Artificial Intelligence driven Traffic Managementand Accident Detection System using CCTV Cameras

Deepak Khandelwal
Computer Science and Engineering
Chandigarh University
Chandigarh, India
1992.deepak.khandelwal@gmail.com

Ashwin George
Computer Science and Engineering
Chandigarh University
Chandigarh, India
ashwindominic1@gmail.com

Farha Khan
Computer Science and Engineering
Chandigarh University
Chandigarh, India
fk338711@gmail.com

Nisha Singh
Computer Science and Engineering
Chandigarh University
Chandigarh, India
Nishasingh00506@gmail.com

***Abstract*— A traffic management system is the cornerstone of a Smart City. In the current problems of the world, urban mobility and traffic accidents are some of the major problems, especially in metropolitan cities. This paper proposes a project that is divided into two parts: dynamic traffic control system and accident detection system. The main focus of dynamic traffic control system is to control the traffic dynamically based on traffic density. While, the accident detection system will get the live CCTV footage from various CCTV cameras that are present on the road and this footage will be processed on a local server by an AI model and will detect the accident. The CCTV footage will be telecasted on a web-page that is accessed by authorized personnel only. Whenever any accident is detected, footage from that particular CCTV camera will be highlighted. It is worth mentioning that both these systems work independently. The proposed model will work in regions that have high internet speed and with a good network of CCTV cameras.**

Keywords—traffic management system, accident detection system, local-server, dynamic, AI model

I. Introduction

Artificial Intelligence refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem solving.[1]

The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal. A subset of artificial intelligence is machine learning, which refers to the concept that computer programs can automatically learn from and adapt to new data without being assisted by humans. Deep learning techniques enable this automatic learning through the absorption of huge amounts of unstructured data such as text, images, or video.[1]

Computer vision is the field of computer science that focuses on replicating parts of the complexity of the human vision system and enabling computers to identify and process objects in images and videos in the same way that humans do. Until recently, computer vision only worked in limited capacity.[2]

The applications of artificial intelligence, machine learning, computer vision and deep learning are endless. The technology can be applied to many different sectors and industries.[1] Two such applications are traffic management system and accident detection system. Deep learning and Computer vision are the key technologies that would help this traffic management and accident detection in real-time.

A. Traffic Managemnet System

Over the years, the traffic management policies have changed and now they are inclining more towards the safety aspect. Now, authorities want a more systemized way to monitor road traffic. To enhance the flow of vehicles on the road and to offer complete safety to the people, an effective traffic management system is what we need.

Proper traffic management can ensure that traffic flows smoothly and efficiently; there is fair access for different transport modes; roads and streets are safe for all users; roads full of motorised traffic do not constitute barriers blocking movement between areas; congestion, local pollution and noise are minimised; neighbourhoods, pedestrian areas and the overall character of localities are protected from the negative impact of high traffic levels; and greenhouse gas is reduced.

This paper proposes a system that will first count the number of vehicles in a particular lane using Computer vision, using which it will calculate the vehicle density and using the vehicle density the system will calculate for how much time the traffic signal should be green. This will reduce the waiting time at the signals and will provide a better flow of traffic at junctions.

B. Accident Detection System:

Accident detection system is an application of artificial intelligence and machine learning, in which, data from various sources like CCTV camera, car's dashboard camera, audio recording from other various sources is used to detect the accident. We are quite fortunate to have well-prepared emergency services that respond to traffic accidents at a moment's notice. But what would happen in an incapacitating accident at 3 am with no bystanders to report it? This paper aims to answer this question by helping our target audience by using Machine Learning and Computer Vision to detect traffic accidents autonomously in a split second*. Detecting otherwise unreported accidents will create safer roads and a more efficient system for the civil defence, devoid of human error.

Our solution is designed to be as effortless and inexpensive as possible to setup, especially since it will simply run-on top of pre-existing, pervasive road CCTV infrastructure. This keeps the costs of this solution very low as it does not require any dramatic paradigm shifts before it can be of use.

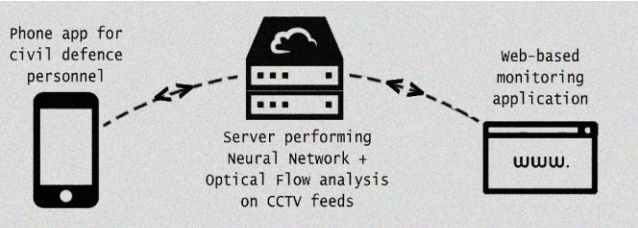

Fig 1: Interaction of website, application and server.

The key features of our solution include an immediate notification system that can autonomously detect potential accidents and can inform relevant civil personnel through a mobile or web alert. We believe this instantaneous feedback system can make all the difference between life and death in even the most severe and critical accidents.

