Deep Learning Based Traffic Accident Detection and Severity Classification in Video Surveillance

Author

SYED ATEEB ALI KAZMI

00000327691

Supervisor

DR. SAIFULLAH AWAN

DEPARTMENT OF ELECTRICAL ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

October 2024

THESIS ACCEPTANCE CERTIFICATE

It is certified that final copy of MS/MPhil thesis written by Mr. Syed Ateeb Ali Kazmi (Registration No. 00000327691) Entry-2020, of (College of E&ME) has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistake and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC member of the scholar have also been incorporated in the said thesis.

Signature: ______________________

Name of Supervisor Dr. Saif Ullah Awan

Date: 02-Oct-2024

Signature (HoD): ______________________

Dr. Qasim Umar Khan

Date: 02 - Oct - 2024

Signature (Dean) ______________________

Brig Dr. Nasir Rashid

Date: 0 2 OCT 2024

DEDICATION

I would like to dedicate my thesis to My beloved Parents

ACKNOWLEDGEMENT

I would like to thank Allah (SWT) who has always showered his countless blessings on me.

A special thanks to my supervisor Dr. SAIFULLAH AWAN for their tremendous supervision, and a very reliable encouragement throughout the course of this thesis. I wouldn't have been here today without his encouragement.

I am also thankful to Dr. Fahad Mumtaz Malik and Dr. Ahmad Rauf Subhani for being on my thesis Guidance and Examination Committee and for guidance.

.

ABSTRACT

Since there has been an increase in the number of cars on the road, there is need to fast identify accident-prone areas and respond to them promptly. This Thesis aims at finding the suitability of employing state-of-the-art deep learning algorithms in identifying traffic accidents and their severity through video surveillance systems. To achieve this, using the large set of traffic videos, we learned several deep learning models which could identify whether an accident occurred based on the video or show how severe the accident was. The main model used in this research, called the Long-term Recurrent Convolutional Network (LRCN), combines two powerful techniques: The first is to distinguish between the fine details of the video at a frame level while the second is used in capturing the sequence of events in the material. Apart from the LRCN model, several other models like LSTM, 3D CNN, and MoVinet were also used for evaluation for better understanding of the proposed method's performance and its variations with other models. The proposed LRCN model was unique by presenting a high degree of accurate identification of incidents and their severity. Processing of the videos gave it a unique chance of capturing both the spatial and temporal aspects, hence seeing patterns, which other models failed on. These findings underscore the ability of LRCN in revolutionizing how we monitor traffic accidents hence making the roads safer and with better response time to the incidents. It provides the way out to realize more sophisticated traffic management systems, which can recognize the cases of the accidents, the urgency of responses, and may subsequently reduce the number of fatalities related to the traffic incidents.

Keywords: Traffic Accident Detection, Accident Severity Classification, Deep Learning, Video-Surveillance, LRCN, LSTM, 3D CNN, Spatial and temporal features, Incident Response, Road Safety, Traffic management Systems.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial intelligence
ML	Machine Learning
IOV	Internet of Vehicles
FRAM	Functional Resonance Analysis Method
FAD	Fine Grained accident Detection
ROC	Receiver Operating Characteristics
AUC	Area Under The curve

CHAPTER 1: INTRODUCTION

1.1 Background and Scope

Growth in the number of vehicles and size of cities brings about an improve of road accidents, and therefore there is need to implement systems that can enable real-time tracking and identification of traffic mishaps. Conventional traffic monitoring was done using sensor-based systems and surveillance that entails human input and thus has some drawbacks like error-proneness and low speed. The new formation of deep learning stimulated extensive interests on automating the process of identifying the occurrence of accidents and their severity with the help of video surveillance [1,2].

Other significant approaches for instance, the CNNs for analyzing spatial features of traffic and RNNs for sequential analysis have come out as useful in traffic monitoring [3]. In particular, Long-term Recurrent Convolutional Network (LRCN) models are capable of processing both spatial and temporal data, giving reliable means of accident detection and classification [4,5].

