Anomaly Detection in Human Activity with LRCN

Paras Sharma1, Shaswat Harsh2, Prabakaran J3,   
1Department of Networking and communications Faculty of Engineering and Technology, College of Engineering and Technology, SRM Institute of Science and Technology, Chennai, Tamil Nadu, 603203, India, sq0207@srmist.edu.in  
2Department of Networking and communications Faculty of Engineering and Technology, College of Engineering and Technology, SRM Institute of Science and Technology, Chennai, Tamil Nadu, 603203, India, sh7973@srmist.edu.in  
3Department of Networking and communications Faculty of Engineering and Technology, College of Engineering and Technology, SRM Institute of Science and Technology, Chennai, Tamil Nadu, 603203, India, prabakaj@srmist.edu.in

*Abstract*—Applications such as industrial safety, surveillance, and healthcare all heavily rely on anomaly detection in the field of human actions. Customized features and heuristic principles are frequently used in traditional approaches, which limits their adaptability to a variety of dynamic and varied situations. Long-term Recurrent Convolutional Networks (LRCNs) are investigated in this research as potential tools for anomaly identification in human activity. The Long Short-Term Memory (LSTM) network model of temporal relationships and the Convolutional Neural Network (CNN) capacity to capture spatial data are combined in LRCN. LRCN can be utilized to identify irregularities in human behavior. Our proposed method concentrates on the model architecture and feature representation. Four distinct categories—fighting, running, gunshot, and covering camera—are identified by the final model as suspicious activities. According to experimental results, the suggested LRCN-based strategy performs better than traditional methods in detecting abnormalities, with an accuracy of 98.86%.

Index Terms—Surveillance, Video Understanding, Suspicious Behavior, Threat Detection.

# Introduction

The field of human activity recognition has gained significant importance and has numerous applications in diverse fields such as security, healthcare, and surveillance. Effective anomaly detection in human activity is one of them, and it is becoming more and more necessary since it can stop or lessen unfavourable situations like fights, explosions, shootings, and burglaries. The Long Short-Term Memory Recurrent Convolutional Network (LRCN) is introduced as a potent tool for this goal and gives an overview of the difficulties and importance of anomaly detection in human activity. Finding odd or unexpected patterns in data is the process of anomaly detection, which is very useful in situations where it's important to preserve the normal flow of events. In the context of human activity recognition, anomalies can be characterized by activities that deviate from the expected behaviour, often indicative of harmful or dangerous situations. The primary aim is to detect and respond to these anomalies in real-time or post-event analysis to ensure the safety and security of individuals and their surroundings.

The necessity for precise and reliable anomaly detection systems is emphasized by the five distinct types of human activities: shooting, explosion, fighting, accidents, and burglaries. The potentially fatal scenarios portrayed in these activities can happen in a variety of locations, such as public areas, homes, industrial sites, and more. For prompt response, public safety, and potential harm prevention, these activities must be accurately identified and categorized. Because human behaviour is inherently complicated and variable, creating anomaly detection systems for human activity presents several significant obstacles. Activities vary in the postures, motions, and interactions that people can perform; the activities under investigation are no exception to this complexity. Fighting, for instance, can take many different forms, from heated arguments to physical aggression, each with its own. In a similar vein, falls, collisions, and other unintentional events with unique characteristics can occur in accidents. It is therefore a difficult and multifaceted task to develop a system that can reliably discern between typical and abnormal human behaviour in a variety of scenarios.

The advancement of deep learning and neural network architectures has provided promising solutions to the challenges of human activity recognition and anomaly detection. The Long Short-Term Memory Recurrent Convolutional Network (LRCN), has drawn interest due to its capacity to extract temporal and spatial information from videos, making it especially well-suited for this kind of work. In order to extract spatial features from video frames and model temporal dependencies within the sequences, LRCN combines Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. With this combination, LRCN can identify and learn intricate relationships and patterns in video data. We investigate the use of LRCN for anomaly detection in human activity in this study. With a focus on the five categories of burglary, shooting, explosion, accident, and fighting, we investigate how this hybrid architecture can be trained to distinguish between typical and anomalous activities. Utilizing LRCN. By using LRCN, we aim to achieve a higher level of accuracy and robustness in detecting these activities, thereby contributing to enhanced safety and security measures in various environments.

