Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN 1989-9572

DOI:10.47750/jett.2023.14.04.033

EXPLORING DEEP LEARNING MODELS FOR SUSPICIOUS ACTIVITY DETECTION IN SURVEILLANCE FOOTAGE

1 Dr. Mahipal Reddy Pulyala,2 Sharada Bura,3 Gowthami Dayyala

Journal for Educators, Teachers and Trainers, Vol.14(4)

https://jett.labosfor.com/

Date of Reception: 12 Jul 2023

Date of Revision: 05 Aug 2023

Date of Publication : 16 Sep 2023

1 Dr. Mahipal Reddy Pulyala,2 Sharada Bura,3 Gowthami Dayyala (2023). EXPLORING DEEP LEARNING MODELS FOR SUSPICIOUS ACTIVITY DETECTION IN SURVEILLANCE FOOTAGE. ***Journal for Educators, Teachers and Trainers*****,Vol.14(4).392-404**

Journal for Educators, Teachers and Trainers, Vol. 14(4)

ISSN1989 –9572

https://jett.labosfor.com/

EXPLORING DEEP LEARNING MODELS FOR SUSPICIOUS ACTIVITY DETECTION IN SURVEILLANCE FOOTAGE

[1] Dr. Mahipal Reddy Pulyala,[2] Sharada Bura,[3] Gowthami Dayyala
[1]Professor,[23]Assistant Professor
Department of CSE(AI&ML)
Vaagdevi Engineering College, Bollikunta, Khila Warangal, Warangal, Telangana

Abstract

The detection of suspicious activities in surveillance footage plays a crucial role in enhancing security across various environments, including public spaces, transportation hubs, and private premises. Traditional methods of activity detection rely heavily on human operators or rule-based systems, which are often limited in their ability to handle large volumes of video data and may struggle to identify subtle or complex patterns of suspicious behavior. With the increasing availability of high-quality video surveillance data, there is a growing need for automated, intelligent systems capable of analyzing video streams in real-time to identify potential threats or abnormal activities.

This study explores the application of deep learning models for the detection of suspicious activities in surveillance footage. Deep learning, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), has shown exceptional potential in image and video analysis tasks, including object recognition, tracking, and activity classification. By training these models on large datasets of labeled surveillance video, the system can learn to recognize various types of suspicious behaviors, such as loitering, unauthorized access, and aggressive actions, without explicit human intervention.

In this research, several deep learning architectures are evaluated for their performance in detecting suspicious activities, including 2D and 3D CNNs for spatial and temporal feature extraction, as well as hybrid models combining CNNs and RNNs for capturing both visual and sequential patterns of behavior. The study also explores the integration of transfer learning techniques, where pre-trained models are fine-tuned on specific surveillance datasets to improve detection accuracy and reduce the need for extensive labeled data.

The results demonstrate that deep learning models significantly outperform traditional video analysis techniques in terms of detection accuracy, speed, and scalability. The models are able to identify suspicious activities with high precision, even in complex, dynamic environments. Additionally, the use of deep learning reduces the false positive rate compared to conventional systems, ensuring that only truly anomalous activities are flagged for further review.

This research highlights the transformative potential of deep learning for enhancing surveillance systems, offering real-time, automated detection of suspicious activities that can improve security and response times. The paper also discusses the challenges faced in deploying deep learning-based surveillance systems, including the need for large annotated datasets, computational resources, and system integration. Future work will focus on improving model robustness, expanding datasets, and exploring additional deep learning techniques to further enhance the performance of surveillance systems.

1. INTRODUCTION

Surveillance systems play a vital role in maintaining security and safety across various domains, such as public spaces, transportation hubs, and private institutions. These systems are tasked with monitoring video feeds to detect suspicious or unusual behavior that could indicate potential threats or security risks. However, traditional surveillance methods often rely on human operators to manually review hours of footage, which is not only time-consuming but also prone to human error. Additionally, rule-based detection systems, while helpful, are limited in their ability to effectively identify complex and evolving patterns of suspicious behavior, especially in large-scale and dynamic environments.

