Real-Time Monitoring and Analysis Through Video Surveillance and Alert Generation for Prompt and Immediate Response

Akshat Kumar[1], Renuka Agrawal[2], Akshra Singh[3], Aaftab Noorani[4], Yashika Jaiswal[5], Preeti Hemnani[6], Safa Hamdare[7]

Department of Computer Science and Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India[1, 2, 3, 4, 5]

Department of Electronics and Telecommunication Engineering, SIES Graduate School of Technology, Mumbai, India[6]

Department of Computer Science, Nottingham Trent University, Nottingham, NG11 8NS, UK[7]

***Abstract*—The efficacy of Closed-Circuit Television systems (CCTV) in residential areas is often linked to the lack of real-time alerts and rapid response mechanisms. Enabling immediate notifications upon the identification of irregularities or aggressive conduct can greatly enhance the possibility of averting serious incidents, or at the very least, significantly mitigate their impact. The integration of an automated system for anomaly detection and monitoring, augmented by a real-time alert mechanism, is now a critical necessity. The proposed work presents an advanced methodology for real-time detection of accidents and violent activities, incorporating a sophisticated alarm system that not only triggers instant alerts but also captures and stores video frames for detailed post-event analysis. MobileNetV2 is utilized for spatial analysis due to its computational efficiency compared to other Convolutional Neural Networks (CNN) architectures, while Visual Geometry Group 16 (VGG16) enhances model accuracy, especially on large-scale datasets. The integration of Bi-directional Long Short-Term Memory (BiLSTM) strengthens temporal continuity, significantly reducing false alarms. The proposed system aims to improve both safety and security by enabling authorities to intervene timely to incidents. Combining rapid computation with high detection accuracy, the proposed model is ideally suited for real-time deployment across both urban and residential settings.**

Keywords*—Rapid response; anomaly detection; MobileNetV2; VGG16; BiLSTM

I. Introduction

The growth in crime and the need for protection especially in the residential compounds has led to the installation of CCTV systems in the compound; the need to enhance safety and control of incidents has also boosted the usage of CCTV systems in residential areas. One typical use of these devices is for real-time monitoring and recording of events, which enable the authorities to respond to any emergent incidents such as violent criminal incidents or accidents occurring in premises [1]. The major drawback with the conventional method of monitoring is that several instances go unnoticed even with the increasing use of the CCTV in commercial and residential areas [2, 3]. Manual monitoring system is vulnerable to human mistakes, weariness, and slow response times. Besides this, there is feedback delay and lost opportunities for timely intervention and prompt action required in case of occurrence of incidents. The necessity for automatic real-time anomaly detection is highlighted by the growing number of CCTV systems installed in homes and in residential societies [4, 5]. Delays in emergency response often stem from the late detection or lack of reporting of violent incidents and traffic accidents in residential premises. This underscores the need for a lightweight, scalable system designed to rapidly detect abnormal events and automatically raise notifications to designated authorities. Implementing such a system would enhance civilian safety and significantly improve traffic management efficiency.

Traditional methods of monitoring, which rely on human operators, are time-consuming and often inefficient, as operators may not be able to keep up with multiple camera feeds or notice critical details in the video. As human fatigue sets in, the effectiveness of these systems further diminishes. Given the vast volume of data generated by modern CCTV networks, manual monitoring is no longer feasible. As a result, automated tools for tasks such as object recognition, classification, and anomaly detection have been developed using deep learning models like Residual Networks (ResNet) and Densely Connected Networks (DenseNet). While these models offer high accuracy, they face challenges in real-time applications due to their computational demands. It has been noted that both ResNet and DenseNet have high accuracy, though, the computing complexity of these two models are massively different. Due to the tensed structure and millions of parameters ResNet cannot be used effectively in real-time scenario and requires a huge amount of RAM and computation power. While DenseNet employs few parameters, it calls for several convolutions per layer and thus, has high processing intensity. These aspects make such models unsuitable for real-time applications since in real-life one has to decide based on the available information, which in turn is valuable in cases where the amount of resources is limited, but the decision has to be made as soon as possible. Earlier automated techniques frequently failed to give real-time alerts and relied on computationally heavy and expensive frameworks. Furthermore, earlier automated systems such as ResNet and DenseNet are not suited for real-time use because of their high processing cost [6]. One of them is ResNet, which helps in achieving high accuracy, but makes the model computationally

expensive. ResNet layers utilize typical convolutions, which demand many multiplications and adds [7]. The deep architecture with many layers translates to many parameters. For instance, ResNet-50 has approximately 25 million parameters. The extensive depth and numerous operations in ResNet contribute to high latency time and computation complexity when employed for anomaly detection [8, 9]. DenseNet connects each layer to every other layer in a feed-forward manner, ensuring maximum information flow between layers [10]. The dense connectivity increases the complexity of the network. Despite the efficient parameter consumption achieved through feature reuse, DenseNet still requires a substantial number of convolutions. The necessity to retain and process feature maps from all previous layers raises memory and computing requirements [11].

The proposed methodology addresses these challenges by providing a scalable and efficient solution for real-time violence detection in residential areas. This system uses MobileNetV2, a convolutional neural network for spatial feature extraction, considerably minimizes the computational overhead while maintaining the accuracy compared to previous models [12]. It uses depth wise separable convolutions, which decompose a standard convolution into two simpler operations: depth wise convolution and pointwise convolution. Furthermore, the model uses BiLSTM networks for temporal sequence analysis for violence detection. This enables the model to better comprehend the context and course of events, minimizing false predictions thus increasing the accuracy of violence detection. By focusing on temporal coherence, it is ensured that the model not only accurately recognizes violent acts but also distinguishes them from nonviolent activity [13].