II. Literature Review

Cheng Wenjie *et al.* introduced a Wireless Sensor based Network (WSN) model form making a dynamic traffic control signal. It proposed an algorithm based on the number of vehicles, speed of individual vehicles, expected duration of vehicles to cover the length, etc. in the waiting queue to find the efficient pattern of Traffic Phase Shifting. This is a sensor-based system that consists of three modules vehicle node, relay node and control node. In this network the vehicle node in each vehicles sends information like vehicle speed and relative position to the relay node and then it relays this data and some calculated parameters to the control node which is directly connected to the traffic light controller. It runs the algorithms with the received parameters and control the traffic lights for the efficient flow of traffic.[3]

P Manikonda *et al.* presented an intelligent traffic management system by using RFID Technology. The system had a passive tag, an RFID reader, a microcontroller, a GPRS module, a high-speed server with a database and a user module. In this system every vehicle has an RFID tag which is unique for every vehicle. Every vehicle also has an infrared sensor that sends signals which is received by RFID Reader module. The average speed of the vehicle is calculated by the time taken by the vehicle to cross two such RFID reader modules located at certain distance. Then this data is sent to a Local High-speed server using a GPRS module. Using this data, the local server calculates the shortest path for the vehicle to reach its destination using Dijkstra's algorithm and is sent to the vehicle using GRPS module.[4]

P S Chakraborty *et al.* primarily focused on determining green light duration. This research paper proposed an algorithm which not only determines green light duration but also handles the emergency vehicle management efficiently. It also handles the deadlock and starvation condition, which causes due to arrival of emergency vehicle in repeated interval of time in traffic intersection. Wireless Sensor Networks technology had been considered as the possible source of input.

Y R Chen *et al.* proposed a Model Predictive Control based Traffic Light Control System (MPCTLCS). MPCTLCS includes a traffic flow prediction model and a traffic timing optimization method. Historical traffic data issued to predict future traffic volumes. An MPC-based traffic light optimization method is proposed to obtain appropriate time settings that can reduce overall congestion. This method also has the ability to dynamically adjust traffic signal timings. It can rapidly respond to real-time traffic conditions to reduce traffic congestion.[5]

W J Chang *et al.* proposed the idea of deep learning-based Internet of Vehicles system called DeepCrash, which includes an in-vehicle infotainment (IVI) telematics platform with a vehicle self-collision detection sensor and a front camera, a cloud-based deep learning server, and a cloud-based management platform.
When a head-on or single-vehicle collision is detected, accident detection information is uploaded to the cloud-based database server for self-collision vehicle accident recognition, and a related emergency notification is provided.[6]

J Singh *et al.* proposed the idea of detection of vehicle accident using Internet of Things (IoT). The proposed system was developed using a vibration sensor to determine the collision impact of an accident and a gyro sensor to determine the x-y displacement of the vehicle. When an accident occurs, the instantaneous coordinates of the vehicle will be captured using a GPS module and transmitted to the emergency response department via a GSM module. The coordinates are visualized on a registered mobile phone at the emergency response department and mirrored to a desktop's Pushbullet application. With that, necessary emergency response units can be deployed to the accident location.[7]

K P Sampoornam *et al.* proposed an intelligent expeditious accident detection and prevention system. The proposed system detects the exact location of the vehicle automatically using GPS and uses GSM to send the data to nearest ambulance vehicle, the police station and family members of that person. The occurrence of accident can be detected using accelerometer sensor and severity of the accident is detected using tilt sensor. It helps to monitor the tilt angle of the chassis. The drowsiness of the driver is detected using drowsiness detector. By measuring the percentage of eyelid closure over the time, the driver's drowsy state is detected and warning is given to the driver through the alarm. By this method we can measure the speed of a vehicle which gives reason for the accident and vehicle GPS location can be view through mobile application for theft detection. All the data can be processed in microcontroller with help of sensor value and send location to cloud.[8]

M Kumar J *et al.* proposed a cost-effective road accident prevention system. The proposed system will be mounted on a car, in front of the driver's seat, which will continuously focus on the person driving the vehicle thereby monitoring the person's actions and give voice messages to the driver, instructing him about the safety procedures every time he comes inside the car. If he found to be drowsy, he is alerted by a voice message telling him that he is drowsy and he should stop the car. If he is found to be under the influence of alcohol or tailgating while driving, the GPS coordinates of the car along with the details of the car, are sent to the nearest control room, so that appropriate action can be taken by the police.[9]

S Eduku *et al.* proposed a design of vehicle accident prevention system by using wireless technology. The proposed system takes in-depth look at vehicle accident prevention system using wireless technology, eye blink sensor and automatic braking system to ensure that the vehicle slows down and comes to a halt when drowsiness is detected, and the system (circuit) is not reset within the threshold period programmed in the microcontroller of the system. The wireless technology which is the backbone of this project work was achieve through radio frequency wave, is then used to send an information (slow down there is halt car ahead) to other vehicles at a transmission distance of wavelength 0.69m with a frequency of 433MHZ. simulation software (Proteus) was used to critically analyzed the model of the design of the vehicle accidents prevention system using wireless technology. [10]

III. Problem Statement

A. *Statement*

Traffic congestion is a serious issue in major developing cities. Traffic congestion problems consist of incremental delay, vehicle operating costs such as fuel consumption, pollution emissions and stress that result from interference among vehicles in the traffic stream, particularly as traffic volumes approach the road's capacity.

Due to the traffic congestion, the accidents are also increasing day by day. This causes the loss of life due to the delay in the arrival of ambulance to the accident spot or from the accident spot to the hospital. So, it necessary to take the accident victim to the hospital as possible.

