e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

DETECTING SUSPICIOUS ACTIVITIES IN SURVEILLANCE VIDEOS USING DEEP LEARNING METHODS

Shreyash Chole[*1], Rishabh Nath Tiwari[*2], Samiullah Siddique[*3], Piyush Jain[*4], Prof. Sagar Mane[*5]

[*1,2,3,4,5]Dept. Of Computer Science & Engineering, NBN Sinhgad School Of Engineering Ambegaon, Pune, India.

DOI : https://www.doi.org/10.56726/IRJMETS33208

ABSTRACT

Video surveillance plays an important role in today's world. Technology has evolved tremendously as artificial intelligence, machine learning and deep learning become mainstream. Using a combination of the above, there are various systems that helps in distinguishing different types of suspicious behavior from live videos. The most unpredictable thing is the behavior of a person and it is very difficult to find out whether it is suspicious or normal. A deep learning approach is used to detect suspicious or unusual activity in the academic environment and send alert messages to the appropriate authorities if suspicious activity is detected. Surveillance is often done using a series of frames captured from a video. All frames are divided into two parts. In the first part, the features are calculated from the video frames, and in the second part, based on the extracted features, the classifier predicts the class as suspicious or normal.

Keywords: Suspicious Activity, Video Surveillance, Deep Learning.

I. INTRODUCTION

It finds many applications in real-world human behavior recognition, intelligent video surveillance, and shopping behavior analysis. Video surveillance has a wide range of applications, especially for indoor and outdoor areas. Surveillance is an integral part of security. Nowadays security cameras are becoming a part of life for safety and security purposes. E-governance is one of the key initiatives of Digital India, a development program of the Government of India. Video surveillance remains a part of it. The advantages of video surveillance include effective surveillance, less labor, cost-effective surveillance capabilities, adoption of new security trends, etc. Now tracking is done by humans. Because we are dealing with a large amount of video data, it is easy for people to feel overwhelmed and manual intervention will also introduce errors. It greatly affects the efficiency of the system. This is solved by automating video surveillance. Currently, it is not possible to monitor all incidents manually on CCTV cameras. Even if the event has already happened, manually searching for the same event in the recorded video is a waste of time. Analyzing abnormal events in video is an emerging topic in the field of automated video surveillance systems.

Human behavior detection in video surveillance systems is an automated way to easily find suspicious objects activity. Airports, Railway Stations, Banks, Offices, Exam Halls etc. There are several effective algorithms to automatically detect human behavior in public spaces such as video surveillance for artificial intelligence, machine learning and deep learning. Artificial intelligence helps computers think like humans. An important component of machine learning is learning from training data and predicting future data. Today, there are GPU (Graphics Processing Unit) processors and large databases, so the concept of deep learning is used.

The combination of computer vision and video surveillance will ensure public safety and security. Computer vision techniques include the following steps: environment modeling, motion detection, moving object classification, tracking, behavior understanding and interpretation, and data fusion from multiple cameras. This method requires a lot of work to extract features in different video sequences. Supervised and unsupervised classification methods. Supervised classification uses manually defined training data, while unsupervised classification is fully computer-driven and does not require human intervention.

