International Journal of Scientific Research and Engineering Development-– Volume 5 Issue 6, Nov- Dec 2022

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

HUMAN SUSPICIOUS ACTIVITY DETECTION FROM SURVEILLANCE VIDEO USING DEEP LEARNING

Sehreen Mir*, Er Simranjit Kaur**

*([1]Pg Scholar, Department of Electronics and Communication Engineering, Sri Sai College of Engineering and Technology, Badhani, Pathankot

Email: Sehr7051@gmail.com)

** (Assistant Professor, Department of Electronics and Communication Engineering, Sri Sai College of Engineering and Technology, Badhani, Pathankot)

--************************----------------------------------

Abstract:

Detecting suspicious activities in public places has become an important task due to the increasing number of shootings, knife attacks, terrorist attacks, etc. taking place in public spaces all throughout the globe. Terrorism, theft, accidents and illegal parking, vandalism, fighting, chain snatching, crime, and other suspicious activities can all be prevented through visual surveillance in sensitive and public areas like bus stations, railway stations, airports, banks, shopping malls, schools and colleges, parking lots, roads, etc. It's next to impossible to keep a constant eye on a public area, so we need sophisticated video surveillance that can track people's movements in real time, classify them as normal or suspicious, and sound an alarm if anything out of the ordinary happens. Various anomalous pursuits are included in the suggested work. Foreground object extraction, object identification using tracking or non-tracking approaches, feature extraction, classification, activity analysis, and recognition are all covered, along with the broader topic of recognising human activity from surveillance movies. This study employs a deep learning strategy to analyse video and picture data for signs of criminal activity using Recurrent Neural Networks (RNN). Here, we do just that, comparing the efficacy of several RNN topologies. We describe the framework of our system that can analyse camera feeds in real time to determine whether or not an event is suspicious. We also provide ideas on how this field of monitoring may advance in the future.

Keywords — **Human activities, Vedios, RNN,attacks.,**

--************************----------------------------------

I. INTRODUCTION

The function of video content analysis is to find meaningful structures and samples from visual data. Video analysis tasks comprise video parsing, content indexing, abstraction, and representation. The task of activity recognition is to overpass the gap among the numerical pixel level data and a high-level abstract activity account. Anomaly detection in video surveillance is a challenging task due to many difficult problems, such as noise, illumination change and deformation in the scenes, diversity of event, and interaction between multiple events. Moreover, the multi-view video sequences are captured frequently under illumination and lighting conditions. Multiple cameras may have like positions, orientations, and zooming factors. From a fundamental point of view, techniques in video investigation are inspired by the need to expand machine learning algorithms that can emulate the abilities of human visual frameworks.

Machine learning approaches are popular in the area of anomaly detection for automated learning and detection which is based on explicit or implicit model that enables classification of the patterns analyzed. This chapter develops an activity recognition system which classifies the abnormal/ normal event from the crowded scenes.

Suspicious human movement acknowledgment from reconnaissance video is a functioning exploration territory of picture preparing and PC visions. Though the visual reconnaissance, human exercises can be checked in delicate and public regions, for example, transport stations, rail route stations, air terminals, banks, shopping centers, school and universities, parking garages, streets, and so on to forestall psychological warfare, burglary, mishaps and unlawful stopping, defacement, battling, chain grabbing, wrongdoing and other dubious exercises. It is hard to watch public places persistently, thusly a shrewd video reconnaissance is required that can screen the human exercises and arrange them as regular and strange exercises; and can create an alarm.

For detecting suspicious human activity, it is important for the model to learn suspicious human poses. For almost 15 years, researchers in the field of computer vision have been trying to solve the difficulty of accurately estimating a human's stance. It has to do with tagging human bodily components so we can monitor their whereabouts. It finds use in virtual reality, motion tracking, video game consoles, etc. The first consoles to detect human movement did so using inexpensive depth sensors (motion sensors). This kind of sensor can only be used inside, and its poor resolution and noisy depth information make it challenging to assess human movement from depth photos. As a result, they shouldn't be considered as a means of identifying anomalous behaviour.

The pose of the individuals in a picture or video may be determined in real time with the help of models like OpenPose[1] and PoseNet[2]. The keypoints of the persons involved are important, but they are not sufficient on their own to determine whether or not an action is suspicious without knowledge of context or the items involved. An example of a strange exercise includes leaving behind equipment for potentially lethal attacks, burglary, jogging in a group, a fight or assault, defacing property, or crossing a border. Common forms of physical activity practised by people in public places include jogging, walking, hand-waving, and clapping. Today, video surveillance is being used to monitor human activity in order to prevent suspicious behaviours.

II. VEDIO PROCESSING

Video processing is defined as the investigation of video content for obtaining an understanding of the scene that it describes. Video surveillance activities can be manual, semi-autonomous or fully- autonomous. Manual video surveillance is the process of analyzing the video content in the video stream by a human. Video processing fused with some form of human intervention is used in Semi-autonomous video surveillance. An example of semi autonomous video surveillance is the system that performs simple motions.