B. *Background*

1) According to the Global Status report on Road Safety 2018, road traffic accidents are a major cause of death and injury across the globe, killing more than 1.35 million people globally, with 90% of those fatalities occurring in developing countries and 11% alone occurring in India. Based on Road Accident Report for 2019, there were 1,51113 deaths and 451,363 injuries attributed to road accidents in India during calendar year 2019. [11]

2) According to the report by the Texas A&M Transportation Institute, the average American commuter wastes 54 extra hours a year in traffic delays. By "extra hours" they mean the extra time spent traveling at congested speeds rather than free-flow speeds. According to the report, approximately 33 percent of total delay occurs in the midday and overnight (outside of the peak hours) times of day when travellers and shippers expect free-flow travel.[12]

3) According to a report by TomTom, a Netherlands-based company, the residents of Bengaluru lose 243 hours on average every year due to traffic. Four Indian cities, including Bengaluru, have been ranked among the top 10 most-congested cities in the Traffic Index prepared by TomTom that covered 416 cities across 57 countries on 6 continents. Mumbai is on number four, followed by Pune on number 5 and New Delhi on Number 8.[13]

C. *Relevance*

The cost of extra fuel spends by an average commuter in America in 2017 because of the traffic was $1,010. According to the report by the Texas A&M Transportation Institute, the average commuter will spend 62 hours in traffic by the year 2025, and the national congestion in America cost will balloon to $200 billion that same year, a 20% increase over the $166 billion related to traffic costs in 2017.[12]

India being one of the most populated countries in the world and a developing nation will face a rapid increase in traffic jams and in death tolls due to accidents. According to World Bank, India accounts for 11 percent of global death in road accidents and the highest in the world. The country accounts for about 4.5 lakh road crashes per annum, in which 1.5 lakh people die. A recent study commissioned by the Ministry of Road Transport and Highways (MoRTH) estimates the socio-economic costs of road crashes at Rs 1,47,114 crore in India, which is equivalent to 0.77 per cent of the country's GDP.[14]

D. *Objectives*

1) This research aims to develop a traffic management system that would control the traffic signals dynamically so that vehicles will have to spend less time on signals.

2) This research also aims to develop an accident detection system which would detect accidents and will notify the nearest emergency service.

IV. Proposed Method

A. *For Traffic Manageemnt System*

The proposed method for traffic management is based on video analytics technique. In particular, deep learning neural network architectures are trained for the detection of vehicle are used. The proposed method comprises four modules that are: 1. Segmentation of video into images
2. Detection of edges of vehicles
3. Calculation of traffic density according to the count of the vehicle
4. Controlling the traffic signals dynamically.

1) *Segmentation of video into images:*

The proposed method will first divide the video input into images using OpenCV. Basically, the video will be divided into small fragments of images in which the background of the image is blurred and the image is in grey-scale format.

2) *Detection of the edges of the vehicles*

Next, the processed image is sent for edge detection. For this, the technology of Computer Vision is used. A rectangle is made around every thing that is visible in the image. But the rectangle whose length and breadth are greater than a particular limit, is detected.

3) *Calculation of traffic density according to the count of the vehicle*

A line is drawn on the image and an offset limit is provided to that line. If the center of the detected rectangle is in the range of offset limit of line, then the counter is increased.

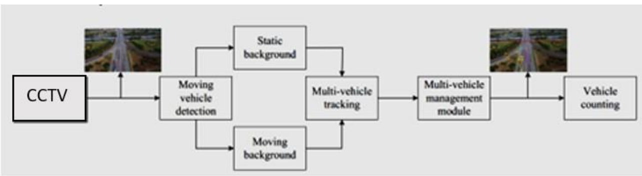

Fig 2: Framework of the proposed method. It consists of vehicle detection, multi-vehicle tracking, multi-vehicle management, and vehicle counting. The CCTV is equipped with a visual sensor. Vehicles are detected by the detector which can handle two situations: static background and moving background. Then, the detected vehicles are tracked by tracking module. By analyzing the results of the tracker, we can count the number of vehicles.

This is how the counting of vehicles is done. After counting the number of vehicles, next step is to calculate the vehicle density. Vehicle density is calculated by the formula:

$$vehicle\ density = \frac{vehicle\ count}{time\ (in\ seconds)} \quad \text{....... equation (1)}$$

4) *Controlling the traffic signals dynamically*

Now, the next task is to control the traffic lights using this calculated vehicle density. This can be achieved by calculating the ratios of vehicle densities of all the roads heading towards junction. Now, the ratio with the greatest value will be given higher priority and signal will be open for more time for that particular road.