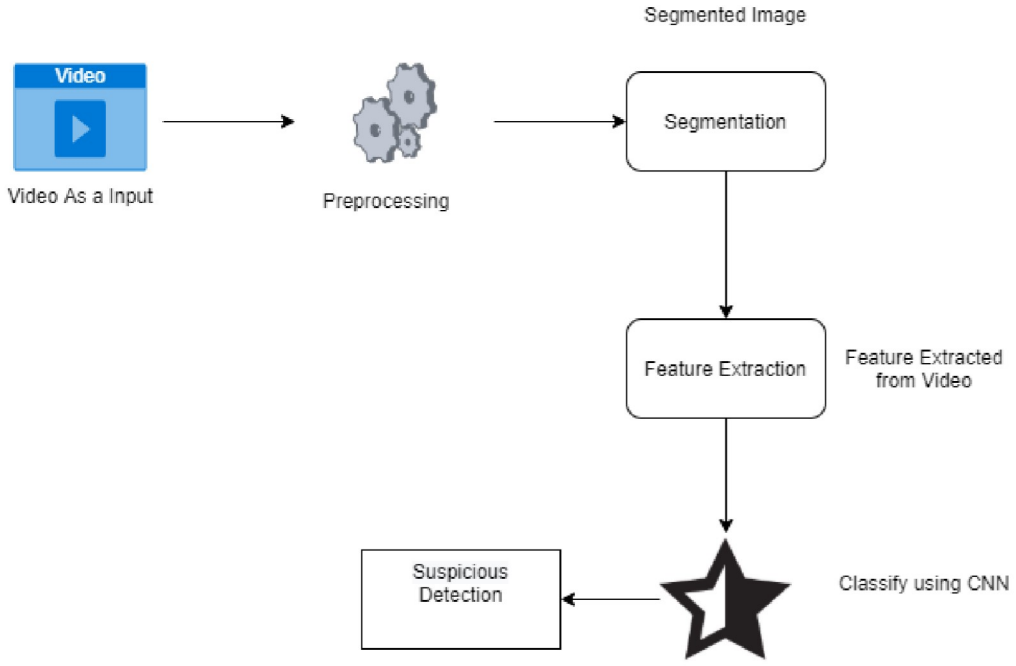
Deep neural networks are the best architectures used to implement complex learning problems. Deep learning models automatically extract features and create high-level representation of image data. This is more common because the feature extraction process is fully automated. A convolutional neural network (CNN) can learn visual patterns directly from image pixels. Long-term memory models (LSTM) in video streaming are capable of

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

learning long-term dependencies. LSTM systems have the ability to store things in memory.

The proposed system will use CCTV footage to monitor the behavior of people on campus and alert them when something suspicious happens. The main components of intelligent video surveillance are event detection and human behavior recognition. Automatic understanding of human behavior is a difficult task. Different areas of the campus should be monitored by video surveillance and various activities. Video footage from the campus was used for testing.

The entire process of developing a monitoring system can be summarized in three stages: data preparation, model preparation and estimation. The framework consists of two neural networks (CNN) and Recurrent Neural Network (RNN). CNN is used to extract high-level features from the image to reduce input complexity. RNN is used to process video streams for classification purposes. The proposed system uses a pre-trained model called VGG-16 (Visual Geometry Group), which is trained on the ImageNet database. A model is now trained to predict behavior from videos. The model can predict the behavior of suspects or normal people in video footage used to assist the surveillance process.



Most systems today use video from CCTV cameras. In the event of a crime or violence, this video will be used for investigative purposes. But if we consider a system that automatically detects unusual or unusual conditions and a mechanism to alert the relevant authorities, it is more interesting and can be used in indoor and outdoor areas. The proposed approach is to design such a system in an academic context.

This paper is organized as follows: Part II describes briefly the work related to behavior analysis to detect suspicious activities. An overview of the proposed method is explained in Section III. Implementation details are described in section IV, followed by conclusions and future work in section V.

II. LITERATURE SURVEY

Related work offers a different approach to detect human behavior through video. The purpose of the work is to detect unusual or suspicious events in video surveillance.

The Advance Motion Detection (AMD) algorithm is used to detect a single input in a restricted area [1]. In the first step, objects are detected by background subtraction and objects are extracted from a sequence of frames. The second step is to detect suspicious activity. The advantage of the system is the performance of the video processing algorithm and the low computational complexity. But the network is limited in terms of service storage and can be done with high-tech video recording in the surveillance area.

[2] proposed a semantic approach. The captured video data is processed and foreground objects are identified by background subtraction. After segmentation, objects are classified as living or non-living according to the

Haar algorithm. Object tracking is done using a Real-Time blob matching algorithm. Fire is found in this paper.

