ALERT GENERATION ON SUSPICIOUS ACTIVITY DETECTION Using Convolutional Neural Network

Suvarna Nandyal[1]
[1]Professor, Department of Computer Science & Engineering
[1]Poojya Doddappa Appa College of Engineering, Kalaburagi, India

Sanjeevkumar Angadi[2]
[2]Assistant Professor, Department of Computer Science & Engineering
[2]Nutan College of Engineering & Research, Pune, India

Abstract—
The technology seeks to enable CCTV cameras to recognize suspicious activity without the need for a person to intervene. The purpose of this research is to spot suspicious behaviour during ATM surveillance and notify higher authorities when it is discovered. Surveillance systems are used in ATM, but it becomes difficult to catch the suspicious event with the existing techniques. Hence, this Therefore, this system seeks to take real-time videos from CCTV as an input and feed it to the CNN model constructed with the aid of transfer learning to detect any type of suspicious event in the ATM environment and notify the higher authorities as soon as it occurs. The primary goal is to create a system that can detect suspicious activity without human interaction and send an alarm, completely revolutionizing the surveillance system in use today.

Keywords—Alert Generation, Suspicious Activity, ATM, CNN
DOI Number: 10.48047/NQ.2022.20.20.NQ109302 **NeuroQuantology2022;20(20):3050-3059**

I. INTRODUCTION

The proposed system, called Activity Detector and Alert Generator (ADAG), is designed to make use of surveillance, which is widely accessible at most ATMs. It attempts to make video surveillance capable of automatically spotting suspicious behaviour. The objective of this research is to develop a real-time alert system for detecting suspicious behaviour in ATM surveillance as it occurs. Many existing techniques have been evolved to solve the problems of intelligent system for suspicious activity detection and notification, without human intervention in ATM surveillance. As a result, there is a compelling need for a system that can detect theft based on suspicious behaviour in ATM.

Real-time surveillance recordings can be used as an input by the developed system, which subsequently extracts video frames and feeds them to the CNN model.This CNN model takes a single frame as input, processes it to identify “Normal activity” or “suspect activity”, such as camera break, ATM break or wearing a helmet in an ATM cabin, and outputs a video with frames that have been labelled. Each output frame includes the probability along with the tags “Normal”,“ Camera break”, “Robbery”, or “ATM break”. The higher authorities receive a warning message whenever the label changes from “Normal” to “ATM break”, “Camera break”, or “wearing helmet”. The message is not sent for frames with the tag “Normal”.

Instead of training the CNN model from scratch, this research leverages transfer learning with pre-trained ImageNet weights to train the model. Extraction of frames from real-time video is the initial step. (For instance, video surveillance).The frame is then given to a trained CNN model in the second step. The anticipated label for each frame is pushed to

Queue in the third step. Repeating step 3 for 'k' frames is the fourth step. The fifth and final phase entails choosing the label that most closely matches the average of the previous 'k' predictions.Display the frame with the predicted class label and send an alarm message if there is a difference of more than 80% between the probabilities of the labels for the other classes and the predicted class; otherwise, display the message "Normal".

As a result, in the modern world, it is essential to have a system that can identify suspicious activity. This system provides the services necessary to combat deceit and forgeries, causing a significant revolution in the surveillance system of today.

The rest of the paper is organized as follows: The second section summarizes related intelligent techniques. The proposed methodology is summarized in Section III. In section IV, the implementation specifics are defined, accompanied by a conclusion and future work in section V.

II. LITERATURE SURVEY

Due to the introduction of various information systems and technologies during the past two decades, the surveillance system has greatly increased and developed [1].Both the surveillance systems themselves and the varied methods in which they are used have undergone significant developments.To attain the highest level of accuracy, a variety of approaches including motion detection [2] [3], object detection [2] [4], object tracking [2], concept of fractal [5], and various clustering techniques [6] are utilized.

Various firms, large and small, have begun to progressively use the managerial database to store the numerous accumulated vast amounts of marketing data, to keep the information sorted in order, but there are still various losses and stealing, robbery, and store break-in are a few to name. To boost the potential for scalability and have a correct track, several management-related tools, and policies such as supply chain management, customer relationship management [7], demand management, and customer demand management etc. started being deployed.This complete set of factors came together to design a more effective system for providing targeted surveillance.