B. *For Accident Detection System*

The proposed method is based on video analysis techniques. In particular, deep learning neural networks architectures are trained to detect the occurrence of a traffic accident are used. Since a video has to be processed, it has to be separated into segments. Therefore, the temporal segmentation of the video required a basic analysis to determine which was the most appropriate scheme to generate the segments, considering a tradeoff between the computational cost of processing the segment and the generation of enough visual characteristics to extract patterns that the network learned. Once the input was defined, the accident event was built as the occurrence in time of a set of visual patterns. For this, the architecture has two parts. The first one extracts a vector of visual characteristics using a modified Inception V4 architecture; this set of characteristics is processed by a recurrent component to extract the temporal component associated with the occurrence of the event. Next, we describe the two stages: temporal video segmentation and automatic detection of traffic accidents

1) *Temporal Video Segmentation Process*

Temporal video segmentation is a problem that has been studied for many years by the scientific community since it is the first step towards the development of more general solutions, such as scene understanding of videos[15]. A video is a sequence of consecutive images with a particular order. When these images are viewed in the correct order and at a specific speed, it is possible to observe the animated event represented by the recorded video [16].

A video camera can capture, with the help of a mechanism, an event that is happening at the moment, in order to store, observe, and process it in the future. Using the same concept of a digital camera, a video camera makes it possible to capture a number of photographs per second, thus allowing the event that is occurring to be digitally recorded. These images, which represent the video, are known as frames. Video cameras allow recording at different numbers of frames per second (FPS). This means that the higher the number of FPS, the more fluid the movement of the objects on screen. The most commonly used FPS values are 30, 60, and 120 [17].

Video segmentation can be divided into two categories: spatial and temporal. Spatial segmentation seeks to visually classify objects of interest in the video in order to spatially locate objects in different frames. This type of segmentation is very useful when tracking objects in a video [18]. On the other hand, time segmentation seeks to solve a different problem. One of these is the reduction in the video time. This can be achieved by dividing the video into multiple fixed time windows in order to transform a long-duration video into a finite number of short-duration videos. However, there are also cases where it is not possible to divide the video into fixed time windows because it may contain multiple events in it, and, if a fixed window division is performed, it is possible that an event will be divided into different segments. In order to be able to discriminate between the multiple situations that the video describes, the scientific community has developed techniques to temporally segment this data, taking into account the scene change frame [16]. This allows a single video to be divided into N temporal segments, where N is the number of scenes that can be observed [16].

Traffic accidents are rare events of short duration. In order to be able to detect them correctly through a video, it is important to preprocess the original recording considering that the video (using a static surveillance camera) will contain many frames with a high similarity index. For this reason, several temporal video segmentation techniques are proposed in order to compare them with the results of the detector model [16].

Three techniques were used to increase the variety of the data segments. The first is based on the metric named the Structural Similarity Index Measure, SSIM (Equation (2)), applied between consecutive frames in order to eliminate those that exceed an empirically defined threshold δss. The threshold is arbitrarily set at a high value ($\delta ss \geq 0.98$), representing a high similarity of the frames, which implies no additional information for the analysis. With this, we could significantly reduce the number of similar frames in the video segment [16].

$$SSIM(F_a, F_b) = \frac{(2\mu F_a \mu F_b + c_1)(2\sigma_{F_a F_b} + c_2)}{(\mu^2_{F_a} + \mu^2_{F_b} + c_1)(\sigma^2_{F_a} + \sigma^2_{F_b} + c_2)} \quad \text{..........equation (2)}$$

where F_a and F_b are two consecutive frames from the video segment, and μ_{Fi} is the intensity mean of all the pixels from frame i. The σ_{FaFb} factor is the covariance between the pixel value from frame a and b, and σ^2_{Fi} is the variance from frame F_i. The $c_1=(k_1L)^2$ and $c_2 = (k_2L)^2$. L is the dynamic range from pixel values ($2^{\#bit\_per\_pixel} - 1$), $k_1 = 0.01$ and $k_2=0.03$. The next technique used consists of a comparison very similar to the previous one. However, this one-use pixel-to-pixel comparison (Equations (2) and (3)) on consecutive frames, thus eliminating those that exceed the empirically defined threshold ($\delta pp \geq 0.9$) [16].

$$PtP(F_a, F_b) = \frac{1}{whk} \sum_{i=1}^{w} \sum_{j=1}^{h} \sum_{c=1}^{k} int\left(1 - \left|F_{aijc} - F_{bijc}\right|\right) \quad \text{......equation (3)}$$

$$int(v) = \begin{cases} 1 & if\ v = 1 \\ 0 & else \end{cases} \quad \text{...... equation (4)}$$

where F_a and F_b are two consecutive frames of w rows, h columns, and k layers. Therefore, F_{qijc} is the pixel value from the q frame in the position (i, j) in layer c. The last technique performs a selection using a fixed skip window. That is, a value is defined for K, which represents the number of frames that must be eliminated before selecting the next candidate to form the segment. If K is equal to 1, it means that the frames to be selected should be those with odd indices (1, 3, 5, 7, 9, ..., etc.). All this continues until the maximum segment number is reached, for which the following values were defined: 10, 15, 30, 45, and 60 [16].

