e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

SUSPICIOUS HUMAN ACTIVITY DETECTION

Ankit Singh*1, Adhyatm Mishra*2, Purva Patil*3, Sahil More*4, Prof. Anita Mhatre*5

*1,2,3,4Student, Department of Information Technology, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India.

*5Assistant Professor, Department of Information Technology, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India.

ABSTRACT

Rise in criminal activities in suburban and metropolitan areas gives a threat to social stability and people's safety. It becomes inevitable to put an end to these practices. In previous times it was a burdensome to monitor these activities. Now with help of Machine Learning and AI, identifying and categorizing suspicious human behavior becomes easier with the help of a CNNLSTM hybrid LRCN model. Conversion of image sequences into labels, probabilities, descriptions as well as learning of visual attributes from video frames etc becomes much easier with the help of this technique. Suspicious activities like fighting, robbery and firing are identified by the system. The proposed system uses Keras in Python, and Tensorflow based on the LRCN model. It gathers video frames and compares them with the trained models for monitoring the problems with an accuracy of approximate 91.5%. Additional features of this system includes the use of AI to raise alarm, send email as an alert to the concerned authorities and create report of detected suspicious activities in a simplified PDF format, all prompting to take necessary actions.

Keywords: Convolutional Neural Network (CNN), Keras, Long Short Term Memory(LSTM), Long Term Recurrent Neural Network(LRCN), Neural Networks, Tensorflow, Alert Generation, Suspicious Activity.

I. INTRODUCTION

In today’s world of rapid technological advancements, finding innovative solutions to real-world problems is crucial and it is very essential to implement inventive solutions for practical issues. For example, an issue that is suspicious activities, which includes violence that takes place in different locations and is harmful. Video monitoring system which monitors the actions of humans are already built. These systems rely solely on manual human observation to detect and identify questionable activity. Such technologies are unable to identify suspicious human activity from surveillance films due to their limited capabilities. In this research, we have built an automatic method that uses video inputs to identify suspicious human activity in the surrounding areas without constant human supervision. This system produces reports on suspected human activity found in an appropriate PDF format, sends out emails and alarms, and generates alerts. We have designed this technique to identify or detect suspicious activity taking place in the designated surrounding areas. Often innocent people suffer as a result of certain dubious human actions that jeopardize numerous lives and valuables. It is a global concern to find effective means of preventing such occurrences. However, human manual monitoring is difficult and demands continuous focus. Automated systems for identifying suspicious activity have become crucial in order to address this. The suggested method classifies suspicious human behavior from CCTV footage by the application of the LRCN model with Keras and TensorFlow. The primary objective of the system is to differentiate between suspicious and typical activity in the data. Manually identifying suspicious events is a laborious and time-consuming task for humans. Hence, it is imperative to monitor the activities that require investigation, such as early warning signals of upcoming attacks. If there is sufficient evidence to take action by filtering, connecting, and categorizing the events. LRCN technique is used by us to transform variable length input to an output that captures complex temporal dynamics to attain visualization. This technique is very effective for large datasets of videos. Inheriting this technique, suspicious activities can be reduced to a certain extent, benefitting the society. Suspicious Human Activity Detection System (SHADS) makes use of machine learning algorithms to identify and react to suspicious human activities in real time. There are sophisticated problems in the security of infrastructure and public areas. Traditional security systems often find themselves at difficulty to stay up to date with potential threats and react accordingly. To solve these threats, SHADS provides an effective solution to improve security in various fields. Development and application of SHADS are thoroughly done in this project. It developed a solid solution by combining techniques of computer vision and

ML. Actions of individuals are identified by the system, and if found suspicious, then the system sends real-time alerts to concerned management.

II. LITERATURE SURVEY

The author Benedict Vinusha V, and with other authors [1] created a smart system that makes use of deep learning and AI, making use of the LRCN system. A combination of CNN and RNN is used in this system. It identifies suspicious activity in various distinct forms, such as assaults, stealing, bullying, and dismissal. This system fails to identify suspected human activity in a real-time scenario.

The system proposed by writer Om M. Rajpurkar with other writers [2] enables CCTV cameras to detect suspicious activity without being compelled to human observers. Additionally, training the model on a particular kind of suspicious activity, it may be subjected to modification to fit other scenarios. One potential area for development could be to identify the behavior of ill-intentional people.

The authors Amudha J., C. Jyotsna, and Amrutha C.V. [3] put out a novel concept with the benefit of stopping criminals in their tracks. It operates by watching and evaluating live CCTV footage continually. The relevant authorities are alerted so they can take measures to prevent any negative outcomes if the analysis indicates that something negative might occur. Despite being restricted to academic environments at the moment, this technology may be used to anticipate suspicious behavior in public or private contexts. Furthermore, by training the model on particular kinds of suspicious activity, it may be modified to fit other scenarios. One potential avenue for development could be to use behavior to identify questionable individuals.