Despite advances in anomaly detection, most existing research focuses solely on detecting accidents or unusual events, often overlooking the crucial step of raising alerts for timely intervention in violent incidents. Many studies primarily enhance the performance of models like CNNs for anomaly detection, but they neglect the integration of alert systems for quick response. For example, Trilles et al. [14] discuss the use of AIoT for anomaly detection but focus on detection rather than alerting. Similarly, Chandrakar et al. [15] improve moving object detection and tracking for traffic surveillance, but their work does not address automated response mechanisms. Ullah et al. [16] combine CNNs and BiLSTM networks for real-time anomaly detection, but their focus is on feature extraction and classification, not alerting systems. Kamble et al. [17] propose a smart surveillance system for anomaly detection but exclude alert mechanisms. Wang et al. [18] utilize DenseNet for anomaly detection, but their work also lacks a real-time alerting system. While these studies focus on detection accuracy, they fail to address the critical need for real-time alerts, a limitation that impedes their practical application in real-world security settings.

The prime objective of the proposed work is to design a lightweight model for detecting similar anomalies in residential areas. Besides an alert mechanism is also incorporated, within the model to raise alert to inform concerned authorities for prompt action and also to mitigate the effect of incidents occurring in residential premises. Hence, to meet the needs that require real-time monitoring, automated systems that can promptly generate the alert to enable authorities respond to occurrences are needed.

This paper aims to compare various models and techniques to determine which offers the best performance for anomaly detection in the form of accidents or violence occurring in residential areas. The rest of this paper is structured as follows: the literature review in Section II, the methodology is detailed in Section III, results findings are shown in Section IV, followed by model comparison and analysis discussion in Section V, and the final conclusion with future research directions are provided in Section VI.

II. Literature Review

The domain of video surveillance and security has seen significant advancements with the use of various AI related Machine Learning or Deep learning methodologies. These technologies have enabled the development of sophisticated systems capable of real-time behavior analysis, object detection, anomaly detection, and activity recognition.

Khan et al. [3] proposed a method combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for behavior analysis, achieving 92% accuracy on the UCF101 dataset. While effective for real-time human action recognition, this approach is constrained by the dataset's limited variety and struggles with complex backgrounds. Similarly, Lindemann et al. [19] utilized Long Short-Term Memory (LSTM) networks for activity recognition, achieving 90% accuracy with the KTH dataset. Although LSTMs excel at distinguishing activities, they face challenges with overlapping actions and require large training datasets.

For anomaly detection, Reddy [10] applied VGG16-based neural networks on the Avenue Dataset, achieving a high Area Under the Curve (AUC) of 88%. However, this method struggles with anomaly diversity and distinguishing between minor and major anomalies. Raiyn and Toledo [13] leveraged Generative Adversarial Networks (GANs) for pattern recognition, effectively augmenting training data using the CIFAR-10 dataset. Despite its efficacy, GANs demand extensive training and may produce unrealistic samples.

In facial recognition, Wang et al. [18] employed self-supervised learning models, achieving 95% accuracy on the LFW dataset. These models perform well without labeled data but are sensitive to occlusions and lighting variations. For image classification, Huang et al. [20] introduced DenseNet-based models that demonstrated 96% accuracy on the ImageNet dataset, showcasing high performance. However, they require substantial computational resources and are prone to overfitting on smaller datasets.

In object detection, Piekarski et al. [21] utilized YOLO (You Only Look Once), achieving 83% mean Average Precision (mAP) on the COCO dataset. While efficient for real-time detection, YOLO struggles with small or overlapping objects and shares the computational demands of more complex models. Hassan et al. [22] explored Faster R-CNN for object detection, achieving 84% mAP on the Pascal VOC dataset. This method offers high precision but requires significant computational power and exhibits slower inference times compared to single-shot detectors.

While these methodologies achieve remarkable results, their limitations highlight the need for improved algorithms to handle complex scenarios, reduce computational overhead, and enhance generalizability. Table I summarizes the methodologies, domains, datasets, performances, and limitations discussed in this section.

TABLE I. WORK DONE ON ANOMALY DETECTION

Ref. No.	Methodology Used	Domain	Data Set Used	Performance	Outcome/Limitations
[3]	Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)	Behavior Analysis	UCF101	92% Accuracy	Real-time human action recognition; limited by dataset variety; performance may degrade with complex backgrounds
[10]	Autoencoders	Anomaly Detection	Avenue Dataset	88% AUC	High accuracy in anomaly detection; limited by anomaly diversity and type; difficulty in distinguishing between minor and major anomalies.
[13]	Generative Adversarial Networks (GANs)	Pattern Recognition	CIFAR-10	Improved data generation	Effective for training data augmentation; requires extensive training; potential for generating unrealistic samples
[18]	Self-Supervised Learning Models	Facial Recognition	LFW	95% accuracy	High accuracy without labeled data; sensitive to occlusions and variations in lighting conditions
[19]	Faster Region-based Convolutional Neural Network (R-CNN)	Object Detection	Pascal VOC	84% mAP	High accuracy for object detection; slower inference time compared to single-shot detectors; requires significant computational resources
[20]	Long Short-Term Memory (LSTM)	Activity Recognition	KTH Dataset	90% accuracy	Accurate activity recognition; struggles with overlapping activities and requires large training dataset
[21]	YOLO (You Only Look Once)	Object Detection	COCO	83% mAP	Enhanced real-time object detection; struggles with small or overlapping objects; requires significant computational resources.
[22]	Residual Network (ResNet)	Image Classification	ImageNet	96% accuracy	Highly accurate image classification; requires large computational resources and training data; potential overfitting on small datasets

III. METHODOLOGY

Using CCTV video data, the suggested traffic anomaly detection system effectively detects and classifies traffic incidents in real-time by utilizing deep learning models. The proposed model for anomaly detection in residential areas aims to detect dual anomalies occurring in residential areas, one is violence detection, and another is accidents occurring in residential areas. The system uses a step-by-step methodology that includes object detection, geographical and temporal feature extraction, and data preprocessing.