Human experts analyze the recorded video. By a fully-autonomous system, a system wherein the input alone is the video sequence taken at the scene where surveillance is performed. The Figure 1. shows that the framework of video surveillance which includes all stages of processing like background estimation, object detection, object tracking, object classification and activity understanding. Moving object segmentation is the basic step for further analysis of the video. It handles the detection of moving objects from stationary background objects. Commonly used techniques for object detection are background subtraction, statistical methods, temporaldifferencing and optical flow.

Motion detection is segmenting the regions, corresponding to moving objects from the rest of static images. The technique used for low level processing is background subtraction, background modeling, temporal differencing, and optical flow. In order to track objects and analyze the behavior, it is necessary to correctly classify the moving objects. The methods used for segmentation are shape based and motionbased segmentation.

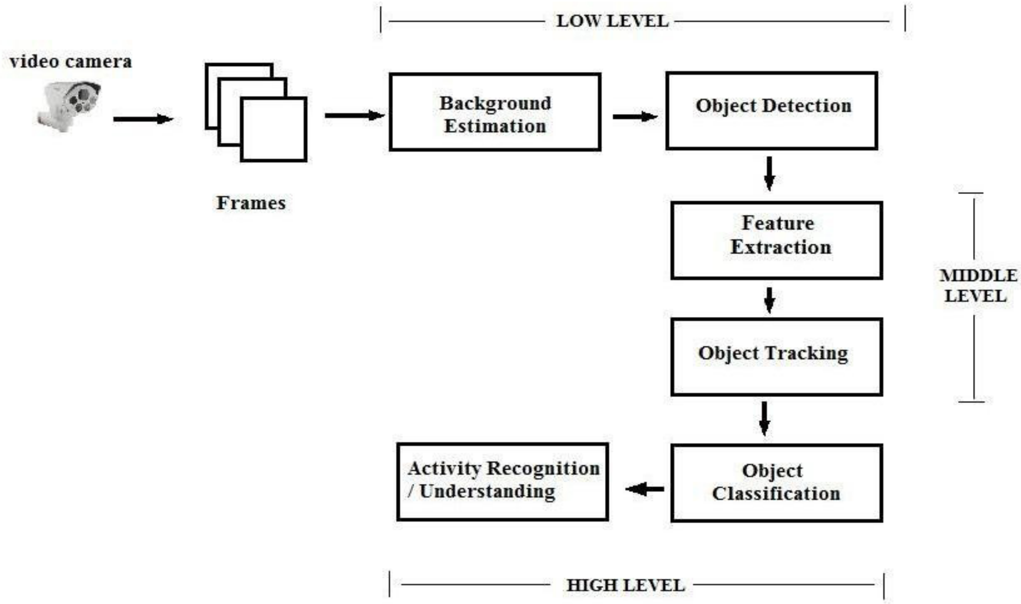


Figure 1: Framework of Video Surveillance

The video tracking algorithm usually has considerable intersection with motion detection during processing. The video processing generally tracks moving objects from one frame to another by region-based, active contour-based and feature based tracking. The final stage of video processing is an activity recognition, which is used for identifying who the objects of interest with the help of predefined operations and algorithm. The machine learning algorithm is used for processing. Some of the learning algorithms are Linear Regression, Logistic Regression, Decision Tree, SVM, Naïve Bayes, KNN, K-Means and Random Forest. In general, the video is processed at different levels, such as low level, middle level, and high level processingwhich is shown in Figure 2

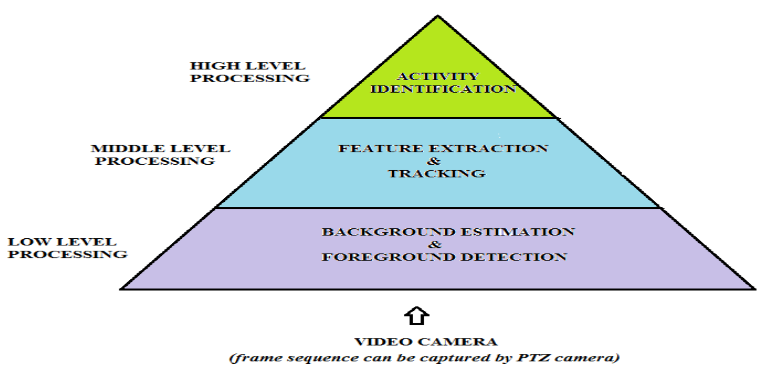


Figure 2 Levels in video processing

Low level processing is mainly focused on estimating thebackground pixel and foreground object detection,which is used by middle level and high level processing.

III. RELATED WORK

From another point of view, there are a number of reasons why human activity recognition has challenging problems such as the presence of many degree of freedom for the human body, no two person being identical, variation in viewpoint, self-occlusion, deformation and so on. This chapter reviews related works on the various levels of processing in video that include background estimation, foreground detection, feature extraction, event detection, behavior detection, and Deep learning approaches.