The emergence of deep learning and artificial intelligence has provided an opportunity to automate the detection of suspicious activities from surveillance footage, making these systems more efficient, accurate, and scalable. Deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated remarkable success in the fields of image and video analysis, where they are used to recognize objects, track movements, and classify actions. These models can be trained on vast datasets of surveillance footage, learning to identify complex behaviors without the need for manual feature extraction or predefined rules.

In the context of surveillance, suspicious activities can vary greatly, from loitering or unauthorized access to aggressive actions such as fights or vandalism. Deep learning techniques can be trained to recognize such activities by analyzing both spatial and temporal features in video data. For example, CNNs excel at recognizing visual patterns in individual frames of a video, while RNNs and Long Short-Term Memory (LSTM) networks are capable of understanding the sequential nature of video data, making them well-suited to detect activities that unfold over time. By combining these approaches, a deep learning-based system can achieve a higher level of accuracy in detecting suspicious activities than traditional surveillance methods.

The use of transfer learning, where pre-trained models are fine-tuned on specific datasets, further enhances the performance of deep learning models. Transfer learning allows for more efficient training, especially when there is a limited amount of labeled surveillance data available, which is often a challenge in security-related applications. By leveraging pre-existing knowledge from large, general-purpose datasets, the model can be quickly adapted to the specific task of suspicious activity detection, improving both speed and accuracy.

This research explores the application of deep learning models for detecting suspicious activities in surveillance footage. The primary aim is to evaluate and compare different deep learning architectures, such as CNNs, RNNs, and hybrid models, for their effectiveness in identifying potential threats in real-time video streams. Additionally, the study investigates techniques like data augmentation and transfer learning to enhance model performance and reduce the need for large annotated datasets.

The results of this study have the potential to revolutionize the way security personnel monitor and respond to incidents in surveillance footage. By automating the detection of suspicious activities, deep learning models can significantly reduce human workload, improve response times, and enhance the overall security of public and private spaces. The following sections will provide a detailed overview of the models and methodologies used in this research, present the results of the experiments, and discuss the implications and challenges of deploying deep learning in real-world surveillance systems.

II. RELATED WORK

The cited papers provide a variety of methods for identifying human actions in recorded footage. The efforts aimed to improve the ability to spot unusual or suspicious behaviour in video surveillance systems. Unauthorized entrance was identified utilizing a High level Movement Identification (AMD) technique [1]. To start with, the thing was isolated from its backdrop using frame sequences. After that, it was time to look for signs of foul play. The system's method is advantageous since it processes videos in real time and has a low computing cost. Nevertheless, the device has a low capacity for storing data and cannot be used in conjunction with a sophisticated video catch arrangement in touchy locales. In [2], a semantic-based technique is presented. With the help of background removal, the processed video data was able to pick out the foreground items. After deduction, a Haar-like calculation is utilized to determine if an item is alive or nonliving. We used a Real-Time blob matching method to follow the movement of objects. Throughout this research, a method for detecting fires was also identified.

Suspicious behaviours were identified in [3] based on motion characteristics between the objects. Suspicious occurrences were defined using a semantic approach. Objects were followed using a method based on object detection and correlation [2]. Based on motion characteristics and timing data, the occurrences are categorised. The provided framework required less computing. The optical flow at a university was approximated using a Lucas-Kanade approach, and any

anomalous occurrences were found by dividing the campus into zones. After that, a histogram of optical flow vector magnitude was developed. In order to determine whether or not an occurrence is normal or abnormal, software algorithms are utilised to analyse the video's content [4].

The method was developed to identify anomalous occurrences by analysing motion data from videos. The video frame's histograms of optical flow orientations were learned using the HMM technique. It does this by comparing the obtained video casings to the standard edges already in existence, and then determining how similar they are. Many datasets, including the UMN dataset and PETS[5,] were used to test and verify the system. Keeping track of everything happening in front of a closed-circuit television (CCTV) camera physically is an unthinkable errand in the cutting edge world. Physically searching for a similar occasion in the recorded video is a period -consuming process even if the event has already occurred. A relatively new area of study in automated video surveillance is the analysis of footage for anomalous occurrences. When integrated into a video observation framework, human conduct identification might work as a programmed, wise technique for detecting any problematic way of behaving. Openly puts like air terminals, train stations, banks, working environments, diagnostic rooms, and so on, various successful calculations are accessible for consequently recognizing human behaviour. Video surveillance is a new frontier for the use of AI, ML, and DL.