One thing is to monitor sales and customer interactions; another is to monitor people's or customers' behaviour.In the past ten years, there has been a significant increase in the deployment of drones and CCTV in public spaces [2].A behaviour surveillance system that ensures public safety has become increasingly popular in the wake of incidents like the 2011 Mumbai terror attack.Additionally, public areas like malls, stadiums for cricket or football, music festivals, and locations where a lot of people assemble should be included.Such locations lack adequate surveillance that ensures people's safety.This paper proposes utilizing the current monitoring system and enhancing it to the point where it can detect questionable human activities.

The purpose of this research is to identify a new strategy to maximizing the accuracy of detecting a suspicious activity rather than utilizing previously tested methods such as SVM [8] [9], genetic algorithm [10], continuous action detection of actions of interest among actions of non-interest [3], video visual analytics system [11], and posture-representation approaches [12], all of which result in forecasting the movement of the customer [10] [5], network concepts [7], random forest algorithm [6], and deep learning based fusion system [13].Numerous studies have been conducted on human tracking, motion detection, and behaviour analysis [14] [15].

Human tracking [16] [17] is another element that aids in understanding a person's behaviour. It serves as a thorough foundation for monitoring human mobility. Essentially, a five-point tracker and a clustering algorithm are employed for human tracking.Let's say five points are marked on the body at various locations, such as the right shoulder, left shoulder, waist, and both knees.The readings are taken on a regular basis.All this analysis would aid in the detection of customer behaviour because a person going quickly and without exhibiting any interest in things would have been coordinated at a distance, while

someone displaying interest in a product and waiting to check it out would form a cluster.A genuine interaction between people and computers is now possible thanks to human gesture and motion detection.Human motion and gesture detection has been used in many applications, which is encouraging the use of the technology more frequently.

This system's implementation along with Human Gesture and Detection presents a comprehensive solution for the store owners and thus illustrates if a customer is interested in a particular product or not [18]. A novel machine learning-based sensing method is employed to check the face detection of a consumer.The same can be utilized for surveillance because a person who doesn't want to buy would make suspicious facial gestures [19], making them easy to spot.The real interaction between humans and machines is made simple using machine learning and deep learning techniques.Thus, this system demonstrates that it is reliable and worthwhile of your time.

Robbery is a global problem. It's an open social problem. Although a growing number of CCTV cameras are beinginstalled at public places (such as airports, banks, shopping malls, etc.), yet the conventional surveillance systems thatrely on human operators are inefficient in detecting rare anomalous events, such as robbery, in real-time. Recently,three-stream C3D+LSTM network [20] have shown promising results in large-scale video analysis tasks. So, a robustand efficient automated surveillance system, able to accurately detect any robbery attempt from the CCTV footage[21], and can respond effectively (e.g., raise alarm, lock the vault, call police, release tear gas, etc.) to foil the plot isneeded. But first a surveillance system to detect robbery in real time is must.

III. SYSTEM OVERVIEW

According to reports, the banking industry loses billions of dollars annually due to theft, robbery, dacoit, burglary.The use of surveillance to observe people's behaviour is possible.Therefore, it is necessary to provide video surveillance with the capability of automatically identifying suspicious activity.The purpose of this research is to assist higher authorities in detecting suspicious behaviour at ATMs in real time.The purpose of this research is to use Computer Vision (CV), a technique extensively utilized to extract relevant information from images and videos.Artificial Intelligence (AI) and Machine Learning(ML) are subsets of CV.CV had extremely limited capabilities before the development of deep learning and neural networks, but now it has advanced significantly.It is evident that CV is superior to humans in several aspects of item labelling and detection.In this paper, we present a Convolution Neural Network-based system that can identify human activities that may be suspicious and alert higher authorities.

Certain suspicious activities in ATM surveillance that are recorded in database, long time duration inside ATM cabin, wearing cap/mask/helmet, sitting near to ATM machine, carrying irrelevant object etc.

To improve system efficiency, ATM transactions can be classified as goaloriented, disoriented, searching for help, or casual. The human whose intention is to withdraw amount from ATM is considered as goal oriented. If any person walks into ATM and peeping here and there for longer duration without performing any kind of operations, this scenario is considered as disoriented. As a result, the system raises a suspicious alert. Searching for help to complete the transaction in ATM results as normal behaviour. It should be emphasized that the intention is determined by how a person enters the ATM cabin.

Therefore, in today's world, it is imperative to provide a system that identifies suspicious activity for surveillance, and as a result, this system supplies such services as handling all such deceit and forgeries and so makes a significant revolution in today's surveillance system.