Based on the characteristics of movement between objects, [3] suspicious activity is detected. A semantic approach is used to identify suspicious events. Object detection and correlation techniques have been used for object tracking [2]. Events based on motion characteristics and temporal data. The computational complexity of the given framework is low.

Abnormal phenomena in the university are detected by dividing them into zones, and the optical flow value in each used zone is evaluated.

Lucas-Kanade method. They then created a histogram of the magnitude of the optical flow vector. Software algorithms are used to analyze video content to classify normal and abnormal events [4].

This system is designed to distinguish motion data from normal phenomena based on video sequence analysis. The HMM method is used to study the optical flow histogram of the video frame. It compares captured video frames with existing normal frames and determines the similarity between these frames. The system has been evaluated and validated on several databases such as the UMN and PETS datasets [5].

Unusual events in video recording can be found by tracking people. People are detected by removing the background from the video. Features are extracted using CNN and fed to DDBN (Discriminatory Deep Belief Network). Tagged videos of various suspicious incidents are provided to DDBN and their features are also extracted. Then, the comparison of features extracted using CNN and features extracted from videos of well-known samples of hidden suspicious activity was done using DDBN, and various suspicious activities were detected from the given video [6].

A violence detection system using deep learning has been developed to prevent crowd or player violence in sports. Frames are captured from real videos in the Spark environment. Alert security staff if the system detects football violence. To prevent violence, the system detects video movements in real time and alerts security forces. The VID dataset was used and obtained an accuracy of 94.5% to detect violence in football stadiums [7].

Anomaly detection consists of different modules for video data processing. Deep architecture has been used to explain human behavior. The Interaction UT database is used in the proposed CNN and LSTM based model. One of the weaknesses of the system is that it is difficult to detect human behavior such as pointing or tapping [8].

Understanding crowd behavior using a deep spatiotemporal approach divides video into the prediction of pedestrians' future paths, destination prices, and crowd behavior. Three different categories. Spatial information in video frames is extracted using convolution layers. LSTM architecture has been used to study or understand the sequence of temporal motion dynamics. The data sets used in the proposed system are PWPD, ETH, UCY and CUHK. The accuracy of the system can be improved by using a deeper architecture [9]. Human daily activities are captured from videos and classification of those videos into household, work, caring and helping images. This is done through deep learning about sports. CNN is used for input features and RNN for classification purposes. They use Inception v3 model with UCF101, Activitynet as database. The achieved accuracy is 85.9% in UCF101 and 45.9% in Activitynet [10].

A system was designed to monitor student behavior using a neural network with a Gaussian distribution. It consists of three different steps: face detection, suspicious state detection and anomaly detection. The learning model determines whether students are in a suspicious state, and the Gaussian distribution determines whether students are behaving in all kinds of anomalies [11]. The accuracy achieved is 97%.

Intelligent video surveillance for crowd analysis has been discussed [12]. This is a review paper covering the relevance of video surveillance analysis in today's world, various deep learning models, algorithms and databases used for video surveillance analysis. Most of the mentioned papers have been done using computer vision using different algorithms or neural networks to infer human behavior analysis from videos. Computer vision techniques require a lot of processing to extract trajectories or motion patterns to understand the evolution of features in video sequences [13]. Furthermore, background reduction is based on the assumption of a static background, which is not often used in real-time scenarios. In the real world, most problems occur in traffic. The methods discussed above are ineffective in crowd management. Based on the literature review, a deep architecture can be modeled to predict suspicious activity using 2D CNN and LSTM, so the accuracy of the system can be improved. In deep learning approaches, most papers only detect suspicious activity. Therefore, an effective mechanism is needed to alert security in case of any suspicious activity.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

III. SYSTEM OVERVIEW

The proposed system will use CCTV footage to monitor student activities on campus and report suspicious incidents to the concerned authorities.