2) Automatic Detection of Traffic Accidents

In order to interpret a video segment to detect whether an event occurs, the data must be exploited in two main ways: visually and temporally. The convolutional-based architectures [19] are the most important techniques for visual analysis of images. These are a significant improvement over traditional artificial neural networks in the performance of image classification solutions. However, convolutional layers do not solve all problems. One of the weaknesses of convolutional layers is that they are not good at extracting temporal features from data. Although convolutional layers are powerful in exploiting the spatial characteristics of the data, recurrent neural networks were designed to exploit the temporal characteristics of the data. Convolutional layers are able to process the data in such a way that the spatial information changes to a more abstract representation saving computational cost. Currently, these architectures are used as automatic extractors of image features due to their performance reducing the dimensionality of the input data. However, spatial data is not everything in a video [16].

Sequential data is of importance in understanding an event that happens over a time span. Recurrent neural networks perform better when processing a sequence over time compared to feed-forward artificial neural networks. There are solutions that use both architectures in order to improve performance in solving video comprehension problems [20]. However, the scientific community has presented a design capable of exploiting both types of data: the Convolutional LSTM (ConvLSTM) layers [16]. These are a special type of architecture where the cells follow the same operations as a Long Short-Term Memory neuron but differ in that the input operations are convolutions instead of basic arithmetic operations. This architecture has shown high performance in problems with video compression. To solve the traffic accident detection problem, the first part of the architecture is designed as an automatic image feature extractor to process each frame of the video segment[16]. Then, this new representation of the data is used as input data in an empirically designed recurrent neural network to extract temporal information from the input data. Finally, a dense artificial neural network block is used to perform the binary classification of detecting an accident, as shown in Figure 4.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

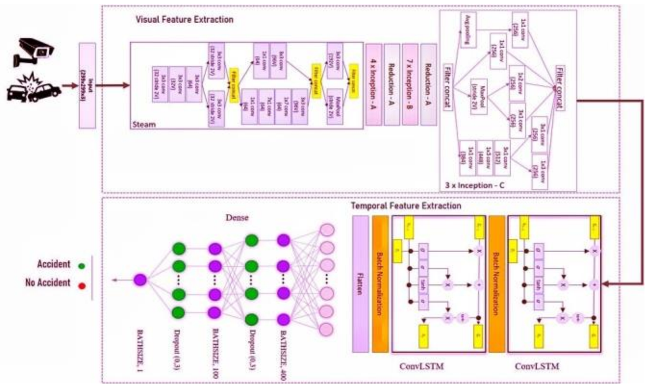

Fig 3: Architecture for video analysis, visual feature extractor based on the InceptionV4 architecture (top) and temporal feature extractor (bottom).

The proposed model consists of three parts: a spatial feature extractor, a temporal feature extractor, and a binary classifier. The first part uses an architecture named InceptionV4, proposed in [21]. This model was trained with the ImageNet dataset, which showed high performance in solving this problem by classifying images among a thousand different categories. However, this pre-trained model does not show good results when detecting a traffic accident in images because the model was trained with a completely different task. Therefore, when applying transfer learning, it seeks to compensate for the acquisition of knowledge in a new task. For this reason, an adjustment is made to the model using a new dataset with examples of traffic accidents in images[16].

To obtain more information from video type data, it is necessary to know the temporal and the spatial features; therefore, it is required to use a model capable of extracting these kinds of characteristics. Therefore, a ConvLSTM layer-based neural network architecture is proposed that receives as input the feature vector computed by the adjusted InceptionV4 architecture[16].

Finally, it is necessary to detect whether the video segment contains a traffic accident. For this, a dense artificial neural network block using regularization methods is proposed so that the final model is able to generalize the solution[16].

V. RESULT

A. For Traffic Manageemnt System

To test the solution approach of the vehicle-detection model a dataset was prepared. The set consists of video of traffic from the CCTVs installed with traffic lights or with at any different position. At first all types of videos were used to check the accuracy in different conditions of the vehicle detection model.

On analysis the result from the whole dataset that there are specific positions and angles at which the model monitors the vehicles more accurately than others. The model uses OpenCV library. This uses edge detection on the frames of the video then, put rectangle shape over the contours, whose dimensions are found to be in range of specific length and breadth. So, more the spacing between the cars the more accurately it can spot each car. So, the scenario in which the detection works best is when cars are evenly spaced i.e., when there is light traffic. For heavy traffic also the module works we have to make some error adjustments to it.

After getting the car count, we get the traffic-time density for that lane. Depending on that value we use mathematical analysis for determine opening time for a single lane. We check if the quantity is abruptly high, then this could be error in vehicle count. By making some adjustments to those values we can still use it.

So, finally we get a module that keeps determining the time for opening traffic lanes using Computer Vision and Python.

B. Accident Detection System

1) Dataset

For the solution proposed, two sets of data were used. The first one consists of images used for the fine tuning of the visual feature vector extractor. The second one consists of videos that present traffic accidents (positive and negative cases) for training the temporal feature extractor [16].

The image dataset was built from scratch, applying the web scraping technique to populate the dataset. For this, a series of logical steps were proposed. First, we identified the sources on the web where the image search was performed. Next, we defined the set of keywords for the searches. For this process, the following keywords were selected: Traffic accidents, Car accidents, Motorcycle accidents, and Truck accidents. Then, the automation stage was performed. The application was developed in the Python programming language together with the Selenium library, which contains useful functions to perform this process. Finally, a manual validation of all the collected images was carried out together with an image transformation in order to standardize the size and format used [16].

