A Major Project report submitted in

SUSPICIOUS HUMAN ACTIVITY DETECTION

Partial fulfillment of the requirement for the award of the Degree of

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING

SUBMITTED

By

MA SHAIK SHOYEB	**21675A0506**
AZMATH ALI	**20671A0528**
THALABOINA RAKESH	**20671A0542**
G. VAMSI	**20671A0513**
AHAMMAD ALI	**20671A0501**

Under the esteemed guidance of

Dr. G. SREENIVASULU

ASSOCIATE PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

J.B. INSTITUTE OF ENGINEERING & TECHNOLOGY

UGC AUTONOMOUS

(Accredited by NAAC & NBA, Approved by AICTE & Permanently affiliated by JNTUH)

Yenkapally, Moinabad Mandal, R.R. Dist-75 (TG)

2020-2024

J.B. INSTITUTE OF ENGINEERING & TECHNOLOGY

UGC AUTONOMOUS

(Accredited by NAAC & NBA, Approved by AICTE & Permanently affiliated by JNTUH)

Yenkapally, Moinabad Mandal, R.R. Dist-75 (TG)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the given Major Project report entitled **“SUSPICIOUS HUMAN ACTIVITY DETECTION”** submitted to the Department of Computer Science and Engineering, J.B Institute of Engineering & Technology, in accordance with Jawaharlal Nehru Technological University regulations as partial fulfillment required for successful completion of Bachelor of Technology is a record of bonafide work carried out during the academic year 2023-24 by,

MA SHAIK SHOYEB	**21675A0506**
AZMATH ALI	**20671A0528**
THALABOINA RAKESH	**20671A0542**
G. VAMSI	**20671A0513**
AHAMMAD ALI	**20671A0501**

Internal Guide	**Head of the Department**
Dr. G. SREENIVASULU	Dr. G. SREENIVASULU
ASSOCIATE PROFESSOR	ASSOCIATE PROFESSOR

External Examiner

J.B. INSTITUTE OF ENGINEERING & TECHNOLOGY

UGC AUTONOMOUS

(Accredited by NAAC & NBA, Approved by AICTE & Permanently affiliated by JNTUH)

Yenkapally, Moinabad Mandal, R.R. Dist-75 (TG)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We hereby certify that given Major Project report entitled **“SUSPICIOUS HUMAN ACTIVITY DETECTION”** carried out under the guidance of**, Dr. G. SREENIVASULU, Professor** is submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering.** This is a record of bonafide work carried out by us and the results embodied in this project report have not been reproduced or copied from any source. The results embodied in this project report have not been submitted to any other university or institute for the award of any other degree or diploma.

Date: / /

MA SHAIK SHOYEB	**21675A0506**
AZMATH ALI	**20671A0528**
THALABOINA RAKESH	**20671A0542**
G. VAMSI	**20671A0513**
AHAMMAD ALI	**20671A0501**

ACKNOWLEDGEMENT

At outset we express our gratitude to almighty lord for showering his grace and blessings upon us to complete this Major Project. Although our name appears on the cover of this book, many people have contributed in some form or the other to this project development. We could not have done this Project without the assistance or support of each of the following.

First, we are highly indebted to **Dr. P. C. KRISHNAMACHARY**, Principal for giving us the permission to carry out this Project Stage-I.

We would like to thank **Dr. G. SREENIVASULU,** Associate professor & Head of the Department of COMPUTER SCIENCE AND ENGINEERING**,** for being moral support throughout the period of the study in the Department.

We are grateful to **Dr. G. SREENIVASULU,** Associate professor & Head of the Department of COMPUTER SCIENCE AND ENGINEERING, for his valuable suggestions and guidance given by him during the execution of this Project work.

We would like to thank Teaching and Non-Teaching Staff of Department of Computer Science and Engineering for sharing their knowledge with us.

MA SHAIK SHOYEB	**21675A0506**
AZMATH ALI	**20671A0528**
THALABOINA RAKESH	**20671A0542**
G. VAMSI	**20671A0513**
AHAMMAD ALI	**20671A0501**

TABLE OF CONTENTS

ABSTRACT

With the increasing prevalence of closed-circuit television (CCTV) systems in public and private spaces, the need for effective surveillance and security measures has become paramount. This study presents an integrated approach for the detection of suspicious activities using CCTV footages, aiming to enhance public safety and security. This project leverages advancements in computer vision and machine learning techniques to automatically analyze video streams and identify anomalous behavior. The primary focus is on domestic settings where privacy concerns and the need for accurate detection are particularly crucial. The research involves the development of a robust model capable of recognizing a diverse range of suspicious activities without compromising individual privacy. The methodology includes the collection of a comprehensive dataset, encompassing various normal and suspicious activities in domestic environments. Frame-level and temporal features are extracted using convolutional neural networks (CNNs) for spatial information and recurrent neural networks (RNNs) for temporal dependencies. The model is trained and fine-tuned using advanced deep learning techniques to optimize performance. The proposed system represents a significant step towards the development of intelligent surveillance solutions that balance the imperative of public safety with the importance of individual privacy.