Fathia .G et al (2021) It's long been used to provide security in sensitive places, but with the advancement of technology, traditional surveillance operations are facing a wide-ranging set of challenges, including the need to deal with a huge volume and a short amount of time, and the risk of information loss that could reveal suspicious behaviour. Video surveillance has received a lot of attention lately. An intelligent surveillance system for identifying anomalous activity that might pose a security concern will be discussed in this article. Walking and running are the two types of human activity that the algorithms are designed to identify. As far as the amount of persons involved and how they were moving, there were no limitations enforced. However, we only allow films shot with a single fixed camera inside in colour. Background subtraction technique is used to identify the moving items in the scene that match to individuals. To classify activities, we rely on centroids' displacement and the size of segmented regions' changes in size as our primary determining factors. Moving objects (suspicious activity) can be detected in video using a series of procedures, including dividing the video into frames, separating the background from objects within the video, and using morphological operations to remove any background noise. The mathematical operations are then used to determine which images contain suspicious activity. In this study, the following procedures were used: The suggested algorithms have a high degree of accuracy in determining the kind of activity.

Ashutosh Rawande et al (2021) Detecting human activity has been a major focus in artificial intelligence and computer vision in recent years. Since the human eye is incapable of correctly detecting suspicious activity, the need for automated monitoring in security cameras has skyrocketed. It is possible to prevent crime before it happens by detecting suspicious activity and automatically reporting it. We have divided human actions into two categories: aberrant and normal. Activities that are considered normal include sitting and walking as well as hand-waving. Kicking, hitting, brandishing a pistol, holding a knife, etc. are all examples of abnormal actions. In order to accomplish this categorization, we make use of convolutional and recurrent neural network architectures. To begin, high-level characteristics from pictures are extracted using a convolutional neural network. For the final prediction, the recurrent neural network is used to process data from a pooling layer, rather than using the convolutional neural network's final classification as input.

Pankaj Bhambri et al (2020): Anomaly detection systems are widely employed in behavioural analysis in combination with machine learning and artificial intelligence to assist identify and forecast the occurrence of abnormalities. Enterprises may use it for anything from intrusion detection to monitoring the health of their systems, as well as for everything from detecting credit card fraud to detecting errors in operating settings. A majority of nations are using accurate anomaly detection systems in order to get closer to a more comfortable zone. In the context of India's 42.38 crime index, the need for anomalydetection frameworks is serious. CCTV systems will not be able to keep an eye on us. In addition to being able to identify myself, these technologies may also be used to forecast odd activity.

Tanzila saba et al (2020): Research and industry are paying close attention to intelligent visual surveillance systems. Intelligent visual surveillance systems may now be developed because to the development of smart surveillance cameras that have more processing capacity than ever before. The safety of individuals may be ensured both at home and in public settings. The goal of this project is to help surveillance systems identify potentially harmful activity. For this, a 63-layer deep CNN model called "L4-Branched-ActionNet" has been proposed and christened. In the proposed CNN structure, AlexNet has been modified and four blanched sub-structures have been added. By executing its training on an object detection dataset named CIFAR-100 using the SoftMax function, the generated framework is turned into a pretrained framework. For feature acquisition, the dataset for suspicious behaviour identification is sent to this pretrained model. An optimization process known as feature subset optimization is used to reduce the size of the deep features. In order to optimise the entropy-based coded features, an ant colony system (ACS) is used on the entropy-coded features. Several classification models based on SVM and KNN use the preset features. In terms of accuracy, the cubic SVM performs best, with a score of 0.9924. Using the Weizmann action dataset, the suggested model achieved an accuracy of 0.9796. The results show that the proposed research is sound.

Adam et al. (2018) introduced a constant non-following based calculation for uncommon movement (for example individual running in a shopping center) discovery which is vigorous and functions admirably in packed scenes. Calculation of this framework screens low level estimations in a bunch of fixed spatial situations as opposed to following to objects. Absence of successive observing is the principle impediment of this calculation.

Wiliem et al. (2018) introduced a programmed dubious conduct finder which uses the logical data. The three fundamental segments, an information stream grouping calculation, a setting space model, and a derivation calculation of the framework; uses logical data to distinguish the dubious conduct. An information stream bunching calculation empowers to the framework to refresh the information ceaselessly from the approaching recordings. Induction calculation consolidates both the logical data and framework information to derive the choice. The framework utilized two datasets-23 clasps of CAVIAR dataset and 2 clasps from Z-Square dataset of Queensland College of Innovation. This framework AUC is 0.778 with 0.144 mistakes.

