IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

Detection Of Suspicious Activity In Video Reconnaissance Framework

Shilpa M[1], Anush S[2], Arun L Jadava Lamani[3], Kiran D[4], Sujan S P[5]

Assistant Professor, Computer Communication Engineering, KSIT, Bengaluru, India[1]

Student, Computer Communication Engineering, KSIT, Bengaluru, India[2]

Student, Computer Communication Engineering, KSIT, Bengaluru, India[3]

Student, Computer Communication Engineering, KSIT, Bengaluru, India[4]

Student, Computer Communication Engineering, KSIT, Bengaluru, India[5]

Abstract: Video observation currently plays a significant role on a global scale. Advancements have led to a significant increase in the integration of manufactured knowledge, machine learning, and deep learning into devices. The application of the combinations and unique frameworks aids in distinguishing various questionable behaviors from the real-time analysis of photographs. The human way of behaving is the most capricious; additionally, it is extremely challenging to decide if it is suspicious or normal. In this work, we have characterized human exercises into two categories: normal and suspicious. Typical exercises incorporate sitting, strolling, running, hand waving, and so forth. Arrest, abuse, shoplifting, and so on are examples of suspicious exercises. We accomplish this arrangement by utilizing convolutional neural networks. First, we use the convolutional neural network to separate significant-level highlights from pictures. We consider the grouping of the convolutional network, eliminate the outcome of the last pooling layer, and create the final forecast. The CIFAR-100 dataset confirmed the recommended model's accuracy of 0.9796 percent.

Keywords: Suspicious action; profound learning; convolutional neural organization

I. INTRODUCTION

As the number of video reconnaissance channels increases, human checking becomes increasingly ineffective. Inconsistency Location expects to mitigate this issue by filtering out average occurrences and only bringing suspicious ones to the attention of human security personnel [4]. Recently, virtual entertainment has witnessed a surge in reports of "anomalous" items or lights captured on reconnaissance film. While a portion of these cases might hold merit, the majority are logically misleading problems or straightforward misinterpretations of typical items or lighting. People have shared these videos, presenting them as proof of supernatural activity. However, in a few instances, it's plausible that a remarkable phenomenon is occurring, which calls for further investigation [6]. Shopping centers commonly fill public places such as air terminals, train stations, and carnivals. As urbanization increases, the number of people passing through these areas continues to rise year over year [3]. This development improves the probability of group charges in open regions. As a result, it is critical to screen groups to immediately identify any unusual events and prevent hazardous situations. Swarm examinations are primarily for wellbeing and security reasons [10]. Human administrators constantly look at visual screens to identify any events of interest, which can be difficult to stay aware of. This has prompted scientists to develop robotized strategies to assist administrators and distinguish unusual behavior in swarms [5].

Every one of these events is interesting, requiring further examination to decide whether something uncommon is going on. Notwithstanding, it is fundamental to stay receptive and try not to rush to make judgment calls about expected results [6]. Until we gather additional evidence, we cannot definitively determine whether any of these events are truly paranormal in nature.

II. OBJECTIVES

- The essential goal is to figure out what's irregular movement with a video by utilizing the transfer's feedback.
- To foster effective and trustworthy techniques for perceiving suspicious way of behaving and simplifying misrepresentation location and speedier.
- The optional objective is to see as any sort of peculiarity in a reconnaissance camera
- It speaks with the Wire application's alarm message when strange action distinguished

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

III. PROBLEM STATEMENT

- In this work we use CNN algorithm for Image recognition, Haar Cascade Face landmark for detecting person, CNN Slow-Fast Algorithm for movement recognition
- DCSASS Dataset from Kaggle website was used to train the model in which the dataset contains videos based on following 13. There is a total of 16853 videos, where 9676 videos are labeled as Normal and 7177 as abnormal in which it provides 73 % for Training and 27% for Testing.
- In this work, we provide an unrecognized video as an input, whether it classifies normal or abnormal activity. It also specifies which activity, i.e., abuse, arrest, assault, accident, burglary, fighting, robbery, shooting, stealing, shopping, and vandalism. Each video is labelled as 0 as normal and 1 as abnormal.
- In this work, we provide a live streaming detection that detects a person's age, categorizes gender as male or female, and detects whether the person is wearing a mask or not. This will greatly lower the number of false positives and speed up security reaction times.
- If it detects abnormal activity then it provides an alert in the mobile application

IV. LITERATURE REVIEW

Xu, D., Yan, Y., Ricci, E., & Sebe, N.[1] In intelligent video surveillance, malicious event detection is of the utmost significance. Right now, most methodologies for the programmed examination of intricate video scenes commonly depend available made appearance and movement highlights. In any case, embracing client characterized portrayals is obviously less than ideal, as learning descriptors well defined for the location of interest is attractive. To adapt to this need, in this paper we propose Appearance and Movement DeepNet (AMDN), a clever methodology in light of profound brain organizations to learn highlight portrayals consequently. To take advantage of the correlative data of both appearance and movement designs, we present a clever twofold combination structure, consolidating the advantages of conventional early combination and late combination methodologies. In particular, stacked denoising autoencoders are proposed to independently learn both appearance and movement highlights as well as a joint portrayal (early combination). Then, in view of the learned highlights, numerous one-class SVM models are utilized to foresee the abnormality scores of each info. At long last, a clever late combination technique is proposed to consolidate the registered scores and identify strange occasions. On publicly accessible video surveillance datasets like UCSD Pedestrian, Subway, and Train, the proposed ADMN is thoroughly evaluated, demonstrating superior performance to current methods.