LIST OF FIGURES

1. INTRODUCTION

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

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

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

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

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

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

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

1.1 EXISTING SYSTEM

A semantic based approach was proposed in [1]. Background subtraction was used to identify foreground objects in the collected video data.
Using a Haar-like method, the objects are identified as living or non-living after subtraction. The Real-Time blob matching technique was used to track the objects. This paper also discovered fire detect.

Based on the analysis of movement information from video sequences, a method was created to discriminate abnormal events from regular events. The HMM approach was used to train the histograms of the video frame's optical flow orientations. It compares the acquired video frames to the existing normal frames and determines their resemblance [2].

People tracking could be used to discover unexpected happenings in video footage. Using the background subtraction method, human persons are spotted in the footage. CNN was used to extract the features, which were then fed into a DDBN (Discriminative Deep Belief Network). The DDBN is also given labelled footage of some questionable situations, and their features are extracted. Then, using a DDBN, features derived using CNN were compared to feature recovered from a labelled sample video of categorized suspicious behaviors, and numerous suspicious activities were discovered from the video [3].

To prevent audience or player violence in sports, a real-time violence detection system based on deep learning was developed. Frames from real time videos were retrieved in a spark environment. If the system detects football violence, security personnel will be notified. The system recognizes video behaviors in real time and warns security forces, preventing violence from occurring in the first place. For detecting violence at football stadiums, the VID dataset was employed, and it had a 94.5 percent accuracy. [4]

Different components for visual data processing make up anomalous event detection. Human behavior was detected by using deep architectures. The UT Interaction dataset was used for the proposed CNN and LSTM models. One of the system's flaws was that it was difficult to distinguish between human activities such as pointing and punching. [5].

Using a deep spatiotemporal approach to crowd behavior, the films are divided into three categories: pedestrian future path prediction, destination estimation, and holistic crowd behavior. There are three distinct categories. With the help of a convolutional layer, spatial information from video frames was retrieved. The sequence of temporal motion dynamics was learned or understood using LSTM architecture. PWPD, ETH, UCY, and CUHK were data sets used in the proposed system. By employing deeper architectures, the system's accuracy can be increased. [6]

An unauthorized entry into a restricted location was detected using the Advance Motion Detection (AMD) method [7]. The object was discovered using background subtraction in the first phase, and the object was retrieved from frame sequences in the second phase. Suspicious activity was discovered in the second phase. The system's advantage was that the method worked in real time and was computationally simple. However, the system's storage capacity was restricted, and it could be used in surveillance regions with a high-tech form of video capture.

For recognizing human behavior analysis from movies, the majority of the publications described above used computer vision employing various algorithms or neural networks. To extract the motion pattern and interpret the evolution of characteristics in a video series, computer vision technologies require a lot of pre-processing. In addition, background removal is based on a static backdrop hypothesis that is frequently inapplicable in
real-world circumstances. The majority of problems in the actual world arise in crowds. When it comes to crowd management, the approaches mentioned above are inefficient. Based on the literature review, a deep architecture for suspicious activity prediction can be modelled using CNN and DDBN, improving the system's accuracy. In deep learning approach, most of the papers detect only the suspicious activity. So an efficient mechanism is
needed to alert the security in the case of any suspicious behavior.

1.2 PROPOSED SYSTEM

In our proposed system, for detecting anomalous behavior, LRCN (Long-term Recurrent Convolutional Network) has 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. 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.
Sequence of 30 frames of the video are extracted and passed to the LRCN Model.

