www.ijarp.com ISSN 2456-9992 Page: 01-12

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 04

ANOMALY DETECTION USING DEEP LEARNING

***Mr. Kandekar Onkar Anil, Mr. Dahale Abhishek Jagdish, Mr. Sudip Somnth Pandore, Mr Yogesh Santosh Rajankar, Prof.Vidhate.S.P**

Department of Computer Engineering, Shri Chhatrapati Shivaji Maharaj College of Engineering, Ahilyanagar Savitribai Phule Pune University, Pune, India.

Article Received: 14 March 2026, Article Revised: 03 April 2026, Published on: 23 April 2026

***Corresponding Author: Mr. Kandekar Onkar Anil**
Department of Computer Engineering, Shri Chhatrapati Shivaji Maharaj College of Engineering, Ahilyanagar Savitribai Phule Pune University, Pune, India.

DOI: https://doi-doi.org/101555/ijarp.9234

ABSTRACT

Ensuring safety using modern systems that mitigate the risk of violence in the workplace, in schools, and online is critical in today's society, as the presence of violence is becoming more ubiquitous. The traditional method of monitoring and surveillance, including the manual recording of events, is inefficient and often riddled with monitoring errors. Therefore, this project presents the automation of violence detection using deep learning methodologies to process and evaluate video inputs for real-time violent activity recognition. The convolutional neural networks (CNNs) and Long Short-Term Memory Networks (LSTM) work collaboratively to capture the spatial and temporal features in video frames in order to effectively classify violent acts such as fighting and assaulting. The model architecture for this project is comprised of video input, frame capture, classification of violent and non-violent acts, and feature learning (spatial and temporal). This model augments, with tuneable parameters, generalization around the core violence detection classification. The model architecture evaluation uses publicly available benchmark datasets. Expected precision, accuracy, and recall thresholds guide the formation of the model, as well as the f1-score. The architecture in this project works to complement current surveillance systems and ensure there is a significant decrease in the human effort that is required to monitor live video feeds. The outcome of this work is to create a safer environment.

KEYWORDS: *Violence Detection, Deep Learning, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Computer Vision, Video Surveillance.*

I. INTRODUCTION

The threat that violence poses in public spaces is serious, and our current system of surveillance leaves much to be desired. There are two significant shortcomings of conventional surveillance: they are manned, which risks human fatigue and energy, and they are reactive. Manual scanning of video feeds runs the risk of human fatigue contributing to missed events that require immediate response. Deep learning in computer science provides an alternative, perhaps the most elegant alternative, that addresses both these shortcomings simultaneously. Deep learning systems can be trained to extract events of interest and classify them to provide immediate alerts to the security response team These events can be identified by the trained system as fights, assaults, or other patterns of potentially violent behavior that require immediate attention.

The rising incidents of violence in public spaces poses a threat to the safety of our society and the maintenance of the socio-political order. The currently implemented systems of surveillance rely on a single monitor watching actors/victims carefully, and this person is tasked with continual observation of a scene for an extended period of time. This continuous observation of a scene by a single person means that the monitor runs the risk of becoming increasingly fatigued as time passes, which impacts the rate and quality of their response to critical events. These shortcomings of conventional video surveillance systems mean that there is a need for new systems that can provide this additional and critical feature: the aid of the system in real-time identification of violent actions.

Deep learning systems can provide a solution to this, and as a subset of artificial intelligence, can be of particular interest to our study. Video surveillance systems and their use of convolutional neural networks (CNNs) are well-established. Video systems can be trained to learn a variety of patterns, how to differentiate violent behavior or activity (such as fighting or assault or overt aggressive behavior) in a system. Models can be trained with a variety of datasets, and such datasets can incorporate both the presence of and the absence of violence in a behavioral activity. This training is essential and provides the systems with the ability to learn and detect motion, and the importance of separating ordinary motion from a more complex behavior resulting in violence.

II. LITERATURE SURVEY

According to [1] explores how deep learning models can enhance anomaly detection across IoT environments. The authors emphasize the growing complexity of IoT ecosystems and the

limitations of traditional detection techniques. Their proposed architecture improves the accuracy and responsiveness of identifying abnormal behaviors. The paper also presents experimental results showing strong performance across multiple IoT datasets. Overall, the research highlights deep learning as a crucial tool for robust IoT security.