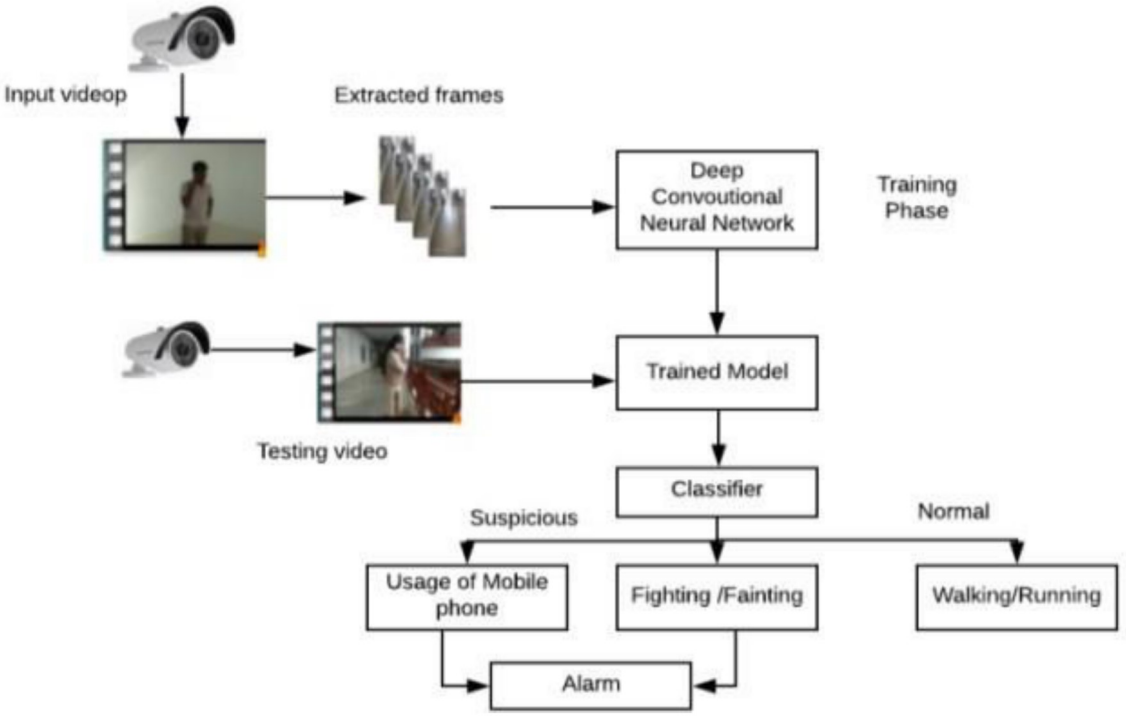
A. System Architecture

The architecture consists of different phases such as video capture, video preprocessing, feature extraction, classification and prediction. An overview of the system architecture is shown in Figure 1. The system divides videos into three categories.