To overcome the challenges of computational complexity and real-time anomaly detection the suggested method employs MobileNetV2, a lightweight CNN architecture designed specifically for deep learning on resource constraint devices. Fig. 1 outlines the methodology steps for both whereas Fig. 2 and Fig. 3 depict violence and accident detection model respectively.

For violence detection, an ensemble model of BiLSTM for temporal feature extraction and MobileNetV2 for Spatial feature extraction is used. For accident detection the ensemble model employed is a combination of VGG16 and MobileNetV2. Further for violence detection 800 samples of violence and nonviolence in the form of video are taken for training and testing ensemble model proposed for detection. On the other hand, accident detection dataset consists of 990 files in the form of frames as images taken from public available dataset for accidents occurring in residential areas. To overcome the challenges of computational complexity and real-time anomaly detection the suggested method employs MobileNetV2, a lightweight CNN architecture designed specifically for deep learning on resource constraint devices. Hence, MobileNetV2 as a model has lesser computational complexity than typical CNNs and therefore ideal for anomaly detection workloads involving real time processing. Due to all of this breakdown, MobileNetV2 emerges as a perfect real-time CCTV footage processing option since it cuts all parameters and computation needs greatly. Another advantage of MobileNetV2 architecture is that it is capable of utilizing pre-trained weights which are trained on large datasets such as ImageNet.

As mentioned earlier, anomaly detection of proposed model includes violence detection and accident detection in residential areas. The first type of anomaly being detected is accidents occurring in residential societies specifically in parking areas. The proposed model can obtain the high-level spatial features from the CCTV feed employing these pre-trained weights and the technique will then be able to effectively identify the outliers such as the violent crime or the traffic accident. Furthermore, the lightweight structure of the model also gives it flexibility in its operational environment and enables its deployment in scenarios that have hardware limitations such as in embedded systems and edge computing systems. The system further improves the MobileNetV2 by adding extra convolutional layers that helps in the feature extraction process. After characteristics of space have been extracted, the model is able to group identified patterns into pre-defined categories. For In traffic monitoring the categories could be “Accident” “No Accident” where the two main categories which will be flagged will be “Violence” and “Non-Violence” for residential areas surveillance. This allows the system to monitor CCTV footage by itself and generate real-time alerts when an anomaly is

detected thus ensuring that the authorities act promptly to prevent further damage from happening. Fig. 2 shows the Methodology involve in violence detection for anomaly detection in residential areas.

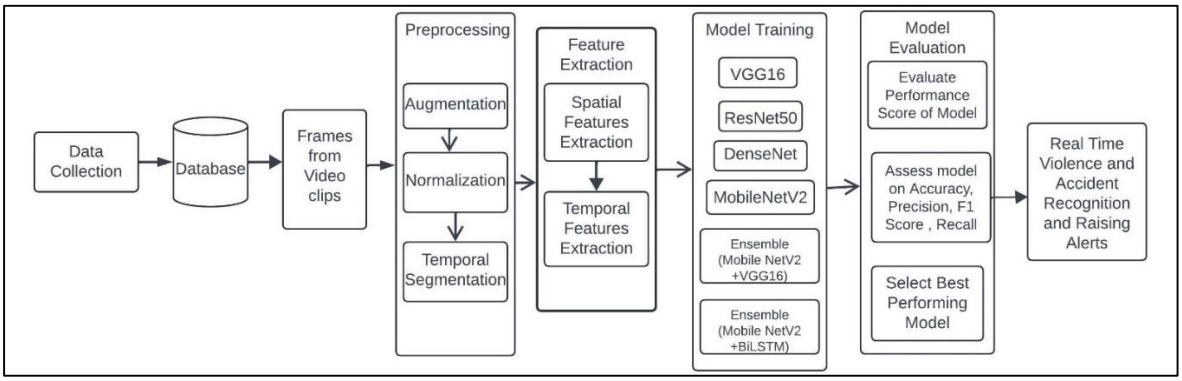

Fig. 1. Methodology of proposed anomaly detection in residential areas.

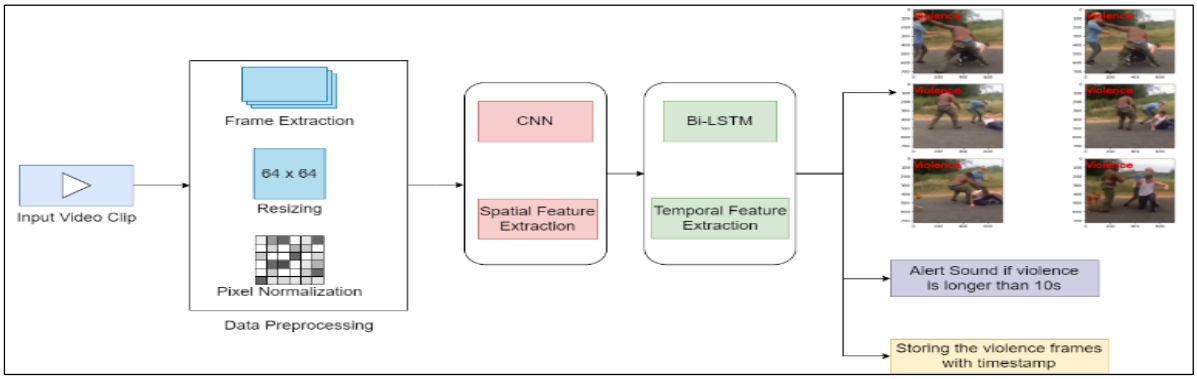

Fig. 2. Architecture of violence detection.

