Shivam Namdev, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

An Intelligent Deep Learning Framework for Violence Detection and Criminal Activity Identification in Smart Surveillance Systems

Shivam Namdev[1], Bhanu Pratap Singh[2]
[1]M. Tech Scholar, [2]HOD
[1,2]Department Computer Science and Engineering
[1,2]Aditya college of technology and science, Satna

Abstract- The increasing growth of urbanization, public gatherings, and security threats has created a strong demand for intelligent surveillance systems capable of automatically detecting violent activities and identifying criminal behavior in real time. Traditional surveillance systems mainly rely on manual monitoring, which often leads to delayed response, human errors, and inefficient threat analysis. To address these limitations, this research proposes an Intelligent Deep Learning Framework for Violence Detection and Criminal Activity Identification in Smart Surveillance Systems using advanced Deep Learning architectures and video analytics techniques. The proposed framework integrates preprocessing, segmentation, feature extraction, and hybrid Deep Learning models including ResNet50, MobileNetV2, LSTM, and the proposed InceptionV3 + LSTM architecture for violence classification. The system processes surveillance videos by extracting spatial and temporal features to accurately distinguish violent and non-violent activities. The dataset consists of two categories, namely Violence and Non-Violence, containing real-world surveillance scenarios such as sports, crowd movement, eating, singing, and violent human interactions. Experimental analysis demonstrates that the proposed InceptionV3 + LSTM model achieved superior performance with improved training accuracy, validation accuracy, and reduced loss values compared with existing models. The framework effectively captures complex motion patterns, human interactions, and abnormal behaviors in surveillance videos while reducing false detections. The proposed intelligent surveillance system can be applied in smart cities, public transportation, educational institutions, stadiums, and crowded environments to improve public safety, automated threat detection, and real-time security monitoring.

Keywords— Violence Detection, Smart Surveillance System, Deep Learning, Criminal Activity Identification, InceptionV3, LSTM, Computer Vision, Video Analytics, Public Safety, Intelligent Security Monitoring

I. INTRODUCTION

The rapid advancement of smart city infrastructure, urban population growth, and increasing public security concerns have significantly increased the importance of intelligent surveillance systems for maintaining public safety and crime prevention. Modern surveillance systems are widely deployed in transportation hubs, shopping malls, educational institutions, public streets, stadiums, and government organizations to monitor suspicious activities and prevent violent incidents. However, traditional surveillance systems primarily depend on human operators to continuously observe surveillance footage, which often results in delayed response times, monitoring fatigue, and reduced operational efficiency.

Violence detection and criminal activity identification have become major research areas in computer vision and artificial intelligence due to the growing need for automated threat analysis and proactive security management. Detecting violent activities from surveillance videos is a highly challenging task because of complex environmental conditions such as dynamic backgrounds, crowd density, illumination changes, occlusion, camera movement, and rapid human interactions. Therefore, intelligent automated systems capable of analyzing visual patterns and identifying abnormal behaviors are essential for improving public safety and emergency response mechanisms.

Recent advancements in Deep Learning, Machine Learning, and Computer Vision have significantly improved the capability of intelligent surveillance systems. Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, transfer learning architectures, and hybrid Deep Learning models have demonstrated strong performance in image classification, object detection, behavioral analysis, and video understanding tasks. These technologies can automatically learn spatial and temporal patterns from surveillance videos and effectively classify violent and non-violent activities. Several researchers have proposed Deep Learning-based violence detection systems using CNNs, 3D-CNNs, ConvLSTM networks, transfer learning methods, and optimization-based feature extraction techniques. Smart surveillance systems also increasingly integrate sensor networks, edge computing, cloud computing, and predictive analytics to improve real-time monitoring capabilities. However, many existing systems still suffer from limitations such as high computational complexity, poor generalization, false detections, limited temporal understanding, and insufficient real-time performance.

To overcome these limitations, this research proposes an Intelligent Deep Learning Framework for Violence Detection and Criminal Activity Identification in Smart Surveillance Systems using a hybrid InceptionV3 + LSTM architecture. The proposed model combines the spatial feature extraction capability of InceptionV3 with the temporal sequence learning capability of LSTM networks to improve violence detection performance. The framework includes preprocessing, segmentation, feature extraction, model training, and real-time prediction generation for automated surveillance analysis.