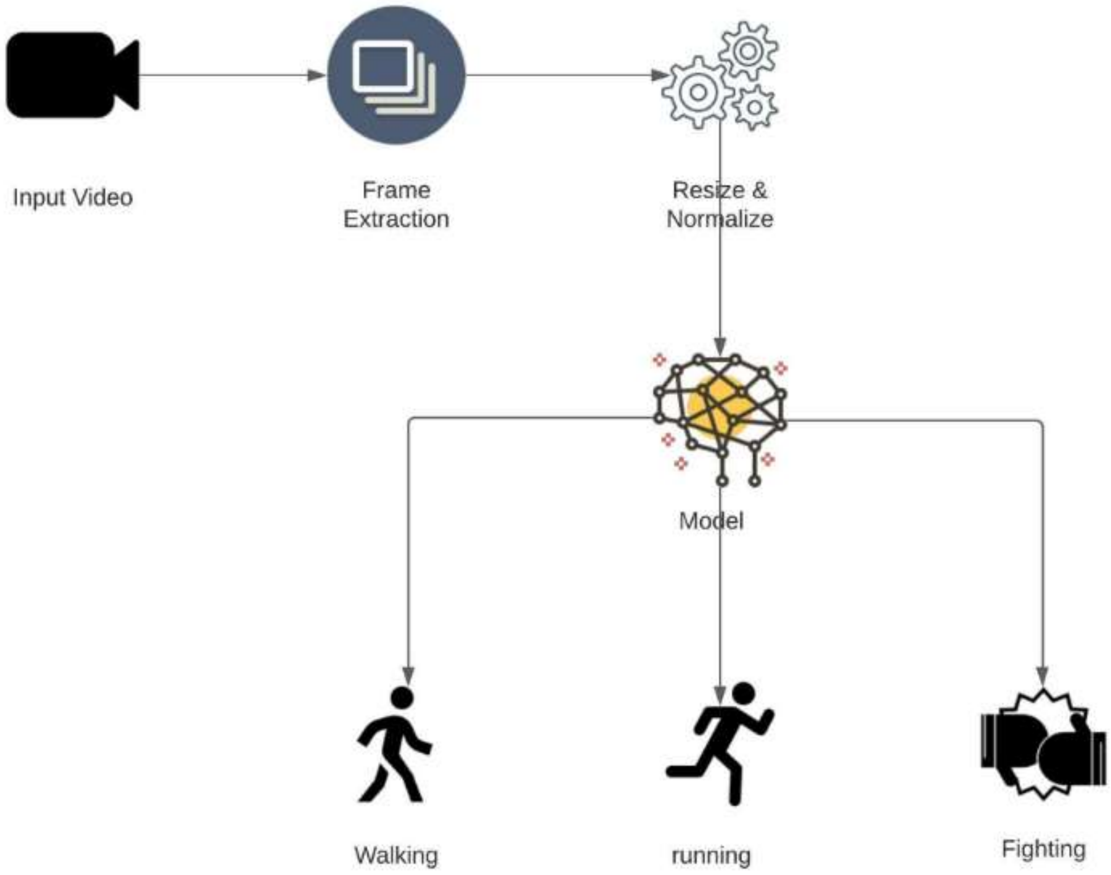

Fig 1.2: Work flow of proposed system

2. LITERATURE SURVEY

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

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

[2] proposed a semantic approach. The captured video data is processed and foreground objects are identified by background subtraction. After segmentation, objects are classified as living or non-living according to the Haar algorithm. Object tracking is done using a Real-Time blob matching algorithm. Fire is found in this paper. Based on the characteristics of movement between objects, [3] suspicious activity is detected. A semantic approach is used to identify suspicious events. Object detection and correlation techniques have been used for object tracking [2]. Events based on motion characteristics and temporal data. The computational complexity of the given framework is low.

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

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

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

Unusual events in video recording can be found by tracking people. People are detected by removing the background from the video. Features are extracted using CNN and fed to DDBN (Discriminatory

Deep Belief Network). Tagged videos of various suspicious incidents are provided to DDBN and their features are also extracted. Then, the comparison of features extracted using CNN and features extracted from videos of well- known samples of hidden suspicious activity was done using DDBN, and various suspicious activities were detected from the given video [6].

A violence detection system using deep learning has been developed to prevent crowd or player violence in sports. Frames are captured from real videos in the Spark environment. Alert security staff if the system detects football violence. To prevent violence, the system detects video movements in real time and alerts security

forces. The VID dataset was used and obtained an accuracy of 94.5% to detect violence in football stadiums [7]. Anomaly detection consists of different modules for video data processing. Deep architecture has been used to explain human behavior. The Interaction UT database is used in the proposed CNN and LSTM based model. One

of the weaknesses of the system is that it is difficult to detect human behavior such as pointing or tapping [8]. Understanding crowd behavior using a deep spatiotemporal approach divides video into the prediction of pedestrians' future paths, destination prices, and crowd behavior. Three different categories. Spatial information in video frames is extracted using convolution layers. LSTM architecture has been used to study or understand the sequence of temporal motion dynamics. The data sets used in the proposed system are PWPD, ETH, UCY and CUHK. The accuracy of the system can be improved by using a deeper architecture [9]. Human daily activities are captured from videos and classification of those videos into household, work, caring and helping images. This is done through deep learning about sports. CNN is used for input features and RNN for classification purposes. They use Inception v3 model with UCF101, Activitynet as database. The achieved accuracy is 85.9% in UCF101 and 45.9% in Activitynet [10].