Zhang, X., Yang, S., Zhang, X., Zhang, W., & Zhang, J. [6] Catching momentary directions utilizing optical stream and portraying them with a histogram-based shape descriptor (shape settings) ,Using a K-NN similarity-based statistical model to find anomalies i.e., Recover K-closest neighbour tests from the preparation set for a given testing test, Utilize the similarities between the K-NN samples to build a Gaussian model. Compute the likelihood of the likenesses between the testing working together example and the K-NN tests under the Gaussian model. Recognize strange occasions assuming the joint likelihood is underneath predefined limits in reality.

Mostafa, T.A., Uddin, J., & Ali, M.H. [10] Choosing a bunch of Focal points (POI) from the video edges and following them across numerous casings , Breaking up the video frames into a number of cubes and looking for spatial-temporal statistical deviations in the motion patterns in each cube.

Ibrahim Salem, F.G., Hassanpour, R., Ahmed, A.A., & Douma, A. [5] Use foundation deduction calculation to recognize moving items (individuals) in indoor variety recordings caught by a fixed camera ,Separate two fundamental highlights for movement order: 1. Uprooting pace of the centroids of the portioned forefront regions 2. Pace of progress in the size of the portioned regions - Partition video into outlines, separate foundation from objects, eliminate commotion utilizing morphological tasks, and perform numerical activities to figure out which casings contain suspicious movement.

Existing solutions for SHA recognition use various approaches, including object tracking and detection by background subtraction, feature extraction, object classification, and suspicious activity detection. Objects are tracked and identified using frame changes, and foreground objects are retrieved. Feature extraction involves extracting motion and shape features from objects. Object classification distinguishes between different objects in the video, using methods like SVM, Bayesian, Haar-classifier, KNN, face recognition, and skin color detection. The final stage is suspicious activity detection, which compares the items in the video stream with several threshold values to check for aberrant behavior [14].

Misha Karim et.al. proposed human actions in a diverse environment (HADE) framework to recognize and categorize human actions in 3Dspace. HADE uses these capabilities to improve 3D action recognition accuracy and reliability.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

HADE Captures a wide spectrum of human movements processed using our novel HADE I and HADE II models [15]. Misha Karim et. al. proposed survey on standard HAR system's architecture, outlining the crucial technological components and their interaction [16]. The authors recommended that connections between HAR and Computer Vision needs to be established to provide accurate values.

Authors[17] performs the comparative analysis of insider threat detection, demonstrating how it benefits significantly from diverse AI-based approaches.

Authors proposed ResDLCNN-GRU Attention Network for creating a sophisticated and efficient system for Violence detection in video streams [18].

Authors proposed ASRNet deep learning model for detecting and classifying various anomalous situations. ASRNet is a model based on CNN architecture with a total of 63 layers [20].

V. EXISTING FRAMEWORKS

There are a wide range of kinds of recognition and distinguishing proof models for oddity identification, the majority of which depend on various factual models or AI calculations[9]. The most widely recognized sort of oddity location depends on a limit model, which is utilized to distinguish values that are more prominent or under a set worth. Different models utilize a circulation model to distinguish peculiarities by recognizing values that are not inside the ordinary dissemination of information. A behaviour model, the third type, uses a set of rules to find unusual behaviour[11].

In existing system video can be categorized as normal or abnormal but in proposed system we categorized as 13 classes ie., abuse, arrest,arson, assault, accident,Burglary , fighting, robbery,shooting,stealing,shoplifting and Vandalism. When it comes live streaming the it detects weapons and sent alert message with telegram application, it detect weather the person is male or female, it detect the age and also detects weather the person is wearing mask or not.

VI. PROPOSED FRAMEWORKS

In videos, the frames consist of two parts: static and dynamic. While we can alter the frame, the static area remains unchanged, whereas the dynamic area undergoes changes in the object, background, and other elements. For instance, during a meeting, the handshakes between two individuals are dynamic and fast-paced, while the background and other objects remain static. This paper introduces SlowFast CNN as a tool for identifying anomalies. We design a slow pathway to capture static information from videos with low frame rates and slow refreshing speed.

The fast pathway captures all dynamic information at high frame rates and a fast refresh speed. The formal pathway is very light-weighted. Lateral connections merge both pathways. In both pathways, the SlowFast network uses the Resnet model and runs 3D convolution operations on it. The slow pathway uses large strides. The stride is defined as the number of frames skipped per second. Generally, we set it to 16. The fast pathway typically allows for two sampled frames per second. The fast pathway uses too small a stride, typically eight. This allows 15 frames per second.

Structure configuration is the sensible model that describes the development, lead, and purpose of a system.