A. Database Collection

The database contains images that are developed during research work andcategorized into 4 classes ATM_Break (86 Images), Camera_break(53 Images), Normal (252 Images) and Robbery (585 Images).

B. Model Training

Instead of building the CNN model from scratch, transfer learning is employed to train the model using pre-trained ImageNet weights.In this phase, ResNet50 is utilized as the model's backbone; up until the Global Average Pooling (GAP) layer, no layers have been altered.The Fully Connected layer was then removed after adding average pooling with a pool size of (7,7), flattening, the Dense layer (512), the Dropout Layer (0.5), and finally the Dense layer.Fig. 1 shows the complete architecture.

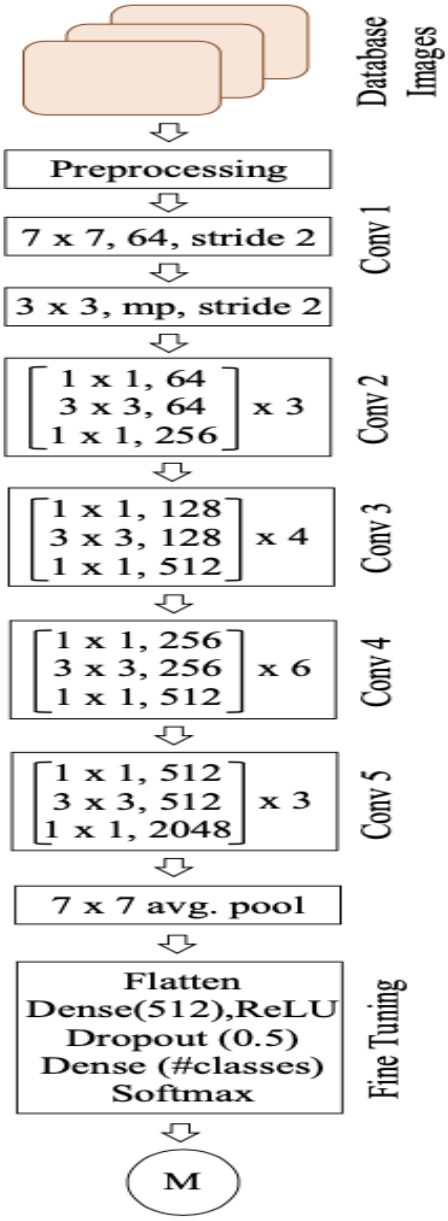


Fig. 1. Fine Tuning of ResNet 50

The CNN is trained with 80 epochs to create a model (M).The performance of the model is assessed in this work.Real-time videos are fed into the system as input in the first step.The frames are taken from a live video stream source (S).The retrieved frame is then pre-processed by RGB channel ordering after being initially downsized (to the size of the image supplied during the training phase of the model).The model receives each frame from the real-time video that has been preprocessed and extracted before predicting its class label.The model outputs the label of the class with the highest probability.All the frames from the real-time video go through the same procedure again.Every input frame's predicted class and probability are shown for each input frame on the output frame.As a result, the output video is a collection of frames, each with a class label and probability.There is a "prediction flickering"

issue with this approach. The output label as a result changes frequently.

C. Principle of Rolling Averaging

The flickering effect is diminished using the rolling average technique.It is often referred to as a moving average or running average.It reduces the volatility in a time series data's output.To apply rolling averaging, a subset of the input time series data is retained in a queue (Q) of size k=64.The mean is computed using the last 'k' predictions from the Q.By choosing the class label (L) with the highest corresponding probability, the output frame is labelled.Fig. 2 shows the mathematical procedure.