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

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

3. SOFTWARE REQUIREMENTS ANALYSIS

3.1 HARDWARE REQUIREMENTS:

• Processor	Dual Core
• Speed	3.1 GHz
• RAM	8 GB or more
• Hard Disk	200 GB

3.2 SOFTWARE REQUIREMENTS:

- **Operating Systems:** Windows 7 SP1 or later (64-bit), x86-64 based.
- **Tools:** Jupyter notebook, google colab.

4. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

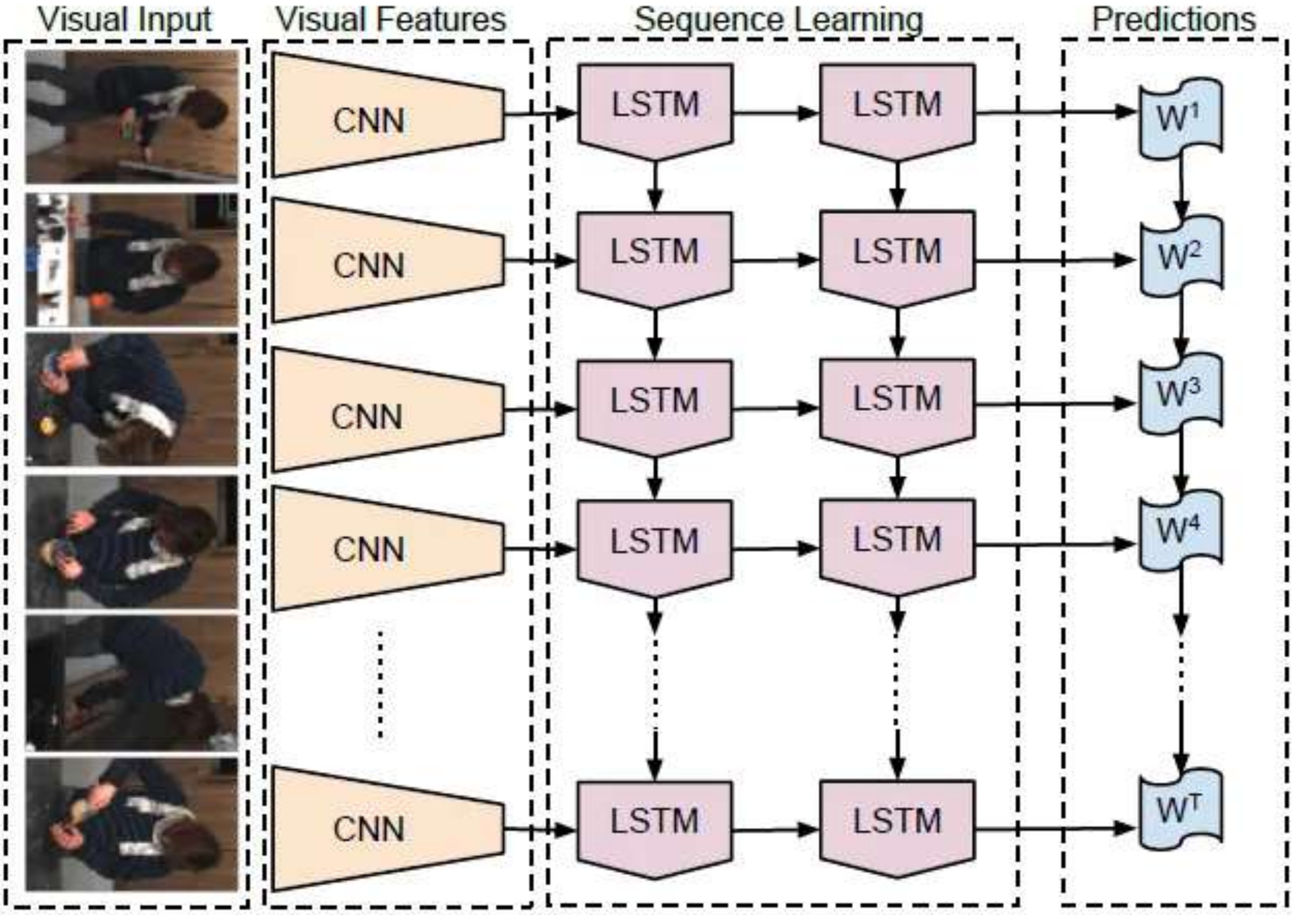

Fig 4.1: Long-term Recurrent Convolutional Network
Source: (https://sh-tsang.medium.com/brief-review-lrcn-long-term-recurrent-convolutional-networks-for-visual-recognition-and-9542bc7e8a79)

LONG-TERM RECURRENT CONVOLUTIONAL NETWORK (LRCN)

This project makes use of (LRCN) to identify unusual behavior.

Recognizing the temporal information in the video is vital for accurate classification of anomalous behaviors. CNN has recently been utilized mostly for extracting important characteristics from each video frame. It is vital for CNN to extract the features in order to effectively categorize the input; as a result, CNN must be able to recognize and extract the required features from the video frames.

The LRCN model is shown in Figure along with the various scaling options that can be used to increase the accuracy of a deep learning model. These options include baseline and scaling like width,

deep resolution,
compound.

Primary principle of LRCN is to combine CNN model to learn visual characteristics from frames and LSTM to convert a series of images into a class label, phrase, and probability. As a result, unprocessed visual input is first sent through a CNN, then fed output to the recurrent sequence model.