A. Assortment of Video Tests

The test tests are the pictures that are taken by the camera and are gathered. The better example matching technique is utilized in the proposed framework to plan the test. The information recordings are selected from various camera recordings. Authentic accounts from cameras are taken and filtered the vital accounts

$X \longrightarrow$ Input image set

During the feature extraction phase, the gray scaled image frame set is defined as X = Xi for i = 0,1,2,...,n. X is the input image set, and Xi is the i th image with n frames.

B. Video preprocessing

There are a number of subcomponents within the video processing component that correct the input image in a variety of ways, including removing noise, outliers, dimensionality, and data, among other things. Do not insert equations as images.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

C. Train the Rundown of abilities of Pictures

Separating the image educational record into planning and testing to cross endorse our pre-arranged model to really take a gander at its precision and perceive given picture is common picture or a surprising picture. Due to the amount of cuts per pixel, the video pictures might not have high goal and contain clamor. In order to get rid of noise and improve the image, the picture is preprocessed by using explicit preprocessing procedures like histogram balance and Middle channel.

X⟶ Pre-processed image set

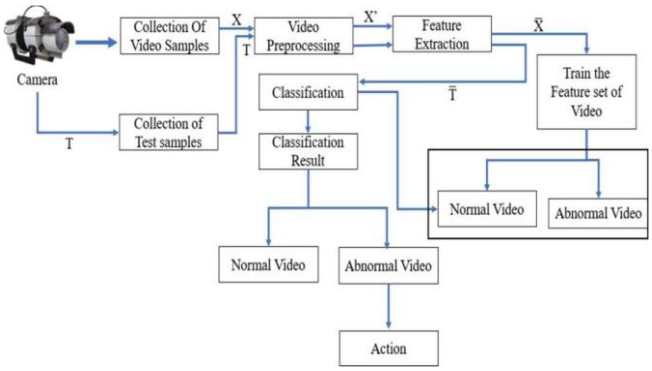

Fig 1. System Architecture

Anomaly detection and classification using the dataset as depicted in the data flow diagram. The information is first pre-handled and cleaned. This is done to get rid of any errors or noise in the data . The anomaly detection algorithm then receives the data. Any out-of-the-ordinary data patterns are found by this algorithm.The distinguished oddities are then given to the recognizable proof calculation. The specific anomalies that have been identified are identified by this algorithm. The user is then presented with the anomaly detection and identification algorithms' outcomes.

- Another name for the DFD is the air pocket outline. A system can be represented graphically using this simple formalism in terms of the input data it receives, the different processing it does on this data, and the output data it generates.
- It is used to demonstrate the components of the framework. These components include the framework cycle, the data streams within the framework, an external substance that interacts with the framework, and the information used by the interaction.
- DFD illustrates how information is changed as it passes through the system through a number of transformations. The data stream and the adjustments made as information flows from contribution are shown graphically.

Moreover referred to as a bubble chart, DFD Any level of reflection on a framework can be addressed with a DFD.

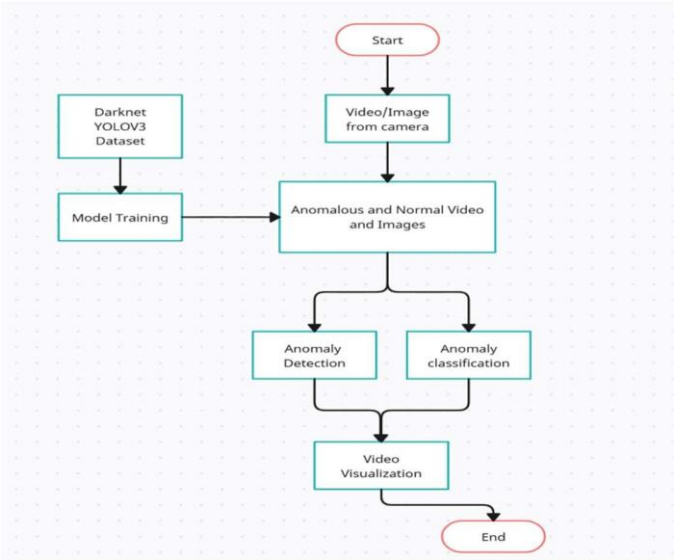

Fig 2. Data Flow Diagram

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

VII. METHODOLOGY

The current review sought to determine how accurate reconnaissance cameras are at identifying bizarre behavior. The study looked at how well the cameras could spot people in the area being monitored who weren't acting the way the script said they should. The classifier will really need to portray the characteristic activities for faster disclosure.

Here we are making use of Median Filter which is a Noise Removal Technique. The median filter is a non-linear digital filtering technique, often used to remove noise from an image or signal. Here 0's are appended at the edged and corners to the matrix which is the representation of the grey scale image. Then for every3*3 matrix, arrange elements in ascending order, then find median/middle element of those 9 elements.

The CIFAR-100 dataset is a well-known dataset used in machine learning and computer vision research. Here are some key details:

Composition: It consists of 60,000 32x32 color images divided into 100 classes 12. Each class contains 600 images.
Labels: Each image has a "fine" label (specific class) and a "coarse" label (superclass). The 100 classes are grouped into 20 superclasses.