The videos dataset was formed from two different data sources. The first one is the CADP dataset [22]. It has a total of 1416 traffic accident videos, i.e., positive examples of the traffic accident problem. This dataset adds up to a total duration of 5.2 h, with an average number of frames of 366. This source was chosen instead of others in the literature, such as [23], due to the number of positive cases that the CADP dataset presents (100%) and the position of the video camera (CCTV), which allows for a third person perspective. The second source used for the video dataset only contains negative cases of the presented problem, i.e., videos where no traffic accidents are presented [24].

2) Temporal Video Segmentation

A video is segmented in order to obtain a greater number of examples with a certain number of constant frames and, in turn, a segment with shorter duration. This is because traffic accidents have a short average duration (10 frames) [23], which allows for processing of the original video in a more efficient way.

In order to select the segmentation technique for the input data, some experiments were performed on the videos taken from the dataset. The four techniques to be evaluated were compared using the same videos in each case. The first technique consists of a segmentation without frame discrimination [16]. Therefore, all consecutive images of the video are selected until the maximum time of the segment is reached. This technique has an average reading time of 0.18 s. The second technique used seeks to skip frames in order to reduce the redundancy that can be observed when using very close images in the video. This is because when the video has been recorded with a traditional camera, the number of similar consecutive frames is very high. For this reason, we experimented by skipping one frame for each frame selected. That is, in this case, the images with an odd index were chosen from the video, until the maximum length of duration established for the segment was reached. The third and fourth techniques presented are based on discriminating consecutive frames with respect to an SSIM [16]. For the third technique, a pixel-to-pixel comparison of two consecutive images is calculated. For decision making, a threshold of 0.9 was set. Therefore, if a consecutive frame exceeds this threshold, the candidate is not chosen and moves to the next frame in the video, for which the same process is performed. Finally, the fourth technique number four shows a similar process to the third technique. However, in this one, the threshold was defined at 0.98, and the matching operation used is the SSIM image-matching metric [25]. A maximum segment length of 45 frames was set for the tests. The comparison between techniques is presented in Tables 1 and 2. The technique chosen was the first described:

Table 1: Comparison between segment generation techniques

Method	**Advantage**	**Disadvantage**
No selection	Low runtime, no data loss	High similarities between adjacent frames
Skip frame (n=1)	Low runtime, with medium/high similarity in adjacent frames	Possible data loss
Pixel Similarity	Low runtime, with medium/high similarity in adjacent frames	Possible data loss
Structural Similarity	Low similarity in adjacent frames	High execution time

Table 2: Results between segment generation techniques

Method	**Frames**	**Execution Time[1]**	**Similarity**
No selection	45	0.919	0.825
Skip frame (n=1)	45	0.972	0.762
Pixel Similarity	45	1.214	0.875
Structural Similarity	45	2.457	0.823

3) Automatic Detection of Traffic Accidents

The solution presented is based on a visual and a temporal feature extractor. The first stage of the model consists of the InceptionV4 architecture (pre-trained with the ImageNet dataset) [21] truncated. That is, all the Inception cells (convolutional layers) were used, eliminating the multilayer perceptron at the end of this architecture [16]. This is to use this part of the model only as a visual feature extractor as in Figure 4, upper part.

However, by performing multiple experiments, it was concluded that the pre-trained model does not differentiate between a vehicle at rest and a vehicle hit by a traffic accident. Therefore, the images dataset was used for training in order to adjust the weights of this pre-trained network. In this process, all the weights of the initial layers of the architecture were frozen, and only those of the last convolutional cell of InceptionV4 were adjusted [16]. To adjust the feature extractor, multiple experiments were performed. This was done using regularization techniques, data augmentation, and hyper-parameter modifications. The results of the tests performed are described in Figure 4.

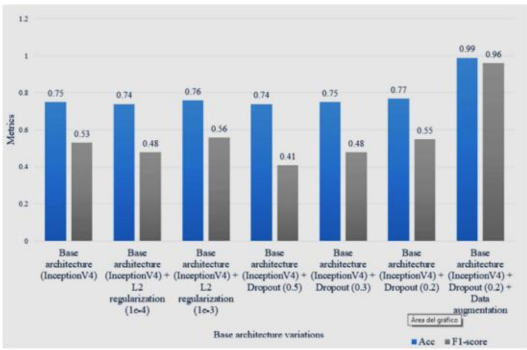

Fig 4: Visual feature extractor experiment

The temporal feature extraction is based on recurrent neural networks. The architecture proposed for this stage consists of two ConvLSTM layers. These were created to extract temporal information in data of more than one dimension, using the convolution operation. Between these layers, a Batch Normalization is added, and the various hyper-parameters are adjusted. The ConvLSTM layers used consist of 64 neurons each, a kernel size of 3 × 3, a dropout of 0.2 and a recurrent dropout of 0.1. The results obtained are presented in Figure 5, while Figure 6 shows the accuracy of the model in the training stage.