# Literature Survey

1. In 2023, Aniruddha Prakash Kshirsagar and H. Azath addressed the substantial challenge for computer vision technology, particularly in the context of human activity recognition (HAR) in the paper “YOLOv3-based human detection and heuristically modified-LSTM for abnormal human activities detection in ATM machine”. Computer vision technology faces a challenge, especially when it comes to human activity recognition (HAR). The primary objective of this research was to create a robust Abnormal Human Activity Recognition (HAR) model in response to suspicious human activity detected by a video surveillance system inside an ATM. It was made easier to distinguish between normal and abnormal activity in the study by using an improved Long Short-Term Memory (LSTM) model, a cutting-edge deep learning technique, and optimising various LSTM parameters.This work presents a new hybrid optimisation algorithm called Hybrid Spider Monkey-Chicken Swarm Optimisation (HSM-CSO). This was a noteworthy invention. This algorithm was crucial for improving the performance of the deep learning-based classification process. The developed ATM HAR model has demonstrated benefits in terms of improving system performance, reducing false alarms, and mitigating illegal activity. These findings demonstrate the model's capacity to increase security.
2. Pedestrian detection, tracking, and suspicious activity recognition are three increasingly important domains in computer vision. In 2023, in the published paper titled “Real-Time Deep Learning Approach for Pedestrian Detection and Suspicious Activity Recognition”, Yogesh Golhar, Kamal Hajari, and Ujwalla Gawande launched the ground-breaking research project. Growing security concerns have made these issues more urgent in recent years. Particularly in densely populated areas, it can be very difficult to maintain constant security and surveillance in both public and private spaces. This is driving up the demand for active video surveillance systems that can monitor pedestrian behaviour in real time. This work presents a novel and robust deep learning system using a special pedestrian dataset that covers a variety of scenarios, including exam fraud, lab equipment theft, student conflicts, and potentially dangerous learning environments. A comparative analysis of the detection, tracking, and identification of suspicious activity in pedestrians using state-of-the-art deep learning techniques is conducted using this benchmark dataset. This work has the potential to advance computer vision significantly and enhance security measures in densely populated areas.
3. In “Suspicious Activity Recognition Using Proposed Deep L4-Branched-Actionnet With Entropy Coded Ant Colony System Optimization” paper, Tanzila Saba, Amjad Rehman, Rabia Latif, Suliman Mohamed Fati, Mudassar Raza, and Muhammad Sharif made a significant contribution to the field of intelligent visual surveillance systems in 2021. Researchers and industry stakeholders have shown a great deal of interest in this field. The introduction of intelligent visual surveillance systems—whose primary objective is to protect people in public and residential spaces—has been largely fueled by the emergence of smart security cameras, which are outfitted with improved processing powers. The main goal of this effort was to create a reliable technique for identifying questionable activity in surveillance settings. In order to accomplish this, the authors suggested an intricate 63-layer deep Convolutional Neural Network (CNN) model, appropriately called "L4-Branched-ActionNet." While incorporating four branched sub-structures, this novel CNN structure was inspired by the architectural foundation of AlexNet, resulting in a potent framework for feature extraction and recognition. The CIFAR-100 object detection dataset and the SoftMax function were used in conjunction with a pre-trained framework to initially refine the model. Then, deep features were extracted from the dataset used to identify suspicious activity using the pre-trained model. After the features were obtained, a thorough feature subset optimisation procedure was carried out, starting with entropy-based coding. An Ant Colony System (ACS) optimisation technique was used to further improve the coded features' efficiency. Following configuration, the features were fed into several classification models, such as Support Vector Machines (SVM) and k-Nearest Neighbours (KNN).
4. The complex problem of anomalous event detection in surveillance videos was tackled in the field of image and video processing by M. Z. Zaheer, A. Mahmood, H. Shin, and S. -I. Lee in 2020 in “A Self-Reasoning Framework for Anomaly Detection Using Video-Level Labels” paper. Their research highlighted the importance of video-level annotations, which are both easy to obtain and likely to contain significant noise. Using only video-level labels for training, they developed a novel weakly supervised anomaly detection framework that makes use of deep neural networks. With this method, the intrinsic noise in the annotations of anomalous videos was effectively reduced by producing pseudo labels through binary clustering of spatiotemporal video features. By promoting cooperative learning between the main network and the clustering mechanism, the suggested framework improved theaccuracy in identifying anomalies. The framework's superiority over current state-of-the-art methods was demonstrated by the authors' evaluation of it using well-known real-world anomaly detection datasets such as UCF-crime, ShanghaiTech, and UCSD Ped2. The potential of self-reasoning and the usefulness of video-level annotations in tackling the difficult issue of anomaly detection in surveillance footage are illuminated by this work, and the implications for real-world image and video processing applications are encouraging.
5. D. Martínez, H. Loaiza, and E. Caicedo in the paper “Algorithm For Early Threat Detection By Suspicious Behavior Representation” presented a ground-breaking early detection algorithm in 2020. It was based on the persuasive idea that reducing the response time for warning alarm generation improves the effectiveness of controlling criminal activity. The study concentrated on creating a representation model for videos that could be used to characterise questionable behaviour using simple actions. The importance of this model rested in its capacity to detect possible threats prior to suspects making physical contact with their intended targets. An innovative technique was incorporated into the algorithm to optimise the early threat detection process by carefully balancing the anticipation level for threats with the possibility of false alarms being generated. Most importantly, the study results, which originate from two separate validation datasets, one of which includes.attacks targeting pedestrians and the other focused on threats against a truck that was parked—clearly demonstrated the effectiveness of the suggested method for early threat identification. The algorithm's high effectiveness and potential to improve security and safety in real-world scenarios were demonstrated by performance metrics that continuously exceeded the 90% threshold.
6. A notable breakthrough in the field of video saliency detection was made in 2020 with the publication of "Motion-Aware Rapid Video Saliency Detection," an article by F. Guo, W. Wang, Z. Shen, J. Shen, L. Shao, and D. Tao in IEEE Transactions on Circuits and Systems for Video Technology. In order to recognise important objects in video sequences, the authors proposed a novel and computationally effective method. Their methodology was based on the fundamental discovery that motion is a more saliency indicator than colour cues, which frequently show notable variations and intricate patterns. By concentrating on dynamic areas of the video, they were able to effectively identify spatiotemporal saliency by utilising this realisation. By analysing the optical flow field, the suggested method efficiently utilises motion information to obtain foreground priors. In addition, it integrates spatial saliency features, such as appearance contrasts and compactness measures, into a multi-cue integration framework to attain temporal consistency. Their method processes video frames in as little as 0.08 seconds per frame, almost 100 times faster than state-of-the-art methods. This is demonstrated by the results of rigorous experiments on challenging datasets such as SegTrackV1, SegTrackV2, and FBMS. The method's excellent performance and quick execution prove its suitability for a wide range of computer vision applications. The research conducted by Guo, Wang, Shen, Shao, and Tao tackles a critical issue in video analysis by providing a novel solution that greatly speeds up processing while also improving accuracy. This discovery holds significant ramifications for fields like object tracking, action recognition, and video summarization, where quick and precise object identification is crucial. Their contribution sets a high bar for future research in the field by offering a reliable and effective method that satisfies the needs of practical applications. It also serves as an example of the ongoing evolution of video saliency detection techniques.
7. The increasing ubiquity of surveillance cameras and the difficulty of real-time analysis of large amounts of video data have propelled notable progress in the field of automated video surveillance in 2023. The scientific community has paid close attention to automated violence detection in surveillance footage as a means of addressing the issue of timely event detection. In this regard, the research landscape has been enhanced by V. D. Huszár, V. K. Adhikarla, I. Négyesi, and C. Krasznay's paper, "Towards Fast and Accurate Violence Detection for Automated Video Surveillance Applications," which was published in IEEE Access. Their work investigates novel methods for detecting violence by utilising the capabilities of smart networks that have 3D convolutions installed. Because these networks are made to recognise temporal as well as spatial structures in the data, they can providing a more thorough comprehension of the dynamic interactions between people and objects in surveillance video. Crucially, by using information from an action recognition model that has already been trained, the authors improve the effectiveness and precision of violence detection. The authors expand and analyse a number of publicly accessible datasets, which include difficult and varied video content, in order to determine the effectiveness of their suggested approaches. Their findings demonstrate that their method outperforms the most advanced techniques, improving accuracy by about 2% while utilising fewer model parameters. Additionally, the research emphasises how resilient their method is to typical compression artefacts, which are frequently seen in applications involving remote server processing. This study adds to the current discussion about automated violence detection and highlights the capabilities of sophisticated machine learning algorithms.in expanding the functionality of video surveillance systems, consequently enabling more successful and efficient real-time event detection in a society that depends more and more on digital video technologies.