IV. PROBLEM IDENTIFICATION

Human activity detection for video surveillance system is an automated way of processing video sequences and making an intelligent decision about the actions in the video. It is one of the growing areas of Computer vision and artificial intelligence. A lot of cameras are installed in many places for surveillance, but the surveillance is done by human, and it is done only if there is a report of anomaly behaviour, otherwise the videos are kept as archives, and never use. Developing algorithms for automatic detection of Human movements, and making appropriate decision when there is any suspicious behaviour, it will result to real time processing of Human activities in public places. It will help in security, and ensuring public safety. Previous Human Activities Recognition approaches were used in classification of activity rather than predicting ongoing activities. The methods were good in recognizing simple actions, but they were not good for complex actions (similar body gestures).

sequential state that model human action as hidden states and create postures to enable recognition ofactions. The shortcomings of the traditional activity recognition approaches are:

a) They are not suitable for predicting real time activities

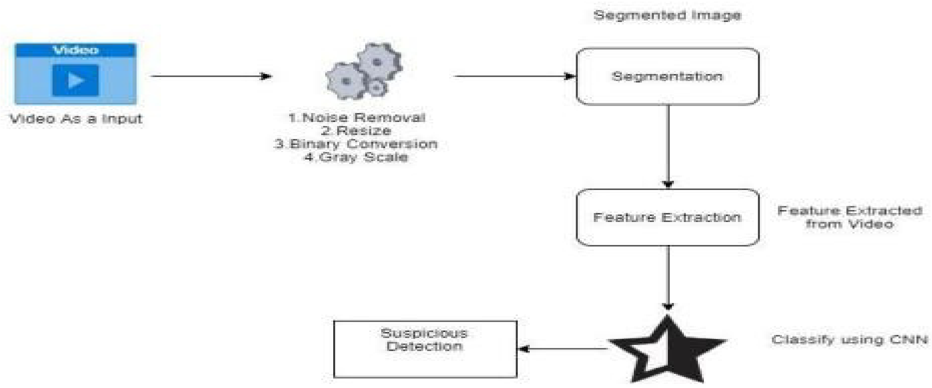
b) They are not suitable for modern high dimension videos

c) They are not suitable for noisy and multiple subjects recognition.

V. METHDOLOGY

Human activity recognition can be useful to a variety of scenarios, and anomaly detection in security systems is one of among them. Seen the increasing demand for security, surveillance cameras have been widely set up as the infrastructure for video analysis. One of the major challenges faced by surveillance video analysis is detecting abnormal activity which requires exhausting human efforts. Fortunately, such a labor-intensive task can be recast as an anomaly detection problem which aims to detect unexpected actions or patterns. Anomaly detection varies from the traditional classification problem in the following aspects:It is very difficult to list all possible negative (anomaly) illustrations.It is a daunting job to collect adequate negative samples due to the rarity.An activity recognition system is projected to identify the basic day to day activities performed by a human being. It is challenging to achieve high rate accuracy for recognition of these activities due to the complexity and diversity in human activities. Activity models required for identification and classification of human activities are constructed based on different approaches specific to the application. The activities of a human being can be generally categorized into normal activities or anomalous activities. A human being's deviation from normal behavior to abnormal causing harm to the surrounding or to himself is classified as an anomalous activity. To achieve anomaly detection, one of the most widespread method is using the videos of normal events as training data to learn a model and then detecting the suspicious events which would do not fit in the learned model. For example, human pose guesstimate is used in applications including video surveillance, animal tracing and actions understanding, sign language recognition, advanced human-computer interaction, as well as marker less motion capturing. Low cost depth sensors consist of limitations like limited to indoor use, and their low resolution and noisy depth information make it difficult to estimate human poses from depth images. Hence, we are to using neural networks to overcome these problems. Anomalous human activity recognition from surveillance video is an active exploration part of image processing and computer visualization.

In our proposed system, for detecting anomalous behavior, the CNN i.e. convolution neural network have been used. For effectively classification of anomalous activities, it is essential to recognize the temporal data in the video. Recently, CNN is mostly used for extracting key features from each frame of the video. CNN is only the algorithm best suited for this purpose. For classifying the given input successful, it is necessary that the features get extracted from CNN, therefore CNN should be capable of knowing and extracting the needed features from the frame of videos.