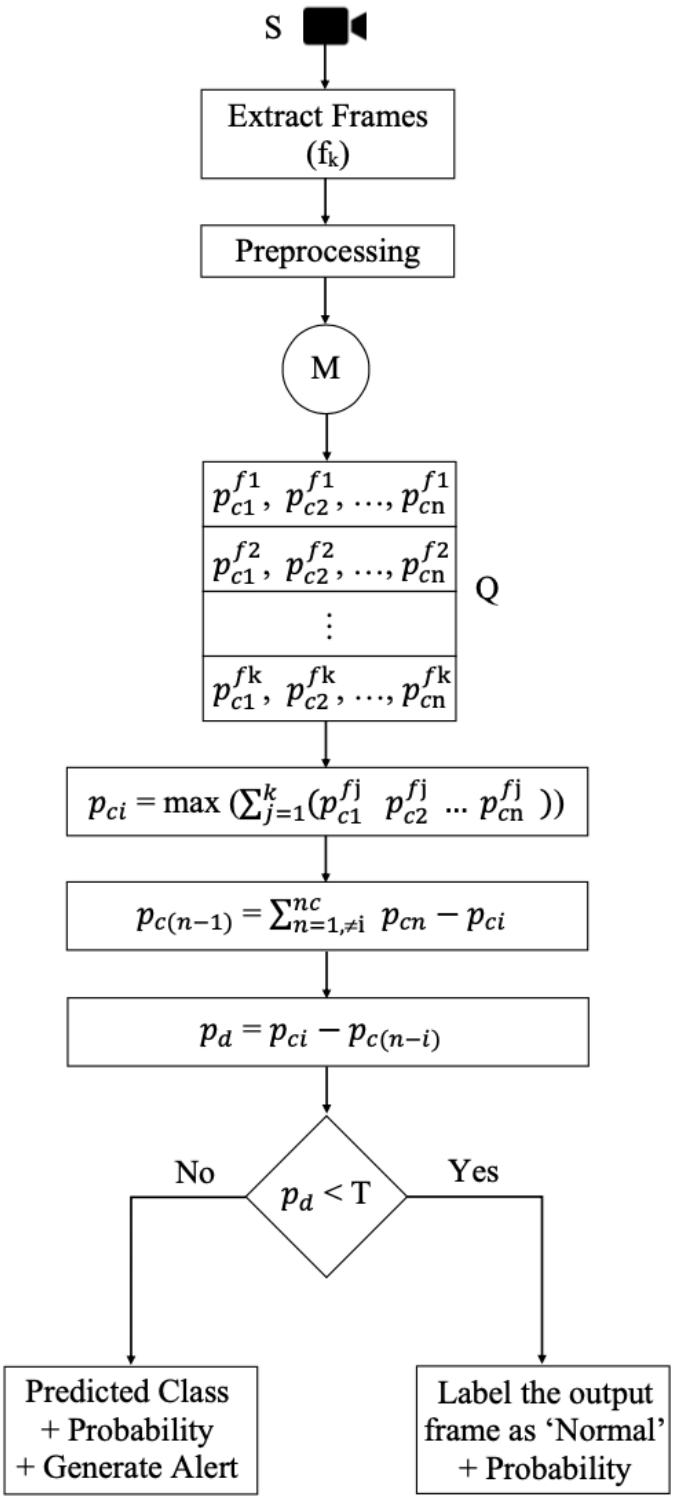
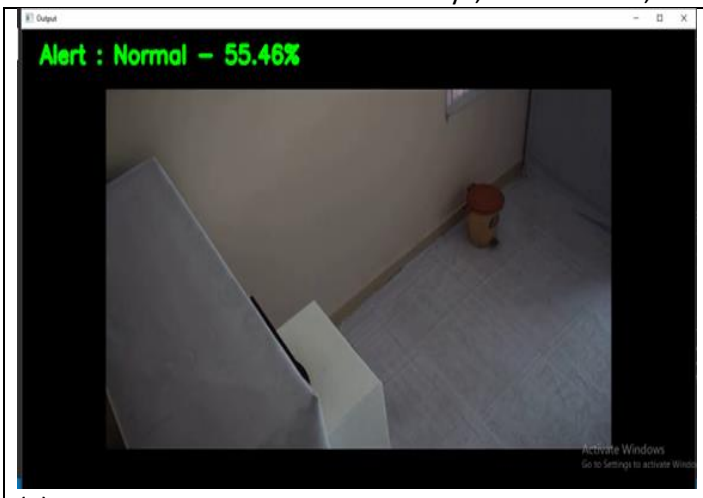


Fig. 2. Mathematical Representation. Pci is the Probability of Suspicious Activity 'ci'

IV. RESULTS AND DISCUSSION

The created system can accept real-time CCTV or pre-recorded videos as input, then extract frames from the video and feed them to the CNN mode.This CNN model takes a single frame as input, processes it to identify a “ATM break”,“Camera break”,“Robbery”, or "Normal" event in an ATM environment, and outputs a video with frames that have been labelled.This CNN model takes a single frame as input, processes it to identify an“ATM break”,“Camera break”,“Robbery” or “Normal” event in an ATM environment, and outputs a video with frames that have been labelled.The higher authorities are notified when the classification changes from “Normal” to “Robbery”,“ATM break”, or “Camera break”.The message is not sent for frames with the ‘Normal’ tag.A simple mail transfer protocol with a 587-port number is used to convey the message to the relevant authorities.