Further, there are those like the 3D CNNs which have been used in analyzing the movement of vehicles in N dimensions in sequences of videos allowing the identification of traffic incidents more professionally [5]. However, these models are computationally costly and at times fail in handling sequences of video data beyond some specific length [6]. To overcome this, models like recurrent models including LSTM (Long Short-Term Memory) networks have been used in estimating accident severity. Compared with other types, LSTMs are suitable for studying the evolution of an accident and mining patterns in different time slices [7].

Besides, classification of accidents according to their severity is now an important and significant element of traffic management systems. As for deep learning, it was used in order to predict the severity of accidents based on the information about the speed and track of the vehicle as well as the environmental conditions of the accident site [6,8]. Such models help in the prioritization of response where the major efforts are focused on aiding the victims of severe accidents as soon as is possible [7][9].

The objectives of this study are to compare well-established deep learning algorithms such as LRCN, LSTM, 3D CNN and Movinet, in order to assess their efficiency when used for traffic accident detection and severity analysis in video-surveillance systems. Through such comparison the research aims to determine the most efficient technique for live traffic tracking with a view of enhancing road safety and minimizing on accident associated causalities [10].

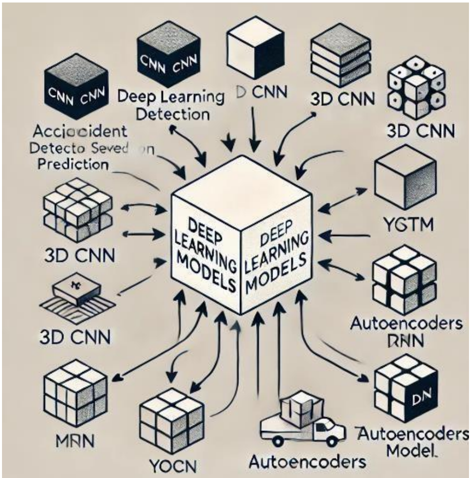


Figure 1 Deep Learning Techniques for Accident Detection.

1.2 Areas of Application

- Traffic management systems.
- Intelligent transportation systems.
- Emergency response services.
- Automotive safety technologies.
- Urban planning and infrastructure development.

1.3 Relevance to National Needs

Pakistan stands among the countries with the highest rate of traffic accidents and estimates indicate that 30,000 people die every year because of this problem; the main causes being the reckless driving, bad quality of roads, and the lack of traffic control. (Republic Policy). But special emphasis should be put onto the number of road accidents, which leads to a loss of life, material losses, and financial losses. This has included better traffic management measures, better development of roads and food for traffic and laws governing the roads but remains challenging with congested roads especially in urban areas and regions with poor roads in the rural areas.

1Furthermore, the real-time data analysis and surveillance systems would also be in accordance with the national strategy on the reduction of the road accidents as outlined by the transport authorities of Pakistan. If these systems are incorporated within the current structures, Pakistan can effectively cope up with increasing traffic density and enhance the measures that need to be taken for safer roads within the cities and the rural areas.

Data from Pakistan:

- Accidents in road traffic have become a major cause of mortality where 67% of this accident age is as a result of human error, 28% due to poor infrastructure. (Pakistan Today).

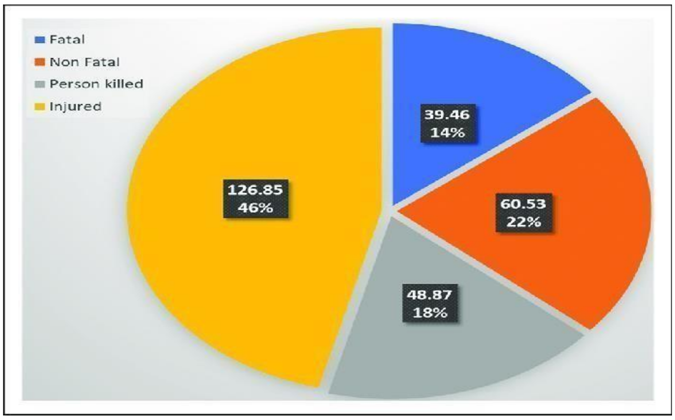


Figure 2 The percentage share of road crashes in Pakistan

1.4 Advantages

- **Accuracy:** Deep learning models can produce high accuracy because the system proactively analyzes a wide variety of data types e.g. the analysis and categorization of several road mishaps, including the newly defined type.
- **Adaptability:** It can also be seen that through the analysis of visual data, one is able to identify accidents in various environmental conditions and scenarios.
- **Scalability:** The solution, which has been developed, can be easily applied and implemented at various types of traffic infrastructure interventions to assist in increasing the general road safety.