8. Researchers have made significant contributions in the field of surveillance and change detection for low-illumination environments in 2023. In their paper "Unsupervised Change Detection in Wide-Field Video Images Under Low Illumination," published in the IEEE Transactions on Circuits and Systems for Video Technology, B. Shi, Z. Jia, J. Yang, and N. K. Kasabov present a novel approach. Their research tackles the difficulties caused by low signal-to-noise ratio and poor visual quality in video images taken in low-light situations. They provide a thorough approach for image change detection (CD) in security footage in order to address this problem. Their method is based on adaptive fusion of difference images (DIs) and optimised clustering of k-medoids. First, they use extremum pixel and log-ratio to create two DIs. operators for pixel ratios. After that, these DIs are adaptively fused using a mix of Laplacian pyramid and local energy analysis methods. To obtain the final DI, the fused DI is subjected to additional processing using improved adaptive median filter and normalisation functions. Finally, an optimised k-medoids clustering algorithm is used to detect the altered image. According to their experiments, their method outperforms existing methods in terms of robustness and accuracy when it comes to detecting subtle changes in eagle eye surveillance imagery under low illumination. In addition, their algorithm is robust against false alarms caused by noise in unaltered scenes and efficient, yielding shorter processing times. Shi et al.'s inventive work adds a great deal to the expanding body of improving surveillance system performance in difficult low-light situations, where conventional approaches frequently fail.
9. In the field of video-based object detection, most methods relied on image-based object detectors and added temporal context through post-processing or feature enhancement from multiple frames. Liu, Liao, and Hu filled a major gap in this area in 2019 by the paper “Perceiving Motion From Dynamic Memory for Vehicle Detection in Surveillance Videos”. Motion-From-Memory (MFM), their innovative contribution, introduced a simple yet powerful module for directly encoding temporal context. In order to produce motion features on each frame, this module used appearance features from a basic Convolutional Neural Network (CNN) and kept a dynamic memory for each input sequence. Surprisingly, this improvement required very little in the way of extra model parameters and computation. In surveillance videos, the MFM module proved to be especially helpful for detecting moving objects. MFM had a significant impact. Upon incorporation into a lighter MobileNet-based Faster RCNN detector, it produced an impressive increase of 13.93% in mean Average Precision (mAP), matching the performance of a more resilient ResNet-50 based detector. The effectiveness and efficiency of MFM in improving object detection capabilities were highlighted by this finding. Additionally, by incorporating MFM into an even faster single-stage detector, the authors achieved a competitive 69.10% mAP compared to the best-performing solution's 69.87%. This detector notably ranked as the second-best among all published works when evaluated against the DEETRAC vehicle detection benchmark. Significantly, their methodology was superior in terms of both speed and performance. utilising 540x960 security footage at a moderate frame rate and operating at 33 frames per second (FPS) Their method beat the competition by a large margin, operating about three times faster than the second-fastest alternative using a commercial GPU (NVIDIA GTX 1080Ti). The potential of directly modelling temporal motion through the MFM module was demonstrated in this groundbreaking work by Liu, Liao, and Hu in 2019. This work opens up a promising path for effective and efficient object detection in surveillance videos and ultimately advances the state of the art in this field.
10. Significant progress in the area of anomaly detection in aerial videos was made in 2022 when P. Jin, L. Mou, G.-S. Xia, and X. X. Zhu published their groundbreaking paper, "Anomaly Detection in Aerial Videos With Transformers." The authors noted that unmanned aerial vehicles (UAVs) are becoming more and more important in a variety of applications, including search and rescue, inspection, and operations that produce enormous volumes of aerial video data. But most of this data consists of regular events, which makes it very difficult to manually separate meaningful information from possible anomalies in long video streams. In order to tackle this problem, the authors presented the DroneAnomaly dataset, which consists of 22 testing and 37 training video sequences that were captured in seven distinct realistic scenarios.showcasing an assortment of unusual occurrences. The 87,488 colour video frames in this dataset were all taken at 30 frames per second, with a frame size of 640x640. Of those, 51,635 were used for training and 35,853 for testing. With their newly created dataset, the authors set a benchmark for anomaly detection in aerial videos after conducting a thorough assessment of current approaches in the field. An additional novel baseline model called "ANomaly Detection with Transformers" (ANDT) was presented by them. This model uses a Transformer encoder to extract feature representations from the series and a decoder to predict the next frame. It treats consecutive video frames as a series of tubelets. Specifically, the model was built to detect normalcy in the training phase and to classify events with non-linear temporal dynamics as anomalies in the testing phase. In order to guarantee a comprehensive evaluation of their suggested technique, the authors not only employed their DroneAnomaly dataset but also incorporated an additional dataset, showcasing the resilience and applicability of their methodology. To further encourage transparency and cooperation among researchers, they have also made their code and dataset publicly accessible. To summarise, the research conducted by P. Jin, L. Mou, G.-S. Xia, and X. X. Zhu has significant implications for the advancement of anomaly detection methods in aerial videos. Additionally, it serves as a valuable resource for future investigations and testing in this field.
11. The paper "Multiscale Fully Convolutional Network for Foreground Object Detection in Infrared Videos," written by D. Zeng and M. Zhu and published in IEEE Geoscience and Remote Sensing Letters in April 2018, makes a substantial contribution to addressing the difficulties involved in precise and quick IR foreground object detection. The task's significance for IR target recognition, precise guidance, and infrared video surveillance is emphasised by the authors. Although the standard technique for identifying foreground objects in a variety of computer vision applications has been background subtraction, the special properties of infrared images have restricted the applicability of current algorithms. The paper presents a novel convolutional neural network (CNN) based method to address these issues. By utilising a large-scale image dataset and a pre-trained CNN model, the authors suggest a multiscale fully convolutional network architecture. The representation of features at multiple scales is made possible by this architecture, which makes use of output features from various CNN layers. This multiscale feature representation combines fine-grained details and semantics at the category level, which is especially important for IR foreground object detection. The experimental results, which are noteworthy for their real-time operation, demonstrate the state-of-the-art performance of the suggested method and are presented in the paper. The authors illustrate the efficacy of their methodology by utilising CNNs' power and customising them to meet the demands of infrared image analysis. This guarantees foreground object detection's applicability in real-world scenarios while also improving its accuracy.
12. The field of intelligent video surveillance saw a noteworthy advancement in 2022 when C. Huang, Z. Wu, J. Wen, Y. Xu, Q. Jiang and Y. Wang published a paper titled "Abnormal Event Detection Using Deep Contrastive Learning for Intelligent Video Surveillance System." Before their work, single pretext tasks like image reconstruction or prediction were the main method used for anomaly detection in surveillance videos. Larger reconstruction errors or inaccurate predictions were used to identify anomalies. These methods, however, had limitations when it came to using temporal context information and discriminative semantics. Furthermore, because the task and the detection goal were not aligned, utilising a single pretext task for anomaly detection was not ideal. The authors suggested a brand-new method known as Temporal-Aware Contrastive Network (TAC-Net) to solve these problems. An unsupervised technique called TAC-Net makes use of deep contrastive technique that uses multiple self-supervised tasks to detect anomalies and makes use of deep contrastive self-supervised learning to capture high-level semantic features. In order to determine an anomaly score, the model uses contrastive similarity and multiple task losses during the inference stage. The experimental results demonstrated that TAC-Net outperformed state-of-the-art techniques on three benchmark datasets. This work highlights the potential for more reliable and efficient anomaly detection systems in intricate urban and industrial environments, underscoring the importance of their contribution to the advancement of intelligent video surveillance. By utilising temporal context and high-level semantics in surveillance videos, the authors' deep contrastive learning approach presents a viable path for improving anomaly detection and meeting the increasing need for intelligent video.

