Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

Criminal Detection by Analysing Live CCTV Footages

Abhishek Kumar
Information Science and Engineering
Dayananda Sagar College of Engineering
Bangalore, India

Ayush Kumar
Information Science and Engineering
Dayananda Sagar College of Engineering
Bangalore, India

Aryan Shrivastava
Information Science and Engineering
Dayananda Sagar College of Engineering
Bangalore, India

Mudit Khandelwal
Information Science and Engineering
Dayananda Sagar College of Engineering
Bangalore, India

Dr. Vaidehi M
Information Science and Engineering Dayananda Sagar College of Engineering Bangalore, India

***Abstract*—**The project known as the Criminal Assault Analysis and Security Project utilizes contemporary surveillance technologies, such as CCTV cameras, to conduct observations. By analyzing video footage, this system can generate alerts in the event of any suspicious activities or acts of violence. Upon detection of criminal behavior, the relevant authorities are promptly notified and can proceed with appropriate actions. The system's functionality involves the identification of human figures, recognition of unlawful actions, and subsequent facial recognition.

To provide the authorities with comprehensive information, the warning or alert system has been designed to relay details regarding the occurrence's location, individuals involved, and the time it took place. Additionally, the system maintains a database of known criminals, automatically updating with new individual images as needed. Authorities also have the capability to manually supplement this database with additional information such as prior records and addresses.

For seamless operation, the system leverages Machine Learning technology and Image Processing techniques.

***Keywords*—** Criminal detection, CCTV footage analysis, Video surveillance, Motion detection, Pattern recognition, Machine learning in video analysis.

I. INTRODUCTION

1.1 Overview:

In recent times, there has been a noticeable rise in offensive and disruptive incidents, leading to an increased emphasis on security measures. Many organizations and sectors now commonly employ CCTV installations to continuously monitor people and their interactions. In a densely populated developed country, individuals are frequently captured by cameras throughout the day, resulting in a significant volume of video footage that is stored for a specific duration. However, it is practically impossible for authorities to constantly monitor these surveillance videos for suspicious events due to the need for a dedicated workforce and their unwavering attention.

A highly accurate automation of this operation is thus urgently needed. It is also critical to pinpoint the precise frames and sections of recordings that contain suspicious or unusual activity. This will enable quicker analyses of suspicious or abnormal situations. The time and effort that would normally be required to manually look through recordings would be saved, and this would help the authorities quickly determine the cause of abnormalities.

We suggest an Anomaly Recognition System, a real-time surveillance programme made to automatically detect and analyse signals of offensive or disruptive actions, to address these issues.
Our method involves using a variety of Deep Learning models to identify and categorise high degrees of

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

movement within frames. Videos are divided into sections, and when a threat is present, a detection alert is set off to identify any suspicious activity that occurred. Threat (anomalous actions) and Safe (regular activities) are the two categories into which the movies are divided.

In addition, 12 particular abnormal acts have been noted: abuse, burglary, explosion, shooting, fighting, shoplifting, road accidents, arson, robbery, theft, assault, and vandalism. The ability to identify these anomalies will greatly improve personal security measures.

We use deep learning algorithms, which excel at activity detection and categorization, to tackle the aforementioned problems. We specifically use CNN and RNN, two different neural networks. In order to simplify the input, CNN is used as the main neural network to extract complex feature maps from the available recordings.

For transfer learning, we utilize InceptionV3, a pre-trained model chosen due to the high parameter count and time required for training modern object recognition models from scratch. The transfer learning approach allows us to initially leverage the knowledge gained from the pre-trained model on a set of classified inputs, such as ImageNet, and subsequently retrain it with new weights assigned to various new classes.

The RNN then receives the output from the CNN. In addition, RNN can predict the following item in a series, serving as a forecasting engine. We want to make sense of the activities or movements that were recorded in the recordings by employing this neural network. In the RNN's primary layer, there is an LSTM cell, followed by hidden layers with the proper activation functions.

The video is finally divided into 13 groups, 12 of which are anomalies and 1 of which is normal, by the output layer. The output of this system is applied to real-time monitoring of CCTV cameras installed in various organisations, assisting in the detection and avoidance of suspicious behaviours. Therefore, there is a large decrease in temporal complexity.