Training and Testing: There are 50,000 training images and 10,000 test images
To demonstrate how standard neural networks do not scale well as image size increases, let's again consider the CIFAR-10 dataset. Each image in CIFAR-10 is 32×32 with a Red, Green, and Blue channel, yielding a total of 32×32×3 = 3,072 total inputs to our network.

A total of 3,072 inputs does not seem to amount to much, but consider if we were using 250×250-pixel images — the total number of inputs and weights would jump to 250×250×3 = 187,500 — and this number is only for the input layer alone! Surely, we would want to add multiple hidden layers with a varying number of nodes per layer — these parameters can quickly add up, and given the poor performance of standard neural networks on raw pixel intensities, this bloat is hardly worth it.

we have been tested in CIFAR-100 dataset as well as DCSASS Dataset in which the dataset contains videos based on following 13 classes that is abuse, arrest,arson, assault, accident, Burglary , fighting, robbery, shooting, stealing, shoplifting and Vandalism. Each video is labeled as normal and abnormal, normal is indicated as 0, abnormal is indicated as 1. lastly the distribution of this dataset is as follows. There is a total of 16853 videos were 9676 video as labeled as normal, and 7177 as abnormal.

- Modules used in methodology

STEPS INVOLVED IN METHODOLOGY

Step 1: Module for Data Collection: For detecting anomalies in surveillance videos, our method outperforms conventional machine learning methods, as demonstrated by our experiments.

Step 2: Pre-processing: The data Pre-processing data is necessary for any machine learning algorithm to function effectively [17].

Step3: Dividing the information into train and test information: Identification of atypical articles or ways of behaving in pictures and recordings gathered from reconnaissance cameras is a significant assignment for security and well being purposes. The test in this errand is to precisely distinguish uncommon articles or ways of behaving involving the LSTM model while disregarding typical varieties in the picture or video information.

Step 4: Casing Extraction: The process of finding and removing objects from a video sequence that do not behave normally or follow a pattern is known as anomalous detection. The proposed approach is assessed on an observation dataset and results show that the proposed approach accomplishes improved results contrasted with conventional techniques.

Step 5: Brutality expectation and location Anomalies can be detected by this model at a rate of 131 frames per second, which is faster than that of other models. Using machine learning techniques, we present a method for spotting unusual behavior in surveillance footage.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

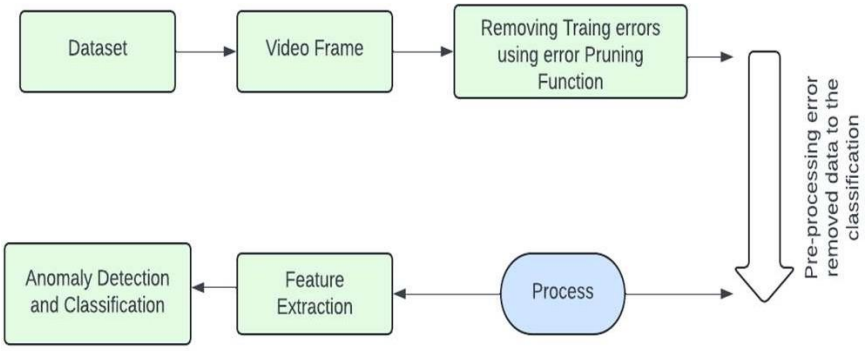

Fig 3. Steps Involved In Methodology

- *CNN Algorithm*

In this part, we portray in short the CNN model. CNN is a convolution brain organization. Its assignment is to separate the significant highlights in the picture. Profound learning comprises of three fundamental layers: the convolution layer, pooling layer, and completely associated layer. CNN incorporates many layers: convolutional layer, max_x0002_pooling layer [2], straightening layer, and full association layer, as displayed in Figure 4. Number of hidden layers = 3, Filter Size=224,224 and Activation Function used is Softmax

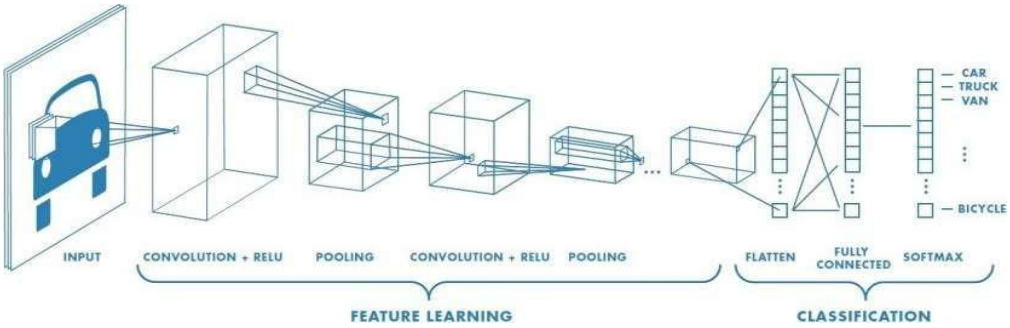

Fig 4. Neural Network With Numerous Convolutional Layer [21].