# Problem Formulation

## Research Problem

The Long Short-Term Memory Recurrent Neural Network (LRCN), a cutting-edge machine learning technique, is the central tool used in this research to detect, model, and react to abnormalities in human behaviour. By automating the anomaly detection process, this project hopes to reduce reliance on feature engineering and static, human-crafted rules. Because of the wide variety of behaviours and the dynamics of their temporal expression, human activities are complex. Anomalies can appear as abrupt, dramatic changes or as minor departures from typical patterns. Traditional manual techniques are less efficient and flexible in changing situations due to these complexities. Thus, there is a chance to improve efficiency, security, and safety in a number of areas by automating the anomaly detection process.

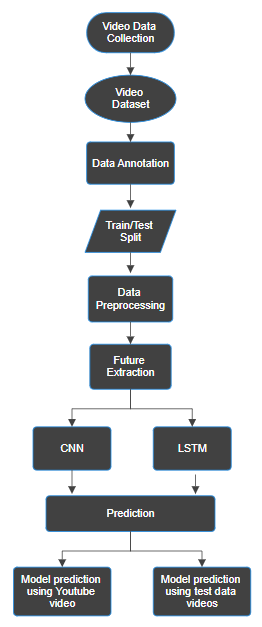
## Significance of the Problem

Automating anomaly detection in human activities is important in many different fields. Automatically identifying suspicious activity, possible threats, and intruders in video feeds is essential in the field of security and surveillance. Airports, public transportation, government buildings, and public areas can all be protected with this. Ensuring public safety and security requires the efficient identification of anomalies. It is a life-saving endeavour to identify abnormalities in healthcare early. Medical conditions can be quickly diagnosed and treated by automating the recognition of abnormal patient behaviours and vital sign patterns. Prompt interventions can lower healthcare costs and greatly improve patient outcomes. Knowing what constitutes unsafe practises is essential to ensuring worker safety and the smooth operation of industrial facilities.