1.2 Problem Statement

The City CCTV Control Room is responsible for receiving video feeds from multiple CCTV cameras situated throughout the city. However, monitoring all the camera feeds in real-time is not feasible. To address this challenge, a solution has been developed that can process and analyze the incoming video feeds for any signs of criminal activity.

1.3 Objectives

The primary goal is to develop an automated surveillance system that utilizes CCTV cameras to monitor and detect criminal activity, thereby minimizing human involvement.

This project introduces specific algorithms capable of promptly notifying the relevant authorities upon the detection of unusual behavior exhibited by individuals. The focus of the project is to design a surveillance program that can autonomously identify gestures or indications of aggression and violence in real-time.

1.4 Motivation

The increasing emphasis on public safety and security has created a demand for efficient detection of criminal activities through the analysis of live CCTV footage. Conventional crime prevention methods primarily rely on investigations after incidents have occurred, making it difficult to respond promptly to ongoing criminal events.

However, by leveraging advanced computer vision and machine learning algorithms, the analysis of live CCTV footage becomes a powerful tool for proactive identification of suspicious behavior, potential threats, and criminal activities in real-time.

This proactive approach holds the potential to significantly improve law enforcement capabilities, deter criminal acts, and safeguard communities, ultimately leading to enhanced public safety and a greater sense of security for all.

II. LITERATURE SURVEY

A crime detection method's importance resides in its capacity to foresee and stop illegal activity. While conventional techniques are useful, they frequently work in isolation. Therefore, a device that could combine the advantages of these traditional procedures would be quite helpful.

Machine learning (ML) techniques were used in a study that used criminal data from Vancouver collected over a 15-year period to predict, detect, and analyse criminal activity. Data gathering, classification, pattern recognition for criminal activity, forecasting, and visualisation were all part of the analysis. The criminal dataset was examined using K-nearest neighbour (KNN) models and enhanced decision trees. A total of 560,000 criminal datasets were examined from 2003 to 2018, and criminal activity was predicted with an accuracy range from 39% to 44%.

The Chicago crime dataset was used to construct ML and data science-based models to predict and identify criminal activity. To find the most precise model for training, many combinations of ML models, including logistic regression, SVM/KNN classification, decision trees, random forests, and Bayesian models, were looked at.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

The classification accuracy that was highest was 78.7% for the KNN algorithm. This study's major goal was to persuade law enforcement organisations to use ML-based techniques to anticipate, identify, and settle illegal activity more successfully, ultimately lowering crime rates in society.

In a different strategy, a deep neural network (DNN)-based feature-level data fusion method was suggested to accurately predict the occurrence of crimes. This approach required fusing environmental background knowledge with multi-model data from many fields. Data from an online crime statistics database (Chicago) as well as meteorological and demographic information were included in the database utilised for crime prediction. For the purpose of predicting crime, various ML models including SVM, regression analysis, and kernel density estimation (KDE) were used. The accuracy of the SVM and KDE models was 67.01% and 66.33%, respectively, whereas the accuracy of the suggested ML model was 84.25%.

III. REQUIREMENTS

3.1 Functional Requirements

- **Acquisition of Live CCTV Footage:** The system should possess the capability to securely acquire real-time footage from multiple CCTV cameras simultaneously.
- **Real-time Analysis:** The system should be able to analyze the live CCTV footage in real-time, enabling the detection and identification of potential criminal activities as they unfold.
- **Anomaly Detection:** By utilizing machine learning algorithms, the system should have the ability to detect anomalies within the CCTV footage, including unexpected movements, suspicious gatherings, or abnormal patterns.
- **Event Notification:** Whenever potential criminal activities are detected, the system should promptly generate alerts or notifications to the appropriate authorities or security personnel.
- **Integration with Existing Systems:** The system should seamlessly integrate with pre-existing surveillance infrastructure and security systems, thereby enhancing their effectiveness and creating a unified security solution.

3.2 Non-Functional Requirements

- **Accuracy:** The system must exhibit a high degree of accuracy in accurately identifying and categorizing potential criminal activities, minimizing both false positives and false negatives.
- **Scalability:** In order to retain peak performance, the system must be scalable to accommodate a sizable number of CCTV cameras and simultaneous video processing.
- **Speed and Efficiency:** Real-time analysis of live CCTV footage should be conducted with minimal delay, enabling swift response and intervention when necessary.
- **Reliability:** The system should be robust and dependable, ensuring uninterrupted operation and minimal downtime to avoid any surveillance gaps.
- **Privacy and Data Security:** Strict adherence to privacy regulations is essential, ensuring the secure handling of sensitive CCTV footage and safeguarding the privacy of individuals involved.
- **Maintenance and Support:** The system should incorporate provisions for regular maintenance, updates, and technical support to address any operational issues that may arise.

