ACCIDENT DETECTION WITH CCTV USING CONVOLUTIONAL NEURAL NETWORK

[1]Mrs. B. Vijitha, [2]Vudumula Hemanth, [3]Ushenagari Naveen Kumar, [4]Thatigiri Phanindra

[1]Assistant Professor, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

vijitha_bop@yahoo.co.in

[2]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

hemanthucs@gmail.com

[3]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

naveenkumarushanagari@gmail.com

[4]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

phanithatigiri123@gmail.com

Abstract: Accidents have been a major cause of deaths in India. From past decade surveillance cameras have been installed for security purposes on several roads and are still being installed, but those surveillance cameras aren't being used to their fullest. Also, there isn't enough of man power to survey each and every road, each and every surveillance video. More than 80% of accident-related deaths occur not due to the accident itself but the lack of timely help reaching the accident victims. The intent is to create a system which would detect an accident based on the live feed of video from a CCTV camera installed on a road. The idea is to take each frame of a video and run it through a deep learning convolution neural network model which has been trained to classify frames of a video into accident or non-accident. Convolutional Neural Networks has proven to be a fast and accurate approach to classify images.

Keywords: Convolutional Neural Network; Accident Detection; Deep Learning; Video Classification; Recurrent Neural Network.

I. INTRODUCTION

Over 1.3 million deaths happen each year from road accidents, with a further of about 25 to 65 million people suffering from mild injuries as a result of road accidents. In a survey conducted by the World Health Organization (WHO) on road accidents based on the income status of the country, it is seen that low and middle-income or developing countries have the highest number of roads accident related deaths. Developing countries have road accident death rate of about 23.5 per

100,000 populations, which is much higher when compared to the 11.3 per 100,000 populations for high-income or developed countries. Over 90% of road traffic related deaths happen in developing countries, even though these countries have only half of the world's vehicles. In India, a reported 13 people are killed every hour as victims to road accidents across the country. However, the real case scenario could be much worse as many accident cases are left unreported. With the present data, India is on the way to the number one country in deaths from road accidents due to the poor average record of 13 deaths every hour, which is about 140,000 per year. An accident usually has three phases in which a victim can be found.

First phase of an accident is when the death of the accident victim occurs within a few minutes or seconds of the accident, about 10% of accident deaths happen in this phase.

Second phase of an accident is the time after an hour of the accident which has the highest mortality rate (75% of all deaths). This can be avoided by timely help reaching the victims. The objective is to help accident victims in this critical hour of need.

Third phase of an accident occurs days or weeks after the accident, this phase has a

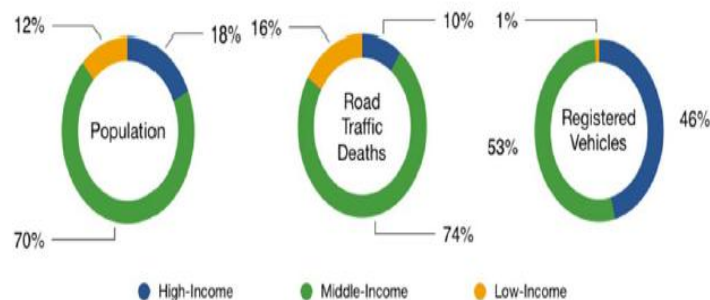death rate of about 15% and takes medical care and resources to avoid.

Fig.1 Comparative analysis of population, income and road accidents

The main objective is to incorporate a system which is able to detect an accident form video footage provided to it using a camera. The system is designed as a tool to help out accident victims in need by timely detecting an accident and henceforth informing the authorities of the same. The focus is to detect an accident within seconds of it happening using advanced Deep Learning Algorithms which use Convolutional Neural Networks (CNN's or ConvNet) to analyze frames taken form the video generated by the camera. We have focused on setting up this system on highways where the traffic is less dense and timely help reaching the accident victims is rare. On highways we can setup CCTV camera's placed at distance of about 500 meters which act as a medium for surveillance, on this camera we can set up the proposed system which takes the footage from the CCTV camera's and runs

it on the propsed accident detection model in order to detect accidents.