In industrial settings, automating anomaly detection can improve worker safety, reduce the risk of accidents, and maximise the efficiency of processes and machinery. Maintaining product quality and minimising defects in manufacturing requires the detection of anomalies in product assembly or production processes. Higher customer satisfaction and cost savings may result from this. Automating anomaly detection in human activities is a complex problem that requires solutions that can grasp a variety of data sources, capture intricate temporal dependencies, and identify subtle spatial patterns. It is in this complex problem space that LRCN exploration as a possible solution becomes crucial and interesting. The LRCN-based approach's efficacy will be clarified in the following sections of this study as they explore the suggested methodology, empirical findings, and implications.

# System Model

The overall architecture of our proposed model is depicted in Fig. 1.



**Fig. 1. The proposed system model for abnormal activity recognition.**

The procedure for detecting anomalies in video data using a hybrid CNN and LSTM model is shown in the flow diagram. First, video data must be gathered and a video dataset must be created. Then, anomalous activities are labelled in the dataset through annotation. The data is then divided into training and testing sets and preprocessed, which includes feature extraction. Using the labelled data, the hybrid CNN and LSTM model is trained. Following training, two distinct scenarios are used to test the model's performance: first, it is used to predict YouTube videos to find anomalies, and then it is applied to test data videos. This flowchart illustrates the entire process of detecting anomalies in videos, from data collection to model application.

# Methods

The proposed model consists of the following steps: data collection and preprocessing (read video and label, splitting into frames to create ssequences, resizing, normalization, storing in NumPy arrays), train test split data, model creation and model training.

## Dataset

"Anomaly Detection in Human Activity with Long-Range Convolutional Recurrent Networks (LRCN)" is a study topic that centres on identifying unusual or suspicious events in video footage, particularly those related to shootings, explosions, crimes, and battles. To make this analysis easier, a large number of datasets from YouTube and other video streaming services have been assembled.

## Data Pre-processing

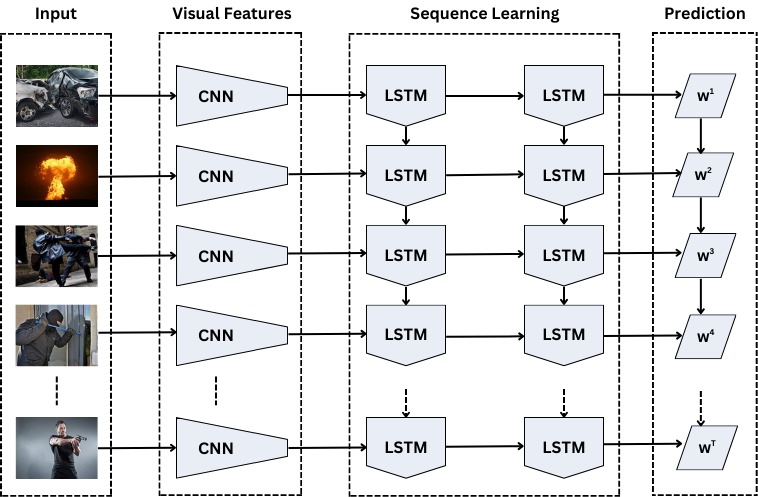
1. Read Video and Label: During the data preprocessing stage, we utilize the OpenCV library to access video files stored in the corresponding class directories. Since it creates the link between video content and class labels, this step is essential to our anomaly detection project. Class labels stand for the different types of human activity that we try to identify, such as shooting, fighting, explosions, and burglaries. We lay the foundation for supervised learning by assigning a corresponding label to every video. As a result, our model is able to learn the unique characteristics of every activity class and generate accurate predictions. We can effectively extract frames from videos and arrange them for additional processing thanks to the OpenCV library's flexible video file handling tools.
2. Splitting into Frames to Create Sequences: Partitioning video clips into frame sequences is the next step in the data preparation process. Every video has 30 frames extracted at evenly spaced time intervals to guarantee a representative and balanced sample. Our model can analyse the sequence of events within each sequence because this method successfully captures the temporal aspects of human activities. This data structure gives the model a chronological viewpoint, which helps it better understand the activities under investigation. This stage is essential for detecting anomalies because variations from regular activity patterns may point to anomalous occurrences.
3. Resizing: For our anomaly detection model to function, the dimensions of the input images must be standardized. To make sure the dataset is consistent, resizing is used. All of the frames are uniformly resized to a given size; this is usually 64 pixels wide by 64 pixels high. Because the model can anticipate receiving input frames of the same size, this consistency makes processing it easier. Furthermore, resizing lessens the computational load on the model, increasing its efficiency. It prevents bias towards higher- or lower-resolution videos in the dataset by bringing all frames to a common scale, which is especially helpful when working with videos of different resolutions. This stage enhances the model's resilience and applicability.
4. Normalization: To improve the training of our anomaly detection model, normalization is an essential step. Pixel values are scaled to aid in the learning algorithm's faster convergence and the extraction of important features from the images. The resized frames are now normalized by dividing each pixel value by 255. The pixel values are scaled to the range of 0 to 1, which makes them consistent between frames through this straightforward but efficient operation. Normalization helps to reduce the problems caused by different pixel value scales that may arise during model training. Additionally, it lessens the model's sensitivity to variations in lighting or image quality, guaranteeing that it can detect anomalies no matter what the original qualities of the input video were.
5. Storing in NumPy Arrays: The resized and normalized frames are then arranged into NumPy arrays so that the model can see them. When handling numerical data, this data structure is very effective, making it a good choice for machine learning applications. NumPy arrays are created from the sequences of 30 frames that were prepared in the preceding steps. This makes it easier for the model to process and analyse the data. The main input for our anomaly detection architecture is the NumPy arrays, which makes the training and assessment of the model easier. By taking this step, data management is streamlined and the model is better equipped to identify and categories human activities in the video data. In order to enable precise anomaly detection, it serves as the last link between the basic machine learning model and the data preprocessing phase.