The dataset used in this work consists of Violence and Non-Violence video categories containing multiple real-world scenarios such as crowd movement, sports, singing, eating activities, and violent interactions. The proposed framework was evaluated using multiple Deep Learning models including ResNet50, MobileNetV2, LSTM, and InceptionV3 + LSTM. Experimental results demonstrate that the proposed hybrid architecture achieved superior classification accuracy, improved validation performance, and reduced loss values compared with traditional Deep Learning models.
The main objective of this research is to develop an intelligent surveillance framework capable of accurately detecting violent activities and supporting automated public safety management in smart environments. The proposed system provides an efficient solution for real-time violence monitoring, threat identification, and criminal activity analysis in modern intelligent surveillance applications.

II. LITERATURE REVIEW

Recent advancements in artificial intelligence, deep learning, and intelligent surveillance systems have significantly transformed violence detection and criminal activity identification in smart environments. Traditional surveillance systems often depend on manual monitoring, which limits efficiency, increases response time, and reduces the accuracy of threat identification. Modern intelligent surveillance frameworks increasingly integrate deep learning architectures, computer vision techniques, sensor networks, transfer learning, and predictive analytics to automate violence detection and improve public safety management. The following sections review major contributions related to deep learning-based surveillance systems, smart monitoring technologies, and intelligent predictive analytics for violence detection.

1. Deep Learning and Computer Vision Approaches for Violence Detection

Deep learning and computer vision techniques have become highly effective for automated violence detection and criminal identification in surveillance systems. Singh, Patil, and Omkar (2018) proposed a real-time drone surveillance system using a

Scatternet hybrid deep learning network for identifying violent individuals. Their framework demonstrated that aerial surveillance combined with deep learning significantly improves violence monitoring in large-scale public environments.

Ilyas, Obaid, and Bawany (2023) analyzed the role of transfer learning and pre-trained models in surveillance-based violence detection. Their findings indicated that transfer learning significantly enhances model generalization and reduces training complexity when large-scale datasets are unavailable. Similarly, Ullah et al. (2019) proposed an action recognition framework using optimized deep autoencoders and CNN architectures for surveillance data streams in non-stationary environments. Their model effectively captured spatiotemporal behavior patterns and improved abnormal activity recognition accuracy.

Tiwari et al. (2023) introduced a hybrid CNN-LSTM model for automated violence detection and classification in surveillance systems. Their framework combined spatial feature extraction with temporal sequence learning, thereby improving violence classification performance in dynamic environments. Sudhakaran and Lanz (2017) further developed a convolutional long short-term memory (ConvLSTM) framework for violent video detection and demonstrated that temporal sequence analysis significantly improves behavioral understanding in surveillance footage.

Accurate facial analysis and behavioral recognition also contribute to intelligent surveillance systems. Shakyawar and Singh (2024) proposed a facial expression recognition model using deep learning for crowd environments, while Prajapati, Patel, and Singh (2023) developed a deep learning-based gender recognition methodology. Patel and Singh (2023) further explored age and gender recognition using deep learning techniques for automated biometric analysis in surveillance applications.

These studies collectively demonstrate that deep learning architectures such as CNNs, LSTMs, ConvLSTMs, transfer learning models, and hybrid frameworks significantly improve violence detection accuracy, facial recognition, and behavioral analysis in intelligent surveillance systems.

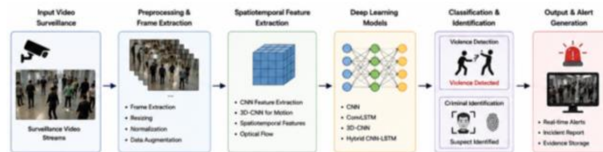

Figure 1: Deep Learning-Based Violence Detection and Criminal Identification Framework

This figure 1 illustrates an intelligent surveillance framework integrating convolutional neural networks (CNNs), ConvLSTM architectures, transfer learning models, facial recognition systems, and spatiotemporal feature extraction for violence detection and criminal activity identification. The framework demonstrates how deep learning techniques process surveillance video streams to recognize suspicious human behavior, detect violent interactions, and identify criminal suspects in real-time monitoring environments.