To have a comprehensive analysis the proposed model which focuses on anomaly detection and raising alerts considered dataset in the form of video clips for Violence detection and in the form of Frames from video for Accident detection. This ensures that system will be able to raise alerts on detection of anomaly, irrespective if the input is in the form of image or video. Further for violence detection the system proposes an ensemble model of MobileNetV2 and BiLSTM for spatial and temporal features extraction [23]. However for accident detection the proposed model utilizes ensemble model of MobileNetV2 and VGG16 for features extraction [19]. The architecture for accident detection models is given in the images Fig. 3:

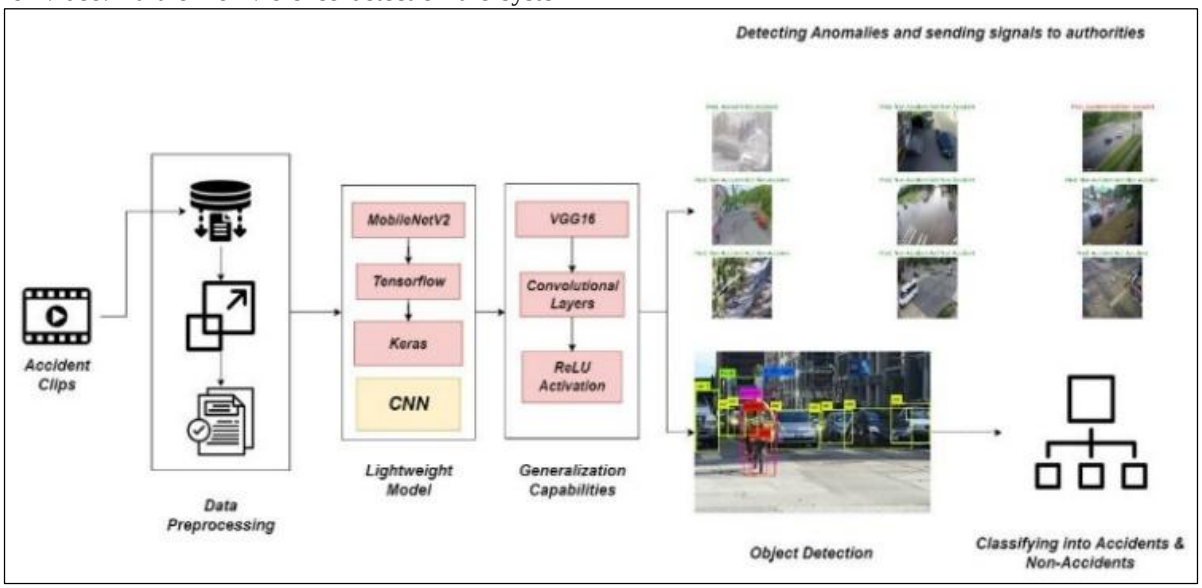

Fig. 3. Architecture of accident detection.

A. Data Preprocessing

The "Real Life Violence Situations Dataset" from Kaggle was used, which contained 800 MP4 files of both violent and non-violent clips with maximum length of 5 seconds. Violent clips were typically real street fight situations with people fighting bare-handed or with items such as sticks and rods. While the Non-Violent clips were various human activities like sports, eating, walking and other ordinary activities. This diversified dataset allowed the model to generalize more reliably and distinguish between violent and nonviolent scenarios. Preprocessing is considered as a critical step in preparing the video data for both model training and testing. By extracting frames at regular intervals, it was ensured that the frames are uniformly distributed across the video. The proposed model extracted 16 frames per video, ensuring it efficiently captures the temporal properties of video content, allowing the model to comprehend the development of activities across time without being overloaded by redundant information.

The first step to process videos into one scene is to extricate frames from the video stream. This is going to be necessary for the later steps of temporal analysis. The first step in the pre-processing of an input image is swapping the color channels, i.e., RGB channel rearranging to make the image compatible with Keras. The next step is the resizing of the input image to a fixed size, i.e., 224 × 224, without taking the aspect ratio of the image into consideration. In the last step in the pre-processing of an image, mean subtraction is applied to the image, where the mean is calculated for all pixel values in the image and the mean is subtracted from the pixel values. The reason behind this step is to make a standardization to which the other parameters such as weights and biases can refer to. The pre-processing of an image makes the image ready to be provided to a CNN model. In the testing phase of this research, the same set of techniques was applied. The testing of the CNN model was performed on videos, so while testing, the frames of the videos were looped, and all the frames were subjected to the same pre-processing as the training images.

Another technique used in the pre-processing phase of training is data augmentation. This technique is very popular while working with images, as it helps create more data, which helps the computation models to be able to generalize better. In this research, multiple data augmentation techniques were used, i.e., shift, rotation, shear, Flip, etc. Data augmentation was not applicable in the testing phase, as there was no need to multiply data and generate more. Raw frames from the video have various shapes each sample provides a set of raw frames with varying size, so before feeding those frames to deep learning models, the frames are resized to have the fixed, standard shape. The pixel values for the frames which have been retrieved are then scaled to fit into an agent ratio which is mostly a range from 0 to 1. This in turn means that consistencies of the intensity values are preserved, enhancing the performance of subsequent CNN based models.