The combination of LSTM technique for quick and exact tracking of known objects with CNN for extracting spatial information is presented by Waqas Iqrar, and others[4] in the HAR system. With the combination of CNN-LSTM method, real-time implementation is made possible. The complexity of model is reduced and the accuracy is improved. The system when implemented on a Raspberry Pi, detects the suspicious activities in real-time and hence optimizes tracking performance.

The author Wenchao Xu with others [5] have created a Convolutional Neural Network (CNN) that uses information gathered from the three-axis accelerometer present in cellphones to identify human activity. It is acknowledged that going upstairs and downstairs, walking, running, sitting, and standing are examples of daily activities. In this case, the CNN was trained using the raw, three-dimensional (3D) accelerometer data without the need for any elaborate preprocessing. When trained and tested using six different types of features retrieved from the raw accelerometer data, the CNN-based method for distinguishing numerous human activities achieved a high accuracy that outperformed the Support Vector Machine (SVM) approach. Consequently, the suggested approach accomplished minimal computational demands while achieving high accuracy in activity identification.

The authors Peshala Liyanage; Pumudu Fernando [6] seek to overcome the shortcomings of video surveillance systems owing to small and imprecise datasets by designing, implementing, and assessing a human crowd suspicious behavior detection approach. Uncertain video samples from a small dataset were used to construct the prototype. With such sparse and unclear datasets, a unique technique that combines MobileNet and LSTM was developed to train video surveillance systems. With the help of this method, four cutting-edge models were tested with different hyperparameters and setups, for a total of 32 models tested.

A class of models known as LRCN models—which are deep in both space and time—was introduced by authors Jeff Donahue, and others[7]. These models provide flexibility for a variety of vision tasks involving sequential data. Their results consistently show that current approaches, which usually concentrate only on learning complicated parameters in the visual domain, can be improved by comprehending sequential dynamics using a deep sequence model. Additionally, their strategy overcomes the output sequence dynamics and rely on a fixed visual representation of the input. Deep sequence model like LRCN is becoming more crucial as as computer vision has have made a mark in involving static input and predictions in order to solve sequential structural challenges. Due to absence of input preprocessing and constructed features, they are optimal solution for displaying concerns for temporal input and sequential output due to their simplicity of integration into current visual recognition pipelines.

In their approach, authors Tejashri Subhash Bora and Monika Dhananjay Rokade [8] use Convolutional neural networks (CNNs) to identify abnormal behavior. For understanding the information in the video is important to categorize abnormal activity with accuracy. CNNs are the best algorithm for this application because they are frequently used for extracting important features from every frame of the video. For classification of behaviors, CNNs are used to identify and extract the required characteristics from video frames for the provided input.

The writers Alavudeen Basha A, and others[9] have explained a method to automatically identify anomalous movements in closed-circuit films. In this method, the movies are first converted into frames, and obtained frames are used to recognize persons by using a background subtraction technique. For feature extraction, a convolutional neural network (CNN) is used. These extracted features are given as input to a Discriminative Deep Belief Network (DDBN). Moreover, the DDBN receives labelled movies showing questionable activity and uses them to extract features. Using the DDBN, the CNN's extracted features are contrasted with those from labelled sample films of classed suspicious activity. This comparison makes it possible to identify a number of questionable activities in the provided video.

The writers, S. A convolutional neural network-based application for identifying suspicious activity is presented by A. Quadri and Komal S. Katakdhond [10]. Instead of utilizing traditional training datasets, they effectively train the suggested model using CCTV footage and use spatiotemporal analysis. The technique they suggest addresses the growing demand for such security systems in light of rising crime rates by using a machine-based method to detect actual criminal behavior in surveillance footage. Their system's output is the identification of anomalies, which intelligent automation replaces with manual operations.

III. PROPOSED METHODOLOGY

We identify anomalous behavior in our proposed system by using a Long-term Recurrent Convolutional Network (LRCN). Accurately identifying such events in the video requires the recognition of patterns across time. Convolutional Neural Networks (CNNs) are frequently employed to extract significant information from individual video frames; however, the successful extraction of these features is contingent upon their effectiveness. To complete this task, 30 frame sequences are taken out of the surveillance videos and can be given as input into the LRCN model. Our method will utilize camera footage as an input for watching nearby surrounding activity and identify the suspicious activity after that. In the process of finding suspicious activity, the system utilizes artificial intelligence (AI) approach to sound an alarm, send alerts to the relevant authorities by email, and create a report on the suspected activity in a simple PDF format. We have developed an approach to identify anomalous behavior in our proposed system by using a Long-term Recurrent Convolutional Network (LRCN). Accurately detecting anomalous behavior in the video requires the recognition of patterns across time. Convolutional Neural Networks (CNNs) are used to take out significant information from individual video frames. To successfully perform extraction of these features with their effectiveness, we take 30 frame sequences out of the movie and feed them into the LRCN model. Because convolutions and pooling procedures enable CNNs to record spatial hierarchies of features, they are very useful for processing spatial data, such as pictures. Sequential data is a good fit for recurrent neural networks (RNNs), such as long short-term memory (LSTM) networks, because they can store information throughout time and identify temporal connections in the data. CNNs and RNNs are combined to create LRCNs, or Long-term Recurrent Convolutional Networks. In order to simulate temporal dependencies over time, it first uses CNNs to bring spatial features from pictures or video frames. These features are then fed into RNNs, which are usually built with LSTM or a similar architecture. Because of this, LRCN is particularly helpful for jobs involving the analysis of video data, where comprehension of the content depends on both spatial and temporal information.