1) **The Convolutional Layer:** The activation role of the convolutional layer is a non-linear function. It has a few kinds; the initiation capability is generally normally utilized. The most usually utilized them are: **ReLU (redressed straight unit)** Its significance is diminishing the quantity of records performed. **Sigmoid**, which is utilized in the result layer.

$$f(x) = \frac{1}{1 + e^{-(x)}}$$

2) **Max-pooling layer:** It gathers the elements extricated from the picture, lessens the aspects, and concentrates the main highlights present in the picture

3) **Flattering layer:** The flattening layer transforms the max-pooling characteristics into a one-dimensional matrix.

4) **Completely associated layer:** it assembles every among the neurons.

- *HAAR CASCADE FACE LANDMARKS*

Presently all potential sizes and areas of every bit is utilized to ascertain a lot of elements. (Can you just imagine how much computing power it requires? Over 160000 features result from even a 24x24 window). For each element computation, we really want to track down number of pixels under white and dark square shapes. To address this, they presented the basic pictures [8]. It reduces the operation of calculating the sum of pixels, or the maximum possible number of pixels, to one involving only four pixels. Isn't it nice? It makes things super-quick.

Multi-scale feature extraction is a technique used in convolutional neural networks (CNNs) to capture and process information at multiple scales or resolutions within an image. This approach is particularly useful in complex surveillance scenarios where objects or activities of interest can vary significantly in size. By using multi-scale feature extraction, the

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

network can effectively detect both small and large objects, as well as those that may be partially occluded or overlapping. However, most of the features we calculated are irrelevant. Take, for instance, the picture below. Two useful features are displayed in the top row. The principal highlight chose appears to zero in on the property that the locale of the eyes is frequently more obscure than the area of the nose and cheeks. The second characteristic that was considered is the fact the darker eyes than the bridge of the nose. However, it is irrelevant to apply the same windows to cheeks or any other location. So *how would we choose the best highlights out of 160000+ elements?* Adaboost achieves this.

So this is a straightforward instinctive clarification of how Viola-Jones face recognition functions. Peruse paper for additional subtleties or look at the references in Extra Assets area.

Viola-Jones algorithm:

a. Haar features, or simple rectangular shapes, are known. This digital picture function can be used to locate people, things, and emotions in images.

b. Integral Image: The idea behind Integral Image enables quick feature recognition. The integral image, an intermediary picture format, enables the rapid calculation of rectangle characteristics. At each pixel (x, y), the integral picture calculates a pixel value in a fast and efficient manner.

c. Yoav Freund and Robert Schapire created the AdaBoost algorithm. To find weak feature selectors and boost their performance, they use this machine learning approach.

d. We use a technique known as cascading classifiers to speed up calculations on areas that resemble faces. This method involves mixing classifiers, which quickly ignore the background windows.

Haar-cascade Detection in OpenCV

OpenCV accompanies a mentor as well as identifier. OpenCV can be used to create a classifier that you can train for any object, including cars, planes, and so on. Its all relevant info are given here: Outpouring Classifier Preparing.

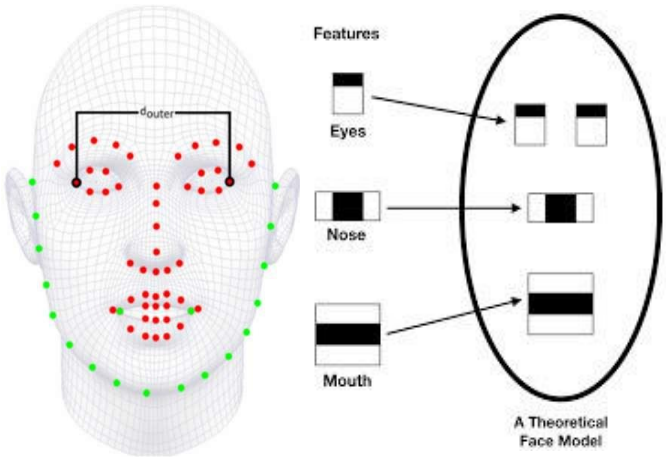

Fig 5. Data Flow Daigramhaar Cascade Landmark For Face Model

- *CNN SLOW-FAST Algorithm*

One of the more well-known Computer Vision tasks is identifying and classifying images of objects. The 2010 ImageNet dataset and competition made this popular. While much headway has been accomplished on ImageNet, an as yet vexing errand is video understanding — dissecting a video fragment and making sense of what's going on within it.

Algorithms today are still far from reaching human-level results, despite recent progress in video understanding. As a result, SlowFast analyzes a video's static content with a slow, high-definition CNN (Fast pathway), while simultaneously analyzing a video's dynamic content with a fast, low-definition CNN (Slow pathway) [7]. The primates' retinal ganglion, in which 80% of the cells (P-cells) operate at a low temporal frequency and recognize fine details and 20% (M-cells) operate at a high temporal frequency and are responsive to rapid changes, serves as a partial model for the method. Likewise, in SlowFast the figure cost of the Sluggish pathway is 4x bigger than that of the Quick pathway.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

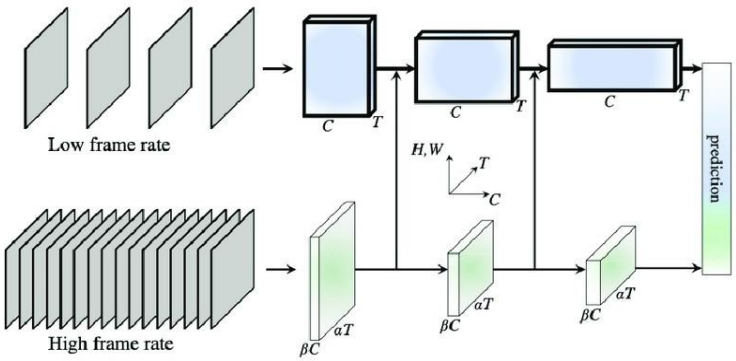

Fig 6. High Level Of Illustration Of Slow-Fast Network [7]

Using a 3D ResNet model, the Slow and Fast pathways simultaneously capture multiple frames and perform 3D convolution operations on them.