3.3 Software Requirements

- **Programming Language:** The software solution should be implemented using the Python programming language, taking advantage of its rich libraries and frameworks for efficient development.
- **Deep Learning Framework:** TensorFlow should be utilized as the core deep learning framework for building and training precise machine learning models.
- **Real-Time Video Processing:** The software should incorporate OpenCV for real-time video processing capabilities, encompassing functions like video acquisition, frame manipulation, and feature extraction.

3.4 Hardware Requirements

- **CCTV Cameras:** The system's functionality relies on the presence of strategically positioned CCTV cameras in the designated surveillance areas. These cameras should possess the capability to capture high-quality video footage with suitable resolution and frame rates.
- **Processing Units:** To ensure smooth operation of the software, it is recommended to have a minimum hardware configuration. This includes a computer or server with at least 4GB of RAM to efficiently handle the video processing tasks.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

- **CPU:** An Intel Core i3 or an equivalent processor is recommended as a baseline for managing the computational requirements of real-time video processing and deep learning algorithms.

- **Storage:** It is necessary to have a minimum of 15GB of available storage space to accommodate the software application, libraries, and any additional data required for the system's operation.

- **Graphics Processing Unit (GPU):** Although not mandatory, the utilization of a dedicated GPU can significantly enhance the performance of deep learning algorithms and video processing tasks. The recommendation for a compatible GPU may vary depending on the system's complexity and deployment scale.

IV. SYSTEM ANALYSIS & DESIGN

4.1 Analysis

The proposed surveillance system makes use of video footage from security cameras to track and spot potentially illegal or suspicious activity in public places. A security system is then informed of the detected actions so that it can take appropriate action.

A number of steps are included in the suggested framework before the actual detection procedure. These steps include feature extraction from the video and detection. This approach tries to show how well the system works to identify and stop criminal activity.

For this system, many security cameras provide the input video. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are used in a deep learning model that is trained using the video's retrieved frames. With the help of a continuous stream of video frames recorded by the various cameras, the trained model is then put to the test in order to identify any suspicious or illegal activity.

4.2 System Design

In order to capture both spatial and temporal information

present in video data, we propose a hybrid architecture that combines convolutional layers for spatial processing and recurrent layers for temporal processing.

The first model in the architecture is a Convolutional Neural Network (CNN) responsible for extracting spatial features from individual frames. These spatial features are then encoded into a feature vector, referred to as the encoder.

The second model in the architecture is a Recurrent Neural Network (RNN) that processes mini-batches of the encoded frames. The RNN incorporates the temporal information present in the sequence of frames and produces the final classification result. This model is referred to as the decoder as it decodes the encoded features into predictions.