Figure 3 shows the results of the proposed approach by considering the video sample. Figure 3 (a)show the input image with normal tag, fig 3 (b) person is present in the input image (normal alert), fig 3 (c) based on head detection, an alert is converted to suspicious after analyzing the consecutive input frames, fig 3 (d) the suspicious alert continues until it reaches the targeted frame, fig 3 (e) output alert notification based on suspicious activity detected inside ATM cabin.

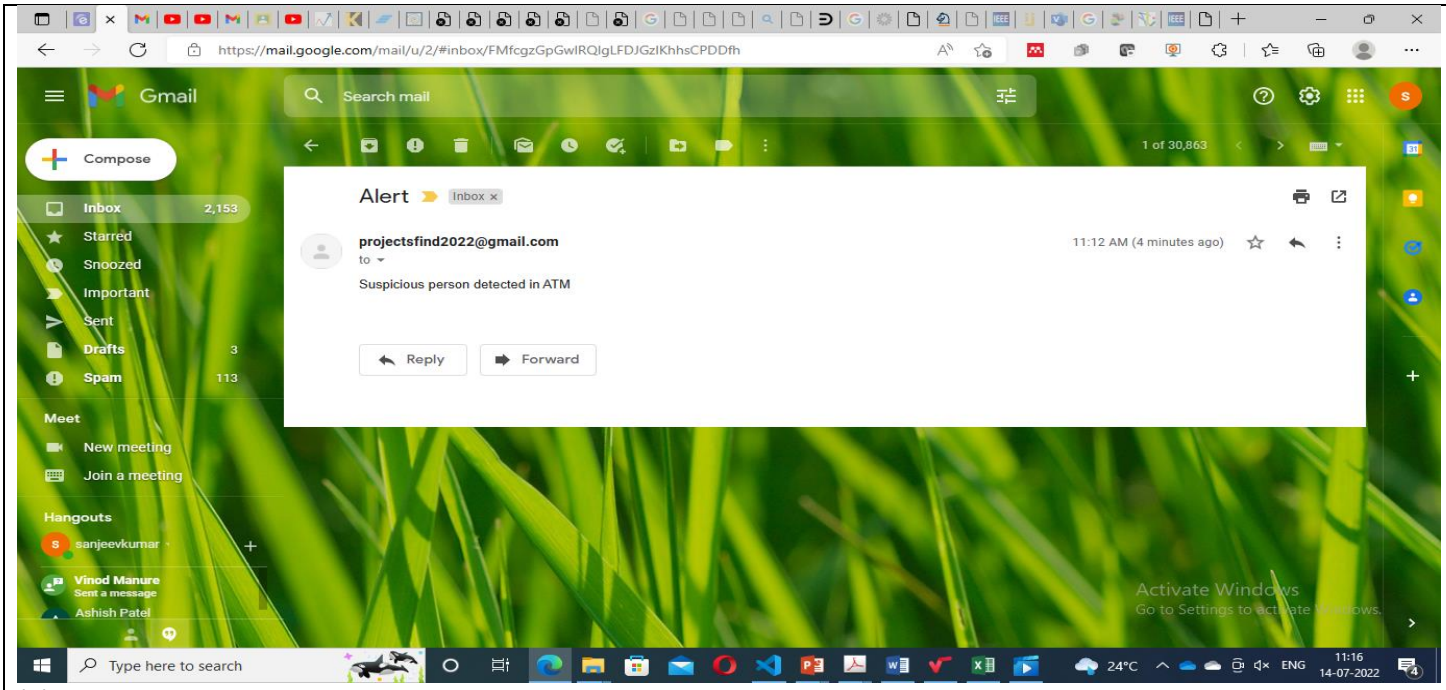


(a)

(b)

(c)

(d)



(e)

Figure 3: Video sample for alert generation a) initial frame of input image (frame no 107), b) person entering ATM cabin (normal tag- frame no 145), c) Head detection (alert raised to suspicious tag after completing 90% of detection- frame no 198), d) Head detection (alert raised to suspicious tag after completing 96% of detection- frame no 234), e) output image (Mail notification with respect to suspicious activity)

3056

The learning curve, which is depicted in Figure 4, indicates how learning performance changes over time as a function of learning experience.Validation Loss, Validation Accuracy, Training Loss, and Training Accuracy are its four parameters.The efficiency of the model is clearly demonstrated by the graph, which shows that both Validation and Training Accuracy rose with the number of epochs.Additionally, it is obvious that the Validation and Training Loss has dropped as the number of epochs increased, proving the model's excellent efficiency.