B. Lightweight Model

MobileNetV2 model is used to get spatial features for the frames that were normalized from the frames extracted from the videos. MobileNetV2 is a light CNN model with high accuracy so it used in real-time applications such as traffic anomaly detection and violence recognition in residential areas. MobileNetV2's architecture is the depth wise separable convolution that reduces the amount of computation required while maintaining the extraction of features accurate. It captures immobile objects in a scene such as cars, traffic flow and other elements of the scene.

Then there is Batch Normalization followed by ReLU as the activation functions after the convolutional layers. This guarantees the model's non-linearities and promotes quicker training convergence: This guarantees the model's non-linearities and promotes quicker training convergence:

$$z = ReLU\ (BN\ Wconv * X)$$

Where Wconv are convolution weights, X is input, and BN refers to Batch Normalization.

C. Generalization using (VGG16)

To measure the temporal features; the features that are spatial are first extracted using mobile NetV2; then the VGG16 is used. They work to provide a perspective of the deeper convolutional layers of the VGG16, to identify how objects within frames of a video move. Dependencies between Space and Time: The weights provided by the VGG16 architecture help in model's ability to detect intricate temporal patterns, which are crucial for detecting such anomalies as mishaps or erratic movements.

$$ht = \sigma\ (WhhHt\text{-}1 + WxhXt)$$

Where Whh are weights for temporal connections, and Wxh captures spatial-temporal dependencies as shown in Fig. 4.

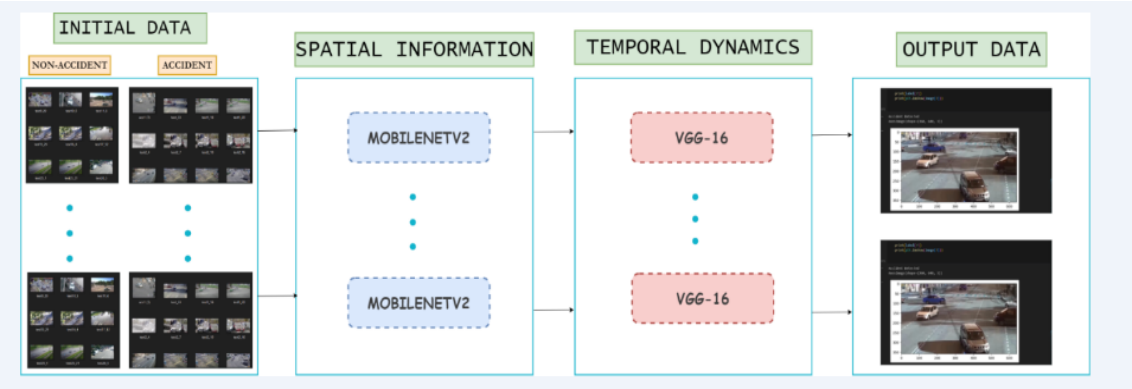

Fig. 4. Extracting temporal and spatial features for anomaly detection.

D. Object Identification

A variety of CNN based object detection algorithms detect objects in each video frame such as vehicles, pedestrians, and others traffic elements. The objects detected have a rectangular frame placed around and the movement of these objects is recorded over time. For this purpose, movement of the observed objects is compared between frames to identify any odd, unexpected behavior which could consist of crash occurrences, traffic congestions, or irregular driving patterns.

1) Accidents Classification and Anomaly Detection: Based on the space and time information gathered all the events depicted in the video as either accidents or non-accidents. In this case, the CNN and the Bi-LSTM models are very important in the identification of traffic pattern such as sudden halts, collisions or erratic movements indicating an accident [20]. The system monitors the occurrence of any irregularities in the flow and activity of the traffic on the road. For this, the technology identifies features that are likely to be abnormal such as times when the car suddenly stops, unpredictably veers or has a crash. Once an irregularity or accident has occurred, a real time alert is made to facilitate the correction of the problem. This may lead to an alert to the right authorities to ensure the situation does not happen again hence ensuring that it act as a stop gap measure.

2) Violence Detection and Time-Based Alerts: For events like violence or aggressive behavior (as depicted in the violence detection figure), if such events last longer than 10 seconds, the system triggers an alert sound to draw attention. The violence frames, along with timestamps, are stored for further analysis or evidence, which could be used in legal or administrative actions.

3) Anomaly Detection in violence cases Temporal features: While extraction of spatial characteristics is helpful in extracting any inconsistencies in a certain frame of a CCTV film, the temporal characteristics of the frames are as important as the former in order to detect anomalies. That is; it might take several frames to gather adequate information in traffic accident detection to help determine that an accident had occurred. As similarly with this, in distinguishing between an aggressiveness and mere touching or coming contact within violent activity detection, the events over time is important.

This difficulty can be addressed by the suggested system Bi-directional Long Short-Term Memory (BiLSTM) networks that are made especially to deal with sequential data. With the help of BiLSTM networks, the model is able to look at temporal sequences in forward as well as in the backward direction thereby adding to the understanding of context and flow of such sequences [19].

IV. Results

A comprehensive evaluation strategy was used to validate the effectiveness of the proposed model for detecting violent activities in residential areas. This involved training and testing the model on the "Real Life Violence Situations Dataset" from Kaggle. An 80-20 ratio was used to divide the dataset into training and testing sets. The testing set was used to assess the model's performance after it had been trained using the training set. The confusion matrix which can be seen in Fig. 5 was used to validate the model's efficiency with a primary focus on accuracy and recall in the Table II.

