhammadi et al., “Artificial Intelligence in 6G Wireless Networks: Opportunities, Applications, and Challenges” Hindawi, 2024
11. Yehia Ibrahim Alzoubi et al., “Research trends in deep learning and machine learning for cloud computing security” Artificial Intelligence Review, Volume 57, article number 132, 2024.
12. Varun Mandalapu et al., “Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions” Vol. 11, 2023.
13. HYUN-BIN KIM et al., “Surveillance System for Real-Time High-Precision Recognition of Criminal Faces from Wild Videos” IEEE Access, Volume 11, 2023.
14. Sarah Bin Hulayyil et al., “Machine-Learning-Based Vulnerability Detection and Classification in Internet of Things Device Security” MDPL,2023.
15. Mrs. Sathya G et al., “facial recognition for criminal detection using deep learning” IRJEDT, Volume: 05 Issue:04,2023.
16. Surbhi Singla et al., “Detecting Criminal Activities from CCTV by using Object Detection and machine Learning Algorithms” IEEE Xplore: 2023
https://ieeexplore.ieee.org/document/10205699
17. Shivalila Hangaragi et al., “Face Detection and Recognition Using Face Mesh and Deep Neural Network” Procedia Computer Science, Elsevier , Volume 218, 2023, Pages 741-749. https://doi.org/10.1016/j.procs.2023.01.054
18. Fangfang Xiong et al., “Research on Face Image Recognition System Based on Computer Artificial Intelligence Technology” IEEE Xplore: 2023
https://ieeexplore.ieee.org/document/10176850
19. K. Kranthi Kumar et al., “Criminal face identification system using deep learning algorithm multi-task cascade neural network (MTCNN)” Materials Today: Proceedings, Elsevier, Volume 80, Part 3, 2023, Pages 2406-2410
https://doi.org/10.1016/j.matpr.2021.06.373
20. Maryam Qasim Gandapur et al., “E2E-VSDL: End-to-end video surveillance-based deep learning model to detect and prevent criminal activities” Image and Vision Computing Volume 123, 2022, 104467.
21. Stephen Khor Wen Hwooi et al., “Deep Learning-Based Approach for Continuous Affect Prediction from Facial Expression Images in Valence-Arousal Space” IEEE Access, Volume 10,2022.
22. Athina Ilioudi et al., “Deep Learning for Object Detection and Segmentation in Videos: Toward an Integration with Domain Knowledge” IEEE Access, Volume 10, 2022.
23. Neil Shah1 et al., “Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention” Visual Computing for Industry, Biomedicine, and Art 2021.
24. Marcus Smith et al., “Biometric Identification, Law and Ethics” IEEE Access ,2021.
25. Ahmed Abdelmoamen Ahmed et al., “Hawk-Eye: An AI-Powered Threat Detector for Intelligent Surveillance Cameras” IEEE Access, 2021.
26. Shruthi et al., “A human behavior analysis model to track object behavior in surveillance videos” Measurement Sensors, Elsevier , Volume 24, 2022, 100454 mhttps://doi.org/10.1016/j.measen.2022.100454
27. Barukula Snehitha et al., “Human Activity Detection from Still Images using Deep Learning Techniques” IEEE Xplore: 2022
https://ieeexplore.ieee.org/document/9730709
28. Velliangiri Sarveshwaran et al., “Investigation on Human Activity Recognition using Deep Learning” Procedia Computer Science, Elsevier , Volume 204, 2022, Pages 73-80
https://doi.org/10.1016/j.procs.2022.08.009\
29. Vandana Japtap et al., “Video-Based Human Activity Detection” IEEE Xplore: 2022
https://ieeexplore.ieee.org/document/9847958

30. ZewenLietal.,
“A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects” IEEE Xplore, Volume: 33 Issue: 12 , 2021
https://ieeexplore.ieee.org/document/9451544
31. Rajib Debnath et al., “A comprehensive survey on computer vision based concepts, methodologies, analysis and applications for automatic gun/knife detection” Journal of Visual Communication and Image Representation, Elsevier, Volume 78, 2021, 103165
https://doi.org/10.1016/j.jvcir.2021.103165
32. Omar Elharrouss et al., “A review of video surveillance systems” Journal of Visual Communication and Image Representation, Elsevier, Volume 77, 2021, 103116
https://doi.org/10.1016/j.jvcir.2021.103116
33. Sheikh Tanjila Naurin et al., “A Proposed Architecture to Suspect and Trace Criminal Activity Using Surveillance Cameras” IEEE Xplore, 2020.
34. Gurlove Singh et al., “Face Detection and Recognition System using Digital Image Processing” IEEE Xplore:2020 https://ieeexplore.ieee.org/document/9074838
35. Serign Modou Bah et al., “An improved face recognition algorithm and its application in attendance management system” Array, Elsevier , Volume 5, 2020, 100014
https://doi.org/10.1016/j.array.2019.100014
36. Harsh Jain Et Al., “Weapon Detection Using Artificial Intelligence and Deep Learning for Security Applications” IEEE Xplore, 2020
37. Virender Singh Et Al., “Real-Time Anomaly Recognition Through CCTV Using Neural Networks” Procedia Computer Science, Elsevier, Volume 173, 2020, Pages 254-263 https://doi.org/10.1016/j.procs.2020.06.030
38. Guruh Fajar Shidik Et Al., “A Systematic Review of Intelligence Video Surveillance: Trends, Techniques, Frameworks, And Datasets” IEEE Access, Volume 7, 2019.
39. Biao Yang et al., “Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images” IEEE Access, Volume 6, 2018.