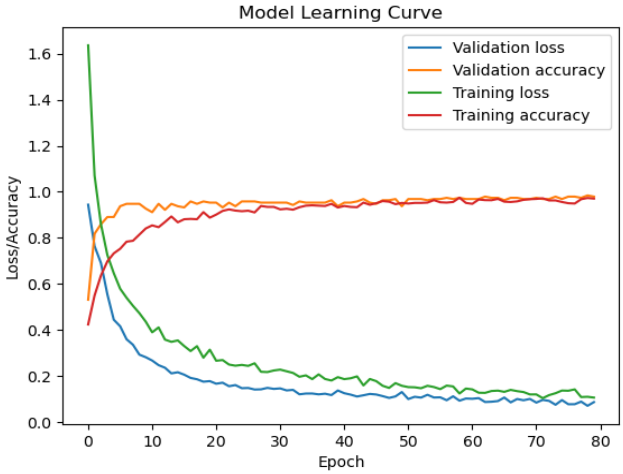
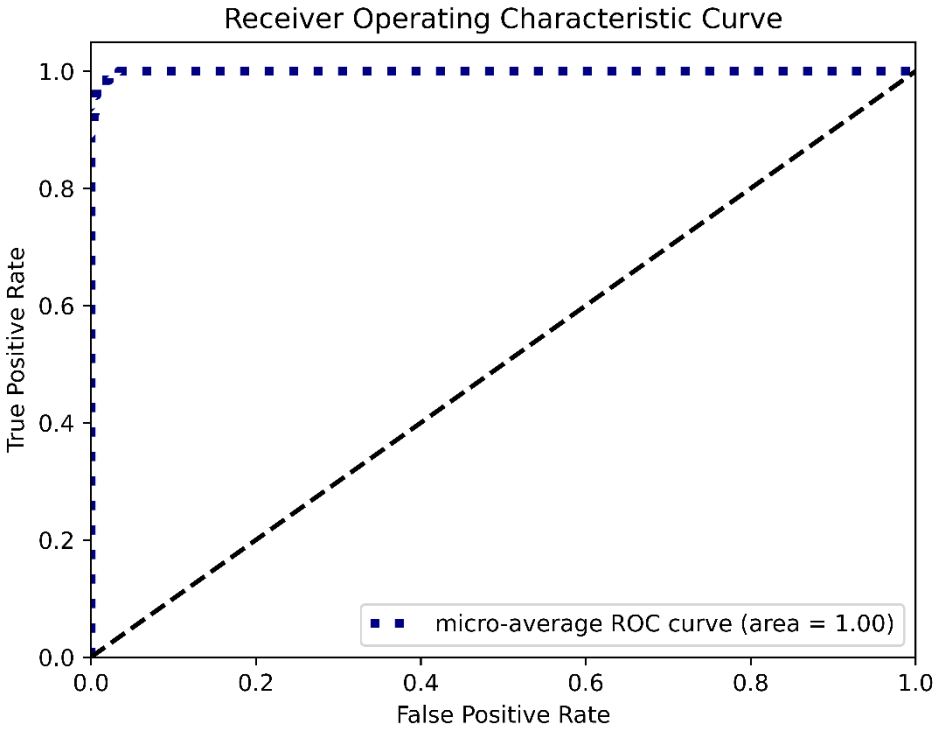


Fig 4: Model Learning Curve

The Precision-Recall curve and Receiver Operating Characteristics (ROC) curve are presented in Figs. 5 and 6, respectively, to assess the results.The accuracy of the model's performance is indicated by the ROC curve's proximity to the top left corner in Figure 4.The model is effective in predicting the classes because the Area Under the Curve (AUC) is 98%.

Figure5: ROC Curve



The Average Precision (AP) score across all classes is 100%, according to the Precision-Recall Curve in Fig. 6. The Confusion Matrix in Fig. 7 depicts the categorization accuracy of the three classes.