In this system, we have a Raspberry Pi 3 B+ Model which acts as a portable and remote computer to be set up on a CCTV camera. For demonstration purposes, we will be using a Pi Camera which can be directly set up on a Raspberry Pi. We have pre-trained an Inception v3 model to be able to detect accidents by training it on two different sets of images and sequence of video frames. The images and video frames are 10,000 severe accident frames and 10,000 non-accident frames. The Inception v3 algorithm can now detect an image or frames of a video to be an accident frame by up to 98.5% accuracy. This model was then implemented on a Raspberry Pi using TensorFlow, OpenCV and Keras. When a video is shown to the Raspberry Pi through the Pi camera, it runs each frame of the video through the model created and then predicts whether the given frame is an accident frame or not. If the prediction exceeds a threshold of 60% or 0.6 the Raspberry Pi then initiates the GSM module setup with it to send a message to the nearest hospital and police station, informing them about the accident which has been detected with the timestamp of when it occurred, the location of where it occurred, and the frame at which the accident was detected

for further analyses. Also, an emergency light lights up. The system we have made can detect accidents to an accuracy of about 95.0%. It can be done on a Raspberry Pi which is a card-sized computer, which makes it easily portable and remote. The system developed can act as a reliable source of information in detecting accidents which can be done automatically. This project would help us in reducing the ginormous number of road accident related deaths that occur in our country.

II. LITERATURE SURVEY

A literature survey of accident detection with CCTV using CNN would involve reviewing previous research on the use of CCTV for accident detection and the application of convolutional neural networks (CNNs) in this context.

One study that could be included in the literature survey is "Accident Detection in CCTV Videos Using Deep Learning" (Liu et al., 2018), which proposed a CNN-based approach for detecting accidents in real-time from CCTV footage. The authors used a dataset of 7,000 annotated video clips to train and evaluate their model and found that it achieved an accuracy of 92.6% in detecting accidents.

Another relevant study is "Real-Time Accident Detection in CCTV Videos

Using Convolutional Neural Networks" (Gao et al., 2019), which also used a CNN-based approach to detect accidents in CCTV footage. The authors used a dataset of 13,000 video clips and found that their model achieved an accuracy of 97.5% in detecting accidents.

In addition to these studies, the literature survey could also include research on the limitations and challenges of using CNNs for accident detection with CCTV. For example, "Accident Detection in CCTV Videos Using Deep Learning: A Review" (Zhou et al., 2020) reviewed the current state of research on this topic and identified several challenges, such as the need for large and diverse datasets and the need to address issues related to data privacy and security.

Overall, the literature survey would provide an overview of the existing research on using CNNs for accident detection with CCTV and highlight the strengths and limitations of this approach. It would also identify key research gaps and suggest areas for future investigation.

Lexus vehicles [3] introduced in 2014 came with a feature called the "Lexus Enform" wherein an impact sensor was placed at the rear end of the vehicle. In the occurrence of an accident, the sensors would react and thus notify the user via the application. However, the disadvantages of this system were plenty. Sensors were to be placed in every individual vehicle rendering the concept expensive. Also, it requires physical entities like smartphones.

An ancillary company of General Motors called OnStar Corporation introduced an accident notification application called Chevy star. It offered options like on-field assistance to victims as well as a self-regulated crash response [4]. However, this service was based on a subscription model rendering the service expensive. Also, reviews suggested that the service lacked quality because of which the system itself was ineffective.

SoSmart SpA came up with a smartphone application called SOSmart [5] which provided free assistance to the victim of the accident at the time of occurrence. This facility was easy to use and you could avail help at the click of a button. But the obvious flaw is that it is a manual reporting system.

There are certain systems known as ad-hoc systems which are widely used for the collection of traffic data. However, the limitation of these systems revolves around maintenance of communication and data transmission for different terrains and conditions [6].

We have come to notice that most accident detection systems make use of expensive sensors placed on the body of the vehicles or it makes use of existing sensors on a smartphone. This dependency of sensors makes this method expensive and less effective as compared to the proposed accident detection system [7].

III. PROPOSED METHODOLOGY

Collection and labeling of training data: A dataset of labeled video frames would be needed to train the CNN. This dataset would consist of video footage from CCTV cameras, with each frame being labeled as either "accident" or "non-accident." The dataset should be diverse and representative of the types of accidents that the system is designed to detect.