1) Students Using Cell Phones on Campus - Suspicious Class

2) Students fighting or passing out in suspicious campus classes

3) Walking, jogging- Regular class



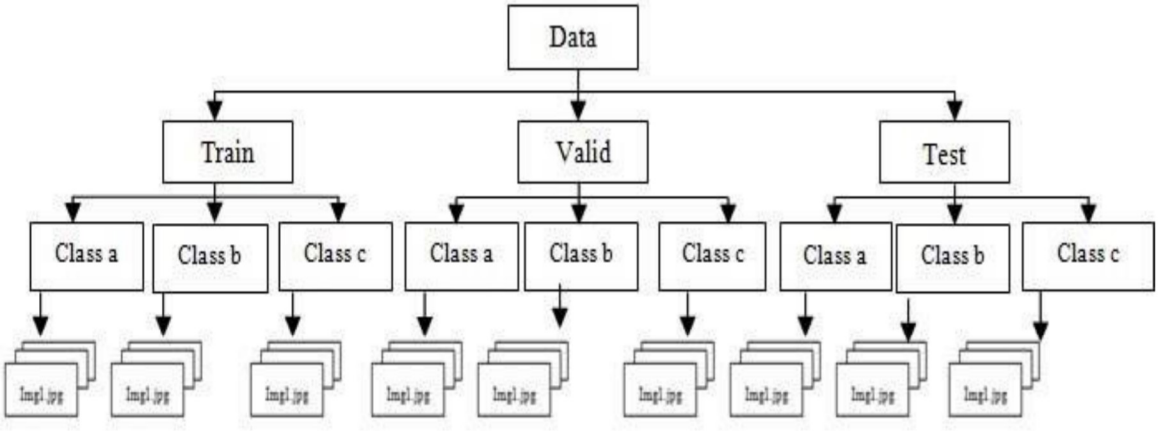
B. Video capture

Installing a CCTV camera and monitoring the captured video is the first step of a video surveillance system. Various videos are recorded from different cameras covering the entire surveillance area. In our implementation, the processing is done using frames, so the video is converted into frames.

C. Dataset Description

The KTH dataset is a standard dataset with 6 sets of motion representation sequences and 100 sequences per motion class. Each sequence contains approximately 600 frames and the video is recorded at 25 fps [14]. This model is trained on this database of typical behavior (running and walking). The CAVIAR dataset, campus videos and YouTube videos were used to study suspicious behavior (cell phones, fighting and socializing on campus). About 7035 frames were captured from several videos. The entire data set was manually labeled and divided into 80% training set and 20% validation set. The directory structure of this database is shown in Figure 2. Our system integrates KTH, CAVIAR, videos and YouTube videos.

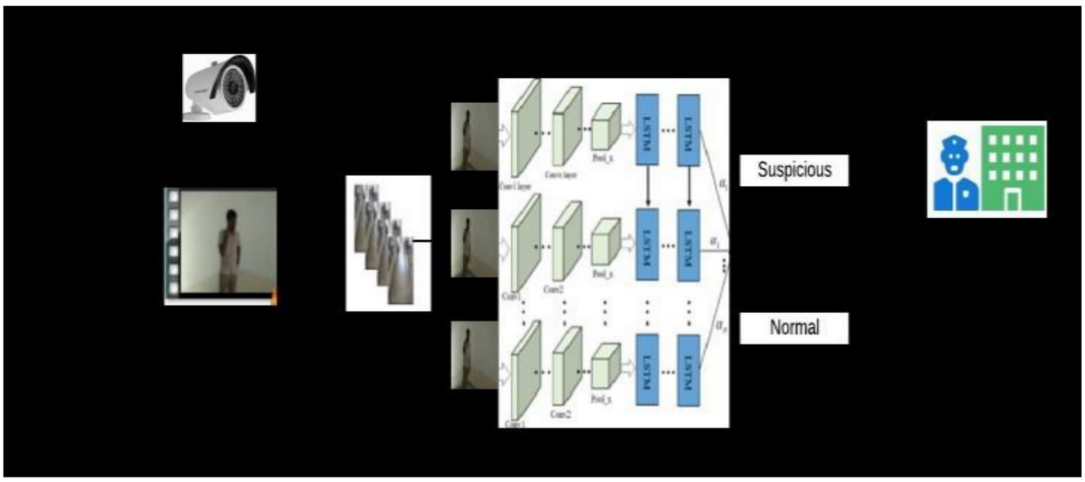
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com



D. Video pre-processing

A deep learning network is used in the proposed system to detect suspicious activities from video surveillance. The accuracy obtained by deep architecture learning can be higher and works better with large databases. A complete design overview is shown in Figure 3.