## Train Test Split Data

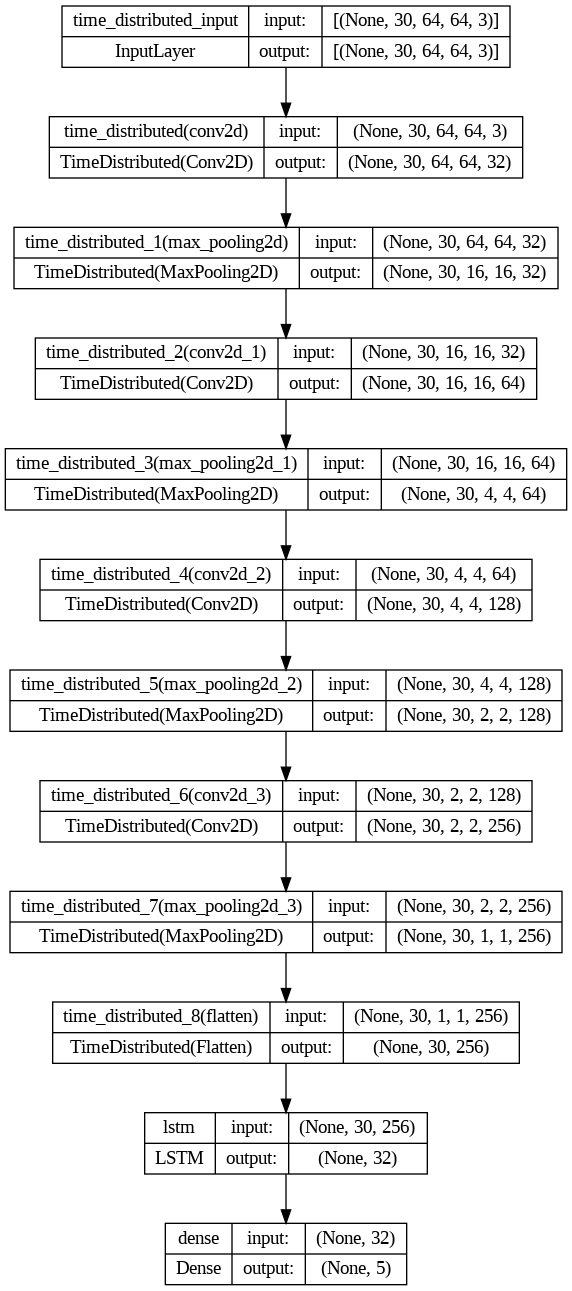
To properly evaluate the effectiveness and generalizability of our anomaly detection model, we must partition our dataset into training and testing subsets during the data preparation stage. In order to do this, we use a train-test split strategy, in which 25% of the data is reserved for testing and the remaining 75% is allocated for training. With this separation, the model is guaranteed to learn from a significant portion of the data, but a separate portion is kept to assess the model's capacity to identify anomalies in scenarios that have not yet been seen. To give the model a baseline understanding, the training set is used to teach it the underlying patterns and traits of typical human activities. Next, the testing set is used to evaluate the model's effectiveness by displaying it data it has not encountered during training. This evaluation helps gauge the model's ability to accurately identify anomalies and distinguish them from routine activities. The train-test split is a critical step in building a robust anomaly detection system, as it enables us to validate the model's effectiveness in real-world applications.

## Model Creation

We use the Long-Range Convolutional Recurrent Network (LRCN), a deep learning network, in the development of our system for suspicious activity detection from video surveillance. LRCN is an advanced architecture that combines the best features of long short-term memory (LSTM) networks with the power of convolutional neural networks (CNNs). (LSTM) networks that possess convolutional neural network (CNN) capabilities. LRCN's main idea is to use CNNs to extract visual features from individual video frames, and then use LSTMs to convert these image embeddings into meaningful outputs, like probability distributions, class labels, or sentences. With this method, we can efficiently examine and decipher the temporal information contained in video data. The raw visual input from video frames is best processed by CNNs because of their remarkable ability to learn visual features. Afterwards, a stack of recurrent sequence models—in this case, LSTMs—are fed their outputs, which represent significant visual features. When dealing with time series data, where there might be gaps of unknown length between important events, LSTMs work well.. They overcome the vanishing gradient problem, a challenge encountered when training traditional Recurrent Neural Networks (RNNs), which makes them particularly effective for tasks requiring memory of past events. By combining the power of CNNs for visual feature extraction and LSTMs for temporal modelling, our LRCN model is capable of comprehensively analysing video data to detect and classify suspicious activities. The architecture of LRCN model is shown in Fig. 2. This architectural fusion allows us to create a robust and accurate system for surveillance-based anomaly detection, effectively addressing the challenges of analysing video data with complex temporal dependencies. The LRCN model created has 12 layers as shown in Fig. 3.



**Fig. 2. Architecture of LRCN Model**



**Fig. 3. Layers of LRCN Model**

## Model Training

Our main goal during the model training phase is to give the model the capacity to predict and categories actions into one of three different classes: fighting, running, and walking. In order to do this, we feed the model the training data, and it uses that data to identify the underlying patterns and characteristics linked to each of these activities.