In time series data, there can be random intervals between occurrences, and LSTM networks are suitable for making predictions, categorizing, and processing based on this type of data. LSTMs have addressed the gradient problem that can arise during the training of conventional RNNs.

CONVOLUTIONAL NEURAL NETWORK (CNN)

A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

LONG-SHORT TERM MEMORY (LSTM)

Standard Recurrent Neural Networks (RNNs) suffer from short-term memory due to a vanishing gradient problem that emerges when working with longer data sequences.

Luckily, we have more advanced versions of RNNs that can preserve important information from earlier parts of the sequence and carry it forward. The two best-known versions are **Long Short-Term Memory (LSTM)** and Gated recurrent units (GRU**).**

The Long Short-Term Memory, as it was called, was an abstraction of how computer memory works. It is "bundled" with whatever processing unit is implemented in the Recurrent Network, although outside of its flow, and is responsible for keeping, reading, and outputting information for the model. The way it works is simple: you have a linear unit, which is the information cell itself, surrounded by three logistic gates responsible for maintaining the data. One gate is for inputting data into the information cell, one is for outputting data from the input cell, and the last one is to keep or forget data depending on the needs of the network.

Thanks to that, it not only solves the problem of keeping states, because the network can choose to forget data whenever information is not needed, it also solves the gradient problems, since the Logistic Gates have a very nice derivative.