4.2.1 System Architecture Diagram

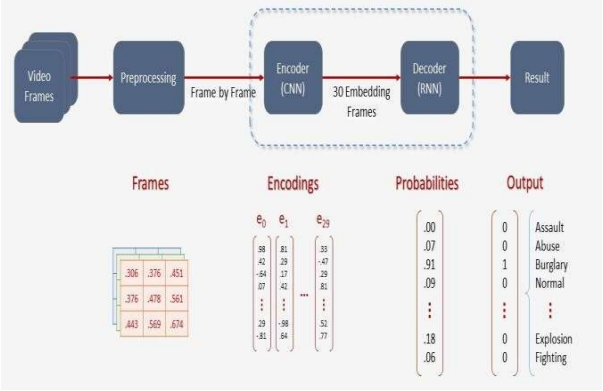

Fig. 1: Model Architecture

4.2.1.1 Data Flow Diagram

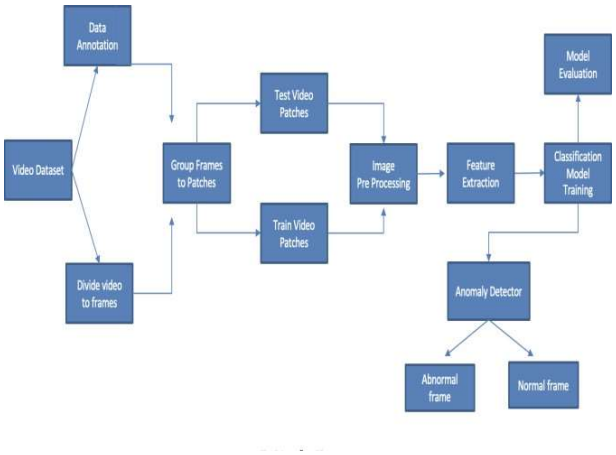

Fig. 2: Data Flow Diagram

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

4.2.1.2 Flow Chart

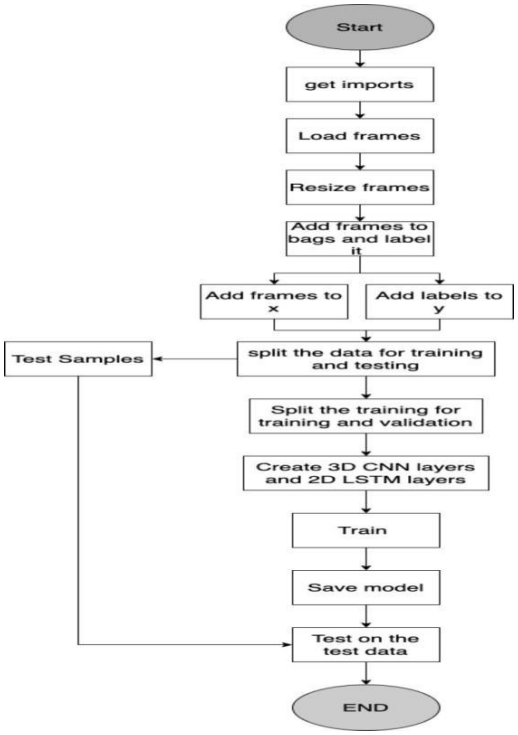

Fig. 3: Flow Chart

4.2.1.3 Use Case Diagram

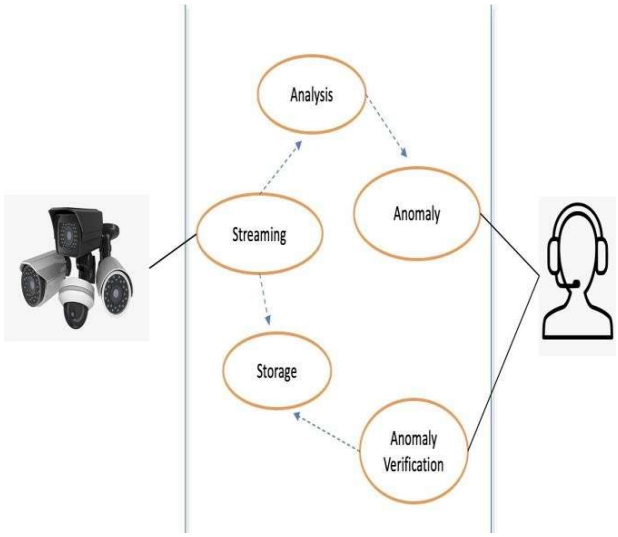

Fig. 4: Use Case Diagram

4.2.1.4 Sequence Diagram

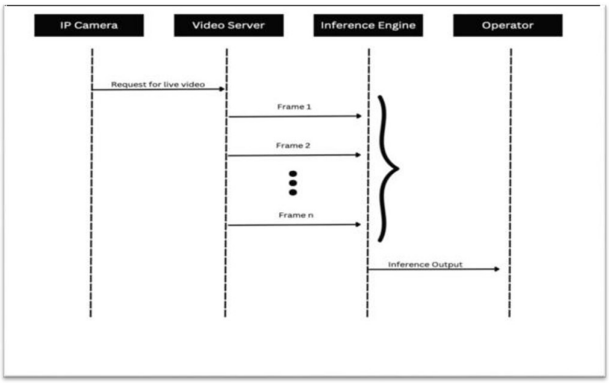

Fig. 5: Sequence Diagram

V. IMPLEMENTATION

5.1 Introduction

The project implementation encompassed several essential stages, including dataset preparation, video processing, and the design of the model architecture. To create a diverse dataset, videos were manually annotated, and clips were extracted to represent normal and anomalous events. Video processing techniques were employed to analyze motion information in batches of frames, ensuring efficient computational processing.