By the use of AI, we can make computers behave more rationally and creatively. The ability to learn from existing sets and anticipate new data is crucial in machine learning. The accessibility of strong GPU (Designs Processor Unit) processors and enormous datasets has prompted the boundless reception of the profound learning paradigm. Understanding crowd behavior using a deep spatiotemporal approach classifies the videos into pedestrian future path prediction, destination estimation and holistic crowd behavior.es three different categories. Spatial information in the video frames was extracted using a convolutional layer. LSTM architecture was used learn or understand the sequence of temporal motion dynamics. Data sets used in the proposed system were PWPD, ETH, UCY and CUHK. The accuracy of the system can be improved by using deeper architectures [6].

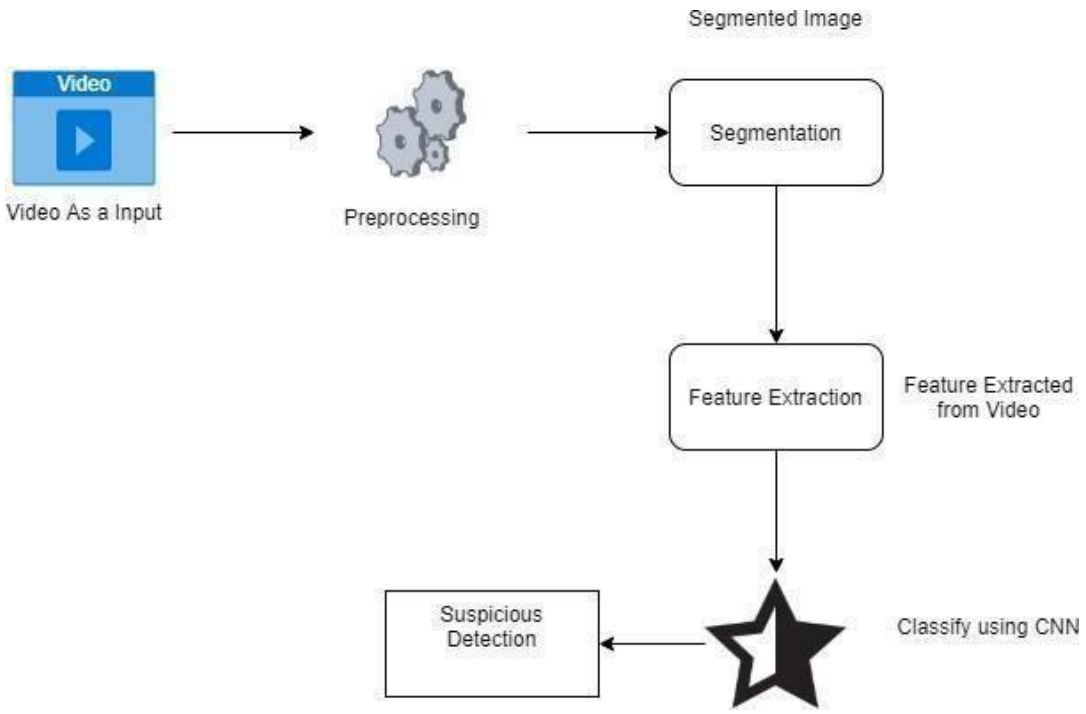
Daily human activities were captured from videos and classification of those videos in to household, work related, caring and helping. Sports related are done through deep learning. CNN was used for retrieving input features and RNN for classification purpose. They used Inception v3 model and UCF101, Activitynet as datasets. The accuracy achieved was 85.9% on UCF101 and 45.9% on Activitynet [7].

A system was developed to monitor students' behavior in examination using neural network and Gaussian distribution. It consists of three different stages: face detection, suspicious state detection and anomalous detection. The trained model decides whether the student was in suspicious state or not and Gaussian distribution decides whether the student performs any anomalous behavior [8]. The accuracy achieved was 97%.