According to [2] introduces an intelligent anomaly detection mechanism for industrial equipment using a feature auto encoder combined with deep learning. The approach focuses on extracting meaningful representations from sensor data to detect early signs of mechanical faults. The system demonstrates effective recognition of subtle deviations in machine operations. Experimental evaluations reveal high detection rates and improved reliability compared to conventional methods. The study underscores the value of deep learning in predictive maintenance.

According to [3] the authors review modern machine learning and deep learning methods used to detect anomalies in IoT networks. They examine emerging research directions, challenges, and performance trends in this domain. The survey compares different detection models and highlights their strengths in addressing IoT-specific threats. Additionally, the paper identifies current gaps and offers insights for future technological advancements. It serves as a comprehensive guide for researchers developing secure IoT systems.

According to [4] article surveys recent progress in using machine learning and deep learning to detect anomalies within cloud networks. The authors outline the growing security demands in cloud computing and the shortcomings of traditional systems. They evaluate a variety of detection techniques, analyzing their accuracy, scalability, and adaptability. The survey also discusses implementation issues and research opportunities. Ultimately, the work presents a detailed overview of state-of-the-art cloud anomaly detection solutions.

According to [5] proposes a video anomaly detection framework that integrates convolutional and recurrent neural networks. The system captures both spatial and temporal patterns to identify suspicious events in video streams. The authors demonstrate that their hybrid architecture outperforms several baseline models. Extensive experiments confirm its robustness across diverse video datasets. The work highlights the potential of deep learning for automated video surveillance.

According to [6] the authors introduce SUSAN, a deep learning framework designed to support anomaly detection in sustainable industrial operations. The system leverages advanced neural models to monitor industrial processes and identify deviations that could affect efficiency or safety. Its performance is validated using real-world industrial datasets, showing strong detection capability. The framework emphasizes energy efficiency and

environmental sustainability. This research demonstrates how AI can contribute to smarter, greener industrial systems.

According to [7] presents a two-stage intrusion detection approach that combines Naïve Bayes for data classification and the elliptic envelope method for identifying anomalies. The hybrid design aims to reduce false alarms while improving the accuracy of intrusion detection. Evaluation results show that the system effectively distinguishes between normal and suspicious activities. The authors argue that their method provides a lightweight and efficient solution for modern security needs. The study contributes to enhancing the reliability of anomaly detection frameworks.

According to [8] the authors deliver a broad survey of deep learning techniques applied to time-series anomaly detection. They review various neural architectures, including autoencoders, RNNs, and transformers, discussing their strengths and practical limitations. The study highlights applications across domains such as IoT, finance, and industrial monitoring. It also outlines major research challenges, including interpretability and scalability. This survey acts as a detailed reference for researchers working with time-series data.

According to [9] research investigates the use of machine learning algorithms to detect anomalies within cloud-based infrastructures. The author analyzes several ML models and compares their performance under different cloud threat scenarios. Results indicate that certain algorithms can efficiently identify unusual network behaviors. The paper also emphasizes the importance of real-time detection capabilities for cloud security. It contributes valuable insights into building more resilient cloud systems.

According to [10] comparative study evaluates multiple anomaly detection techniques used to protect IoT systems from emerging cyber threats. The study uses adaptive machine learning methods to assess their effectiveness under dynamic IoT conditions. Findings demonstrate that adaptive approaches outperform static models in identifying diverse attack patterns. The paper highlights key factors influencing detection accuracy and operational efficiency. It provides a useful benchmark for improving IoT security strategies. performances, reliability, and most importantly, real-time applicability, the metrics of standard acceptance among the artificial intelligence world has been adopted: accuracy, precision, recall, and F1-score.

Table 1: Comparison table for the literature review.

Author/Year	Technique / Model Used	Description / Approach	Result / Accuracy
Yaseen, Asad (2023) [11]	Machine Learning	Network anomaly detection for cybersecurity	Summarized ML techniques; improved detection accuracy
			reported
SHEKERBEK, AINUR, M. Svoboda (2024)[12]	Adaptive ML Systems	Enhancing cybersecurity through adaptive anomaly detection	Improved threat detection in cyber-physical systems
Moriano, Pablo, et al. (2025) [13]	Adaptive Anomaly Detection	Identifying attacks in cyber-physical systems; systematic review	Highlights state-of-the-art adaptive detection methods
Zhang, Hanqing, et al. (2025) [14]	Deep Learning	Real-time data quality assessment and anomaly detection in large-scale data streams	High accuracy and real-time performance
Jeffrey, Nicholas, Qing Tan, José R. Villar (2023) [15]	Review	Survey of anomaly detection strategies for cyber-physical systems	Comparative analysis and identification of effective strategies