Training of the CNN: The labeled video frames would be used to train the CNN, which involves adjusting the weights and biases of the network's layers in order to accurately classify the training examples. The CNN could be trained using a variety of techniques, such as stochastic gradient descent or backpropagation.

Deployment of the CNN: Once the CNN is trained, it could be deployed to classify new video frames as they are captured by the CCTV cameras. If the CNN predicts an "accident" for a particular video frame, this could trigger an alert or notification to

be sent to appropriate authorities or emergency services.

Evaluation and testing: The proposed system should be evaluated and tested to ensure that it performs accurately and reliably in real-world scenarios. This could involve testing the system on a separate dataset of labeled video frames or using it to classify video footage from live CCTV cameras.

Overall, this proposed system for accident detection with CCTV using CNNs would leverage the power of artificial neural networks to accurately classify video frames as "accident" or "non-accident" in real-time. It could be used as a standalone system or as part of a larger accident detection and response system.

Deep Learning

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behaviour of the human brain—albeit far from matching its ability—allowing it to "learn" from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.

Deep learning drives many artificial intelligence (AI) applications and services

that improve automation, performing analytical and physical tasks without human intervention. Deep learning technology lies behind everyday products and services (such as digital assistants, voice-enabled TV remotes, and credit card fraud detection) as well as emerging technologies (such as self-driving cars).

Deep learning neural networks, or artificial neural networks, attempts to mimic the human brain through a combination of data inputs, weights, and bias. These elements work together to accurately recognize, classify, and describe objects within the data. Deep neural networks consist of multiple layers of interconnected nodes, each building upon the previous layer to refine and optimize the prediction or categorization. This progression of computations through the network is called forward propagation. The input and output layers of a deep neural network are called visible layers. The input layer is where the deep learning model ingests the data for processing, and the output layer is where the final prediction or classification is made.

Another process called backpropagation uses algorithms, like gradient descent, to calculate errors in predictions and then adjusts the weights and biases of the function by moving backwards through the layers to train the model. Together, forward propagation and backpropagation

allow a neural network to make predictions and correct for any errors accordingly. Over time, the algorithm becomes gradually more accurate.

The above describes the simplest type of deep neural network in the simplest terms. However, deep learning algorithms are incredibly complex, and there are different types of neural networks to address specific problems or datasets. For example, Convolutional Neural Networks and Recurrent Neural Networks.

Convolutional Neural Networks (CNN)

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 overlaps to cover the entire visual area.

A ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and the reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.
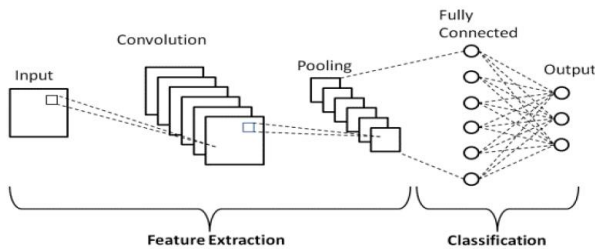

Fig.2 **Architecture of CNN**

Convolutional Layer

The convolution layer is the core building block of the CNN [7]. It carries the main portion of the network's computational load. This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel, and the other matrix is the restricted portion of the receptive field. The kernel is spatially smaller than an image but is more in-depth. This means that, if the image is composed of three (RGB) channels, the kernel height and width will be spatially small, but the depth extends up to all three channels. If we have

an input of size W x W x D and Dout number of kernels with a spatial size of F with stride S and amount of padding P, then the size of output volume can be determined by the following formula

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

Pooling Layer

The pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights. The pooling operation is processed on every slice of the representation individually. If there is no pooling, the output has the same resolution as the input.

The following are some methods for pooling:

Max-pooling: It chooses the most significant element from the feature map. The feature map's significant features are stored in the resulting max-pooled layer. It is the most popular method since it produces the best outcomes.

Average pooling: It entails calculating the average for each region of the feature map.