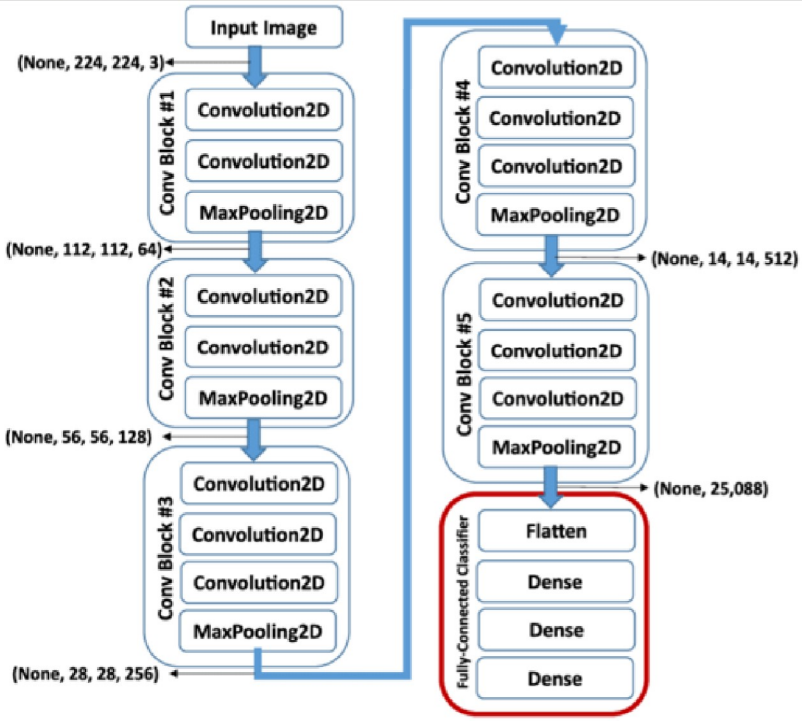
Video access is obtained from existing and established databases. Frames are extracted from captured videos as part of preprocessing. Based on the video, three labeled folders are created and save the frame inside. The entire video is converted to 7035 frames and the frames are saved as images in jpg format. Each frame is then scaled and stored at 224 × 224 to fit the 2D CNN architecture. The test video is converted to a frame with a size of 224 × 224 and stored in a folder. The OpenCV library in Python is used for video preprocessing.



To extract image features, a pre-trained CNN model called VGG-16 is trained on the ImageNet database. The architecture of VGG-16 is shown in Figure 4. The VGG-16 neural network [15] has 3 × 3 convolutional layers, 2 × 2 max-print layers, and finally a fully connected layer, which is the deep learning architecture used here with a total of 16 layers . The input image should be 224×224×3 RGB format. Representation of different layers including traction layer, ReLU (Corrected Linear Unit) layer, ie activation function, maximum compression layer, fully connected solid layer and regularization layer. The model can be adjusted as desired and the last layer of the model is removed. The model is then trained on the LSTM architecture. An LSTM network is a type of RNN that can learn sequence dependencies in sequence prediction problems. We have a ReLU activation function, drop layer and fully connected solid layer. The number of neurons in the last layer is the same as the number of classes we have, so the number of neurons here is three.

The system classifies videos as suspicious (students using cell phones, fighting, passing out) or normal (walking, running). SMS (Short Message Service) will be sent to the appropriate authorities in case of suspicious activity.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com



IV. RESULT ANALYSIS

The aim of the project is to use CCTV videos for suspicious activities on campus and to alert security in case of suspicious incidents. This is done by extracting features from the frame using CNN. After extraction, LSTM architecture is used to classify frames as suspicious or normal class. Figure 5 shows the sequence of suspicious and normal videos.

Collecting video sequences from CCTV footage, removing frames from films, pre-processing the images, creating training and validation sets from the datasets, and training and testing are the steps involved in establishing the entire system. The system notifies the appropriate authority through SMS when it detects suspicious behaviour. Python was used to develop the system on an open source platform. By registering a Twilio account and installing the Twilio library in Python, you can send SMS messages. Twilio enables programmatic sending and receiving of text messages, as well as making and receiving phone calls.