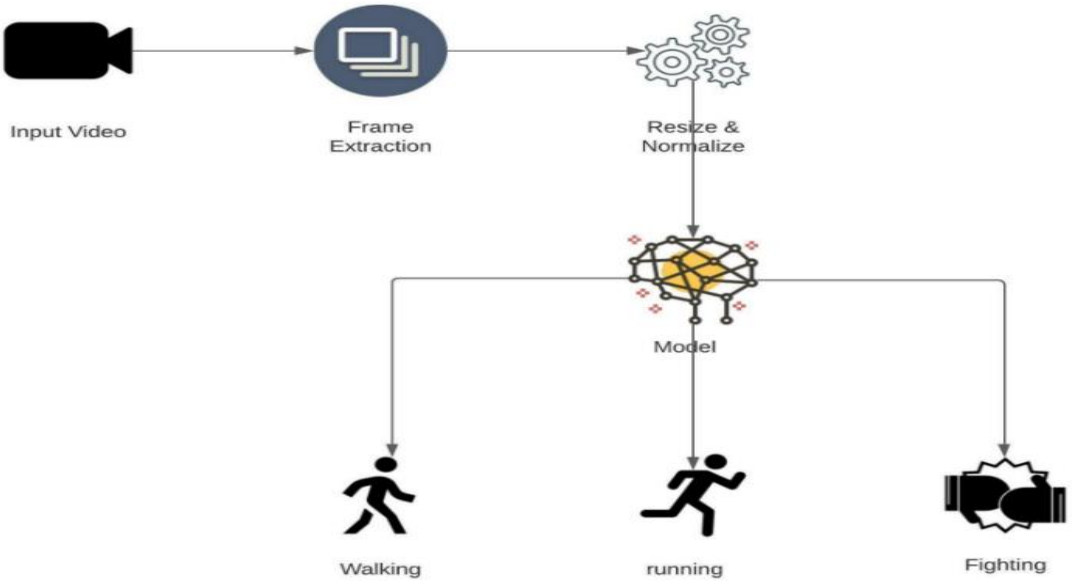


Figure 1: Approach

IV. PROPOSED SOLUTION

1. Downloading and Visualizing of the Data with its Labels

In the initial stage, download and visually explore the dataset along with its labels to understand its contents better. UCF50 Action Recognition Dataset is used. The videos in this dataset are sourced from YouTube. The videos of this dataset are real-life recordings, these are not acting performed by actors. Hence this dataset proves out to be unique.

2. Preprocessing the dataset

The dataset is pre-processed by reading the video files and making them uniform. To help in reducing the computational load, the frames are resized to a standard height and width. Furthermore, the pixel values are adjusted between a range of 0 to 1 by dividing them by 255. The training process is accelerated by this step to ensure faster convergence of the neural network.

3. Splitting of the dataset into train and test datasets

Now divide the data into training and testing sets. Before splitting, arrange the dataset randomly to prevent any potential partiality and ensure that the splits are accurate to represent the overall distribution of the data. This step helps to create balanced training and testing datasets, which are crucial for training and evaluating the model effectively.

4. Implementing the LRCN approach

Integration of CNN and LSTM layers is done in this step which collectively is called as LRCN approach. Spatial video frames are extracted by the Convolution layers and then are passed to the LSTM layers for temporal sequence modelling for each time step. Hence using this approach spatial-temporal features are unified framework and the model becomes more vigorous.