The model architecture adopted a hybrid approach by combining convolutional and recurrent layers to capture both spatial and temporal information in the video sequences. For the encoder model, transfer learning was implemented using pre-trained weights from the InceptionV3 network. On the other hand, the decoder model, based on LSTM architecture, was trained from scratch, leveraging the captured features from the encoder.

5.2 Implementation (Video Processing)

The video is composed of multiple frames that are displayed rapidly to create the illusion of motion. Our primary objective is to analyze the motion of objects within the video frames and determine whether it is normal or abnormal. However, processing the entire video at once can be computationally expensive. To address this, we adopt a batch processing approach where we analyze smaller groups of frames, or clips, to understand the motion information within them and classify them as normal or anomalous.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

To enable real-time video monitoring, an efficient video streaming pipeline is crucial. It should be capable of handling the continuous video stream from CCTV cameras, while simultaneously performing inference on the frames using the models. Our video streaming engine leverages multiprocessing to create two parallel processes. One process manages the video streaming, capturing the video feed from the CCTV or any other input device. The other process handles the inference engine, analyzing the frames in parallel to classify them.

By utilizing multiprocessing and parallel processing, we can effectively handle the video streaming and inference tasks, optimizing the utilization of available hardware resources and minimizing latency in the system.

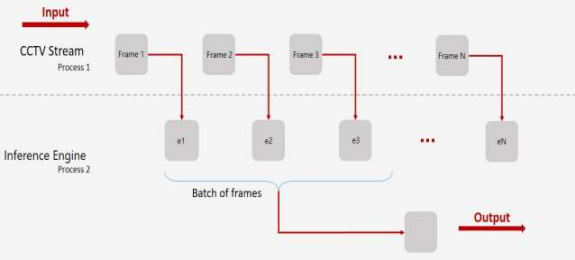

Fig. 6: Video Processing Architecture

5.3 Overview of System Implementation

5.3.1 System Implementation

Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are combined in the hybrid architecture used in the suggested system to process video input. We can record the spatial and temporal information that is present in the video frames using this method.

The first model, known as the encoder, utilizes a CNN to extract spatial features from each frame of the video. The input size for the encoder model is [240 × 320 × 3], indicating the width, height, and number of channels (RGB) of the frames. The CNN transforms these frames into an encoded feature vector. This encoding step helps to capture important spatial information from each frame.

The encoded feature vectors from the encoder are then fed into the second model, called the decoder. The decoder employs an RNN to process mini-batches of encoded frames. Each batch contains 30 encoded frames, resulting in an input size of [30 × 2048]. The RNN processes this temporal sequence of encoded frames and generates the final classification result. The decoder model is responsible for decoding the temporal information captured by the RNN and making predictions based on it.

The output of the decoder model is a vector of probabilities for different classes, indicating the likelihood of each class being present in the video. The output size is [10,] which corresponds to the number of classes or categories that the system is trained to detect. Based on these probabilities, the system can make predictions about the presence of certain activities or events in the video.

For the encoder model, transfer learning is employed. The weights of a pre-trained CNN model, specifically InceptionV3, trained on the ImageNet dataset, are used as a starting point. By leveraging the pre-trained weights, the encoder model can benefit from the learned features of the ImageNet dataset, which consists of a large number of diverse images. To prevent the weights from being updated during the training process, the weights are frozen or kept constant. The fully connected layers at the top of the InceptionV3 network are removed, and average pooling is applied to aggregate the feature maps along the temporal dimension.

By using transfer learning and the hybrid architecture of CNNs and RNNs, the system can effectively extract spatial and temporal features from the video frames, enabling accurate classification and prediction of different activities or events.

5.3.1.1 Programming Language: Python

5.3.2 Libraries Used:

In the implementation of the project, several key libraries and frameworks were utilized to facilitate different tasks related to computer vision, video processing, and deep learning. These libraries provided essential functionalities and tools that contributed to the success of the project. Here are some of the main libraries and frameworks used:

- **OpenCV:** OpenCV is a widely-used library for computer vision and image processing tasks. It offers a comprehensive set of functions and algorithms for tasks such as video streaming, frame extraction, manipulation, and various image processing operations. OpenCV played a crucial role in efficiently processing and analyzing video streams, enabling tasks such as video acquisition, frame manipulation, and feature extraction.