The Sluggish pathway utilizes an enormous fleeting step (for example number of edges skipped each second) τ, normally set at 16, considering around 2 examined outlines each second. The Fast pathway makes use of a much shorter temporal stride called τ/α which is usually set to 8 and allows for 15 frames per second. By employing a significantly smaller channel size (i.e., convolution width; number of filters used), typically set at 1/8 of the Slow channel size, the Fast pathway is kept light. The channel size of the Quick pathway is set apart as β. The result of the more modest channel size is that the Quick pathway requires 4x less register than the Sluggish pathway regardless of having a higher transient recurrence.

The model achieves balance between the recall and decision threshold ie..,Threshold adjustment, by varying the weight given the given to certain classes during the training so it reduces the false positive and improving the quality and relevance of feature can reduce frequency of deceptive patterns that result in false positive.

VIII. RESULT & DISCUSSION

Fig 7. “Home Page”A Web Application For Recognizing Suspicious Activity Is Depicted In The Image. It Has Two Sections: "Analyse" For File Uploads And "Live Streaming" With a Streaming Button.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

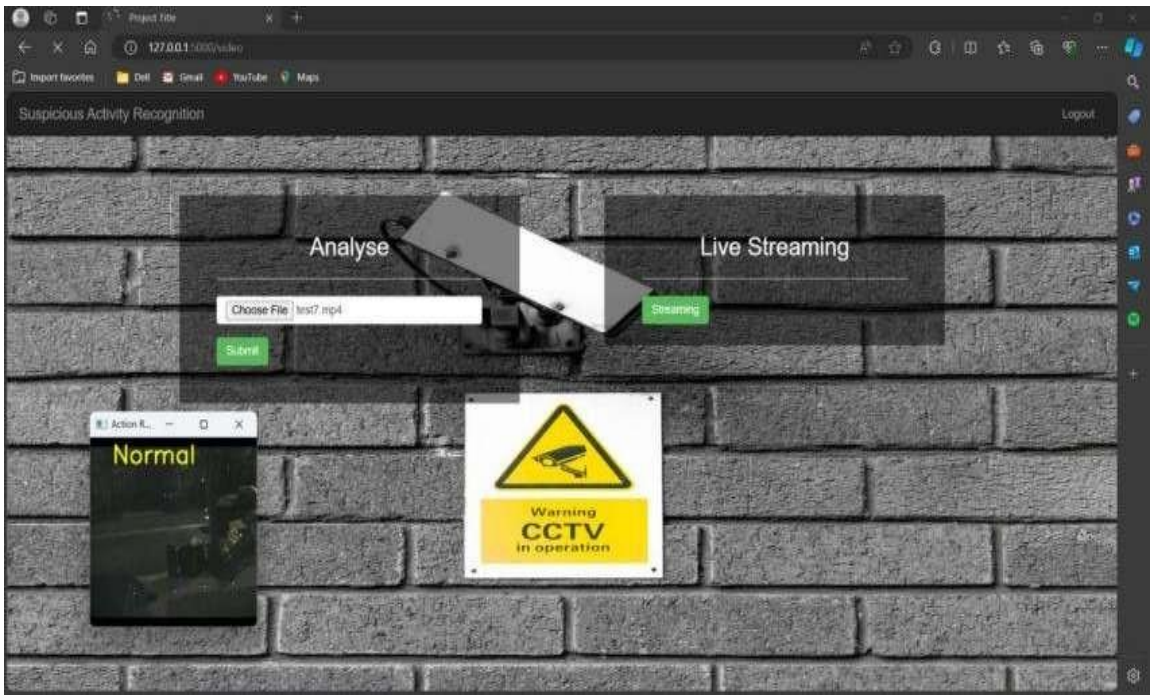

Fig 8. "NORMAL ACTIVITY" The picture shows a web application for Normal movement acknowledgment with segments for document investigation and live streaming

Fig 9. Anomalous Activity The Picture Shows a Web Application For Anomalous Movement Acknowledgment With Segments For Document Investigation And Live Streaming