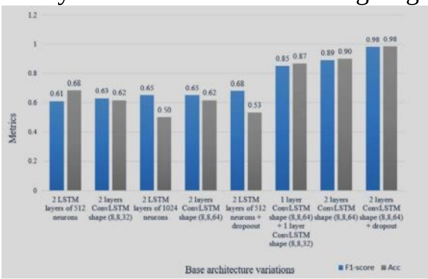

Fig 5: Experimenting with the temporal feature extractor

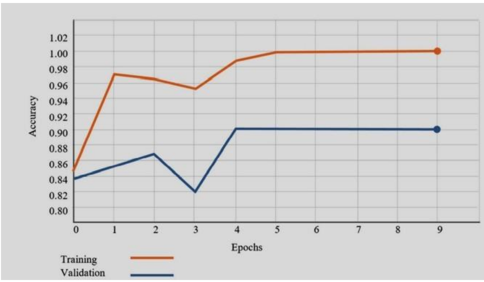

Fig 6: Behavior of the model's accuracy by epochs with the training set and the validation set.

The last stage of the accident detection process is given by a densely layered block. The proposed neural network consists of a total of three hidden layers and one output layer, plus a regularization technique called dropout with a value of 0.3 [16]. The distribution of the neurons in the mentioned layers is as follows: four hundred, one hundred, and one neuron, where the first two layers use the hyperbolic tangent activation function while the last layer (output layer) uses the sigmoid activation function in order to perform a binary classification (accident or non-accident). The training and validation results are presented in Table 3 (Note: the dataset is distributed as 54% accident and 46% non-accident). The established hyper-parameter values are presented in Table 4. The model was trained on a computer with a 5th generation I7 4820k@3.70GHz processor, 64 GB of RAM memory, and two Nvidia 1080TI video cards with 11 GB of GDDR 5X RAM at 405 MHz [16].

Table 3 Confusion matrix of the complete model on the validation set [16].

		Predicted Class	
		No Accident	**Accident**
	No Accident	0.98	0.011
	Accident	0.02	0.96

Table 4 Hyper-parameter of the proposed neural network [16]

Hyper-Parameter	**Value**
Input size	45 frames
Batch size	4
Loss function	Binary cross-entropy
Optimizer	Adam optimization
Weight initialization	Xavier initialization
Learning rate	0.001
Number of epochs	10

VI. Conclusion and Future Work

For particularly specific issues, pre-trained neural networks are unable to generate a vector with relevant information. As a result, the weights of these models must be adjusted using instances of the problem to be solved.

Because the similar values between the segments of the techniques with frame selection present negligible differences between them, while the computational cost, processing time, and accuracy in accident detection present better results by not conditioning the selection of frames to a metric, the technique that best represents a temporal segment of a traffic accident does not eliminate any data.

Artificial vision has made significant progress in comprehending video scenes. Artificial neural networks are one of the most effective strategies. To extract as much information as possible from the input data, many of these models use architectures built of convolutional and recurrent layers. The suggested solution is built on this type of architecture and performs well in detecting traffic incidents in movies, with an F1 score of 0.98 and a 98 percent accuracy.

For video traffic accident detection, the proposed model performs well. However, the conditions under which the model operates are constrained due to the scarcity of such datasets in the scientific community. Due to the small number of cases available, the solution is limited to automobile crashes, omitting motorbikes, bicycles, and pedestrians. Furthermore, the model makes mistakes when determining accident segments with poor illumination (such as films taken at night) or low resolution and occlusion (low quality video cameras and locations).

Managing the traffic lights according to the collected data includes profiling the traffic and storing the collected data. A traffic study can be conducted with the profiling and determine when and where traffic density changes throughout the day, week, month and year. Additionally, this system can help in detecting accidents which would be very beneficial for emergency services as there would be very less time wastage in getting the exact co-ordinates of the accident spot.

Right now, image processing gives its best performance on high end PCs and devices. So, video quality is sacrificed for better frame rate. If Computer Vision technologies get optimized in future, we can process better quality video on lower end machines. This would help in increase accuracy of the module.

This is a decentralized model. Both parts of the model work independently. So, even more AI or Machine learning modules can be added to this setup to increase the functionality of this system. Like we can a module that could detect standby vehicles, and if they block the traffic it could alarm to get attention of its owner or gets it license plate number and register it for an e-Challan (e-fine). On the contrary we could also make this system centralized which have its own possibilities to scale it up.

References

[1] 'How Artificial Intelligence Works', *Investopedia*. https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp (accessed May 13, 2022).

[2] I. Mihajlovic, 'Everything You Ever Wanted To Know About Computer Vision. Here's A Look Why It's So Awesome.', *Medium*, Sep. 24, 2021. https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e (accessed May 13, 2022).

[3] Wenjie Chen, Lifeng Chen, Zhanglong Chen, and Shiliang Tu, 'A Realtime Dynamic Traffic Control System Based on Wireless Sensor Network', in *2005 International Conference on Parallel Processing Workshops (ICPPW'05)*, Oslo, Norway, 2005, pp. 258–264. doi: 10.1109/ICPPW.2005.16.

[4] P. Manikonda, A. K. Yerrapragada, and S. S. Annasamudram, 'Intelligent Traffic Management System', p. 4.