2. Smart Surveillance, Sensor Networks, and Real-Time Monitoring Systems

Smart surveillance systems increasingly utilize sensor networks, drone surveillance, big data analytics, and intelligent monitoring systems to improve real-time violence detection and public safety management. Simpson (2021) proposed a real-time drone surveillance framework for violent crowd behavior monitoring using unmanned aircraft systems and human autonomy teaming. Their research demonstrated that drone-based surveillance significantly improves large-area security monitoring and emergency response efficiency.

Baba et al. (2019) introduced a sensor network approach for violence detection in smart cities using deep learning. Their framework integrated sensor intelligence with automated surveillance systems to improve crowd monitoring and public safety operations in urban environments. Rezaee et al. (2024) conducted a survey on deep learning-based real-time crowd anomaly detection in distributed video surveillance systems. Their findings highlighted the importance of scalable intelligent surveillance architectures for handling large-scale real-time monitoring tasks.

Fenil et al. (2019) proposed a real-time violence detection framework for football stadium environments using big data analysis and bidirectional LSTM networks. Their study emphasized that integrating big data analytics with sequential deep learning improves crowd violence prediction and security management in densely populated areas.

Weaver, Zelenkauskaite, and Samson (2012) analyzed violent content trends in web videos and emphasized the increasing need for automated violence content filtering systems. Choi and Kim (2024) further studied cognitive, emotional, and social development in children's YouTube content, highlighting the importance of intelligent content monitoring systems for online platforms.

Recent intelligent surveillance applications also extend to biometric identification and object recognition systems. Gupta et al. (2023) developed a human face mask recognition system using ResNet-152, while Baghel, Pahadiya, and Singh (2022) proposed face mask identification using OpenCV and deep learning techniques. Bamne et al. (2020) further introduced transfer learning-based object detection frameworks using convolutional neural networks for intelligent surveillance applications.

These studies collectively indicate that smart surveillance systems integrating sensor networks, drone monitoring, crowd analytics, biometric recognition, and intelligent video processing significantly improve public safety management and real-time violence monitoring capabilities.

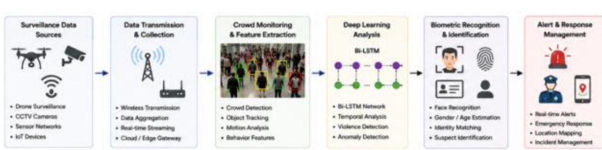

Figure 2: Smart Surveillance and Real-Time Crowd Violence Monitoring Framework

This figure 2 presents a smart surveillance framework integrating drone surveillance systems, sensor networks, crowd monitoring analytics, bidirectional LSTM models, biometric recognition systems, and intelligent CCTV monitoring for real-time violence detection. The framework demonstrates how smart city surveillance technologies improve crowd behavior analysis, emergency response coordination, and automated public safety management in densely populated environments.

3. Intelligent Predictive Analytics, Optimization Techniques, and AI-Based Surveillance Systems

Recent advancements in intelligent predictive analytics and optimization techniques have significantly improved automated surveillance efficiency, violence prediction, and security decision-making. Jebur et al. (2022) reviewed deep learning approaches for anomaly event detection in video surveillance and emphasized that predictive analytics and intelligent event recognition significantly improve automated threat analysis and surveillance performance.

Khan et al. (2024) proposed VD-Net, an edge vision-based surveillance system for violence detection. Their framework demonstrated that edge computing combined with intelligent analytics improves real-time violence detection efficiency while reducing computational latency and bandwidth dependency. Naik and Gopalakrishna (2022) introduced an optimization-oriented feature selection framework using spider monkey-grasshopper optimization and deep neural networks for automated crowd violence detection. Their findings indicated that optimization algorithms significantly improve feature selection efficiency and reduce classification complexity in surveillance systems.

Nagar and Singh (2025) developed ProSoccerTrainer, an AI-powered penalty detection and recommendation system using advanced video analytics. Although focused on sports analytics, the framework demonstrated the effectiveness of intelligent video analysis and predictive AI models for automated activity recognition. Similarly, Ray and Singh (2023) applied Hybrid Task Cascade Region-Based CNN models for cricket score analysis, highlighting the applicability of advanced computer vision techniques in dynamic visual environments.