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

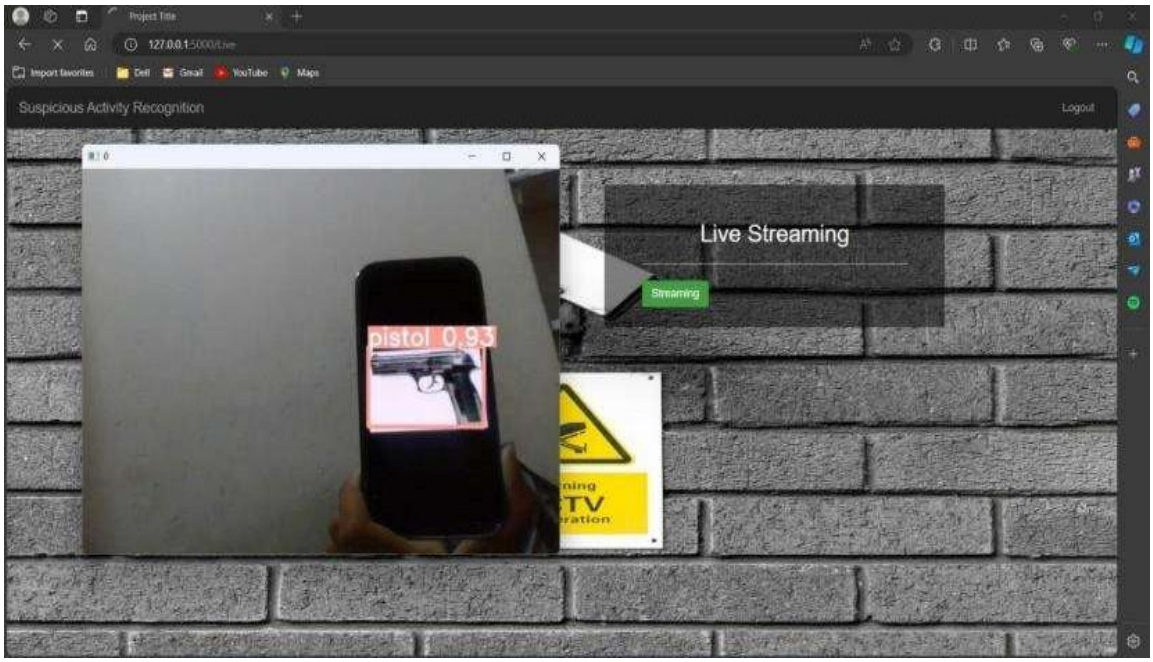

Fig 10. “Live Streaming” Depicting The Real Time Recognition Of a Pistol Using The Application

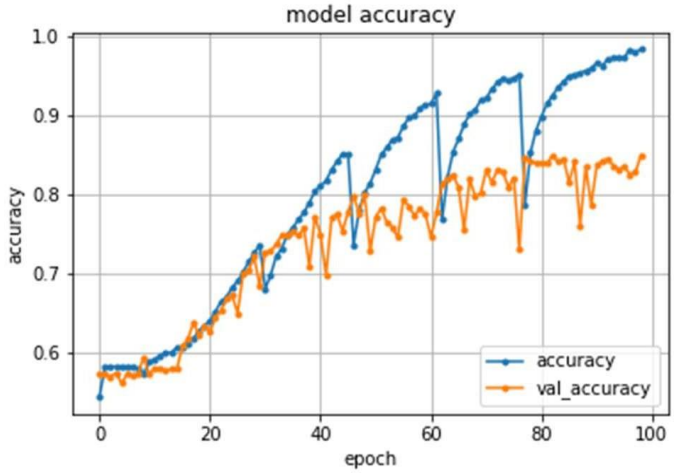

Fig 11. Validation Accuracy Varies And Plateaus,Suggested Possible Overfitting The Training Accuracy Increases

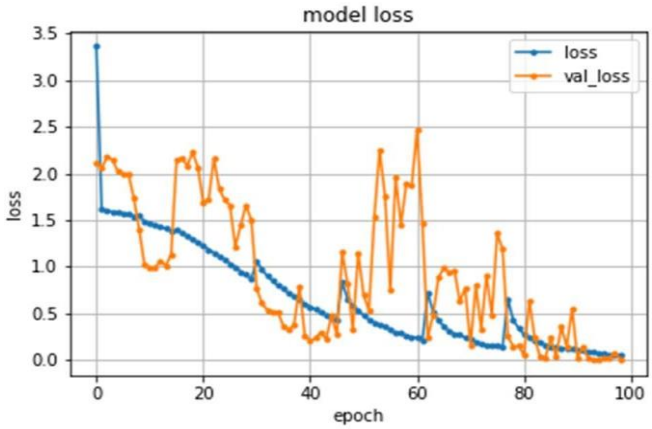

Fig 12. The Model Perform Good On Training Data When Compared To Validation Reliably

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 7, July 2025

DOI: 10.17148/IARJSET.2025.12710

Evaluation Metrics:

I. Accuracy: 0.97
II. Precision: 0.58
III. Recall: 0.98
IV. F1-Score: 0.71

IX. CONCLUSION

To decrease monetary and human misfortunes, this work centers around naturally perceiving suspicious exercises progressively. The CIFAR-100 object detection dataset served as the training ground for the 63-layer CNN network that we propose. This pre-prepared model, named L4-Fanned ActionNet, is utilized to recognize five explicit suspicious exercises. An entropy-coded ACS method is used to further reduce the dimensionality of the features that this network extracts. In order to find the best configuration, we run five experiments with varying numbers of features (100, 250, 750, and 1000) during the feature selection phase. The Whelp SVM classifier, while utilizing 100 highlights, accomplishes an exactness of 0.9844. With 1000 features and a precision of 0.9924, the Whelp SVM classifier achieves the highest performance. All experiments show that the Whelp SVM classifier is the most effective. The robustness and validity of our method are also demonstrated by comparing the findings with those of earlier studies and validating them on the Weizmann dataset. As previous studies have demonstrated promising results, feature integration from other CNN-based pre-trained networks could be the subject of future research. We likewise suggest examining new element determination procedures and profound learning models to additional improve execution in this space. Applying techniques such as genetic algorithms and recursive feature elimination to identify and select the most relevant features from the video can be incorporated to improve the performance.