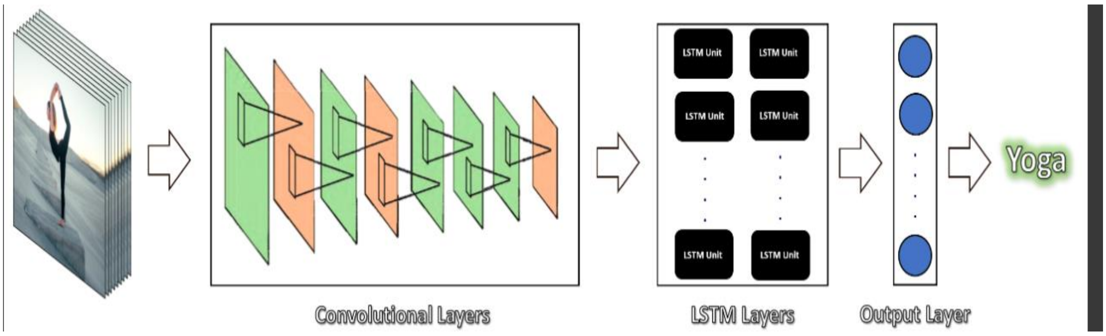
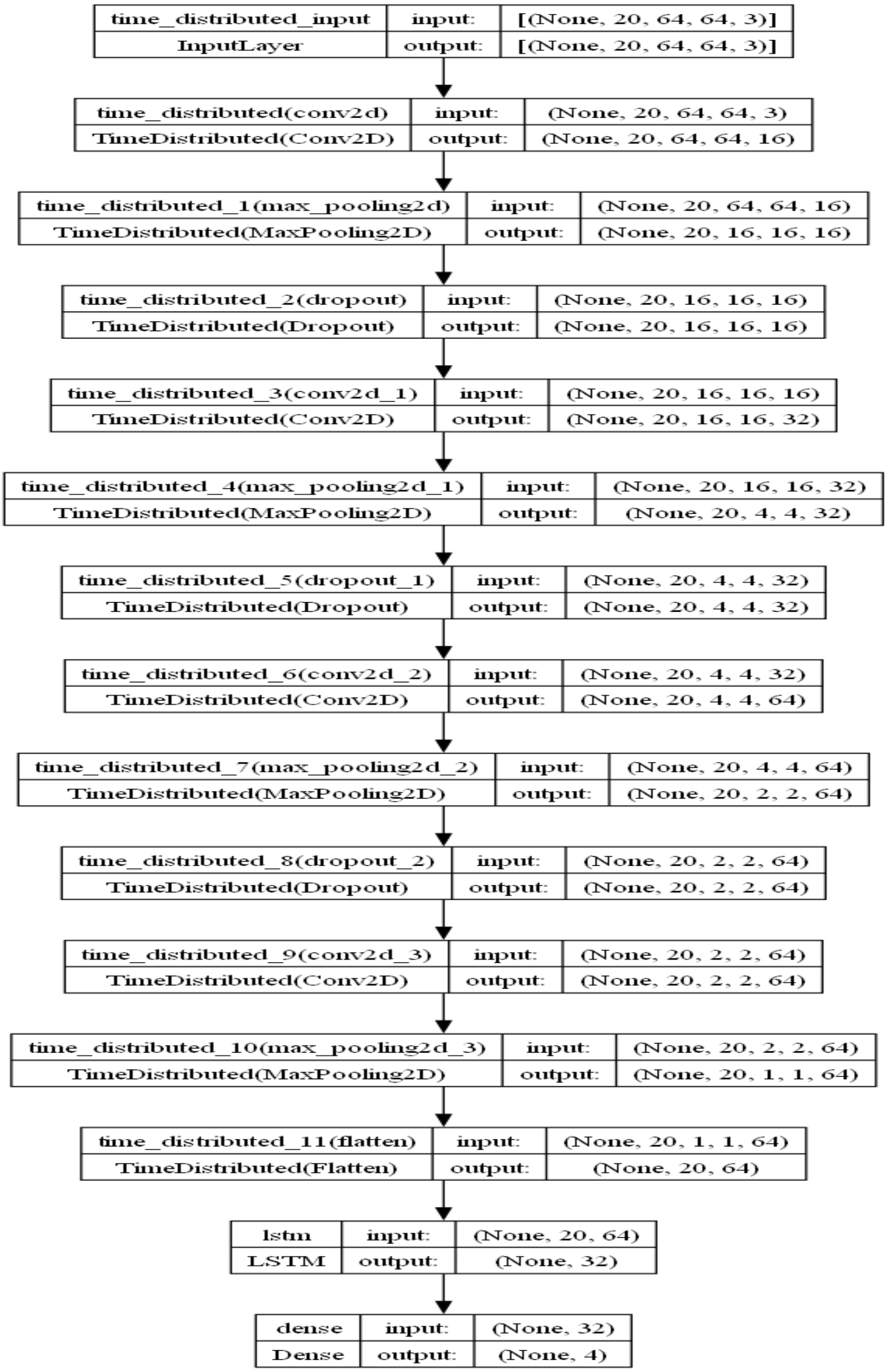


Figure 2: LCRN Implementation(layers)

4.1. Implementing the LRCN approach

Time-distributed Conv2D layers, MaxPooling2D and Dropout layers are utilized to construct a LRCN architecture. Conv2D layers will take features, these features will be flattened by Flatten layer and lastly the layers are fed into an LSTM layer. The LSTM layer will process the temporal information, and its output will be used by a Dense layer with soft-max activation to predict the action being performed.