Recent intelligent systems also integrate predictive analytics into healthcare and cybersecurity applications. Vishwakarma et al. (2023) utilized EfficientNetB3 for brain tumor segmentation and detection, while Ranjan et al. (2022) applied random forest and deep learning techniques for cancer prediction. Waskle, Parashar, and Singh (2020) proposed an intrusion detection system using PCA with random forest, and Nihale et al. (2020) introduced LSTM-based network traffic prediction systems. These studies collectively demonstrate the adaptability of intelligent predictive analytics and deep learning models across multiple domains.

Despite significant progress in intelligent surveillance research, existing literature still lacks a unified framework integrating deep learning architectures, transfer learning, optimization techniques, sensor intelligence, edge computing, predictive analytics, and automated decision-support systems for real-time violence detection and criminal activity identification. Furthermore, limited studies evaluate how integrated intelligent surveillance ecosystems improve predictive threat analysis, emergency response efficiency, and large-scale public security management in smart city environments.

Therefore, there remains a strong need for comprehensive intelligent surveillance frameworks capable of combining deep learning, predictive analytics, optimization algorithms, sensor networks, and automated public safety management systems to improve violence detection reliability and support smart surveillance operations in modern urban environments.

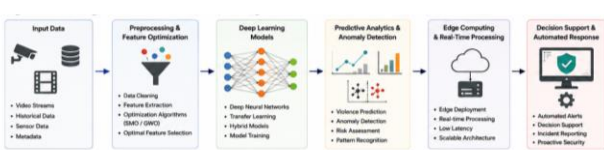

Figure 3: Intelligent Predictive Surveillance and Optimization-Based Violence Detection Framework

This figure 3 depicts an intelligent predictive surveillance framework integrating edge computing, optimization-oriented feature selection, deep neural networks, anomaly detection systems, predictive analytics, and automated decision-support mechanisms for violence detection. The framework illustrates how intelligent analytics and optimization techniques improve classification accuracy, predictive threat analysis, automated security monitoring, and public safety decision-making in modern smart surveillance systems.

Recent studies by U. Singh and co-authors have focused extensively on artificial intelligence, deep learning, machine learning, computer vision, cybersecurity, and healthcare applications. Nagar and Singh (2025) proposed an AI-powered secure soccer penalty detection and performance recommendation system, while Shakyawar and Singh (2024) developed a deep learning model for facial expression recognition in crowds. Prajapati et al. (2023) and Patel and Singh (2023) worked on gender and age recognition using deep learning techniques. Gupta et al. (2023) and Baghel et al. (2022) focused on human face mask recognition using ResNet and OpenCV-based deep learning approaches. Vishwakarma et al. (2023) introduced an EfficientNetB3-based brain tumor segmentation and detection framework, whereas Singh and Songare (2022) developed a GoogLeNet-based monkeypox detection model. Research by Ranjan et al. (2022) explored cancer prediction using Random Forest and deep learning techniques. In recommendation systems, Singh et al. (2023) proposed a job recommendation model using machine learning and deep learning, while Taiwade et al. (2022) implemented a hierarchical K-means clustering approach for friend recommendation systems. Ray and Singh (2023) applied Hybrid Task Cascade Region-Based CNN for cricket score analysis. In cybersecurity and networking, Waskle et al. (2020) proposed an intrusion detection system using PCA with Random Forest, and Nihale et al. (2020) developed an LSTM-based network traffic prediction model. Additionally, Saxena et al. (2020) presented a CNN-based glaucoma detection system, while Bamne et al. (2020) explored transfer learning-based object detection using convolutional neural networks. Rajput et al. (2023) also conducted a comparative study of cloud providers including Azure, Amazon, and Oracle based on service availability and pricing.

III. PROPOSED WORK

1. Proposed Architecture Description

The proposed architecture 4 is an end-to-end deep learning framework for violence detection from image and video data. The system begins with a dataset containing collected surveillance images and videos. These inputs are passed through data cleaning, where noise reduction and artifact removal are performed to improve input quality. After preprocessing, segmentation techniques such as thresholding and region growing are applied to isolate important regions from video frames.

Next, feature extraction is performed to obtain meaningful visual and shape-based features. The processed features are divided into training and testing datasets. The training data is used to train multiple deep learning models, including MobileNetV2, ResNet101, LSTM, and the proposed InceptionV3 + LSTM model. The trained model is then used for prediction, classifying surveillance input into Violence or Non-Violence. Finally, model performance is evaluated using loss, accuracy, validation loss, and validation accuracy.