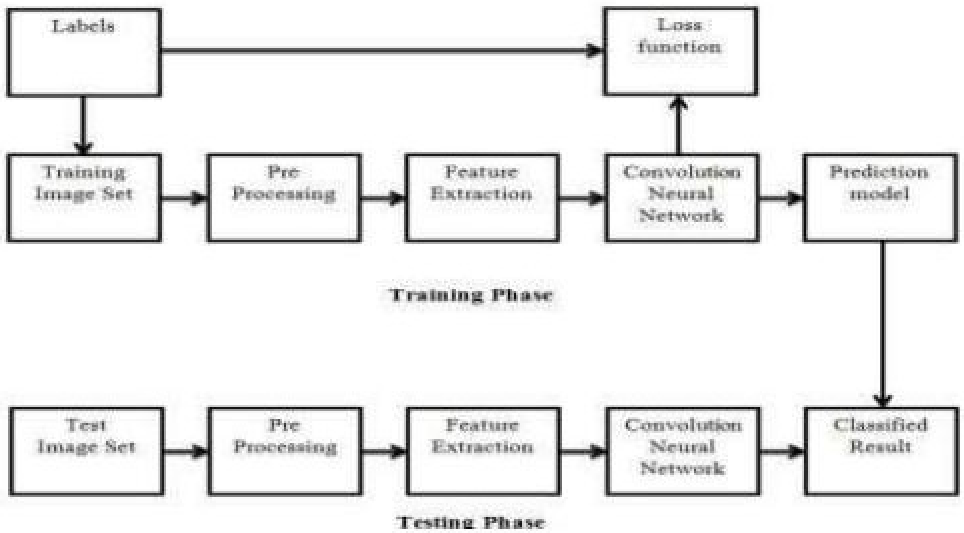
VI. PROPOSED WORK OF MODEL

Data Collection: First of all, the information for different Websites and Social Media applicationsbased on certain parameters is extracted data.

Preprocessing: Then we will apply various pre-processing steps such as Noise removal, resizing, binaryconversion and gray scaling in order to make our dataset proper.
Noise removal: Noise is removed from the input video. In image processing, the key process for denoising is filtering. Generally average filters, median filters, Wiener filters and Kalman filters are utilized to reduce noise.
Resizing: Image resizing is necessary when we need to increase or decrease the total number of pixels, whereas remapping can be done when we are adjusting for lens distortion or rotating an image.
Binary conversion: A binary image is one that holds the pixels that can have any one of precisely two colors, classically black and white. Binary images are also well known as bi-level or as two-
level. This means that each and every single pixel is put in storage as a solitary bit – i.e. in value of 0and1.
Gray scaling: Gray-scaling is the method of transforming a continuous-tone image to an image thata computer can manipulate effortlessly.
Segmentation: Image segmentation is the significant process in which isolation of a digital image into multiple segments is carried out i.e. (sets of pixels, also recognized as image objects).6. Data
Training: We compile artificial as well as real time using online news data and provide training with anymachine learning classifier.
Feature extraction: Feature extraction is a part of the dimensionality decrease procedure, in which, aninitial set of the raw data is separated and compact to more controllable groups.
Classification: Classification is the method of sorting and labeling groups of pixels or vectors within an image based on definite rules and instruction
Data Training: We gathered artificial as well as real time using social media data and provide trainingwith any machine learning classifier.
Testing with machine learning: We give testing dataset to system and apply machine learning algorithm to detect the activity accordingly.

Analysis: We determine the accuracy of proposed system and estimate with other existing systems.

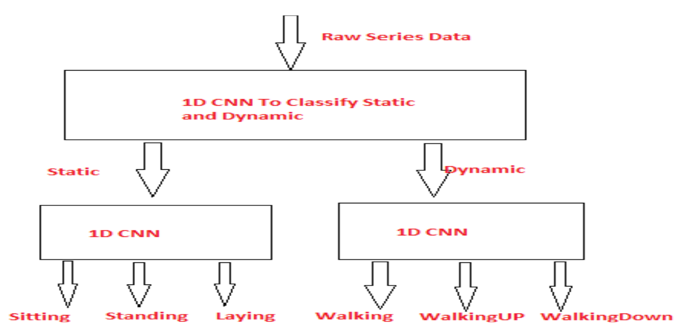
VII. BLOCK DIAGRAM FOR PROPOSED MODEL CNN