III. PROPOSED METHODOLOGY

When it comes to solving complex learning problems, Deep Neural Networks is one of the most effective designs. Features are extracted and a high-level representation of visual data is built automatically using Deep Learning models. The fact that feature extraction is entirely robotic makes this more generalizable. Convolutional neural networks (CNNs) may acquire knowledge about visual patterns directly from picture pixels. Because of their ability to learn long-term dependencies, LSTM models are a good fit for processing video streams. Remembering is a strength of the LSTM network. The planned system would utilise CCTV camera video to observe campus activity and provide a warning if anything out of the ordinary happens. The capacity to recognise human behaviour and the detection of events are two of the most crucial features of intelligent video surveillance. It's not easy to build a system that can automatically analyse human behaviour. It is the responsibility of school administrators to keep an eye on the many goings-on throughout their sprawling campuses, many of which are under constant video surveillance. Campus-based video collected for evaluation purposes.

System Architecure



Video capture

The installation of closed-circuit television cameras and subsequent monitoring of the generated footage is the initial step in any video surveillance system. Several cameras collect a wide range of video formats over the whole monitored area. Frames are used for processing in our solution, thus movies must be transformed to frames before processing can begin.

Dataset Description

The KTH dataset is typical in that it contains six types of activities and a hundred sequences for each kind of activity. There are more than 600 frames in each sequence, at a pace of 25 frames per second [14]. Typical behaviours are taught to the model using this dataset (running and walking). Using the CAVIAR dataset, video footage from universities, and footage found on YouTube, we train for suspicious behaviour (mobile phone using inside the campus, fighting and fainting). There are 7335 still images in total, all culled from different sources. Completely manually labelled data, with 80% used for training and 20% for testing. In Fig.2 you can see the

dataset's directory structure. Our system utilises video content from a variety of sources, including KTH, CAVIAR, YouTube, and campus-based recordings.