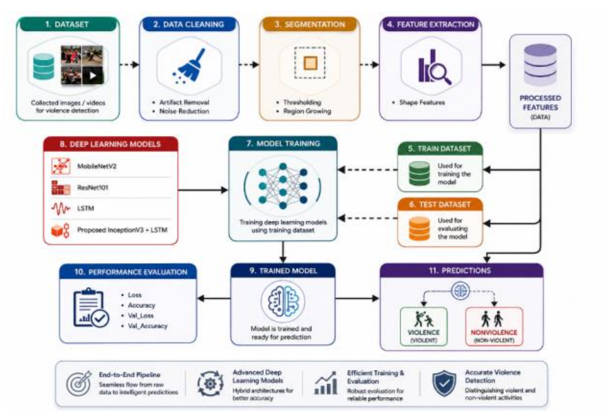

Figure 4. Proposed Architecture

2. Proposed Algorithm Steps with Equations

Proposed Algorithm: Violence Detection Using InceptionV3 + LSTM

Step 1: Input Dataset

Let the input video dataset be:

$$D = \{V_1, V_2, V_3, \ldots, V_n\}$$

where V_irepresents the i^thvideo sample.

Step 2: Frame Extraction

Each video is divided into frames:

$$V_i = \{F_1, F_2, F_3, \ldots, F_t\}$$

where F_tis the frame at time t.

Step 3: Data Preprocessing

Each frame is resized and normalized:

$$F_t' = \frac{F_t}{255}$$

where F_t^'is the normalized frame.

Step 4: Feature Extraction Using InceptionV3

The InceptionV3 model extracts spatial features from each frame:

$$X_t = f_{InceptionV3}(F_t')$$

where X_trepresents the extracted feature vector.

Step 5: Sequence Learning Using LSTM

The extracted frame features are passed to LSTM for temporal learning:

$$h_t = LSTM(X_t, h_{t-1})$$

where h_tis the hidden state at time t.

Step 6: Classification Layer

The final hidden representation is passed through a dense layer with sigmoid activation:

$$\hat{y} = \sigma(Wh_t + b$$

where:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Step 7: Output Prediction

The final class is predicted as:

$$Class = \begin{cases} Violence, & \hat{y} \geq 0.5 \\ Non\text{-}Violence, & \hat{y} < 0.5 \end{cases}$$

Step 8: Loss Calculation

Binary cross-entropy loss is used:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y_i}) + (1 - y_i)\log(1 - \hat{y_i})]$$

Step 9: Performance Evaluation

Accuracy is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3. Hyperparameter Tuning Parameters

Table 1. Hyperparameter Tuning Parameters

Parameter	Description	Suggested Values
Input Frame Size	Size of video frames given to the model	224 × 224
Batch Size	Number of samples processed per training batch	16, 32, 64
Epochs	Number of complete training cycles	25, 50, 100
Optimizer	Algorithm used for weight optimization	Adam, SGD, RMSprop
Learning Rate	Step size for updating model weights	0.001, 0.0001, 0.00001
Dropout Rate	Prevents overfitting by randomly dropping neurons	0.2, 0.3, 0.5
LSTM Units	Number of hidden units in LSTM layer	64, 128, 256
Dense Units	Number of neurons in fully connected layer	64, 128
Activation Function	Function used in hidden layers	ReLU
Output Activation	Function used for binary classification	Sigmoid
Loss Function	Error function used for training	Binary Cross-Entropy
Train-Test Split	Ratio for training and testing data	80:20
Model Architectures	Deep learning models used for comparison	MobileNetV2, ResNet101, LSTM, InceptionV3 + LSTM
Evaluation Metrics	Performance measures for model validation	Accuracy, Loss, Val_Accuracy, Val_Loss

IV. IMPLEMENTATION AND RESULT

1. Hardware and Software Requirements

The proposed violence detection framework was implemented using a high-performance computing environment suitable for deep learning and real-time video analytics applications. The system was developed using Python programming language due to its extensive support for machine learning, deep learning, and computer vision libraries. The implementation utilized TensorFlow and Keras frameworks for designing and training deep learning models such as MobileNetV2, ResNet101, LSTM, and the proposed InceptionV3 + LSTM architecture. OpenCV was used for video frame extraction, preprocessing, segmentation, and image enhancement operations. NumPy and Pandas libraries were used for numerical computations and dataset handling, while Matplotlib was employed for visualization and performance analysis.