[5] Y.-R. Chen, Keng-Pin Chen, and Pao-Ann Hsiungy, 'Dynamic traffic light optimization and Control System using model-predictive control method', in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, Nov. 2016, pp. 2366–2371. doi: 10.1109/ITSC.2016.7795937.

[6] W.-J. Chang, L.-B. Chen, and K.-Y. Su, 'DeepCrash: A Deep Learning-Based Internet of Vehicles System for Head-On and Single-Vehicle Accident Detection With Emergency Notification', *IEEE Access*, vol. 7, pp. 148163–148175, 2019, doi: 10.1109/ACCESS.2019.2946468.

[7] J. Singh, V. Velu, and U. Nirmal, 'Vehicle Accident Detection System using Internet of Things (VADS - IoT)', in *2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, Malaysia, Apr. 2021, pp. 353–359. doi: 10.1109/ISCAIE51753.2021.9431813.

[8] K. P. Sampoornam, S. Saranya, S. Vigneshwaran, P. Sofiarani, S. Sarmitha, and N. Sarumathi, 'Intelligent Expeditious Accident Detection and Prevention System', *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1059, no. 1, p. 012012, Feb. 2021, doi: 10.1088/1757-899X/1059/1/012012.

[9] J. M. Kumar, R. Mahajan, D. Prabhu, and D. Ghose, 'Cost effective road accident prevention system', in *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, Greater Noida, India, Dec. 2016, pp. 353–357. doi: 10.1109/IC3I.2016.7917988.

[10] S. Eduku and M. O. Alhassan, 'Design of Vehicle Accident Prevention System Using Wireless Technology', vol. 7, no. 10, p. 8, 2017.

[11] 'RA_Uploading.pdf'. Accessed: May 15, 2022. [Online]. Available: https://morth.nic.in/sites/default/files/RA_Uploading.pdf

[12] A. W. CNN, 'Commuters waste an average of 54 hours a year stalled in traffic, study says', *CNN*. https://www.cnn.com/2019/08/22/us/traffic-commute-gridlock-transportation-study-trnd/index.html (accessed May 15, 2022).

[13] 'Bengaluru people spent 243 hours on average in traffic in 2019, time they could use to watch 215 episodes of GoT', *Business Today*. https://www.businesstoday.in/latest/trends/story/bengaluru-people-spent-243-hours-on-average-in-traffic-in-2019-time-they-could-use-to-watch-215-episodes-of-got-243012-2020-01-30 (accessed May 15, 2022).

[14] PTI, 'India accounts for 11 per cent of global death in road accidents: World Bank'. https://www.thehindubusinessline.com/news/india-accounts-for-11-per-cent-of-global-death-in-road-accidents-world-bank/article33834556.ece (accessed May 15, 2022).

[15] H. Jiang, G. Zhang, H. Wang, and H. Bao, 'Spatio-Temporal Video Segmentation of Static Scenes and Its Applications', *IEEE Trans. Multimed.*, vol. 17, no. 1, pp. 3–15, Jan. 2015, doi: 10.1109/TMM.2014.2368273.

[16] S. Robles-Serrano, G. Sanchez-Torres, and J. Branch-Bedoya, 'Automatic Detection of Traffic Accidents from Video Using Deep Learning Techniques', *Computers*, vol. 10, no. 11, p. 148, Nov. 2021, doi: 10.3390/computers10110148.

[17] A. Mackin, F. Zhang, and D. R. Bull, 'A Study of High Frame Rate Video Formats', *IEEE Trans. Multimed.*, vol. 21, no. 6, pp. 1499–1512, Jun. 2019, doi: 10.1109/TMM.2018.2880603.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, 'You Only Look Once: Unified, Real-Time Object Detection'. arXiv, May 09, 2016. Accessed: May 16, 2022. [Online]. Available: http://arxiv.org/abs/1506.02640

[19] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, 'Object Recognition with Gradient-Based Learning', in *Shape, Contour and Grouping in Computer Vision*, vol. 1681, Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 319–345. doi: 10.1007/3-540-46805-6_19.

[20] W. Lim, D. Jang, and T. Lee, 'Speech emotion recognition using convolutional and Recurrent Neural Networks', in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Jeju, South Korea, Dec. 2016, pp. 1–4. doi: 10.1109/APSIPA.2016.7820699.

[21] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, 'Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning', p. 7.

[22] 'CADP:A Novel Dataset for CCTV Traffic Camera based Accident Analysis'.

https://ankitshah009.github.io/accident_forecasting_traffic_camera (accessed May 16, 2022).

[23] F.-H. Chan, Y.-T. Chen, Y. Xiang, and M. Sun, 'Anticipating Accidents in Dashcam Videos', 2016. doi: 10.1007/978-3-319-54190-7_9.

[24] S. Lyu *et al.*, 'UA-DETRAC 2018: Report of AVSS2018 & IWT4S Challenge on Advanced Traffic Monitoring', p. 6.

[25] I. González-Díaz, T. Martínez-Cortés, A. Gallardo-Antolín, and F. Díaz-de-María, 'Temporal segmentation and keyframe selection methods for user-generated video search-based annotation', *Expert Syst. Appl.*, vol. 42, no. 1, pp. 488–502, Jan. 2015, doi: 10.1016/j.eswa.2014.08.001.