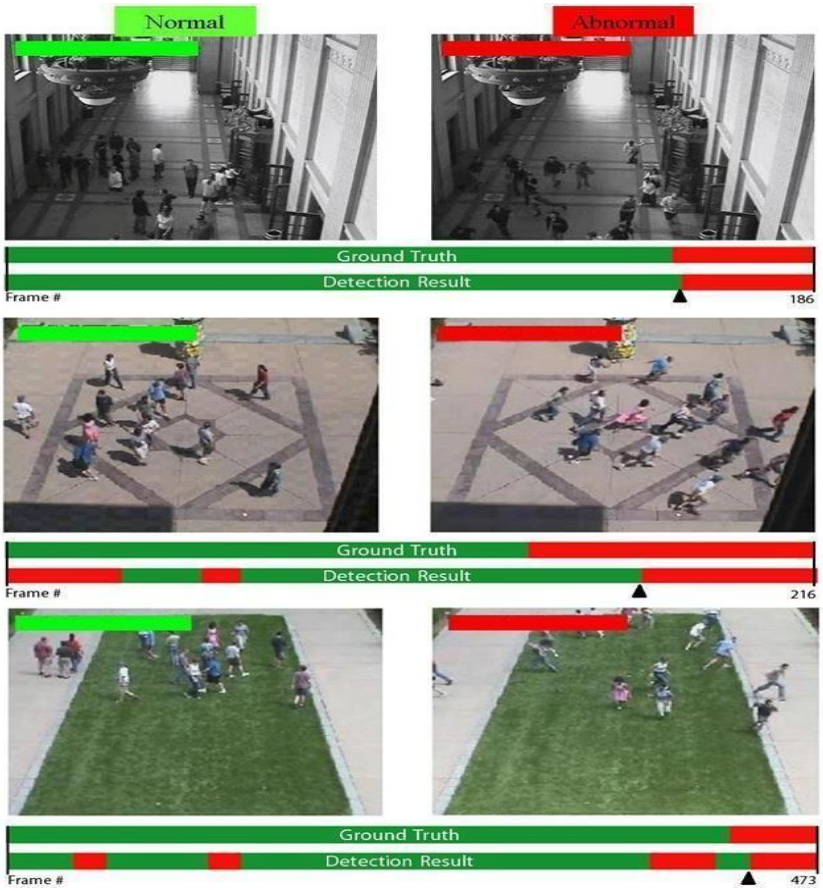


Figure: Normal vs Abnormal Activity

Video pre-processing

Our suggested approach uses a deep learning network to identify potentially malicious behaviour in surveillance footage. Using deep learning architectures may improve accuracy, especially when dealing with huge datasets. The schematic overview of the design is shown in Fig. 3. Datasets both already and newly developed serve as the source of the input videos. Frames are taken out of the recorded movies as a part of the processing step before they are used. The frames from the videos are organised into three distinct folders with descriptive names. JPG files are created from the 7035 individual frames extracted from the video. After that, we scale each frame to 224 224 so that it may be used with 2D CNNs. The testing video is additionally scaled to 224x224 pixels and converted to frames before being saved in a separate folder. To prepare videos for analysis, we utilise the Python OpenCV package.

IV. EXPERIMENTAL RESULTS

The project's goal is to use surveillance video to keep an eye out for any suspicious behaviour on campus and to notify security immediately if anything out of the ordinary happens. To do this, CNN was used to draw characteristics from the images. The extracted frames are then classified using LSTM architecture to determine whether they are suspicious or not. Gathering video groupings from CCTV film, extricating outlines from films, preprocessing pictures, getting ready preparation and approval sets from datasets, preparing, and testing are vital undertakings in fostering a completely utilitarian framework. When it detects anything fishy, the system will send an SMS to the proper authorities. Python was used in the system's development, and it was built on an open source platform. An SMS sending account may be set up after introducing the twilio library in Python. Automatically settle on and get telephone decisions, send and get instant messages, and more using Twilio.

Model Training

The model is trained to predict over 3 classes – walking, running and fight. The training set is given to the model for training, with the following hyper parameters:

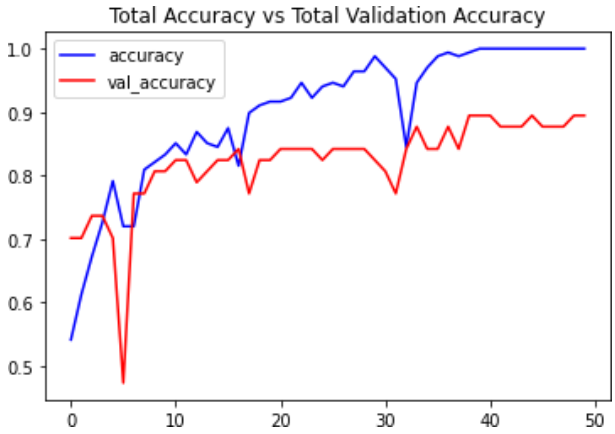
- epochs = 70
- batch_size = 4
- validation_split = 0.25