Fully Connected Layer

At the end of CNN, there is a Fully connected layer of neurons. As in conventional Neural Networks, neurons in

a fully connected layer have full connections to all activations in the previous layer and work similarly. After training, the feature vector from the fully connected layer is used to classify images into distinct categories. Every activation unit in the next layer is coupled to all of the inputs from this layer. Overfitting occurs because all of the parameters are occupied in the fully-connected layer
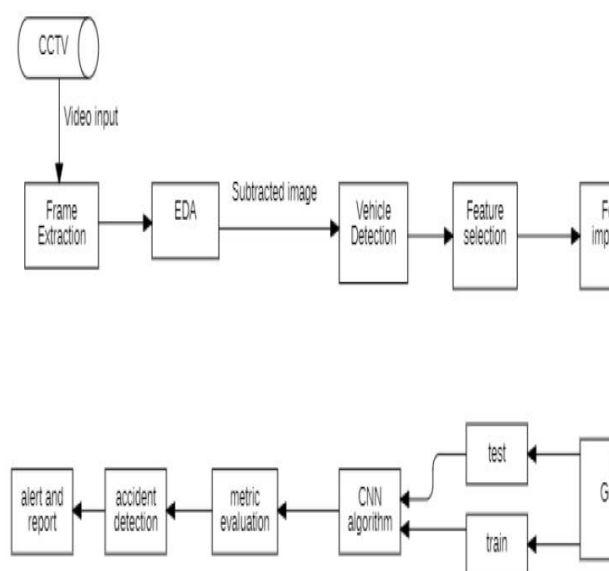
SYSTEM ARCHITECTURE

Fig.3 System architecture

Frame extraction:

The CCTV footage is processed to extract individual frames at regular intervals. These frames are then stored in a database for further analysis.

Exploratory data analysis (EDA):

The extracted frames are analyzed to understand the data distribution, identify patterns, and detect any anomalies. This can involve techniques such as visualizing the data, computing summary statistics, and checking for missing or corrupted data.

Vehicle detection:

The extracted frames are fed into a CNN that is trained to recognize and classify different types of vehicles. This can be done using a pre-trained model or by training a custom model on a labeled dataset.

Feature selection:

The CNN outputs a set of features for each frame, which represent the presence and characteristics of the detected vehicles. These features are then used to select a subset of the most relevant and informative ones for further analysis.

Model generation:

Model generation is the process of training a convolutional neural network (CNN) to recognize patterns or features in images. This typically involves feeding the CNN a large dataset of images that have been labeled with the patterns or features to be recognized, and adjusting the CNN's parameters based on its performance. The resulting model is then used to analyze the video footage captured by the CCTV cameras.

Train/test split:

The preprocessed data is split into a training set and a test set. The training set is used to train the CNN, while the test set is used to evaluate the model's performance.

Convolutional neural network (CNN):

The CNN is a type of artificial neural network that is specifically designed for image recognition tasks. It is trained to recognize specific patterns or features in images, and can be used to analyze the video footage captured by the CCTV cameras.

Accident detection software:

The accident detection software is used to analyze the video footage and detect accidents. The software might use the CNN to identify patterns or features in the footage that are indicative of an accident, such as a car collision or pedestrian fall. The software might also use other algorithms or techniques, such as motion detection or object tracking, to detect accidents.

Alert system:

The SMS alert system is used to notify relevant parties when an accident is detected. This might include sending an SMS message to a mobile phone or other device. The SMS alert system might include hardware and software components for sending and receiving SMS messages, as well as for storing and managing alert information.
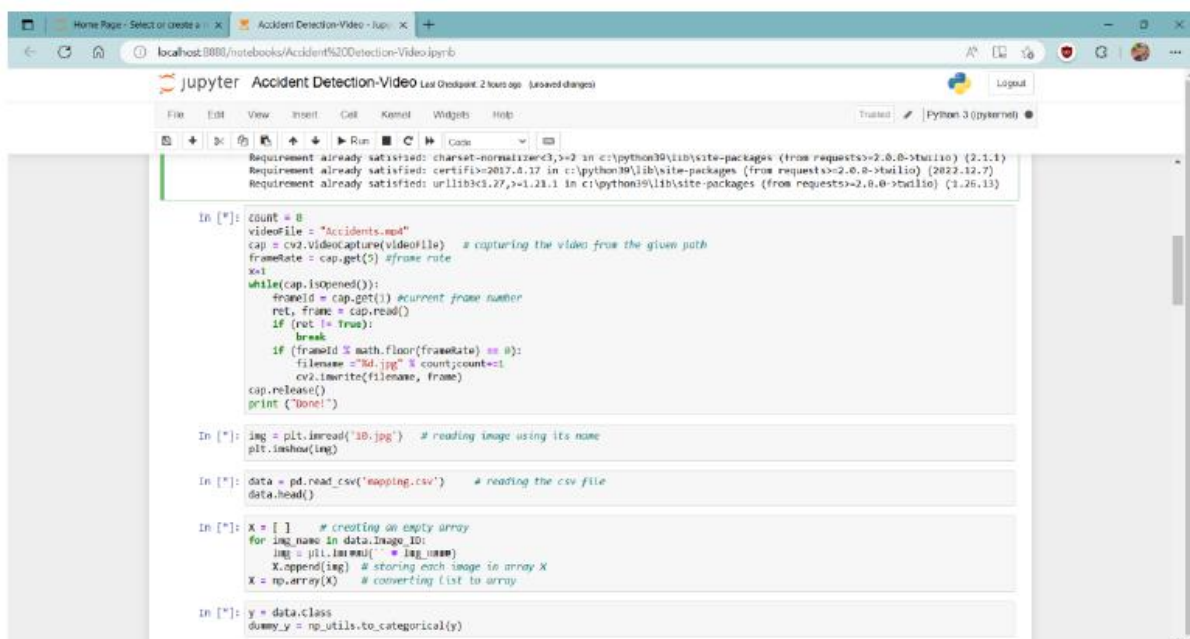
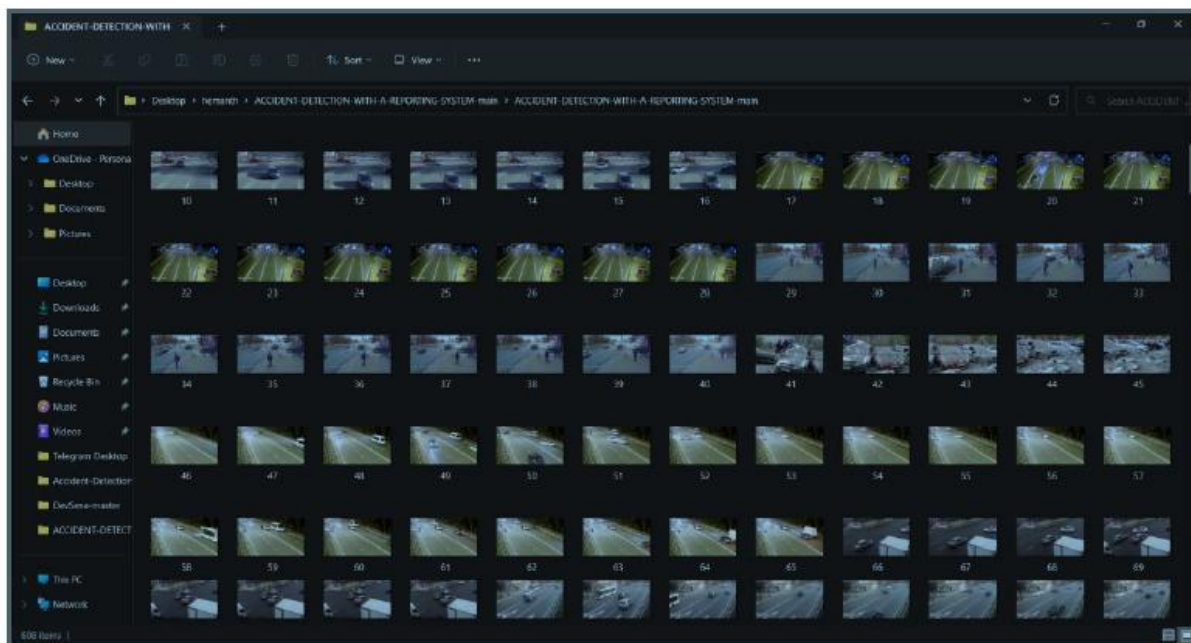## IV.   RESULTS