time_distributed_input	input:	[(None, 20, 64, 64, 3)]
InputLayer	output:	[(None, 20, 64, 64, 3)]
time_distributed(conv2d)	input:	(None, 20, 64, 64, 3)
TimeDistributed(Conv2D)	output:	(None, 20, 64, 64, 16)
time_distributed_1(max_pooling2d)	input:	(None, 20, 64, 64, 16)
TimeDistributed(MaxPooling2D)	output:	(None, 20, 16, 16, 16)
time_distributed_2(dropout)	input:	(None, 20, 16, 16, 16)
TimeDistributed(Dropout)	output:	(None, 20, 16, 16, 16)
time_distributed_3(conv2d_1)	input:	(None, 20, 16, 16, 16)
TimeDistributed(Conv2D)	output:	(None, 20, 16, 16, 32)
time_distributed_4(max_pooling2d_1)	input:	(None, 20, 16, 16, 32)
TimeDistributed(MaxPooling2D)	output:	(None, 20, 4, 4, 32)
time_distributed_5(dropout_1)	input:	(None, 20, 4, 4, 32)
TimeDistributed(Dropout)	output:	(None, 20, 4, 4, 32)
time_distributed_6(conv2d_2)	input:	(None, 20, 4, 4, 32)
TimeDistributed(Conv2D)	output:	(None, 20, 4, 4, 64)
time_distributed_7(max_pooling2d_2)	input:	(None, 20, 4, 4, 64)
TimeDistributed(MaxPooling2D)	output:	(None, 20, 2, 2, 64)
time_distributed_8(dropout_2)	input:	(None, 20, 2, 2, 64)
TimeDistributed(Dropout)	output:	(None, 20, 2, 2, 64)
time_distributed_9(conv2d_3)	input:	(None, 20, 2, 2, 64)
TimeDistributed(Conv2D)	output:	(None, 20, 2, 2, 64)
time_distributed_10(max_pooling2d_3)	input:	(None, 20, 2, 2, 64)
TimeDistributed(MaxPooling2D)	output:	(None, 20, 1, 1, 64)
time_distributed_11(flatten)	input:	(None, 20, 1, 1, 64)
TimeDistributed(Flatten)	output:	(None, 20, 64)
lstm	input:	(None, 20, 64)
LSTM	output:	(None, 32)
dense	input:	(None, 32)
Dense	output:	(None, 4)

Figure 3: LCRN model structure plot

4.2. Compiling and training the model

After checking the structure, compile and start training the model.