TABLE II. Model Evaluation Metrics for Violence Detection

Classes	Precision	Recall	F1 Score
Non Violent	0.98	0.89	0.94
Violent	0.94	0.99	0.91

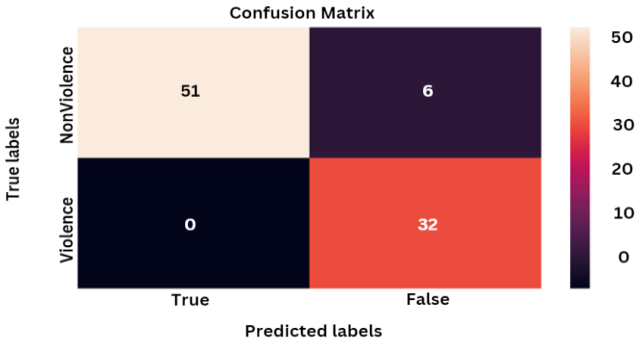

Fig. 5. Confusion matrix for violence classification.

Performance Parameters obtained from ensemble model of MobileNetV2 and VGG16 for accident classification is shown in Table III and the confusion matrix for calculation of performance for the same is shown in Fig. 6.

TABLE III. Model Evaluation Metrics for Accident Detection

Classes	Accuracy	Precision	Recall	F1Score
Non-Accident	0.96	0.96	0.84	0.93
Accident	0.96	0.95	0.91	0.89

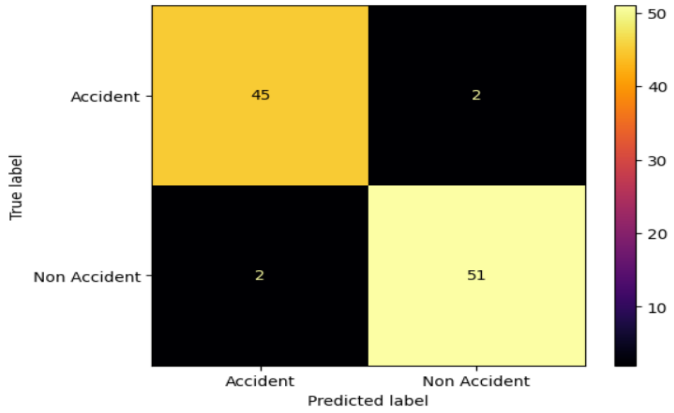

Fig. 6. Confusion matrix for accident classification.

The accuracy depicts the overall accuracy of the model in classifying the video clips correctly. A plot of Loss and Accuracy of the ensemble model for anomaly detection is shown in Table IV. The model's accuracy improved consistently throughout epochs, and the loss consistently decreased with the number of epochs.

TABLE IV. ACCURACY AND LOSS OF THE PROPOSED MODEL OVER EPOCHS

Epoch	Loss	Accuracy	Validation Accuracy
1	0.62	0.55	0.50
3	0.50	0.70	0.62
5	0.40	0.82	0.69
7	0.32	0.89	0.75
9	0.25	0.92	0.80
11	0.15	0.96	0.82
13	0.10	0.96	0.85
15	0.06	0.98	0.86
17	0.04	0.98	0.88
19	0.02	0.99	0.89

For real-time validation, the model was combined with OpenCV and evaluated on live video feeds to imitate real-world settings. The cv2 library was implemented to analyze the video frames. The model demonstrated high accuracy and low loss values for violent/non-violent action recognition, making it suitable for real-time applications.

V. DISCUSSION

The proposed model demonstrates exceptional performance in identifying violent and non-violent activities, achieving an overall test accuracy of 93.25%. Its Recall score of 0.99 for the violent class underscores its effectiveness in identifying nearly all instances of violence, a critical feature for anomaly detection systems in safety-critical environments. This aligns with prior research emphasizing the importance of high recall for detecting relevant instances of violence [22,23].

The model was further validated in real-time using OpenCV, where the cv2 library facilitated frame-by-frame video analysis to simulate real-world scenarios. This integration tested the model's ability to detect violent behavior in live settings, ensuring the classification process remained accurate and prompt. The high accuracy, low loss values, and strong classification metrics achieved in real-time settings validate its practical applicability and reliability for violence/non-violence recognition, as supported by prior findings in similar validation studies [24].

Fig. 7 compares the ROC curves of three models: BiLSTM, MobileNetV2, and the ensemble approach. The BiLSTM model, with an AUC of 0.88, effectively captures temporal features but falls short in spatial feature representation, limiting its ability to distinguish classes [25]. In contrast, MobileNetV2, which specializes in spatial feature extraction, achieves a superior AUC of 0.95. The ensemble model leverages the strengths of both, achieving a near-optimal AUC of 0.98. This performance reflects the ensemble model's ability to comprehensively extract temporal and spatial features, ensuring high classification accuracy across diverse scenarios[26].

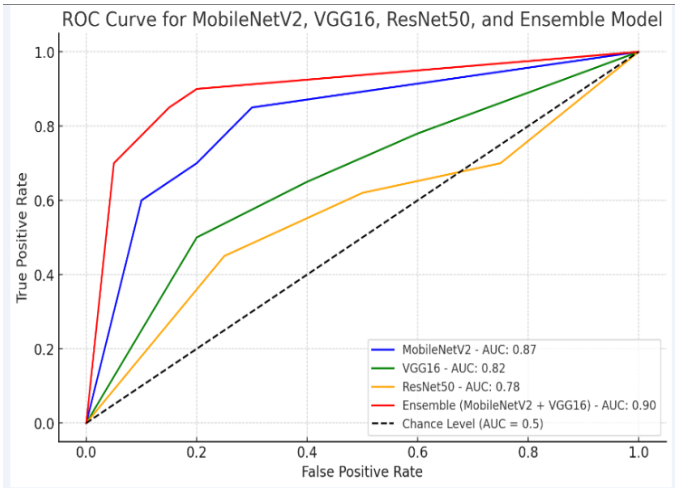

Fig. 7. ROC Curve for comparison of different models.