4.1.1USE CASE DIAGRAM

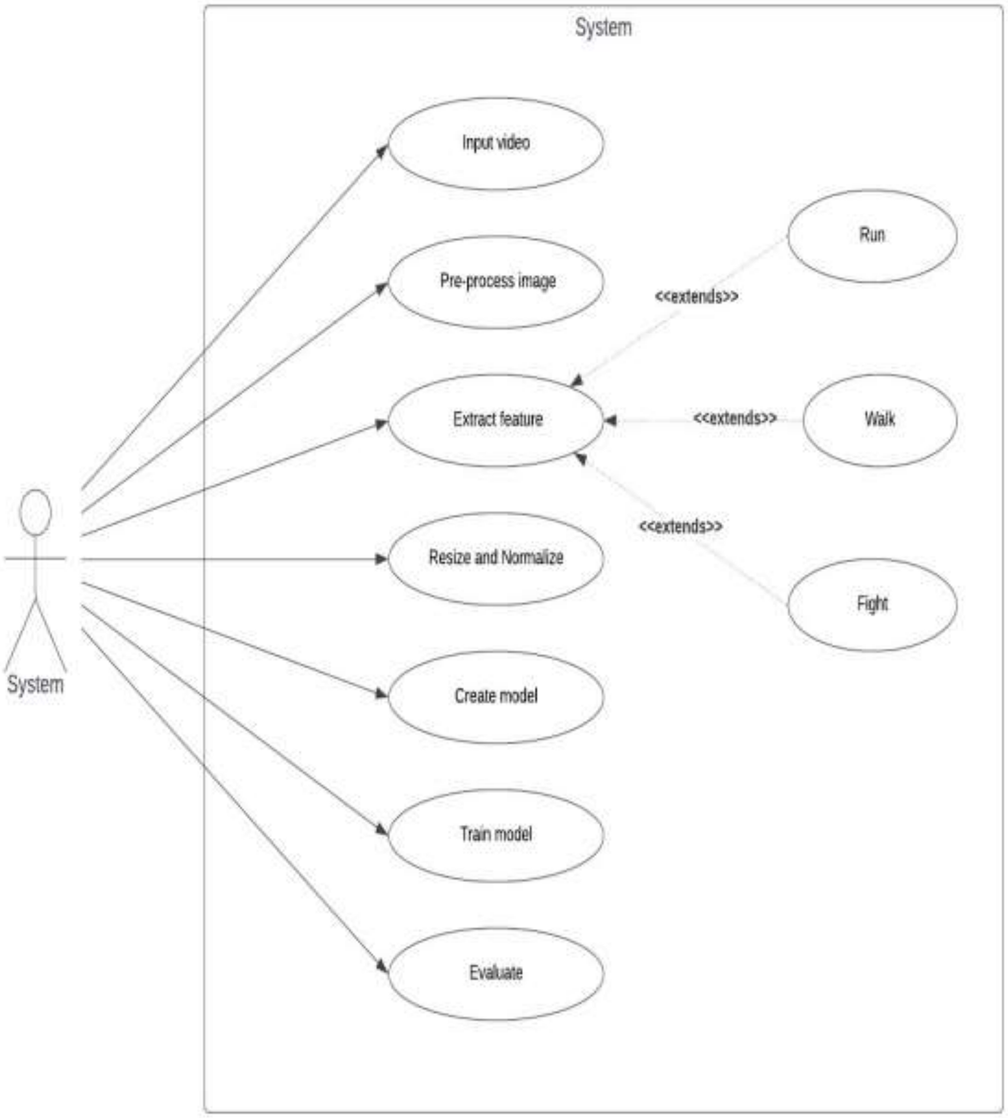

Fig 4.1.1: Use case diagram

Use case diagram is to capture the dynamic aspect of a system. However, this definition is too generic to describe the purpose, as other four diagrams (activity, sequence, collaboration, and State chart also have the same purpose. A use case involves the interaction of actors and the system or other subject. An actor represents a coherent set of roles that users of use cases play when interacting with these use cases. Actors can be humans, or they can be automated systems. The actors are outside the boundary of the system, whereas the use cases are inside the boundary of the system.