III. METHODOLOGY

The proposed system for detecting violence engages with a fully defined Deep Learning pipeline for the purpose of detecting violence in video streams. The first step is gather input video streams from both benchmark data sets and resourceful input from the real world. The first stage of preprocessing involves converting the video stream into individual frames, followed by normalization and resizing, and filtering to remove noise and careful redundancies. Data augmentation steps, in the form of flipping, rotation and scaling, are to be used to improve model generalization. After data augmentation comes the extraction of spatial features from the individual frames. This is done by means of Convolutional Neural Networks (CNN), known for their use of visual patterns (for example, detecting human posture, intensity of motion, and data patterns of interaction with each other). The resultant features from this operation are transferred to a Recurrent Neural Network (RNN), and more specifically towards a Long Short-Term Memory (LSTM) network, in order to learn the temporal dependencies and motion having to do with a sequence of frames. The main idea of combining CNN and LSTM is to model both spatial and temporal features of violent actions. The model employs a supervised learning mechanism and is trained with violence and non-violence labelled data video clips. During this training, it is common practice to use some form of optimization, whether with the Adam optimizer and/or cross-entropy loss. The model

is finally able to classify previously unseen video data and determine whether the associated activities are violent, or if they do not exhibit violence. In order to gauge the system's

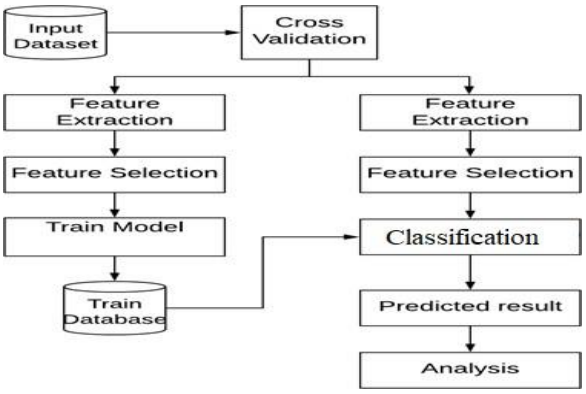

Figure 1.1 Research Methodologies.

A complete machine learning/deep learning process has been developed for violence detection, as shown in the diagram.

Input Dataset

The collection of video data, both violent and non-violent, initiates the process.

This input dataset builds the foundation for both training and testing the model.

Cross Validation

The dataset is divided into multiple segments (folds).

It helps improve the model's accuracy, and prevent overfitting as the model is tested on different subsets.

Training Phase Feature Extraction

From the video frames, relevant details are captured.

In your project, this is accomplished by using CNN to seize spatial features (objects, movements).

Feature Selection

The most pertinent features are only selected.

This improves model performance, and reduces complexity.

Train Model

The features selected are then used to train the deep learning model (CNN + LSTM).

The model learns the patterns of violent and non-violent behaviors.

Train Database

The model once again learns and retains parameters.

This serves as a knowledge base for predictions in the future.

Testing Phase Feature Extraction

The same feature extraction process is applied to a new (unseen) video input.

Feature Selection

In training, relevant features are selected, and the same applies here.

Classification

The trained model classifies the input into the following categories:

- Violent
- Non-violent

Algorithm: Violence Detection System Input: Video Dataset V (violent, non-violent)

Output: Predicted Label (Violent / Non-Violent)

Step 1: Load Dataset Read video dataset V

Assign labels (0 = Non-Violent, 1 = Violent) **Step 2:** Preprocessing

For each video v in V:

Extract frames F = {f1, f2, f3, ... fn}

Resize each frame to fixed size (e.g., 224x224) Normalize pixel values (0 to 1)

Apply data augmentation (flip, rotation, zoom)

Step 3: Feature Extraction (CNN) For each frame fi:

Pass fi through CNN model Extract spatial feature vector Xi

End For

Create sequence X = {X1, X2, X3, ... Xn} **Step 4:** Temporal Modeling (LSTM)

Input sequence X into LSTM network Learn temporal dependencies across frames Generate sequence representation H