REFERENCES

[1]. Xu, Dan, Yan Yan, Elisa Ricci, and Nicu Sebe. "Detecting anomalous events in videos by learning deep representations of appearance and motion." Computer Vision and Image Understanding 156 (2017): 117-127.
[2]. Bora, T. Subhash, and Monika Dhananjay Rokade. "Human suspicious activity detection system using CNN model for video surveillance." vol 7 (2021): 2021-2021.
[3]. Tripathi, R. K., Jalal, A. S., & Agrawal, S. C. (2018). Suspicious human activity recognition: a review. Artificial Intelligence Review, 50, 283-339.
[4]. Ouivirach, K., Gharti, S., & Dailey, M. N. (2013). Incremental behavior modeling and suspicious activity detection. Pattern recognition, 46(3), 671-680.
[5]. Ibrahim Salem, F.G., Hassanpour, R., Ahmed, A.A., & Douma, A. (2021). 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, 794-801.
[6]. Zhang, X., Yang, S., Zhang, X., Zhang, W., & Zhang, J. (2018). Anomaly Detection and Localization in Crowded Scenes by Motion-field Shape Description and Similarity-based Statistical Learning. ArXiv, abs/1805.10620.
[7]. Agarwal, M., Parashar, P., Mathur, A., Utkarsh, K., & Sinha, A. (2022). Suspicious Activity Detection in Surveillance Applications Using Slow-Fast Convolutional Neural Network. In Advances in Data Computing, Communication and Security: Proceedings of I3CS2021 (pp. 647-658). Singapore: Springer Nature Singapore.
[8]. An, M. L. (2020). Social Perspective of Suspicious Activity Detection in Facial Analysis. Artificial Intelligence Paradigms for Smart Cyber-Physical Systems, 87..
[9]. Vorapatratorn, S. (2024). Enhancing monitoring of suspicious activities with AI-based and big data fusion. PeerJ Computer Science, 10..
[10]. Mostafa, T.A., Uddin, J., & Ali, M.H. (2017). Abnormal event detection in crowded scenarios. 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), 1-6.
[11]. Amrutha, C. V., Jyotsna, C., & Amudha, J. (2020, March). Deep learning approach for suspicious activity detection from surveillance video. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 335-339). IEEE..
[12]. Marx, G. T. (2002). What's New About the" New Surveillance"? Classifying for Change and Continuity. Surveillance & Society, 1(1), 9-29.
[13]. Acharya, B. R., & Gantayat, P. K. (2015). Recognition of human unusual activity in surveillance videos. International Journal of Research and Scientific Innovation (IJRSI), 2(5), 18-23.
[14]. Gupta, N., & Agarwal, B. B. (2023). Recognition of Suspicious Human Activity in Video Surveillance: A Review. Engineering, Technology & Applied Science Research, 13(2), 10529-10534.
[15]. Karim, M., Khalid, S., Aleryani, A., Tairan, N., Ali, Z., & Ali, F. (2024). HADE: Exploiting Human Action Recognition Through Fine-Tuned Deep Learning Methods. IEEE Access, 12, 42769-42790.

DOI: 10.17148/IARJSET.2025.12710

[16]. Karim, M., Khalid, S., Aleryani, A., Khan, J., Ullah, I., & Ali, Z. (2024). Human action recognition systems: A review of the trends and state-of-the-art. IEEE Access.

[17]. Khalid, S., Wu, S., Alam, A., & Ullah, I. (2021). Real-time feedback query expansion technique for supporting scholarly search using citation network analysis. Journal of Information Science, 47(1), 3-15.

[18]. Yilmaz, Erhan, and Ozgu Can. "Unveiling Shadows: Harnessing Artificial Intelligence for Insider Threat Detection." Engineering, Technology & Applied Science Research 14, no. 2 (2024): 13341-13346.

[19]. Dey, Arnab, Samit Biswas, and Laith Abualigah. "Efficient Violence Recognition in Video Streams using ResDLCNN-GRU Attention Network." ECTI Transactions on Computer and Information Technology (ECTI-CIT) 18, no. 3 (2024): 329-341.

[20]. Arshad, Qurat-Ul-Ain, Mudassar Raza, Wazir Zada Khan, Ayesha Siddiqa, Abdul Muiz, Muhammad Attique Khan, Usman Tariq, Taerang Kim, and Jae-Hyuk Cha. "Anomalous situations recognition in surveillance images using deep learning." Computers, Materials and Continua 76, no. 1 (2023): 1103-1125.

[21]. https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