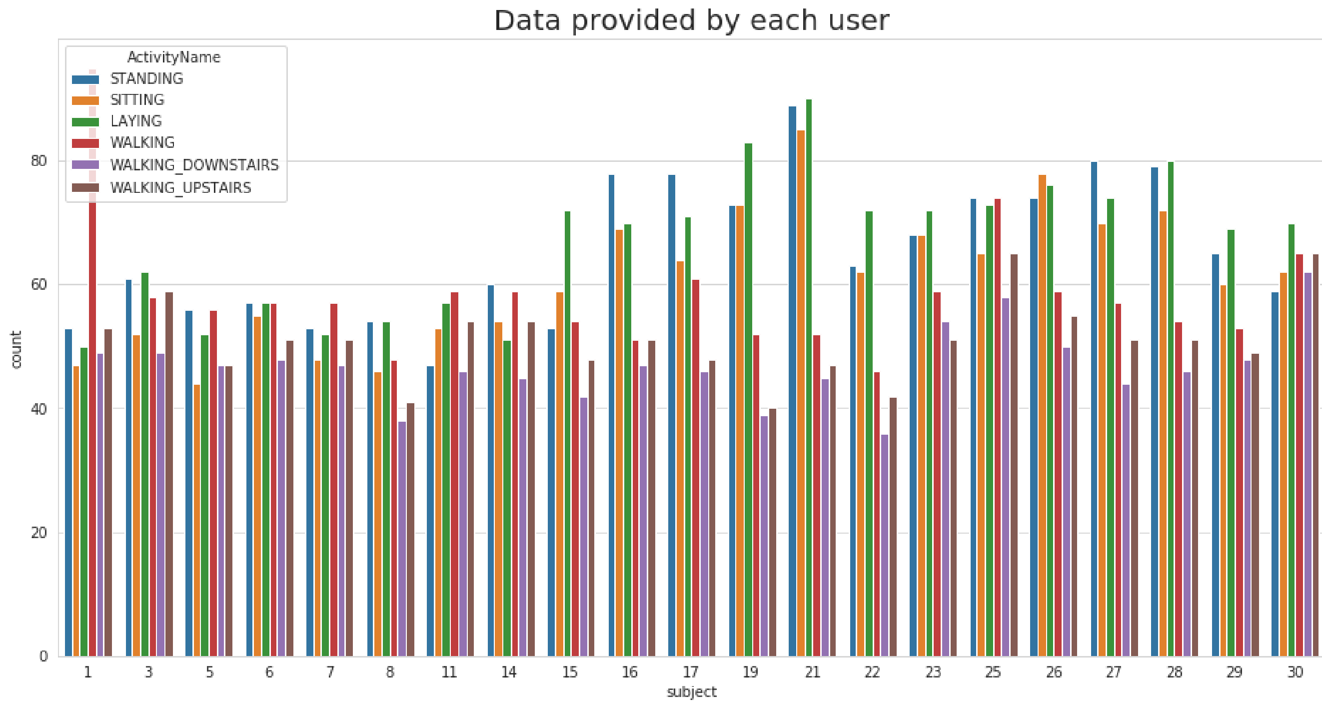
VIII. Algorithm Design

Algorithm: Convolution Neural Network(CNN) Step 1: Input is given as image / video.
Step 2: Then many different filters are applied to the input to create a feature map.Step 3: Next a ReLU function is applied to increase non-linearity.
Step 4: Then applies a pooling layer to each and every feature map.
Step 5: The algorithm compresses the pooled images into one long vector.
Step 6: In next step, inputs the vector to the algorithm into a fully connected artificial neuralnetwork.
Step 7: Processes the features via the network. At the end fully connected layer delivers the "voting"of the classes.
Step 8: In this last step trains through forward propagation and back propagation for numerous epochs. This repetition occurs until we have a well-defined neural network with trained weights and feature detectors.

.