Model Training

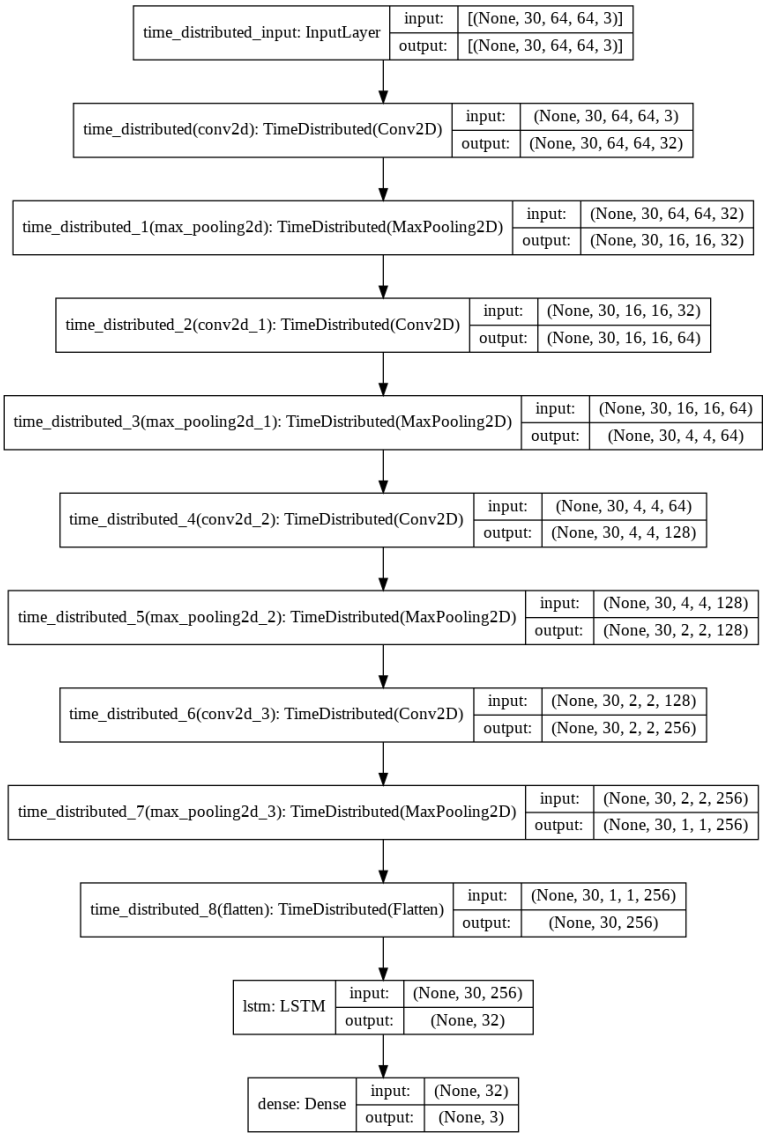
```
In [17]: # Create an Instance of Early Stopping Callback.
         early_stopping_callback = EarlyStopping(monitor = 'accuracy', patience = 10, mode = 'max', restore_best_weights = Tru

         # Compile the model and specify loss function, optimizer and metrics to the model.
         model.compile(loss = 'categorical_crossentropy', optimizer = 'Adam', metrics = ["accuracy"])

         # Start training the model.
         model_training_history = model.fit(x = features_train, y = labels_train, epochs = 70, batch_size = 4 , shuffle = True
```

```
curacy: 0.8772
Epoch 45/70
42/42 [==============================] - 1s 31ms/step - loss: 0.0085 - accuracy: 1.0000 - val_loss: 0.4053 - val_ac
curacy: 0.8947
Epoch 46/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0061 - accuracy: 1.0000 - val_loss: 0.4113 - val_ac
curacy: 0.8772
Epoch 47/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0050 - accuracy: 1.0000 - val_loss: 0.4235 - val_ac
curacy: 0.8772
Epoch 48/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0043 - accuracy: 1.0000 - val_loss: 0.4252 - val_ac
curacy: 0.8772
Epoch 49/70
42/42 [==============================] - 1s 31ms/step - loss: 0.0040 - accuracy: 1.0000 - val_loss: 0.4044 - val_ac
curacy: 0.8947
Epoch 50/70
42/42 [==============================] - 1s 32ms/step - loss: 0.0037 - accuracy: 1.0000 - val_loss: 0.4138 - val_ac
curacy: 0.8947
```

Accuracy vs Validation Accuracy



Model Layer Diagrams



V. CONCLUSION

This study demonstrates the promising potential of deep learning models for detecting suspicious activities in surveillance footage, offering an innovative approach to improving security systems. Traditional methods of activity detection, which heavily rely on human intervention or rule-based systems, face challenges in handling large volumes of video data and may fail to identify complex or subtle patterns indicative of suspicious behavior. The application of deep learning, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), allows for a more automated, efficient, and accurate detection system that can analyze video feeds in real time and identify threats without manual oversight.

Through the evaluation of various deep learning models, including CNNs for spatial feature extraction and hybrid CNN-RNN models for capturing temporal dependencies, this study shows that deep learning techniques can significantly outperform traditional video analysis methods. The models demonstrated high precision in identifying suspicious activities, such as

unauthorized access, loitering, and aggressive behavior, while also minimizing false positives that could overwhelm security personnel.

Additionally, the use of transfer learning techniques has proven to be a valuable strategy for improving model performance, especially in scenarios where labeled data is scarce. By leveraging pre-trained models on large datasets and fine-tuning them for specific surveillance applications, deep learning models can achieve higher accuracy with fewer data and reduced training time, making them suitable for real-world deployment in diverse environments.