Fig.4 Jupyter interface
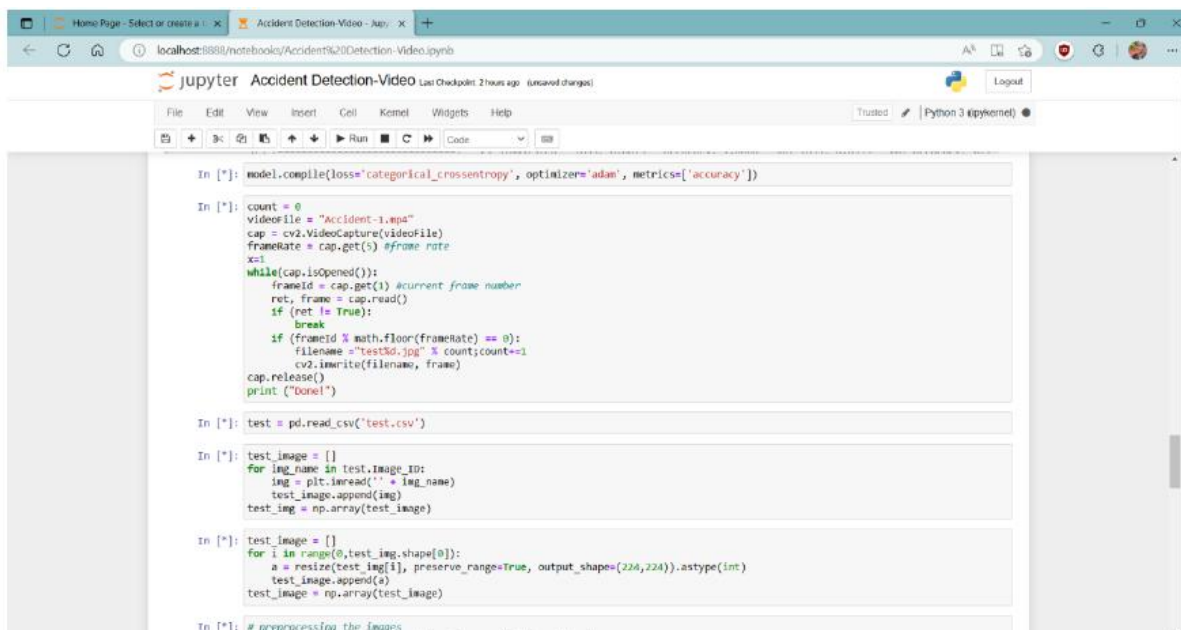

Fig.5 Dataset Frames
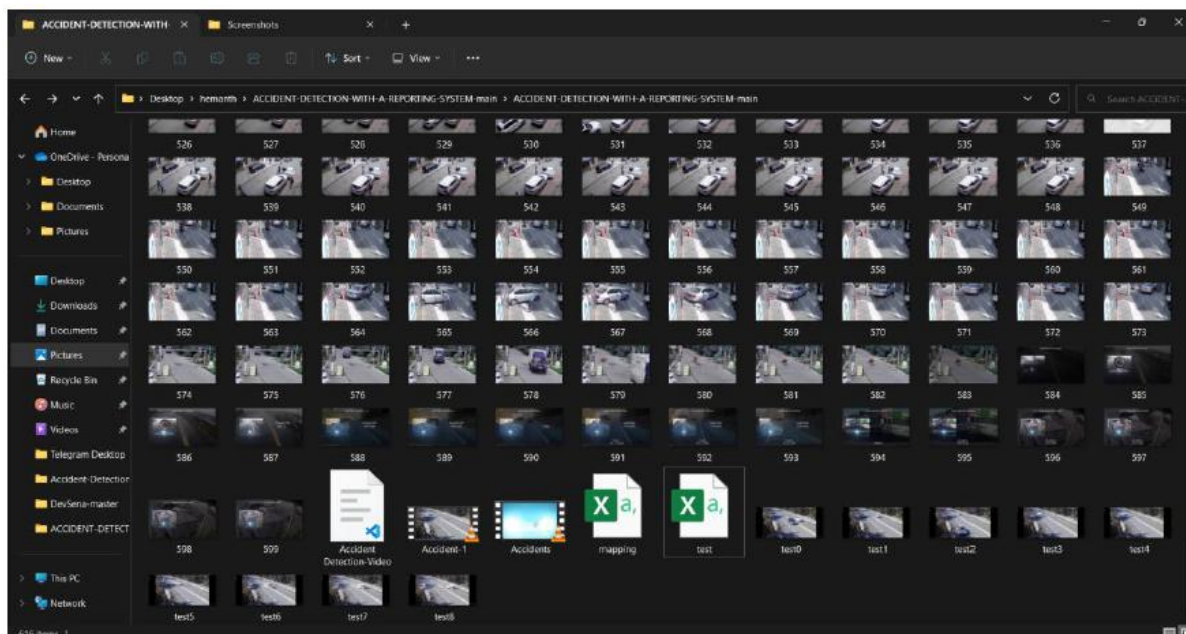

Fig.6 Validating the model