The comparative analysis in Table V highlights the superiority of the proposed model over existing solutions. Competing models, such as Faster R-CNN with RegNet+, EfficientNetB3 with Cascaded Convolution, and CNN-BiLSTM autoencoder, exhibit limitations, including reliance on low-quality datasets, inability to handle edge cases, and lack of real-time alert mechanisms. The proposed model surpasses these challenges by achieving 96% accuracy and raising alerts within 10 seconds of anomaly detection. Furthermore, its reliance on publicly available datasets enhances its adaptability, addressing limitations seen in prior research.

The dataset was stratified into an 80% training set and a 20% test set to ensure robust model training and unbiased performance evaluation. The confusion matrix reveals an accuracy of 96% and a recall score of 0.90 for accident detection, underscoring the model's efficiency in identifying activities associated with accidents. By raising alerts post-detection, the model further enhances its utility for real-time safety applications.

Thus, the proposed ensemble model effectively combines temporal and spatial feature extraction capabilities, achieving state-of-the-art performance in violence and accident detection. Its high recall and accuracy metrics, superior ROC performance Fig. 7, and unique real-time alert mechanism position it as a robust solution for enhancing community safety. Comparative insights from Table V affirm its versatility, addressing the shortcomings of existing systems and establishing it as a critical tool for anomaly detection in real-world settings. These findings align with prior research, validating its potential as a scalable and reliable anomaly detection system.

TABLE VI. COMPARATIVE ANALYSIS OF PROPOSED MODEL WITH OTHERS

Ref No.	*Dataset Used*	*Model Used*	*Data Balancing*	*Alerts Raised*	*Accuracy (In %)*	*Limitation*
[27]	Animated Customized dataset for Violence detection	Faster RCNN, RegNet+	Yes- Adam Optimzer	No	80.8	Dataset quality can be improved as it consists of animation videos. Also inference time can be further reduced
[28]	MVTec-AD, Railway Track Foreign Object Detection (TFOD)	Cascaded Convolution Self Attention Efficient NetB3	NA	No	99	Anomaly detected on Railway tracks but the model do not work well on edge points.
[29]	unique RGB+D dataset for Bank ATM anomaly detection	CNN-BiLSTM autoencoder framework	No	No	91	Used for Anomaly detection of ATM without raising alerts.
Proposed Model	Violence Recognition from Videos- Kaggle and Accident detection from CCTV footage Dataset .	Ensemble lightweight model for Violence and accident detection	Yes- SMOTE	Raise alerts after 10 sec of anomaly detection	96	Besides detection of anomaly as accidents or violence activity in residential areas, post detection alerts are also raised. Moreover, datasets used are those available in public domain so effective in varied situations.

VI. CONCLUSION

The use of AI-driven machine learning and deep learning techniques has significantly advanced the field of video surveillance and security. These technologies enable near real-time behavior analysis, object recognition, anomaly detection, and activity recognition, enhancing overall system efficiency. The proposed methodology focuses on detecting accidents and violent activities, generating and storing frames for post-event analysis, and raising real-time alerts for timely intervention. This dual functionality—anomaly detection and alert generation—addresses the critical need for proactive safety measures, particularly in residential areas where CCTV systems are primarily installed to detect and respond to emergencies. Unlike many existing anomaly detection models, which are computationally expensive, lack real-time responsiveness, and fail to trigger alerts, the proposed system is lightweight, efficient, and practical for real-world applications. By integrating an alert mechanism, the model ensures that authorities are notified promptly, reducing the impact of undesirable activities and improving community safety.

Future work will focus on improving the system to handle more complex situations, like detecting early signs of violence or predicting accidents before they happen. The model will also be improved to work better in different environments, such as low lighting or when objects are blocking the view. Additionally, using more types of data, like combining video with sound or sensor information, could make the system better at detecting unusual activities. Lastly, the system will be optimized for use on smaller devices, ensuring it can run quickly and efficiently without needing too much computing power.

REFERENCES

[1] Piza, E.L., 2018. The crime prevention effect of CCTV in public places: A propensity score analysis. Journal of crime and justice, 41(1), pp.14-30. doi.org/10.1080/0735648X.2016.1226931

[2] Piza, E.L., Welsh, B.C., Farrington, D.P. and Thomas, A.L., 2019. CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis. *Criminology & public policy*, *18*(1), pp.135-159. doi.org/10.1111/1745-9133.12419.

[3] S. W. Khan *et al.*, 'Anomaly Detection in Traffic Surveillance Videos Using Deep Learning', *Sensors*, vol. 22, no. 17, Sep. 2022, doi: 10.3390/s22176563.

[4] Kumar, K.K. and Venkateswara Reddy, H., 2022. Crime activities prediction system in video surveillance by an optimized deep learning framework. Concurrency and Computation: Practice and Experience, 34(11), p.e6852. https://doi.org/10.1002/cpe.6852

[5] Karunarathne, L., 2024. Enhancing Security: Deep Learning Models for Anomaly Detection in Surveillance Videos.

[6] Elmetwally, A., Eldeeb, R. and Elmougy, S., 2024. Deep learning based anomaly detection in real-time video. Multimedia Tools and Applications, pp.1-17. https://doi.org/10.1007/s11042-024-19116-9.