- **NumPy:** The foundational Python library for scientific computing is called NumPy. It offers assistance for manipulating arrays and performing effective numerical operations. NumPy's array data structure was extensively used for processing and manipulating video frames, facilitating operations such as resizing, normalization, and feature extraction.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

It offered efficient and optimized functions for numerical computations, enhancing the performance of various data processing tasks.

- **Pandas:** Pandas is a potent library for handling and analysing data. It facilitates the efficient management and manipulation of structured data through the use of a high-level data structure called DataFrame. The dataset was organised and preprocessed using Pandas, which was also used to manage annotations and make data exploration easier. It made it simpler to deal with the dataset by offering useful capabilities for data manipulation, aggregation, filtering, and merging.

- **Decords:** Decords is a Python library specifically designed for handling video datasets. It provides functionalities for reading and manipulating video files, allowing efficient dataset preparation and processing. Decords facilitated tasks such as extracting clips from videos based on annotated timestamps, enabling the creation of a well-structured dataset for training and evaluation. It offered efficient video decoding capabilities, ensuring smooth and reliable access to video data.

- **TensorFlow:** For creating and training neural networks, many people utilise the well-known deep learning framework TensorFlow. It provides a wide range of tools and features for putting deep learning models into practise. TensorFlow was important in the project, making it easier to complete tasks like model creation, training, and inference.

VI. RESULTS

During the training process, mini-batch gradient descent was employed with a batch size of 32. The training dataset consisted of 672 samples per epoch, and each batch contained 32 alternate frames from the videos. These frames were passed through the pretrained InceptionV3 network to obtain feature vectors, which were then fed to the RNN model for classification.

The pretrained weights of InceptionV3, trained on the ImageNet dataset, were used as a starting point for the encoder model. The model was fine-tuned during the training process to adapt to the specific task of anomaly detection in videos.

To evaluate the performance of the trained model, a separate testing dataset consisting of approximately 5000 videos was used. This dataset was set aside before training to ensure an unbiased evaluation. The model was applied to the testing videos, and the validation accuracy was calculated.

The results showed an increase in validation accuracy with an increasing number of epochs. The highest achieved accuracy was 89.6% with epoch 37.

The RNN model was trained for a total of 40 epochs, and it contained around 2.5 million parameters. The number of parameters reflects the complexity of the model and its ability to learn and capture relevant information from the video sequences.

These training and evaluation details provide insights into the model's performance, its capacity to learn from the dataset, and the achieved accuracy in detecting anomalies in videos.