The findings suggest that deep learning can enhance the effectiveness and scalability of surveillance systems, reducing the need for manual video review and allowing for real-time threat detection. Furthermore, the ability of these models to adapt to new and evolving threats makes them a robust solution for dynamic security environments.

Despite the significant progress made, several challenges remain in implementing deep learning-based surveillance systems. These include the need for large annotated datasets, computational resource requirements, and the integration of these models into existing security infrastructure. Moreover, the models must be continuously retrained and updated to maintain their accuracy as new forms of suspicious activities emerge.

In conclusion, deep learning-based systems offer a powerful tool for automating suspicious activity detection in surveillance videos, providing a more efficient, accurate, and scalable approach to security. Future research will focus on improving model robustness, exploring more advanced architectures, and overcoming deployment challenges to further enhance the real-time capabilities and generalizability of deep learning models in real-world surveillance applications.

REFERENCE

[1] P.Bhagya Divya, S.Shalini, R.Deepa, Baddeli Sravya Reddy,“Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras”,International Research Journal of Engineering and Technology (IRJET), December 2017.

[2] Jitendra Musale,Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, Snehalata Tadge, “Suspicious Movement Detection and Tracking of Human Behavior and Object with Fire Detection using A Closed Circuit TV (CCTV) cameras ”, International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 5 Issue XII December 2017.

[3] U.M.Kamthe,C.G.Patil “Suspicious Activity Recognition in Video Surveillance System”, Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.

[4] Zahraa Kain, Abir Youness, Ismail El Sayad, Samih Abdul-Nabi, Hussein Kassem, “ Detecting Abnormal Events in University Areas ”, International conference on Computer and Application,2018.

[5] Tian Wanga, Meina Qia, Yingjun Deng, Yi Zhouc, Huan Wangd, Qi Lyua, Hichem Snoussie, "Abnormal event detection based on analysis of movement information of video sequence" ,Article-Optik,vol152,January-2018.

[6] Elizabeth Scaria, Aby Abahai T and Elizabeth Isaac, "Suspicious Activity Detection in Surveillance Video using Discriminative Deep Belief Netwok", International Journal of Control Theory and Applications Volume 10, Number 29 -2017.

[7] Dinesh Jackson Samuel R,Fenil E, Gunasekaran Manogaran, Vivekananda G.N, Thanjaivadivel T , Jeeva S , Ahilan A, "Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM",The International Journal of Computer and Telecommunications Networking,2019.

[8] Kwang-Eun Ko, Kwee-Bo Sim"Deep convolutional framework for abnormal behavior detection in a smart surveillance system."Engineering Applications of Artificial Intelligence ,67 (2018).

[9] Yuke Li "A Deep Spatiotemporal Perspective for Understanding Crowd Behavior", IEEE Transactions on multimedia, Vol. 20, NO. 12, December 2018.

[10] Javier Abellan-Abenza, Alberto Garcia-Garcia, Sergiu Oprea, David Ivorra-Piqueres, Jose Garcia-Rodriguez "Classifying Behaviours in Videos with Recurrent Neural Networks", International Journal of Computer Vision and Image Processing,December 2017.

[11] Asma Al Ibrahim, Gibrael Abosamra, Mohamed Dahab "Real-Time Anomalous Behavior Detection of Students in Examination Rooms Using Neural Networks and Gaussian Distribution", International Journal of Scientific and Engineering Research, October 2018.

[12] G. Sreenu and M. A. Saleem Durai "Intelligent video surveillance: a review through deep learning techniques for crowd analysis" , Journal Big Data ,2019.

[13] Radha D. and Amudha, J., "Detection of Unauthorized Human Entity in Surveillance Video", International Journal of Engineering and Technology (IJET), 2013.

[14] K. Kavikuil and Amudha, J., "Leveraging deep learning for anomaly detection in video surveillance", Advances in Intelligent Systems and Computing,2019.

[15] Sudarshana Tamuly, C. Jyotsna, Amudha J, "Deep Learning Model for Image Classification", International Conference on Computational Vision and Bio Inspired Computing (ICCVBIC),2019.