Step 5: Classification

Pass H through Fully Connected Layer Apply Softmax activation

Predict output label Y (0 or 1) **Step 6:** Training

Initialize model parameters For each epoch:

Forward pass through CNN + LSTM Compute loss (Cross Entropy)

Back propagate error

Update weights (Adam optimizer) End For

Step 7: Evaluation

Test model on unseen data Compute metrics:

Accuracy Precision Recall F1-score

Step 8: Prediction

Input new video stream Repeat Steps 2–5

If Y == 1:

Trigger Alert ("Violence Detected") Else:

Continue Monitoring End Algorithm

APPLICATION

The suggested violence detection system has a multitude of applications in areas where ongoing observation and increased layers of security are paramount. Using automated, deep learning, and violent activity detection, human intervention is minimized, and response velocities are improved. The solution is designed for public and institutional digital safety and surveillance, and it is adaptable to a myriad of contemporary security concerns. The solution's ability to evaluate the spatiotemporal elements of violent activities and impacts enhances the system's predictive capacity and public safety.

- CCTV monitoring is integrated into public transport (buses, trains).
- Systems for monitoring, detecting, and filtering violent content on social media.
- CCTV integration into security systems for smart cities.
- Providing assistance to law enforcement for crime prevention and detection.
- Safety management, monitoring, and control in workplaces and industry.
- Monitoring and control of violence and bullying in schools.
- Safety monitoring in malls, airports, and railway stations.
- Emergency and incident management, including real-time alerts.

IV. LIMITATION

The violence detection system has many limitations that can hinder its effectiveness. The model's ability to predict accurately is largely determined by the quality or lack of quality omnipresent in the training dataset. In the training dataset, the model predicts in real world/diverse scenarios that it has never encountered. Detection accuracy can also be highly impacted by the environment. An example of the environment is lighting or how crowded the space is where the model is deployed. The system relies heavily on real-time performance. This requires a lot of computing power and can prohibit the deployment of the system on cost-effective devices. The system can be deployed on devices that are high cost but are not realistic. The true root of the problem is the lack of explainability and contextual

understanding, which cause the greatest emergency of trust in the system.

The following elements can help stave off limitations in performance as they relate to datasets as the primary focus.

- Obstacles for processing complex environments such as crowds, obstructions, and low-light situations
- High processing power required for computation in real time
- It is difficult to remove false negative and false positive results
- Limited contextual and intentional understanding
- Extensive and large datasets that are labeled required for training
- A general lack of ability to generalize to new and varied environments Surveillance technologies generate privacy issues.

V. RESULTS

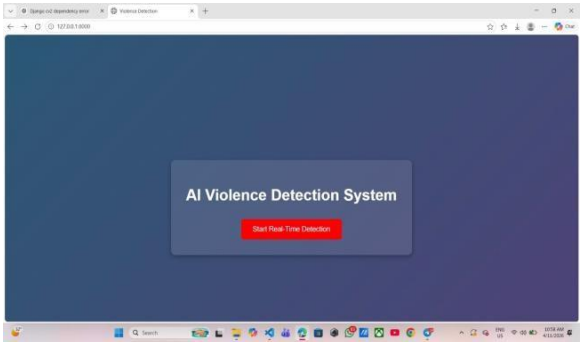

Figure 2: Home Page.

Figure 2 Home Page illustrates the primary design of the interface for the violence detection system. It represents the first point of contact for users engaging with the application. Standard features on this page include video upload functionality, live camera surveillance, and access to various system components. It has been designed with the user in mind, guaranteeing that users who do not have technical backgrounds can easily navigate and utilize the system. The interface may include features that communicate the system status, the operational instructions, and controls for the detection process. The Home Page is designed to facilitate the operational interaction of the user and the underlying deep learning model.

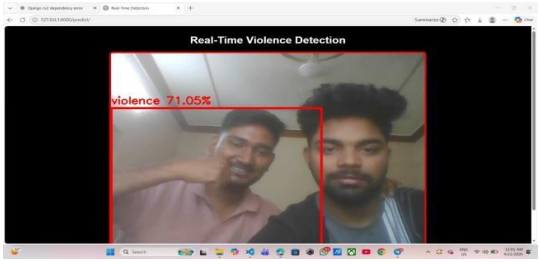

Figure 3: Real Time Violence Detection.