The experiments were conducted on a system equipped with Intel Core i7 processor, NVIDIA GPU with CUDA support, 16 GB RAM, and 1 TB storage capacity to efficiently process large video datasets and accelerate model training. The software environment included Windows/Linux operating system, Python 3.10, TensorFlow 2.x, Keras, OpenCV, and Jupyter Notebook for implementation and experimental evaluation. GPU acceleration significantly reduced training time and improved computational performance during deep learning model execution.

2. Dataset Description

Dataset

The dataset used in this research consists of two major folders namely NonViolence and Violence. The dataset was designed to support binary classification for identifying violent and non-violent activities in surveillance videos.

The NonViolence folder contains 1,001 videos representing normal human activities and everyday scenarios without violent behavior. These videos include activities such as eating, sports, singing, walking, conversations, and other regular crowd interactions. These samples help the model learn normal behavioral patterns and distinguish them from abnormal or violent activities.

The Violence folder contains 1,000 videos representing various violent incidents and aggressive human interactions captured in different environments and scenarios. These videos include fighting, physical attacks, crowd violence, assault activities, and other abnormal violent behaviors.

The dataset provides a balanced and clean separation between violent and non-violent classes, enabling effective training and evaluation of deep learning models for violence detection. Each video was converted into multiple frames during preprocessing, and relevant spatial-temporal features were extracted for model learning and classification.

The dataset was divided into training and testing sets using an 80:20 ratio. The training dataset was used for learning model parameters, while the testing dataset was used to evaluate model performance and generalization capability in real-world surveillance conditions.

3. Result Analysis Tables

Table 2 Training and Validation Performance Analysis of Deep Learning Models

Model	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss	Performance Analysis
ResNet50	92.62%	80.65%	0.2541	0.5053	Good feature extraction capability but suffers from overfitting and lower validation performance in complex surveillance videos.
LSTM	96.72%	83.87%	0.2433	0.4301	Captures temporal dependencies effectively but lacks strong spatial feature extraction capability.
MobileNetV2	97.54%	87.10%	0.1647	0.6314	Lightweight and computationally efficient model suitable for real-time applications but exhibits higher validation loss.
Proposed InceptionV3 + LSTM	99.86%	92.48%	0.1397	0.4335	Achieved the best performance due to combined spatial and temporal feature learning with improved robustness and generalization capability.

Table 2 presents the comparative training and validation performance of different Deep Learning models used for violence detection. The proposed InceptionV3 + LSTM model achieved the highest training and validation accuracy with the lowest training loss, demonstrating superior learning capability and better generalization compared with existing models. The hybrid architecture effectively combines spatial and temporal feature extraction, resulting in improved violence classification performance.

Table 3 Comparative Analysis of Model Accuracy and Computational Efficiency

Model	Accuracy	Precision	Recall	F1-Score	Computational Complexity	Real-Time Suitability
ResNet50	High	Moderate	Moderate	Moderate	High	Medium

LSTM	High	High	Moderate	High	Medium	Medium
MobileNetV2	Very High	High	High	High	Low	High
Proposed InceptionV3 + LSTM	Excellent	Excellent	Excellent	Excellent	Optimized Medium	Very High

Table 3 compares the classification accuracy, computational complexity, and real-time suitability of various Deep Learning architectures. The proposed hybrid InceptionV3 + LSTM model demonstrated excellent performance across all evaluation metrics while maintaining optimized computational efficiency. The model is highly suitable for intelligent surveillance systems requiring accurate real-time violence detection.