4.1.2 CLASS DIAGRAM

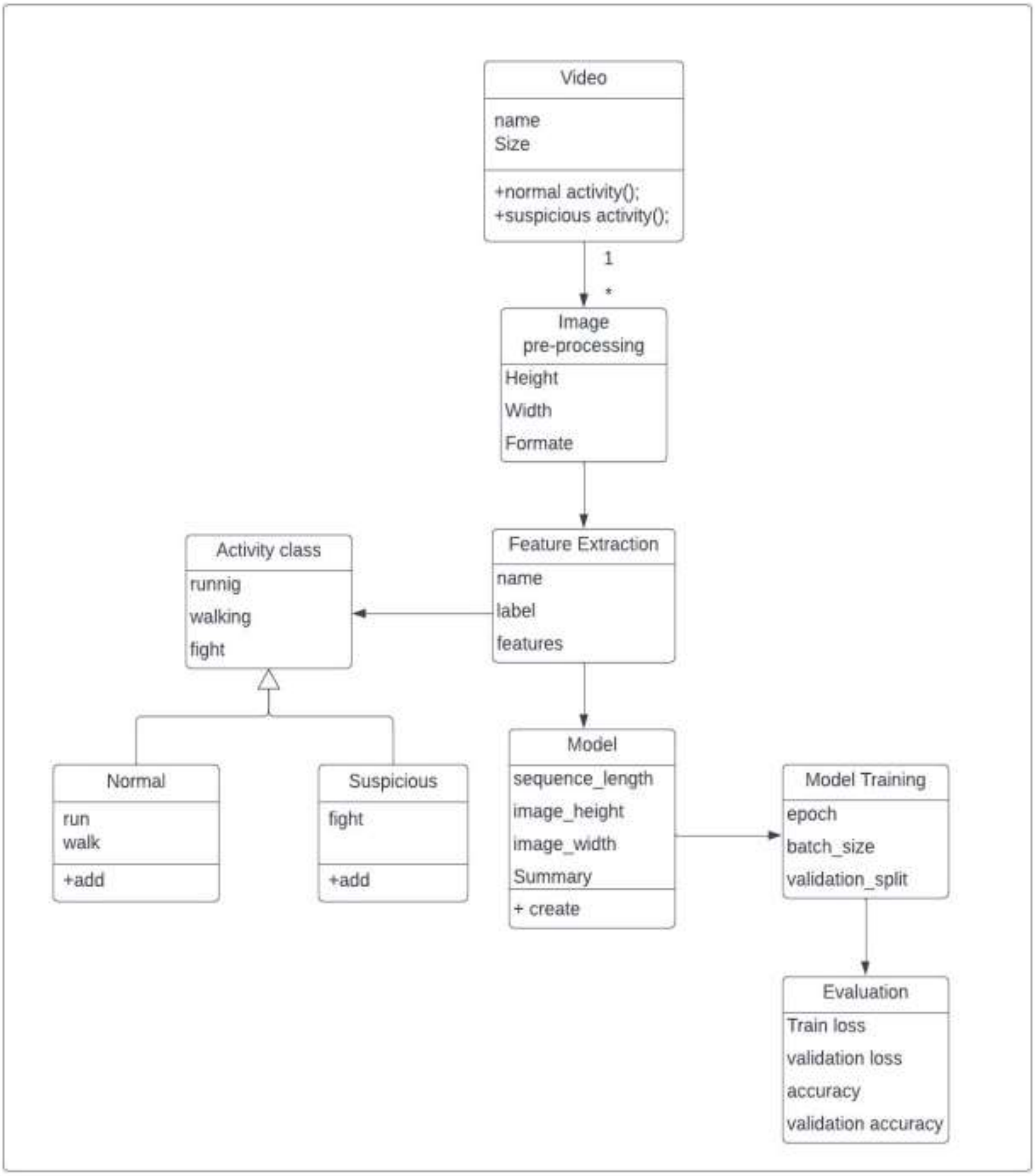

Fig 4.1.2: Class Diagram

A Class diagram is to model the static view of an application. Class diagrams are the only diagrams which can be directly mapped with object-oriented languages and thus widely used at the time of construction.

Modeling a system involves identifying the things that are important to your view. These things form the vocabulary of the system you are modeling. A class is an abstraction of the things that are a part of your vocabulary. A class is not an individual object, but rather represents a whole set of objects. The class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints.