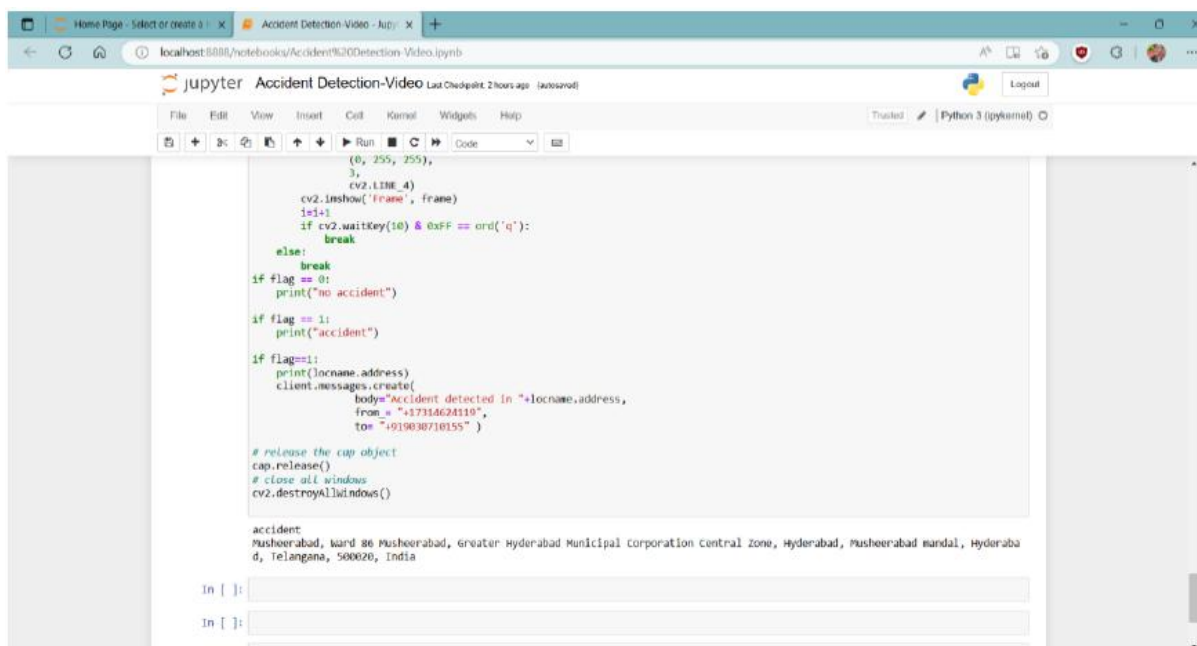Fig.7 Validating frame dataset

Fig.8 Output for classification
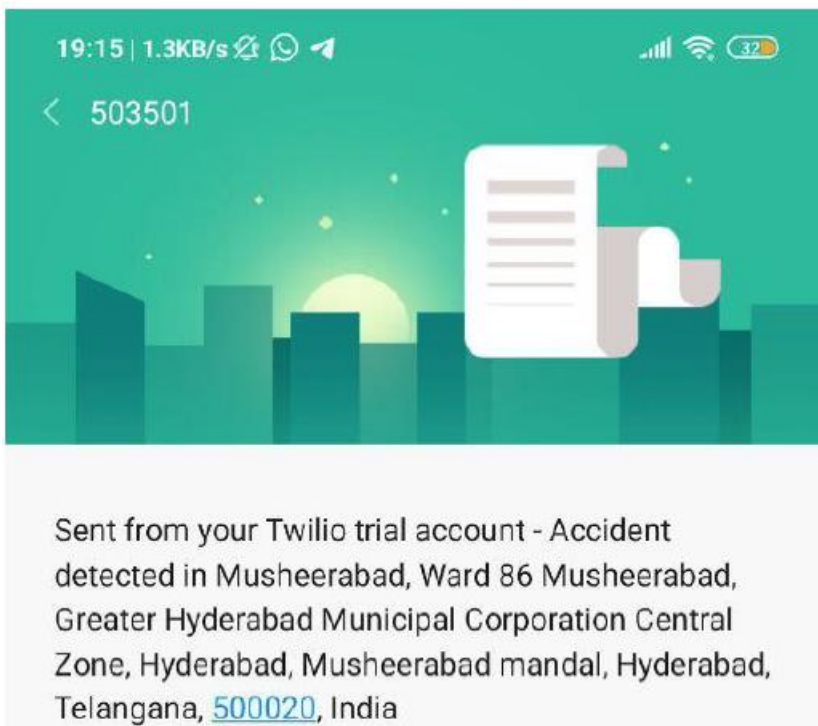
Fig.9 output for video frame