Table 4 Violence Detection Prediction Analysis

Input Video Type	Predicted Class	Confidence Score	Detection Status	System Response
Sports Activity	Non-Violence	0.859	Correct Detection	Normal Monitoring
Fighting Scene	Violence	0.999	Correct Detection	Alert Generated
Crowd Movement	Non-Violence	0.812	Correct Detection	Monitoring Continued
Physical Assault	Violence	0.993	Correct Detection	Emergency Alert Triggered
Running Activity	Non-Violence	0.884	Correct Detection	No Threat Detected

Table 4 illustrates the prediction analysis of the proposed violence detection system on different surveillance video scenarios. The system successfully classified violent and non-violent activities with high confidence scores. The proposed framework effectively generated alerts for violent situations while minimizing false detections in normal activities.

Table 5 Comparative Validation Performance of Proposed Model

Performance Metric	ResNet50	LSTM	MobileNetV2	Proposed InceptionV3 + LSTM
Validation Accuracy	80.65%	83.87%	87.10%	92.48%
Validation Loss	0.5053	0.4301	0.6314	0.4335
Overfitting Reduction	Moderate	Good	Poor	Excellent
Temporal Learning	Limited	High	Limited	Excellent
Spatial Feature Extraction	Excellent	Limited	Good	Excellent

Table 5 presents the comparative validation performance of different Deep Learning models used in the proposed surveillance system. The proposed InceptionV3 + LSTM hybrid model achieved the highest validation accuracy and demonstrated strong spatial-temporal learning capability. The model effectively reduced overfitting and improved detection robustness in complex surveillance environments.

V. CONCLUSION & FUTURE WORK

This research successfully developed an intelligent Deep Learning-based violence detection framework for smart surveillance systems using a hybrid InceptionV3-LSTM architecture. The proposed model effectively combines the spatial feature extraction capability of InceptionV3 with the temporal sequence learning ability of LSTM networks to accurately classify surveillance videos into Violence and Non-Violence categories. The complete system workflow included dataset collection, preprocessing, segmentation, feature extraction, model training, testing, and real-time prediction generation.

Experimental analysis demonstrated that the proposed hybrid model achieved superior performance compared with existing Deep Learning models such as ResNet50, MobileNetV2, and standalone LSTM architectures. The proposed InceptionV3-LSTM model obtained higher classification accuracy, lower loss values, improved validation performance, and better generalization capability in complex surveillance environments. The hybrid approach effectively captured both spatial and temporal information from video sequences, resulting in more accurate violence detection and reduced false positive rates.

The system also showed strong robustness under varying environmental conditions such as dynamic backgrounds, crowded scenes, and illumination variations. Real-time violence prediction capability and optimized computational performance make the proposed framework suitable for intelligent surveillance systems, smart city monitoring, crowd management, and public safety applications. The research confirms that hybrid Deep Learning architectures can significantly improve automated violence detection performance and support advanced AI-based security solutions.

Future work can focus on improving the proposed violence detection system using advanced Deep Learning models such as Transformers and Attention-based networks for higher accuracy. The framework can be extended for multi-class violence recognition, real-time edge deployment, and integration with IoT-based smart surveillance systems. Future research may also incorporate object detection, explainable AI, and larger real-world datasets to improve prediction reliability, scalability, and practical implementation in smart city security applications.

REFERENCES

1. Singh, D. Patil, and S. N. Omkar, "Eye in the sky: Real-time drone surveillance system (DSS) for violent individuals identification using Scatternet hybrid deep learning network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 1629–1637.
2. T. Simpson, "Real-time drone surveillance system for violent crowd behavior unmanned aircraft system (UAS)–human autonomy teaming (HAT)," in 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), 2021, pp. 1–9.
3. Ilyas, S. Obaid, and N. Z. Bawany, "Deep learning for violence detection in surveillance: The role of transfer learning and pre-trained models," in 2023 24th International Arab Conference on Information Technology (ACIT), 2023, pp. 1–8.
4. Ullah, K. Muhammad, I. Ul Haq, and S. W. Baik, "Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments," Future Generation Computer Systems, vol. 96, pp. 386–397, 2019.
5. S. A. Jebur, K. A. Hussein, H. K. Hoomod, L. Alzubaidi, and J. Santamaría, "Review on deep learning approaches for anomaly event detection in video surveillance," Electronics, vol. 12, no. 1, p. 29, 2022.
6. R. G. Tiwari, H. Maheshwari, A. K. Agarwal, and V. Jain, "Hybrid CNN-LSTM model for automated violence detection and classification in surveillance systems," in 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART), 2023, pp. 169–175.
7. K. Rezaee, S. M. Rezakhani, M. R. Khosravi, and M. K. Moghimi, "A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance," Personal and Ubiquitous Computing, vol. 28, no. 1, pp. 135–151, 2024.
8. J. Weaver, A. Zelenkauskaite, and L. Samson, "The (non) violent world of YouTube: Content trends in web video," Journal of Communication, vol. 62, no. 6, pp. 1065–1083, 2012.