A. Training and Testing

The videos used as input are from the CAVIAR dataset, the KTH dataset, YouTube videos, and on-campus videos. 300 videos of various suspicious and regular conduct have been gathered. Frames are retrieved from the recorded videos as part of the pre-processing procedure. VGG-16 is the pre-trained model that our system uses, and we employ its lessons to solve the challenge. Based on our needs, the final layer of this model was deleted, and LSTM architecture is employed for classification. It has been trained using our dataset. For testing, our campus's CCTV video footage of various circumstances is taken and rendered into frames. The trained model receives the stored frames, and then the classifier determines whether the video depicts suspicious or typical activity.

B. Results

For the first 10 epochs, the training phase's accuracy is 76%. Increasing the number of iterations will increase the model's accuracy. For testing purposes, the frames are taken from videos and kept in a single folder. The

system classifies the frames as typical (walking, running) or suspect (mobile phone use inside the campus, fighting, or fainting) based on our trained model. A communication with the expected class will be forwarded to the appropriate authority in the event of suspicious behaviour. The achieved accuracy is 87.15%. Table I contains the confusion matrix.

	Prediction M	Prediction F	Prediction N
Actual M	45	3	2
Actual F	2	18	1
Actual N	2	3	30

V. CONCLUSION

In today's world, almost everyone understands the importance of CCTV videos, but most of these videos are used for research purposes after a crime/incident has occurred. The proposed model has the benefit of stopping crime before it happens. CCTV video is tracked and analyzed in real time. If the results of the investigation show that an incident occurred outside the control of the relevant authorities, it is an order to take action. So it can be stopped.

Although the proposed system is limited to the academic field, it can be used to predict more suspicious behavior in public or private settings. The model can be used in any scenario that needs to be trained with suspicious activity that matches that scenario. The model can be improved by identifying suspicious people from suspicious behavior.

VI. REFERENCES

[1] S. Karuppuswami, M. I. M. Ghazali, S. Mondal, and P. Chahal, "Wireless eas sensor tags for volatile profiling in food packages," in 2018 IEEE 68th Electronic Components and Technology Conference (ECTC), pp. 2174–2179, 2018.

[2] D. D. M. Dinama, Q. A'yun, A. D. Syahroni, I. A. Sulistijono, and A. Risnumawan, "Human detection and tracking on surveillance video footage using convolutional neural networks," in 2019 International Electronics Symposium (IES), pp. 534–538, 2019.

[3] M. Popa, L. Rothkrantz, Z. Yang, P. Wiggers, R. Braspenning, and C. Shan, "Analysis of shopping behavior based on surveillance system," in 2010 IEEE International Conference on Systems, Man and Cybernetics, pp. 2512–2519, 2010.

[4] N. Dawar and N. Kehtarnavaz, "Continuous detection and recognition of actions of interest among actions of non-interest using a depth camera," in 2017 IEEE International Conference on Image Processing (ICIP), pp. 4227–4231, 2017.

[5] C.-H. Chuang, J.-W. Hsieh, and K.-C. Fan, "Suspicious object detection and robbery event analysis," in 2007 16th International Conference on Computer Communications and Networks, pp. 1189–1192, 2007.

[6] Y. Kaneko, "Fractal analysis of a grocery store shopping path," in 2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), pp. 1–7, 2015.

[7] H. Valecha, A. Varma, I. Khare, A. Sachdeva, and M. Goyal, "Prediction of consumer behaviour using random forest algorithm," in 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), pp. 1– 6, 2018.

[8] Y. Zuo, K. Yada, T. Li, and P. Chen, "Application of network analysis techniques for customer in-store behavior in supermarket," in 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1861–1866, 2018.

[9] Y. Zuo and K. Yada, "Using statistical learning theory for purchase behavior prediction via direct observation of in-store behavior," in 2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), pp. 1–6, 2015.

[10] S. Peker, A. Kocyigit, and P. E. Eren, "An empirical comparison of customer behavior modeling approaches for shopping list prediction," in 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp.1220–1225,2018.