4.3. Plotting Model's Loss & Accuracy Curves

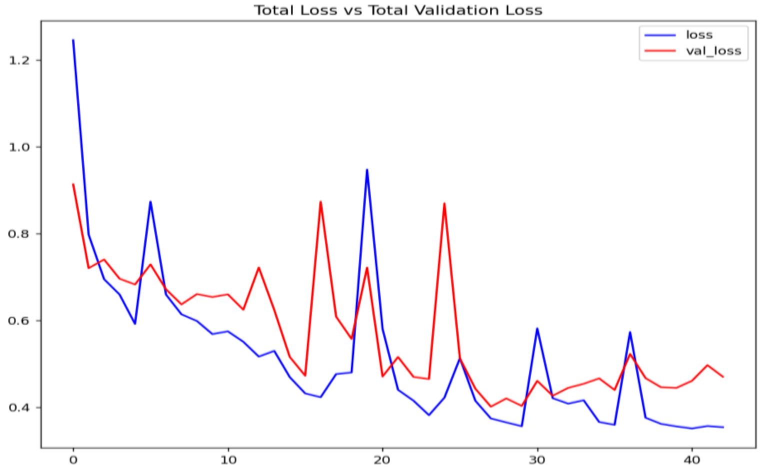


Figure 4: Total Accuracy vs Total Validation Accuracy

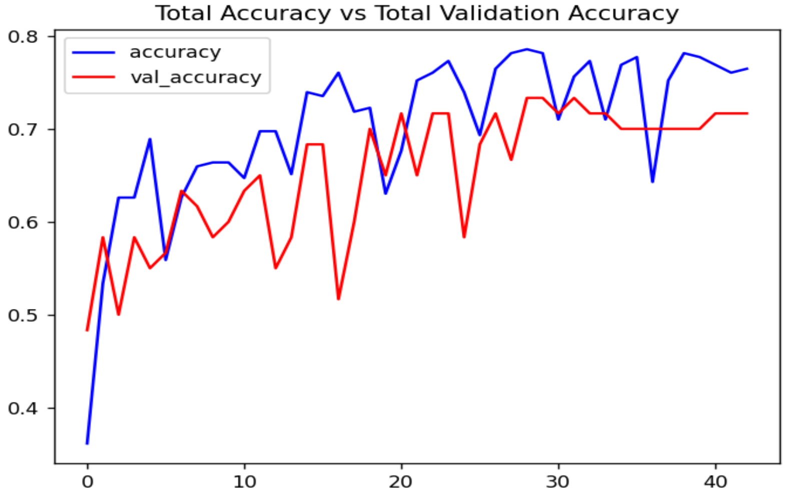
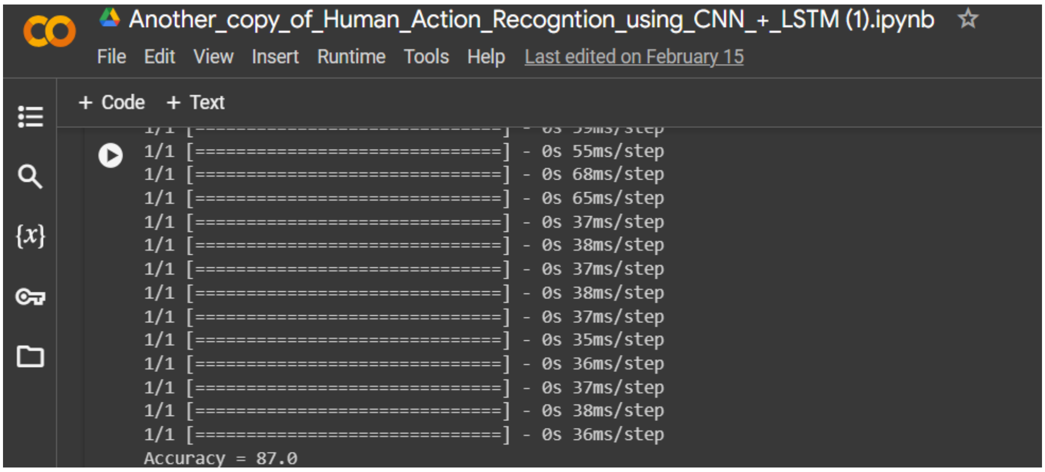


Figure 5: Total accuracy vs Total Validation accuracy

Testing the model on videos

Predicting a single action on single video at first and then predicted multiple actions in a single video.

5.1 Calculated the test accuracy on test dataset which came out to be 87% and predicted single actions


Another_copy_of_Human_Action_Recogntion_using_CNN_+_LSTM (1).ipynb
File
Edit
View
Insert
Runtime
Tools
Help
Last edited on February 15
+ Code
+ Text

```
1/1 [==============================] - 0s 59ms/step
1/1 [==============================] - 0s 55ms/step
1/1 [==============================] - 0s 68ms/step
1/1 [==============================] - 0s 65ms/step
1/1 [==============================] - 0s 37ms/step
1/1 [==============================] - 0s 38ms/step
1/1 [==============================] - 0s 37ms/step
1/1 [==============================] - 0s 38ms/step
1/1 [==============================] - 0s 37ms/step
1/1 [==============================] - 0s 35ms/step
1/1 [==============================] - 0s 36ms/step
1/1 [==============================] - 0s 37ms/step
1/1 [==============================] - 0s 38ms/step
1/1 [==============================] - 0s 36ms/step
Accuracy = 87.0
```