9. Y. J. Choi and C. Kim, "A content analysis of cognitive, emotional, and social development in popular kid's YouTube," International Journal of Behavioral Development, 2024. DOI: 10.1177/01650254241239964.
10. M. Jain and M. Kumar, "A review of violence detection techniques," in 2024 2nd International Conference on Computer, Communication and Control (IC4), 2024, pp. 1–6.
11. M. Khan, A. El Saddik, W. Gueaieb, G. De Masi, and F. Karray, "VD-Net: An edge vision-based surveillance system for violence detection," IEEE Access, vol. 12, pp. 43796–43808, 2024.
12. K. Nagar, U. Singh, "ProSoccerTrainer: A secure AI-powered Penalty Detection and Performance Recommendation System," in 2025 5th Intelligent Cybersecurity Conference (ICSC), 2025, pp. 248–255.
13. P. Shakyawar, U. Singh, "Facial Expression Recognition Using Deep Learning Model on Crowd," in 2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET), 2024, pp. 1–6.
14. Prajapati, N. Patel, U. Singh, "Gender Recognition using Deep Learning Methodology," in 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), 2023, pp. 798–804.
15. Gupta, P. Tripathi, B. P. Singh, U. Singh, "Human Face Mask Recognition using ResNet 152 Model," in 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), 2023, pp. 1503–1510.
16. D. Singh, N. Patel, U. Singh, "Method for Job Recommendation based on Machine Learning and Deep Learning Model," in 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), 2023, pp. 875–883.
17. M. Vishwakarma, P. Tripathi, B. P. Singh, U. Singh, "Segmentation and Detection of Brain Tumors using EfficientNetB3 Model," in 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), 2023, pp. 1463–1470.
18. S. Ray, U. Singh, "Hybrid Task Cascade Region-Based Convolutional Neural Network is Used to Analyse Cricket Scores," in 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), 2023, pp. 19–26.
19. Rajput, P. Gupta, P. Ghodeshwar, S. Varma, K. K. Sharma, U. Singh, "Study of Cloud Providers (Azure, Amazon, and Oracle) According To Service Availability and Price," in 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), 2023, pp. 1177–1188.
20. M. Patel, U. Singh, "Age and Gender Recognition using Deep Learning Technique," in 2023 3rd International Conference on Smart Data Intelligence (ICSMDI), 2023, pp. 238–245.
21. U. Singh, L. S. Songare, "Analysis and Detection of Monkeypox using the GoogLeNet Model," in 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), 2022, pp. 1000–1008.
22. Taiwade, N. Gupta, R. Tiwari, S. Kumar, U. Singh, "Hierarchical K-Means Clustering Method for Friend Recommendation System," in 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 89–95.
23. R. Baghel, P. Pahadiya, U. Singh, "Human Face Mask Identification using Deep Learning with OpenCV Techniques," in 2022 7th International Conference on Communication and Electronics Systems (ICCES), 2022, pp. 1051–1057.
24. M. Ranjan, A. Shukla, K. Soni, S. Varma, M. Kuliha, U. Singh, "Cancer Prediction Using Random Forest and Deep Learning Techniques," in 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 227–231.
25. S. Waskle, L. Parashar, U. Singh, "Intrusion Detection System Using PCA with Random Forest Approach," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 803–808.
26. S. Nihale, S. Sharma, L. Parashar, U. Singh, "Network Traffic Prediction Using Long Short-Term Memory," in 2020 International Conference on Electronics and Sustainable

Communication Systems (ICESC), 2020, pp. 338–343.
27. Saxena, A. Vyas, L. Parashar, U. Singh, "A Glaucoma Detection using Convolutional Neural Network," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 815–820.
28. Bamne, N. Shrivastava, L. Parashar, U. Singh, "Transfer learning-based Object Detection by using Convolutional Neural Networks," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 328–332.