IX. RESULTS AND DISCUSSIONS



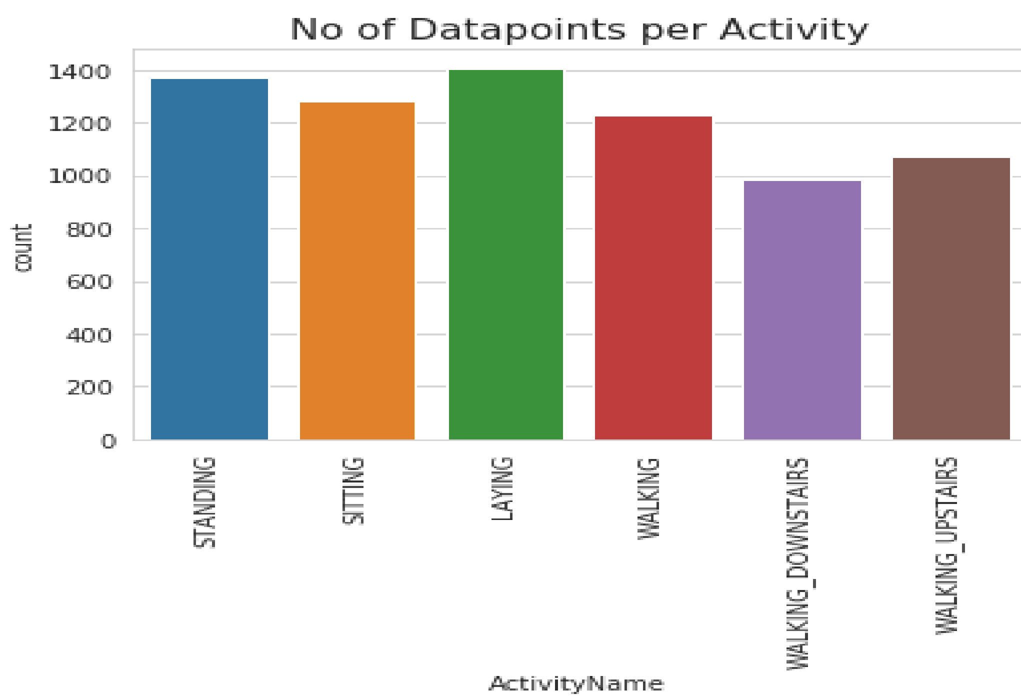


Figure 4: Data as per activity

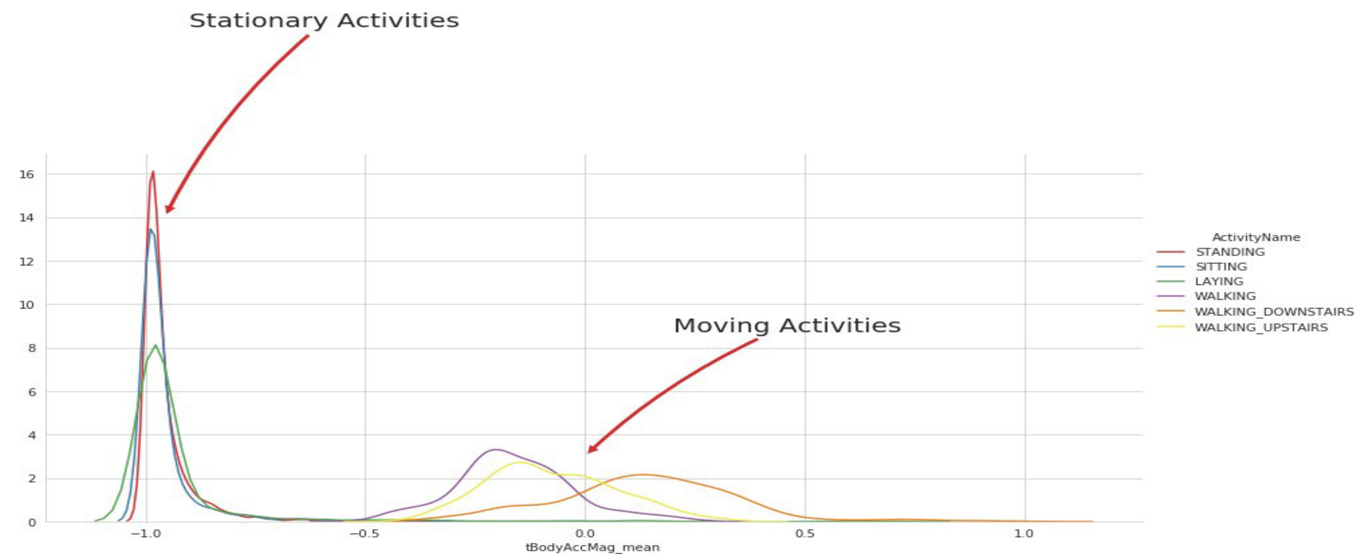


Figure 4(a): Total activities

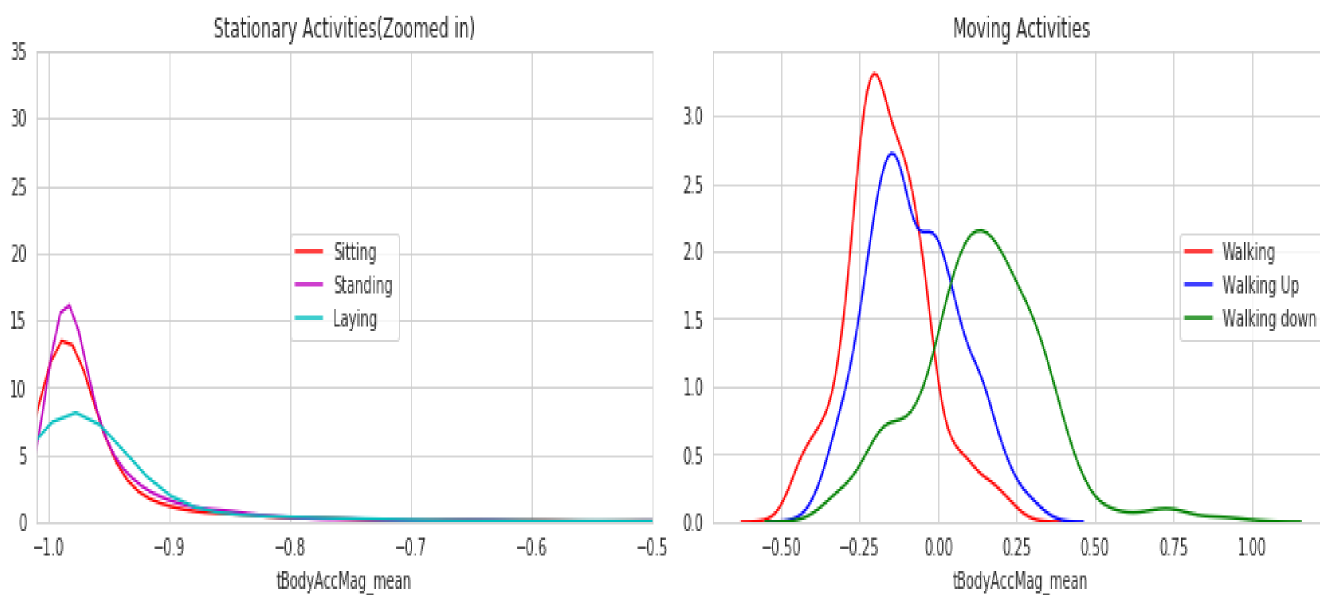


Figure 4(b): Classification of activities as per input data

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Accuracy = (TP+TN)/total = 93.37%