4.1.3 STATE DIAGRAM

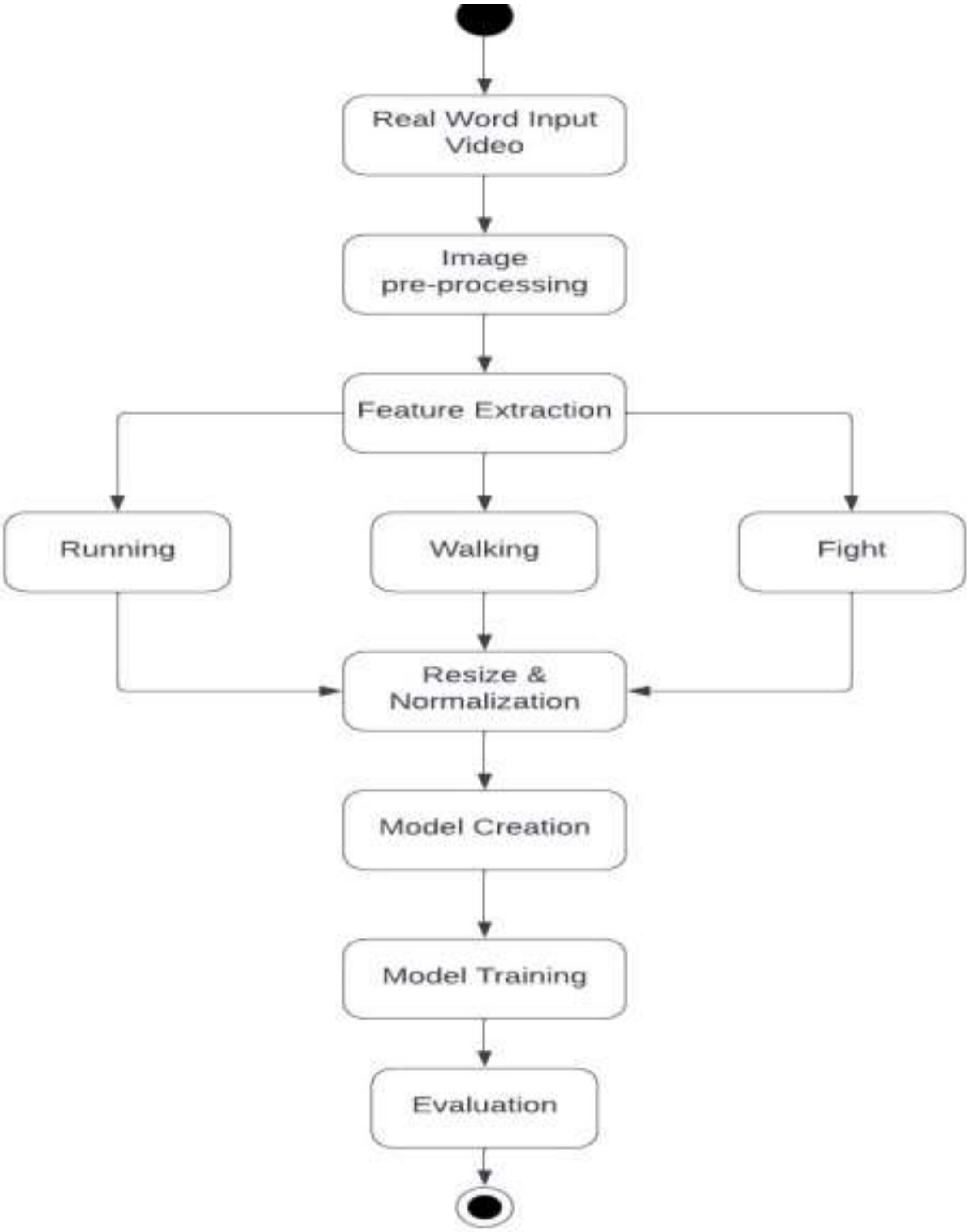

Fig 4.2.3: State Diagram

5. REFERENCES

1. https://medium.com/mlearning-ai/suspicious-human-activity-detection-95b870dae688

2. https://www.csc.kth.se/cvap/actions/

3. https://www.kaggle.com/datasets/odins0n/ucf-crime-dataset

4. https://www.kaggle.com/datasets/naveenk903/movies-fight-detection-dataset

5. P.Bhagya Divya, S.Shalini, R.Deepa, Baddeli Sravya Reddy,“Inspection of suspicious human activity in the crowdsourced areas captured in surveillance cameras”,International Research Journal of Engineering and Technology (IRJET), December 2017.

6. Jitendra Musale,Akshata Gavhane, Liyakat Shaikh, Pournima Hagwane, Snehalata Tadge, “Suspicious Movement Detection and Tracking of Human Behavior and Object with Fire Detection using A Closed Circuit TV (CCTV) cameras ”, International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 5 Issue XII December 2017.

7. Elizabeth Scaria, Aby Abahai T and Elizabeth Isaac, “Suspicious Activity Detection in Surveillance Video using Discriminative Deep Belief Netwok”, International Journal of Control Theory and Applications Volume 10, Number 29 -2017.

8. Dinesh Jackson Samuel R,Fenil E, Gunasekaran Manogaran, Vivekananda G.N, Thanjaivadivel T , Jeeva S , Ahilan A, “Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM”,The International Journal of Computer and Telecommunications Networking,2019.

9. Kwang-Eun Ko, Kwee- abnormal behavior detection in a smart surveillance system."Engineering Applications of Artificial intelligence ,67 IEEE Transactions on multimedia, Vol. 20, NO. 12, December 2018. [7] P. Bhagya Divya, Shalini, R. Deepa, Baddeli Sravya crowdsourced areas captured in surveillance cameras", International Research Journal of Engineering and Technology (IRJET), December 2017.

10. Tian Wanga, Meina Qia, Yingjun Deng, Qi Lyua, Hichem Snoussie, eventdetection based on analysis of movement information -Optik, vol152, January-2018.

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

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

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

14. C. V. Amrutha, C. Jyotsna, J. Amudha (2020) Deep learning Approach for suspicious activity detection from surveillance video, Publisher IEEE Bangalore www.ieeexplore.ieee.org/document/9074920 (Original work published 2020).

15. Sik-Ho Tsang (2022) LRCN: Long-term Recurrent Convolution Networks[Python] https://sh-tsang.medium.com/brief-review-lrcn (Original work published on 2022.

16. Dinesh Jackson, suspicious activity detection in surveillance video using discriminative deep belief network, International Journal of Control Theory and Applications, Volume 10, Number 29 - 2017.