[7] Rezaee, K., Rezakhani, S.M., Khosravi, M.R. and Moghimi, M.K., 2024. A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. *Personal and Ubiquitous Computing*, *28*(1), pp.135-151.

[8] Joshi, M. and Chaudhari, J., 2022. Anomaly Detection in Video Surveillance using SlowFast Resnet-50. *International Journal of Advanced Computer Science and Applications*, *13*(10).

[9] Qian, H., Zhou, X. and Zheng, M., 2020. Abnormal Behavior Detection and Recognition Method Based on Improved ResNet Model. *Computers, Materials & Continua*, *65*(3).

[10] G. Venkata Rami Reddy, 'DETECTING ABNORMALITIES USING VGG 16 NEURAL NETWORKS: AN ANOMALY DETECTION FRAMEWORK', pp. 1484–1491, 2022, doi: 10.48047/nq.2022.20.4.nq22380.

[11] Rahman, M.M., Afrin, M.S., Atikuzzaman, M. and Rahaman, M.A., 2021, December. Real-time anomaly detection and classification from surveillance cameras using Deep Neural Network. In 2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1-6). IEEE.

[12] Sivakumar, G., Mogesh, G., Pragatheeswaran, N. and Sambathkumar, T., 2024. Video Anomaly Detection in Crime Analysis using Deep learning Architecture-A survey. *Journal of Trends in Computer Science and Smart Technology*, *6*(1), pp.1-17.

[13] J. Raiyn and T. Toledo, 'Real-Time Road Traffic Anomaly Detection', *J Transp Technol*, vol. 04, no. 03, pp. 256–266, 2014, doi: 10.4236/jtts.2014.43023.

[14] Trilles, S., Hammad, S.S. and Iskandaryan, D., 2024. Anomaly detection based on artificial intelligence of things: A systematic literature mapping. Internet of Things, p.101063. https://doi.org/10.1016/j.iot.2024.101063

[15] Chandrakar R, Miri R, Kushwaha A (2022) Enhanced the moving object detection and object tracking for traffic surveillance using RBF-FDLNN and CBF algorithm. Expert Syst Appl 191:116306, ISSN: 0957-4174. https://doi.org/10.1016/j.eswa.2021.116306

[16] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M. and Baik, S.W., 2021. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimedia tools and applications*, *80*, pp.16979-16995.

[17] Kamble, K., Jadhav, P., Shanware, A. and Chitte, P., 2022. Smart Surveillance System for Anomaly Recognition. In *ITM Web of Conferences* (Vol. 44, p. 02003). EDP Sciences.

[18] G. Wang, Z. Guo, X. Wan, and X. Zheng, 'Study on Image Classification Algorithm Based on Improved DenseNet', in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jun. 2021. doi: 10.1088/1742-6596/1952/2/022011.

[19] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich, 'A survey on long short-term memory networks for time series prediction', in *Procedia CIRP*, Elsevier B.V., 2021, pp. 650–655. doi: 10.1016/j.procir.2021.03.088.

[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, 'Densely connected convolutional networks', in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Institute of Electrical and Electronics Engineers Inc., Nov. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[21] M. Piekarski, J. Jaworek-Korjakowska, A. I. Wawrzyniak, and M. Gorgon, 'Convolutional neural network architecture for beam instabilities identification in Synchrotron Radiation Systems as an anomaly detection problem', Measurement (Lond), vol. 165, Dec. 2020, doi: 10.1016/j.measurement.2020.108116.

[22] N. Hassan, A. S. M. Miah, and J. Shin, 'A Deep Bidirectional LSTM Model Enhanced by Transfer-Learning-Based Feature Extraction for Dynamic Human Activity Recognition', Applied Sciences (Switzerland), vol. 14, no. 2, Jan. 2024, doi: 10.3390/app14020603.

[23] C. Zeng, D. Zhu, Z. Wang, M. Wu, W. Xiong, and N. Zhao, 'Spatial and temporal learning representation for end-to-end recording device identification', *EURASIP J Adv Signal Process*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13634-021-00763-1.

[24] C Yixin Tang, Yu Chen, Sagar A.S.M. Sharifuzzaman, Tie Li, An automatic fine-grained violence detection system for animation based on modified faster R-CNN, Expert Systems with Applications, Volume 237, Part C,2024, 121691, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.121691.

[25] Agrawal, R., Singh, J., Ghosh, S.M. (2020). Performance Appraisal of an Educational Institute Using Data Mining Techniques. In: Iyer, B., Deshpande, P., Sharma, S., Shiurkar, U. (eds) Computing in Engineering and Technology. Advances in Intelligent Systems and Computing, vol 1025. Springer, Singapore. https://doi.org/10.1007/978-981-32-9515-5

[26] Raja, R., Sharma, P.C., Mahmood, M.R. and Saini, D.K., 2023. Analysis of anomaly detection in surveillance video: recent trends and future vision. Multimedia Tools and Applications, 82(8), pp.12635-12651.

[27] Sahay, K.B., Balachander, B., Jagadeesh, B., Kumar, G.A., Kumar, R. and Parvathy, L.R., 2022. A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques. *Computers and Electrical Engineering*, *103*, p.108319.

[28] Liu, R., Liu, W., Duan, M., Xie, W., Dai, Y. and Liao, X., 2024. MemFormer: A memory based unified model for anomaly detection on metro railway tracks. Expert Systems with Applications, 237, p.121509.

[29] Khaire, P. and Kumar, P., 2022. A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments. Forensic Science International: Digital Investigation, 40, p.301346.