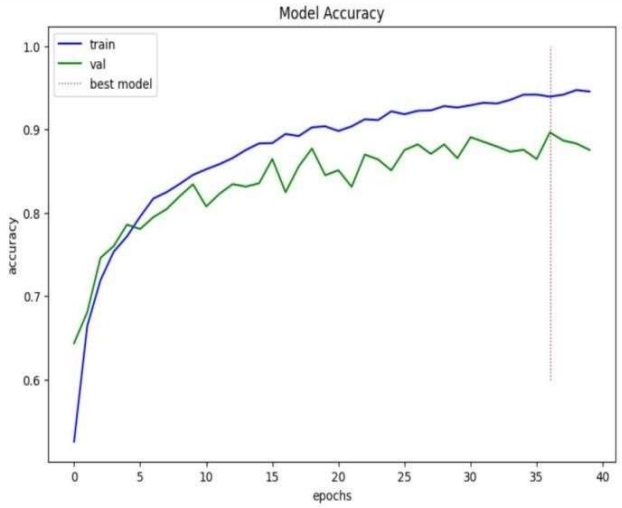

Fig 7: Accuracy v/s epochs

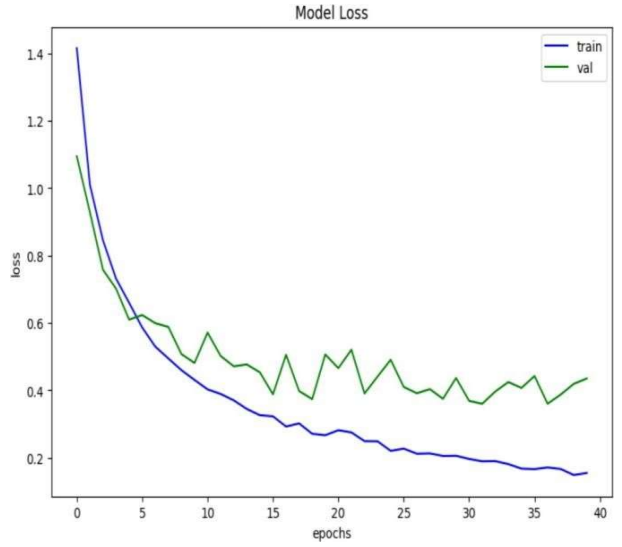

Fig 8: loss v/s epochs

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

Fig 9: Predicted Road Accident

Fig 10: Predicted Abuse

VII. CONCLUSION AND FUTURE SCOPE

In conclusion, the developed software solution has successfully demonstrated the effectiveness of computer vision and deep learning techniques in detecting criminal activities through live CCTV footage analysis. By leveraging technologies such as Python, TensorFlow, and OpenCV, the system has achieved real-time video processing, object recognition, and anomaly detection capabilities, enabling proactive identification of suspicious behavior.

The project has emphasized the importance of high-quality CCTV cameras and a suitable hardware configuration for smooth operation. However, there are opportunities for further advancements and expansions in the future. Some potential areas for development include:

1. Advanced Video Analytics: By incorporating advanced techniques like facial recognition, behavior analysis, and object tracking, the system can improve its accuracy and precision in detecting and identifying criminal activities.

2. Integration with AI-powered Systems: The system can benefit from integration with other AI-powered systems, such as automated surveillance systems or predictive analytics tools. This integration would enhance the system's capabilities and provide more comprehensive insights for crime prevention.

3. Cloud-Based Architecture: Moving towards a cloud-based architecture would enable scalability, flexibility, and easier integration with other systems. It would also facilitate remote access and collaboration among multiple stakeholders, including law enforcement agencies.

4. Collaboration with Law Enforcement Agencies: Building partnerships with law enforcement agencies would allow for better coordination and utilization of the system's capabilities in real-world scenarios. This collaboration could involve sharing data, integrating with existing infrastructure, and aligning with law enforcement protocols and procedures.

5. Predictive Analytics: By analyzing historical data and patterns, the system can develop predictive models to anticipate potential criminal activities and proactively alert authorities, enabling preventive measures.

VIII. REFERENCES

[1] Waqas Sultani, Chen Chen, Mubarak Shah. "Real-World Anomaly Detection in Surveillance Videos", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[2] K.A. Joshi, D.G. Thakore, A survey on moving object detection and tracking in video surveillance system, Int. J. Soft Comput. Eng. 2 (3) (2012) 44–48.

[3] MalavikaNair, Mathew Gillroy, Neethu Jose and Jasmy Davies in their “i- survelliance Crime Monitoring and Prevention Using Neural Networks”, International Research Journal of Engineering and Technology(IRJET), e- ISSN:2395-0056Volume: 05 03rd March 2018.

[4] Bharati, R.A.K. Sarvanaguru, Crime prediction and analysis using machine learning, Int. Res. J. Eng. Technol. 5 (9) (2018) 1037–1042.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

[5] H.W. Kang, H.B. Kang, Prediction of crime occurrence from multi-modal data using deep learning, PLoS One 12 (4) (2017), e0176244.

[6] S. Kim, P. Joshi, P.S. Kalsi, P. Taheri, Crime analysis through machine learning, 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE 2018, November, pp. 415–420.

[7] C.S. Sung, J.Y. Park, Design of an intelligent video surveillance system for crime prevention: applying deep learning technology, Multimed. Tools Appl.80 (26) (2021) 34297–34309.

[8] Ravinarayana, Shakhin, Shetty Pooja Jayaram, Shreya Shetty, Yashodha ambig, Crime Activity Detection Using Machine Learning, Crime Activity Detection Using Machine Learning, Volume 10 Issue VII July (2022) 2321- 9653

[9] Virender Singh, Swati Singh, Dr. Pooja Gupta, Real-Time Anomaly Recognition Through CCTV Using Neural Networks, ICITETM2020