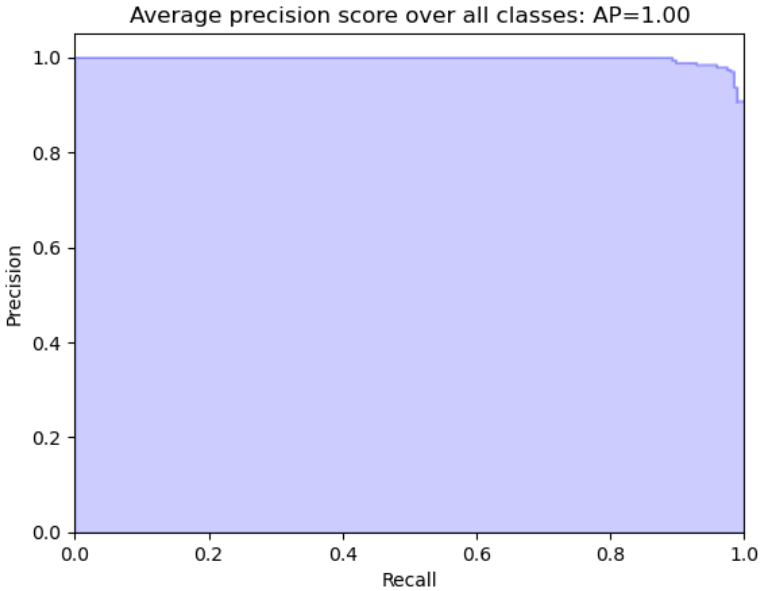
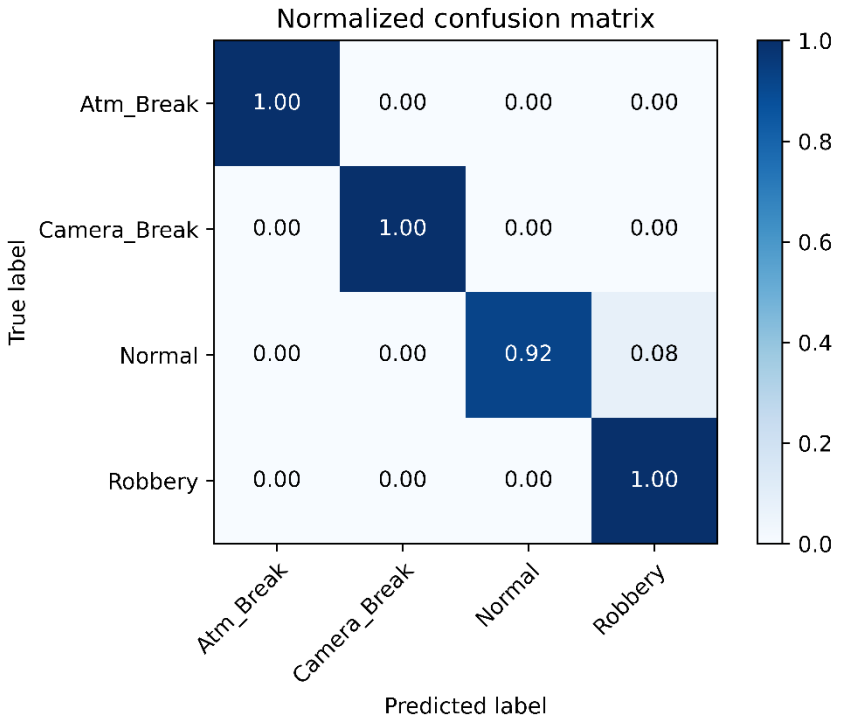


Figure6: Precision-Recall Curve



3058

Figure7: Normalised Confusion Matrix

V.CONCLUSION

The technology described in this paper enables video surveillance to identify suspicious activity without the need for human involvement.This paper's objective, which was to generate an alert when suspicious activity was discovered.It is accomplished by using ATM real-time videos from security cameras as an input, passing them to the CNN model, which then predicts "Normal," "Suspicious," "Camera break," or "ATM break" in an ATM environment and immediately notifies the appropriate authorities. The created database is used. Input videos are converted into frames for further process. They are split into training and testing set in a ratio of 80:20. An accuracy of 98% on the testing set wasseen in the result. Various other measures like precision, recallwas also considered. The system's flickering effect is its only drawback, which can be further reduced by carefully choosing a subset of frames from the queue.The model demonstrated greater accuracy than the results of the prior tests and can be used to identify human activity that might be suspicious based on the overall observed results.Finally, it is determined that creating a system that can assess human behaviour and identify suspicious activity without the need for human interaction is a significant advancement in the surveillance system in use today.