The following hyperparameters are used during the training process:

1. Epochs: The training epoch count was set to 70. One full iteration of the training dataset is represented by an epoch. To reduce training loss and improve prediction accuracy, the model iteratively adjusts its internal parameters over 70 epochs. This large number of epochs enables the model to fully understand and adjust to the subtleties of the data.
2. Batch Size: For efficient training, we use a batch size of 4. Batch size determines the number of data samples processed in each iteration. A smaller batch size, such as 4, is chosen to facilitate better convergence during training and can help avoid memory limitations in GPU resources. It also adds a degree of stochasticity, aiding the model's generalization.
3. Validation Split: During training, we allocate 25% of the training data for validation, indicated by a validation split of 0.25. This portion of the data is not used for training but rather for evaluating the model's performance during each epoch. Validation helps prevent overfitting by providing an independent dataset for model evaluation. It allows us to monitor how well the model generalizes to unseen data and ensures that the training process leads to a well-performing model.

These hyperparameters are carefully selected to strike a balance between model training efficiency, generalization, and the ability to capture intricate patterns in the data. The combination of the chosen number of epochs, batch size, and validation split enhances the model's capacity to accurately classify activities as walking, running, or fighting, making it a robust and reliable system for real-world application in surveillance and anomaly detection.

# Experiments and Results

The outcomes of our suggested model mark an important turning point in the field of anomaly detection in video surveillance. Our main goal was to reliably identify anomalous behaviour, and using our specially constructed dataset, we achieved an amazing accuracy rate of 98.86%. We shrunk our frames from 224 pixels to 64 pixels in order to further maximize the performance of the model. In addition to saving memory, this resizing expedited the input frame processing. The model's overall efficiency was enhanced by the reduction in frame size, which ensures that our system can function properly even in environments with limited resources. The accuracy of the model was not compromised. Our findings demonstrate the practical applicability of our model in addition to its accuracy. Frame resizing, a varied dataset, and decreased computational complexity all show our dedication to developing a model that can accurately and quickly identify anomalies in video surveillance. Our system is further enhanced by its real-time detection capabilities, which highlight its readiness for deployment in security and surveillance applications. The results' images provide a visual depiction of the model's capabilities and demonstrate how well it can differentiate between various activities. Finally, we can say that our suggested model outperformed our previous VGG-16 model, achieving an accuracy rate of 98.86% on our personal dataset. Our system is now more suitable for real-time detection and requires less time due to the switch to the LRCN architecture and optimizations like frame resizing. The model's capacity to recognize suspicious activity is improved by the incorporation of a large variety of videos that show both typical and abnormal behaviour. Our efforts have produced an anomaly detection system that is strong, useful, and adaptable and is prepared to improve security and surveillance in real-world situations. Fig. 4 shows the output of the model after detecting various actions.





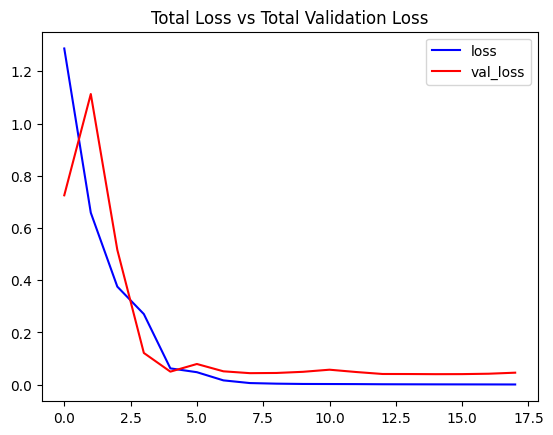




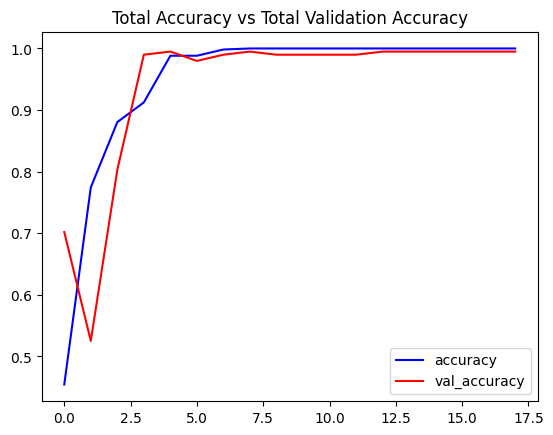


**Fig. 4. Model Prediction of Anomaly Activities**

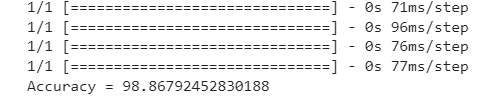
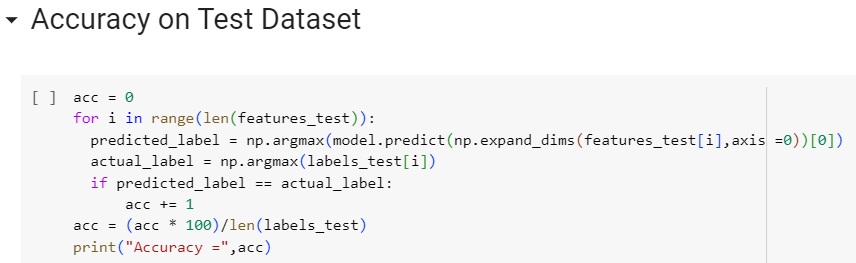
The total loss vs total validation loss plot, total accuracy vs total validation accuracy plot and accuracy of the model can be show by Figure 3.5, 3.6 and 3.7 respectively.