The figure 3 demonstrates the real-time violence detection module of the system. Here, video input is taken via a stream or a video recording. The system analyzes the input frame-by-frame using a trained CNN-LSTM model to capture the relevant spatial and temporal characteristics. When violence is detected, the system marks the event (e.g., with bounding boxes or alerts) and issues a notification. The output is updated continuously, incorporating real-time monitoring and fast action. The module is vital for real-world deployment, as it assures prompt detection and decreases the time taken to respond to violent acts, thus, increasing public safety.

VI. Conclusion

The purpose of this project is to develop a significant deep learning system capable of detecting violence in footage from surveillance cameras. Through the use of integrated Convolutional Neural Networks (CNN) for the extraction of spatial features, and Long Short-Term Memory (LSTM) networks for the analysis of features in the video as a function of time, the model being described is fully equipped to obtain the relevant visual and motion features present in video footage showing violent events. This system lessens the need for manual supervision of videos and offers a flexible answer to the problem of real-time video surveillance. The results of the model's scaling and the use of data augmentation and transfer learning to further improve the system's robustness and capability of generalization is demonstrated in the evaluation of the model where violent and non-violent activities were successfully differentiated. sThe system allows for the early detection of violent events and therefore the quicker mobilization of appropriate responses, ultimately increasing the level of safety of the general public. It can be used in electronic surveillance, pan-tilt-zoom (PTZ) cameras, and online monitoring at public locations. For the future, this model needs to be adaptable to rapidly changing detection conditions, use in combination with audio analysis, and require less time to be formatted for real-time usage.

References

1. Abusitta, Adel, et al. ”Deep learning-enabled anomaly detection for IoT systems.” Internet of Things 21 (2023): 100656.
2. Ahmed, Imran, et al. ”A smart-anomaly-detection system for industrial machines based on feature autoencoder and deep learning.” Micromachines 14.1 (2023): 154.
3. Rafique, Saida Hafsa, et al. ”Machine learning and deep learning techniques for internet of things network anomaly detection—current research trends.” Sensors 24.6 (2024): 1968.
4. Abdallah, Amira Mahamat, et al. ”Cloud network anomaly detection using machine and deep learning techniques—recent research advancements.” IEEE Access 12 (2024): 56749-56773.
5. Qasim, Maryam, and Elena Verdu. ”Video anomaly detection system using deep convolutional and recurrent models.” Results in Engineering 18 (2023): 101026.
6. G´omez, ´Angel Luis Perales, et al. ”SUSAN: A Deep Learning based anomaly detection framework for sustainable industry.” Sustainable Computing: Informatics and Systems 37 (2023): 100842.
7. Vishwakarma, Monika, and Nishtha Kesswani. ”A new two-phase intrusion detection system with Na¨ıve Bayes machine learning for data classification and elliptic envelop method for anomaly detection.” Decision Analytics Journal 7 (2023): 100233.
8. Zamanzadeh Darban, Zahra, et al. ”Deep learning for time series anomaly detection: A survey.” ACM Computing Surveys 57.1 (2024): 1-42.
9. Gudelli, Venkata Ramana. ”Anomaly detection in cloud networks using machine learning algorithms.” African Journal of Artificial Intelligence and Sustainable Development 4.1 (2024).
10. Alsalman, Dheyaaldin. ”A comparative study of anomaly detection techniques for IoT security using adaptive machine learning for IoT threats.” IEEE Access 12 (2024): 14719-14730.
11. Yaseen, Asad. ”The role of machine learning in network anomaly detection for cybersecurity.” Sage Science Review of Applied Machine Learning 6.8 (2023): 16-34.
12. SHEKERBEK, AINUR, and M. Svoboda. ”ENHANCING CYBERSECURITY WITH ADAPTIVE ANOMALY DETECTION SYSTEMS THROUGH MACHINE LEARNING.” 2 (2024): 177-189.
13. Moriano, Pablo, et al. ”Adaptive anomaly detection for identifying attacks in cyber-physical systems: A systematic literature review.” Artificial Intelligence Review 58.9 (2025): 283.

14. Zhang, Hanqing, et al. ”Deep Learning-Based Real-Time Data Quality Assessment and Anomaly Detection for Large-Scale Distributed Data Streams.” International Journal of Medical and All Body Health Research 6.1 (2025): 01-11.
15. Jeffrey, Nicholas, Qing Tan, and Jos´e R. Villar. ”A review of anomaly detection strategies to detect threats to cyber-physical systems.” Electronics 12.15 (2023): 3283.