Precision = TP/predicted:yes = 83.19%

Recall = TP/Actual:Yes = 89.34%

F-score = (2 * precision * recall) / (precision + recall) = 0.861553

X. CONCLUSION

From this research work, this proposal method is possible to detect Suspicious Activity from the watched person's behavior, and to measure the degree of risk of Suspicious Activity due to finding the detecting point. This surveillance camera system can identify unsafe oversight area and suspicious person, and bring observer's attention to them using this proposal method. And, an observer is relieved of the burden of mind and body occurred from the matter, which is an observer must watch enormous quantity of image data shot by multiple Web cameras constant monitoring of remote control. Then, this system occurs serious problem, which an observer misses important predictor of crime in area under surveillance. This proposal method pinpoints the detecting point of Suspicious Activity, and finds the degree of risk of Suspicious Activity so that observer can lessen physical and mental burden in monitoring

Refrences

1. F. G. Ibrahim Salem, R. Hassanpour, A. A. Ahmed and A. Douma, "Detection of Suspicious Activities of Human from Surveillance Videos," 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, 2021, pp. 794-801, doi: 10.1109/MI-STA52233.2021.9464477.
2. Pawade, A., Anjaria, R., Satpute, V.R. (2021). Suspicious Activity Detection for Security Cameras. In: Kumar, R., Dohare, R.K., Dubey, H., Singh, V.P. (eds) Applications of Advanced Computing in Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-33-4862-2_22
3. T. Saba, A. Rehman, R. Latif, S. M. Fati, M. Raza and M. Sharif, "Suspicious Activity Recognition Using Proposed Deep L4-Branched-Actionnet With Entropy Coded Ant Colony System Optimization," in IEEE Access, vol. 9, pp. 89181-89197, 2021, doi: 10.1109/ACCESS.2021.3091081.
4. R.Venkatesh Babu, Patrick Perez, PatrickBouthemy, Robust tracking with motion estimation and local Kernel-based color modeling, Image and Vision Computing, Vol. 25, ppI205-1216, 2007.
5. Qiang Zhou, Limin Ma, David Chelberg, Adaptive object detection and recognition based on a feedback strategy, Image and Vision Computing, Vol. 24, pp80-93, 2006.
6. S.J.McKenna, H.Nait-Charif, Tracking human motion using auxiliary particle filters and iterated likelihood weighting, Image and Vision Computing, Vol. 25, pp852-862, 2007.
7. Paul Brasnett, LyudmiJa Mihaylova, David Bull, Nishan Canagarajah, Sequential Monte Carlo tracking by fusing multiple cues in video sequences, Image and Vision Computing,Vol. 25, ppI217-1227,2007.
8. Rama Bindiganavale, Norman I. Badler, Motion Abstraction and Mapping with Spatial Constraints. Nadia Magnenat - Thalmann, Springer - Verlag Berlin Heidelberg,CAPTECH 98, LNAI 1537, pp70-82, 1998.
9. Ankur Agarwal, Bill Triggs, Tracking Articulated Motion Using a Mixture of Autoregressive Models, Springer - Verlag Berlin Heidelberg, ECCV 2004, LNCS 3023, pp54-65, 2004.
10. Shiming Xiang, Changshui Zhang, Xiaoping Chen, Naijiang Lu, A New Approach to Human Motion Sequence Recognition with Application to Diving Actions, Springer - Verlag Berlin Heidelberg,MLDM 2005, LNAI 3587, pp487-496, 2005.
11. Feiyue Huang, Huijun Di, Guangyou Xu, Viewpoint Insensitive Posture Representation for Action Recognition, Springer - Verlag Berlin Heidelberg,AMDO 2006, LNCS 4069, pp143- 152,2006.
12. Kenichi Morooka, Junya Arakawa, Hiroshi Nagahashi, Face Detection by Generating and Selecting Features Based on Kullback - Leibler Divergence, Transactions on Fundamentals of Electronics, Information and Communication Engineers D Springer - Verlag Berlin Heidelberg,Vol. J89-D, No.3, pp530- 540,2006.
13. Rui Ishiyama, Masahiko Hamanaka, Shizuo Sakamoto, Face Recognition under Variable Pose and Illumination Conditions Using 3D Facial Appearance Models, Transactions on Fundamentals of Electronics, Information and Communication Engineers D-II, Vol. J88-D-II, No. 10, pp2069-2080, 2005.
14. Howard, Jeremy, and S Gugger. 2020. Deep Learning for Coders with fastai and

PyTorch: AI Applications Without a PhD. O'Reilly Media, Inc.
15. Howard, Jeremy, and Sylvain Gugger. 2020. "fastai: A Layered API for Deep Learning." arXiv.
16. Paszke, Adam, Sam Gross, and Others. 2019. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In Advances in Neural Information Processing Systems 32, 8024--8035. Curran Associates, Inc.
17. Cao, Zhe, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields." arXiv.
18. Cipolla, Alex Kendall, Matthew Grimes, and Roberto. 2015. "PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization." arXiv.
19. Chollet, François. 2015. "Keras." GitHub repository (GitHub). https://github.com/fchollet