**Fig. 5. Total Loss vs Total Validation Loss Plot**



**Fig. 6. Total Accuracy vs Total Validation Accuracy Plot**



**Fig. 7. Accuracy of the Model**

##### Conclusion

To sum up, we have developed a strong, effective, and remarkably accurate model as the result of our journey towards creating an anomaly detection system for video surveillance. We switched from the computationally demanding VGG-16 model to the simpler LRCN architecture, placing a great focus on improving real-time detection capabilities. This calculated decision increased processing speed while maintaining a high degree of accuracy, yielding an astounding accuracy rate of 98.86% on our customized dataset. Furthermore, without compromising performance, we resized frames from 224 pixels to 64 pixels in order to optimize the model's memory usage. Our dataset has been expanded to include a wide range of activities, including both typical activities like walking and running. and abnormal behaviour like fighting, which strengthens our model's capacity to discriminate between various classes. This joint effort is a major advancement in the field of video surveillance, providing a workable, real-world method for identifying questionable activity in ever-changing settings. In addition to proving the usefulness of our model, our findings highlight its suitability for use in security and surveillance applications, where it is critical to quickly and accurately identify anomalies. Our proposed system, which is dedicated to efficiency, accuracy, and practicality, is expected to improve safety and security across multiple domains and contribute to the ongoing advancement of video surveillance technology.

##### References

1. Aniruddha Prakash Kshirsagar, H. Azath (2023) “YOLOv3-based human detection and heuristically modified-LSTM for abnormal human activities detection in ATM machine,” in *Elsevier B.V.* Volume 95, September 2023, 103901.
2. Ujwalla Gawande , Kamal Hajari , Yogesh Golhar (2023) “Real-Time Deep Learning Approach for Pedestrian Detection and Suspicious Activity Recognition” in *Elsevier B.V.* Volume 218, 2023, Pages 2438-2447.
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. M. Z. Zaheer, A. Mahmood, H. Shin and S. -I. Lee, "A Self-Reasoning Framework for Anomaly Detection Using Video-Level Labels," in IEEE Signal Processing Letters, vol. 27, pp. 1705-1709, 2020, doi: 10.1109/LSP.2020.3025688.
5. D. Martínez, H. Loaiza and E. Caicedo, "Algorithm For Early Threat Detection By Suspicious Behavior Representation," in *IEEE Latin America Transactions*, vol. 18, no. 05, pp. 825-832, May 2020, doi: 10.1109/TLA.2020.9082909.
6. F. Guo, W. Wang, Z. Shen, J. Shen, L. Shao and D. Tao, "Motion-Aware Rapid Video Saliency Detection," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 12, pp. 4887-4898, Dec. 2020, doi: 10.1109/TCSVT.2019.2906226.
7. V. D. Huszár, V. K. Adhikarla, I. Négyesi and C. Krasznay, "Toward Fast and Accurate Violence Detection for Automated Video Surveillance Applications," in IEEE Access, vol. 11, pp. 18772-18793, 2023, doi: 10.1109/ACCESS.2023.3245521.
8. B. Shi, Z. Jia, J. Yang and N. K. Kasabov, "Unsupervised Change Detection in Wide-Field Video Images Under Low Illumination," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 4, pp. 1564-1576, April 2023, doi: 10.1109/TCSVT.2022.3216457.
9. W. Liu, S. Liao and W. Hu, "Perceiving Motion From Dynamic Memory for Vehicle Detection in Surveillance Videos," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 12, pp. 3558-3567, Dec. 2019, doi: 10.1109/TCSVT.2019.2906195.
10. P. Jin, L. Mou, G. -S. Xia and X. X. Zhu, "Anomaly Detection in Aerial Videos With Transformers," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-13, 2022, Art no. 5628213, doi: 10.1109/TGRS.2022.3198130.
11. D. Zeng and M. Zhu, "Multiscale Fully Convolutional Network for Foreground Object Detection in Infrared Videos," in IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 4, pp. 617-621, April 2018, doi: 10.1109/LGRS.2018.2797538.
12. C. Huang, Z. Wu, J. Wen, Y. Xu, Q. Jiang and Y. Wang, "Abnormal Event Detection Using Deep Contrastive Learning for Intelligent Video Surveillance System," in IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5171-5179, Aug. 2022, doi: 10.1109/TII.2021.3122801.
13. X. Tian, P. Zheng and J. Huang, "Robust Privacy-Preserving Motion Detection and Object Tracking in Encrypted Streaming Video," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 5381-5396, 2021, doi: 10.1109/TIFS.2021.3128817.
14. Z. A. Khan and W. Sohn, "Abnormal human activity recognition system based on R-transform and kernel discriminant technique for elderly home care," in *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1843-1850, November 2011, doi: 10.1109/TCE.2011.6131162.
15. J. Yin, Q. Yang and J. J. Pan, "Sensor-Based Abnormal Human-Activity Detection," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082-1090, Aug. 2008, doi: 10.1109/TKDE.2007.1042.
16. Chhavi Dhiman, Dinesh Kumar Vishwakarma, “A review of state-of-the-art techniques for abnormal human activity recognition,” in *Engineering Applications of Artificial Intelligence*, Volume 77, 2019, Pages 21-45, ISSN 0952-1976.
17. Shaik Jameer, Hussain Syed, “A DCNN-LSTM based human activity recognition by mobile and wearable sensor networks,” in *Alexandria Engineering Journal*, Volume 80, 2023, Pages 542-552, ISSN 1110-0168.