Figure 6: Accuracy of LCRN Model

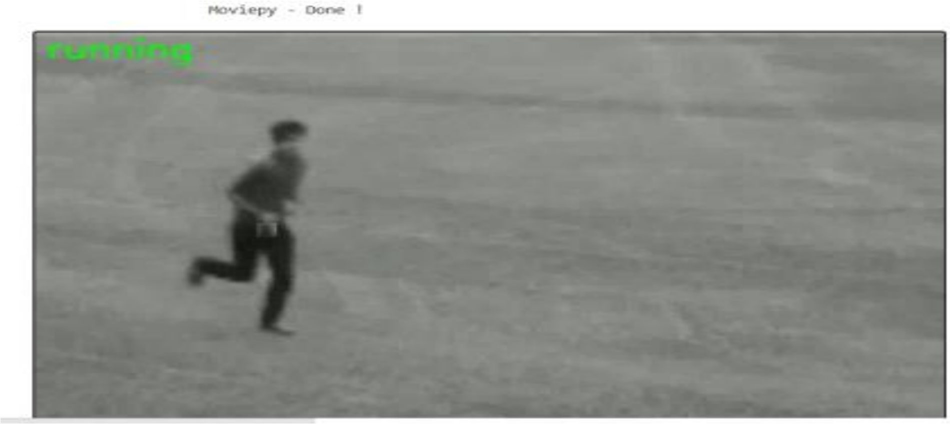


Figure 7: Single Action Prediction - Running

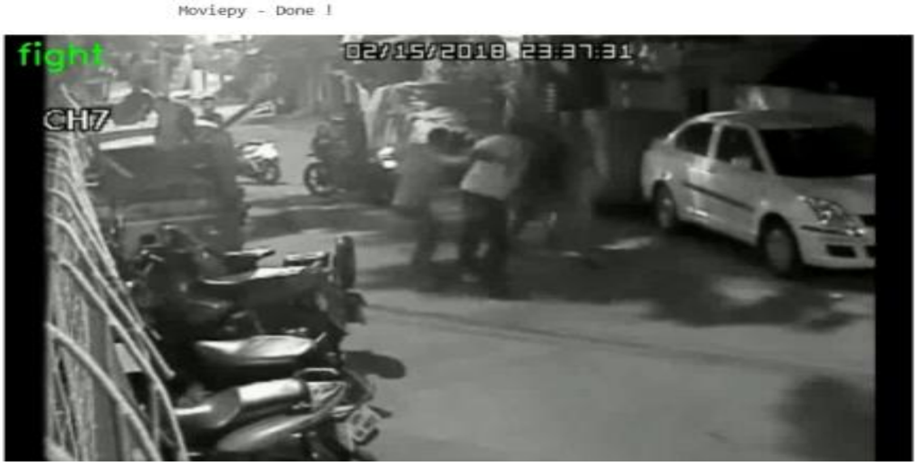


Figure 8: Single Action Prediction - Fighting

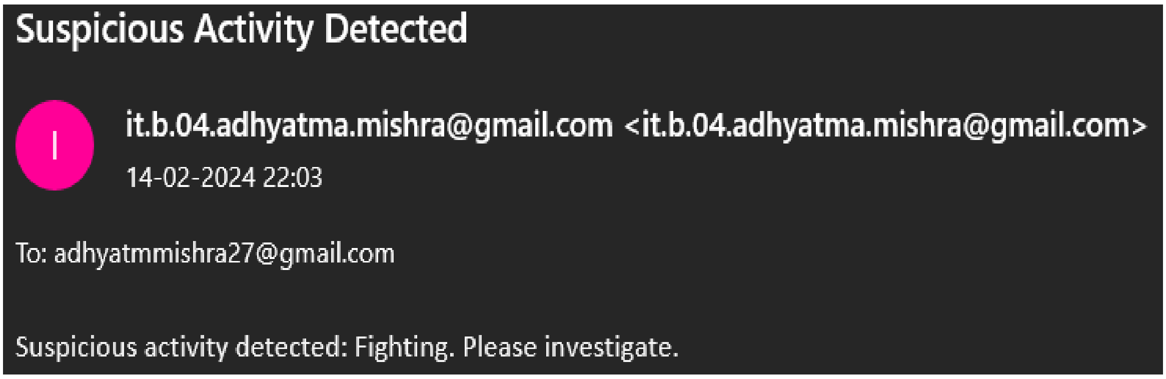
5.2 Multiple actions have been predicted in a single video and saved as a separate video file

6. Alarm Creation

If any suspicious activity is detected, then an alarm will be ranged in which an AI voice reads out a warning to the system administrator to take necessary actions.

7. Alert Generation

On detection of a suspicious activity, an alert will be generated in the form of a message which will be sent via email to the concerned authorities prompting to take necessary action.



Suspicious Activity Detected

I it.b.04.adhyatma.mishra@gmail.com <it.b.04.adhyatma.mishra@gmail.com>
14-02-2024 22:03

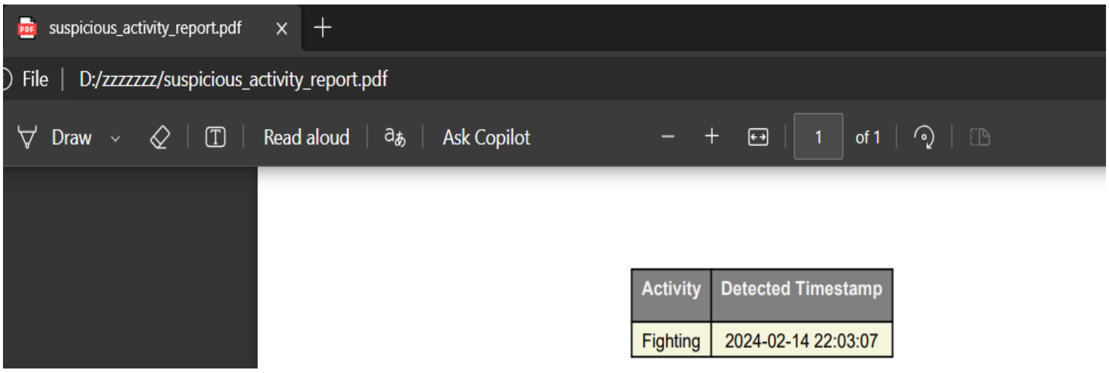
To: adhyatmmishra27@gmail.com

Suspicious activity detected: Fighting. Please investigate.