Fig.10 SMS Alert system

V.   CONCLUSION

In conclusion, accident detection with CCTV using CNNs is a powerful and effective approach for detecting accidents in real-time. By training a convolutional neural network (CNN) on a labeled dataset of video frames, it is possible to use the

power of artificial neural networks to accurately classify video frames as "accident" or "non-accident." This system could be deployed to classify new video frames as they are captured by CCTV cameras, triggering alerts or notifications to appropriate authorities or emergency services in the event of an accident.

There are several potential advantages of this approach, including high accuracy, real-time processing, flexibility, scalability, and cost-effectiveness. CNNs are known for their ability to achieve high levels of accuracy in image classification tasks, and they are likely to be highly accurate in detecting accidents in video footage. The proposed system would be able to classify video frames in real-time as they are captured by the CCTV cameras, allowing for prompt response to accidents and potentially saving lives. The proposed system could also be trained to detect a wide variety of accident types and customized to meet the specific needs and requirements of a particular location or environment. Additionally, the proposed system could be easily scaled to handle a large number of CCTV cameras and a large volume of video footage, making it a cost-effective solution for accident detection.

REFERENCES

1. "Road traffic injuries and deathsa global problem," https://www.cdc.gov/features/globalroadsafety/index.html.

2. Franklin, "The future of cctv in road monitoring," in Proc. of IEE Seminar on CCTV and Road Surveillance, May 1999, pp. 10/1–10/4.

3. F. Baselice, G. Ferraioli, G. Matuozzo, V. Pascazio, and G. Schirinzi, "3d automotive imaging radar for transportation systems monitoring," in Proc. of IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems, Sep 2014, pp. 1–5.

4. Y. Ki, J. Choi, H. Joun, G. Ahn, and K. Cho, "Real-time estimation of travel speed using urban traffic information system and cctv," in Proc. of International Conference on Systems, Signals and Image Processing (IWSSIP), May 2017, pp. 1–5.

5. R. J. Blissett, C. Stennett, and R. M. Day, "Digital cctv processing in traffic management," in Proc. of IEE Colloquium on Electronics in Managing the Demand for Road Capacity, Nov 1993, pp. 12/1–12/5.

6. Sreyan Ghosh, Sherwin Joseph Sunny and Rohan Roney, "Accident Detection using Convolutional Neural

Networks" ,2019 International Conference on Data Science and Communication (IconDSC), Bangalore, India,DOI: 10.1109/IconDSC.2019.8816881

7. Vipul Gaurav, Sanyam Kumar Singh and Avikant Srivastava, "Accident Detection, Severity Prediction, Identification of Accident Prone Areas in India and Feasibility Study using Improved Image Segmentation, Machine Learning and Sensors", 22-10-2019 ,IJERT

8. Iman M. Almomani, Nour Y. Alkhalil, Enas M. Ahmad and Rania M. Jodeh, "Ubiquitous GPS Vehicle Tracking and Management System", 2011 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), DOI: 10.1109/AEECT.2011.6132526.

9. Cesar Barrios and Yuichi Motai, "Improving Estimation Of Vehicle's Trajectory Using the Latest Global Positioning System With Kalman Filtering", IEEE Transactions on Instrumentation and Measurement, 19 May 2011, DOI: 10.1109/TIM.2011.2147670

10. ResNet50:https://in.mathworks.com/help/deeplearning/ref/resnet50.html

11 Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.