REFERENCES

[1] 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", International Electronics Symposium (IES), pp. 534–538, 2019.

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

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

[4] C.-H. Chuang, J.-W. Hsieh, and K.-C. Fan, "Suspicious object detectionand robbery event analysis," 16thInternational Conferenceon

Computer Communications and Networks, pp. 1189–1192,2007.
[5] Y. Kaneko, "Fractal analysis of a grocery store shopping path," 2nd Asia-Pacific World Congress on Computer Science andEngineering (APWC on CSE), pp. 1–7, 2015.
[6] H. Valecha, A. Varma, I. Khare, A. Sachdeva, and M. Goyal, "Predictionof consumer behaviour using random forest algorithm," 5th IEEE Uttar Pradesh Section International Conference onElectrical, Electronics and Computer Engineering (UPCON), pp. 1–6, 2018.
[7] Y. Zuo, K. Yada, T. Li, and P. Chen, "Application of network analysis techniques for customer in-store behavior in supermarket," IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1861–1866, 2018.
[8] Y. Zuo and K. Yada, "Using statistical learning theory for purchasebehavior prediction via direct observation of in-store behavior," in2015 2nd Asia-Pacific World Congress on Computer Science andEngineering (APWC on CSE), pp. 1–6, 2015.
[9] S. Peker, A. Kocyigit, and P. E. Eren, "An empirical comparison ofcustomer behavior modeling approaches for shopping list prediction,"41st International Convention on Information and CommunicationTechnology, Electronics and Microelectronics (MIPRO),pp. 1220–1225, 2018.
[10] X. Chen, Y. Li, and T. Hu, "Solving the supermarket shopping routeplanning problem based on genetic algorithm," in 2015 IEEE/ACIS14th International Conference on Computer and Information Science(ICIS), pp. 529–533, 2015.
[11] A. H. Meghdadi and P. Irani, "Interactive exploration of surveillancevideo through action shot summarization and trajectory visualization,"IEEE Transactions on Visualization and Computer Graphics, vol. 19,no. 12, pp. 2119–2128, 2013.
[12] W. Lao, J. Han, and P. H. De With, "Automatic video-based humanmotion analyzer for consumer surveillance system," IEEE Transactionson Consumer Electronics, vol. 55, no. 2, pp. 591–598, 2009.
[13] N. Dawar and N. Kehtarnavaz, "Action detection and recognition incontinuous action streams by deep learning-based sensing fusion,"IEEE Sensors Journal, vol. 18, no. 23, pp. 9660–9668, 2018.
[14] W. Liang, Z.Wu, J. Cao, and J. Gu, "Understanding customer behaviorin shopping mall from indoor tracking data," in 2018 IEEE 22nd International Conference on Computer Supported Cooperative Workin Design ((CSCWD)), pp. 648–653, 2018.
[15] Y. Chen, J. Zhang, M. Guo, and J. Cao, "Understanding customerbehaviour in urban shopping mall from wifi logs," in 2017 IEEE InternationalConference on Pervasive Computing and CommunicationsWorkshops (PerCom Workshops), pp. 50–53, 2017.
[16] J. Kaewyotha and W. Songpan, "A study on the optimization algorithm for solving the supermarket shopping path problem," 3rd International Conference on Computer and Communication Systems (ICCCS), pp. 11–15, 2018.
[17] S. K. Teoh, V. V. Yap, and H. Nisar, "A non-overlapping view human tracking algorithm using hsv colour space," InternationalConference on Green and Human Information Technology (ICGHIT), pp. 97–102, 2019.
[18] N. Zerrouki, F. Harrou, Y. Sun, and A. Houacine, "Vision-basedhuman action classification using adaptive boosting algorithm," IEEESensors Journal, vol. 18, no. 12, pp. 5115–5121, 2018.
[19] M. B. Ayed, S. Elkosantini, S. A. Alshaya, and M. Abid, "Suspiciousbehavior recognition based on face features," IEEE Access, vol. 7,pp. 149952–149958, 2019.
[20] Z. Yahya and M. M. Ullah, "Classification and temporal localization of robbery events in cctv videos through multi-stream deep networks," IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT), pp. 028–032,2019.
[21] R. Kakadiya, R. Lemos, S. Mangalan, M. Pillai, and S. Nikam, "Ai based automatic robbery/theft detection using smart surveillancein banks,"3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 201–204, 2019.