Figure 9: Alert Generation

8. Report Generation

A report of all the suspicious activities if detected is stored in a simplified pdf format along with the timestamp of detection.


suspicious_activity_report.pdf
File
D:/zzzzzzz/suspicious_activity_report.pdf
Draw
Read aloud
Ask Copilot
1
of 1

Activity	Detected Timestamp
Fighting	2024-02-14 22:03:07

Figure 10: Report Generation

V. RESULTS AND ANALYSIS

We previously used convLSTM (CNN) model that had the following shortcomings:

1. Took a long to train, even on high-speed end devices.
2. Took a long time to predict the activities.
3. The accuracy of the model came out to be 77%.
4. Not suitable for real-life scenarios.
5. Captures only spatial dependencies/features in sequential data.

Upgradations we did to our model: Created a LRCN model.

1. Took less time to train.
2. Took less time to predict the activities.
3. The accuracy of the model came out to be 87%.
4. Totally suitable for real-life scenarios.
5. Captures both spatial and temporal features

Table 1: Comparison of CNN model and LRCN model

Model	Time for training.	Time for prediction.	Accuracy	Near Real-time	Feature capturing
Conv LSTM (CNN)	Long Time	Long Time	77%	NO	Spatial Features
LRCN	Short Time	Short Time	87%	YES	Both Spatial and Temporal Features

VI. CONCLUSION

Utilizing the LCRN concept, we have created a system that can identify questionable human behavior and send out email alerts in response. Moreover, this system uses deep learning and artificial intelligence principles based on the LRCN model to provide a report of suspicious activity observed in an abridged PDF format by merging CNNs and LSTMs. Several types of suspicious activity can be identified by this technology. With the application of computer vision and machine learning technologies, the Suspicious Activity Detection System (SHADS) project has created a flexible system that can identify and react in real-time to potentially suspicious human actions. SHADS has proven to be very accurate and responsive through its successful development and thorough performance tests, making it relevant in a variety of fields. Upcoming research topics include improving databases, resolving privacy issues, and maximizing efficiency in contexts with limited resources. All things considered, SHADS has made a substantial contribution to security technology by providing proactive answers to challenging security problems.

VII. FUTURE SCOPE

The Suspicious Human Activity Detection System (SHADS) project can influence the future ahead of it, with plenty of chances for advancement and creativity in the security and surveillance technology space. Further improvements in machine learning, namely in deep learning and reinforcement learning, could lead to even higher precision and effectiveness in identifying questionable behavior. The accuracy and resilience of the system can be further increased by multi-modal sensor fusion, which combines data from multiple sources, including microphones, cameras, and environmental sensors. It will be crucial to address privacy issues and create solutions that preserve privacy in order to safeguard sensitive data and enable efficient activity recognition. Moreover, the system will be more effective and scalable for Internet of Things (IoT) applications if SHADS is optimized for edge computing environments.

VIII. REFERENCES

[1] Benedict Vinusha V, V Indhuja, Medarametla Varshitha Reddy, Nagalla Nikhitha, Priyanka Pramila, " Suspicious Activity Detection using LCRN ", 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2023.

[2] Om M. Rajpurkar, Siddesh S. Kamble, Jayram P. Nandagiri and Anant V. Nimkar, " Alert Generation On Detection Of Suspicious Activity Using Transfer Learning ", in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT),2020.

[3] Amrutha C.V, C. Jyotsna, Amudha J., " Deep Learning Approach for Suspicious Activity Detection from Surveillance Video ", in 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA),2020.

[4] Waqas Iqrar, Malik ZainUl Abidien, Waqas Hameed, Aamir Shahzad, " CNN-LSTM Based Smart Real-time Video Surveillance System ", 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 2020.

[5] Wenchao Xu, Yuxin Pang, Yanqin Yang, Yanbo Liu, " Human Activity Recognition Based On Convolutional Neural Network ", 2018 24th International Conference on Pattern Recognition (ICPR), 2018. . Suspicious Human Crowd Behaviour Detection - A Transfer Learning Approach Learning Approach.

[6] Peshala Liyanage, Pumudu Fernando, " Suspicious Human Crowd Behaviour Detection - A Transfer Learning Approach Learning Approach ", 2021 21st International Conference on Advances in ICT for Emerging Regions (ICter),2021.

[7] Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, Trevor Darrell, " Longterm Recurrent Convolutional Networks for Visual Recognition and Description ", IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 39, Issue: 4, 01 April 2017), 2017.

[8] Tejashri Subhash Bora , Monika Dhananjay Rokade, " Long-term Recurrent Convolutional Networks for Visual Recognition and Description ", International Journal of Advance Research and Innovative Ideas in Education, 2021.

[9] Alavudeen Basha A., Parthasarathy P., Vivekanandan S., " Detection of Suspicious Human Activity based on CNN-DBNN Algorithm for Video Surveillance Applications ", 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), 2019. [10] S. A. Quadri, Komal S Katakdhond, " Suspicious Activity Detection Using Convolution Neural Network ", Journal of Pharmaceutical Negative Results, 2022.

[10] S. A. Quadri, Komal S Katakdhond, " Suspicious Activity Detection Using Convolution Neural Network ", Journal of Pharmaceutical Negative Results, 2022.