1.5 Reason/Justification for the Selection of the Topic

Although this is a cutting-edge research area that will be discussed-which pertains to deep learning-based traffic accident detection and severity classification-the motivation behind this topic lies in many critical challenges faced all over the world and particularly very strongly in countries like Pakistan. Reasons for selecting this topic are mentioned below:

1. **Increasing Traffic Accident Rates in Pakistan:** Pakistan has severe road traffic accidents with more than 30,000 deaths every year; making it among the countries with the highest rates of road deaths in the world. Therefore, the requirement of such systems to pinpoint accidental occurrence in real-time and rate the danger levels is crucial to averting such mishaps and guaranteeing a swift response to crises. This research topic covers a topic of national safety importance in regard to investigation.

2. **Importance of Real-Time Monitoring:** Static analysis of traffic is insufficient for real-time traffic monitoring as well as for immediate response. They also discussed how by implementing deep learning into video surveillance systems, the accidents can be detected along with identification of the severity of the accident at real time. This provides a way of improving the existing traffic controls in a very efficient and scalable manner especially for city centers that face major traffic jungles.

3. **Technological Advancements in AI:** Supporting the theory that deep learning is a rapidly advancing field, current developments highlighted for the topical area solely include object detection and sequential methods. Fault detection employing models such as LRCN, YOLO, and 3D CNN is still unique, and at the same time, is in accordance with the trends in AI in smart cities around the world.

4. **Addressing Infrastructure and Resource Gaps:** Poor Road structure and lack of traffic management are one of the major reasons for high accident rates in Pakistan, This research demonstrates how AI-based solutions could help fill the gaps of human capabilities and infrastructures, thereby making this an overly appropriate and timely solution for both national and global needs.

5. **Potential to Save Lives:** Correct identification and classification of accident severity can lead to reducing emergency response times thereby minimizing causality. Quick detection leads to rapid dispatch of medical and rescue services to ensure the saving of lives and minimizing extreme consequences of accidents

1.6 Motivation

The present thesis finds its motivation in the increasing need to improve road safety through smarter, more efficient accident detection and response systems. In these days and ages, with an increasingly high rate of accidents occurring on roads, especially in countries like Pakistan, the current methods of accident detection and response have become slow and inefficient, thus leading to preventable deaths and injuries. That is to say, through deep learning technologies in video surveillance, accidents can automatically be found along with their severity. Fast responses to emergencies may thus be recorded in life-saving. Pursuing this research, its objective is meant to improve the degree of security on roads by integrating AI models such as LRCN, LSTM, and 3D CNN into it so that accident predictions and their classifications are made correct.

1.7 Objectives

The objectives of this thesis work are as follows,

- Collect and preprocess a diverse dataset of real videos and images depicting road accidents.
- Train and optimize deep neural networks to accurately detect and classify various types of road accidents, based on their severity levels.
- Have to Make dataset for training the model of severity mainly we will divide severity into 2 classes i.e. Severe and non-severe (based on the impact of crash i.e. medium and low) and will self-annotate the data that will be used for training of the model.
- Evaluate the performance of the proposed system through extensive experiments and comparisons with existing methods.

1.8 Thesis Outline

In this Thesis the deep learning-based techniques are used for the detection and severity classification of traffic accidents using video surveillance. Improvements are applied to the framework of the model, which involves LRCN, LSTM, 3D CNN, and Movinet models for the real-time differentiation of traffic accident scenarios. This research thesis can be segmented into the following chapters:

Chapter 1: Introduction
This chapter introduces the research topic, saying that an increasing need for accurate and real-time accident detection on roads. The background and motivation are further outlined by stating the objectives of the research and giving an overview of the thesis structure.

Chapter 2: Literature Review
This chapter surveys the exiting literature on deep learning applications in traffic monitoring, accident detection, and severity classification. It reviews the earlier methods and checks for the gaps in the current state of affairs of the system, proving that there is a need for models such as LRCN, LSTM, 3D CNN, and Movinet for the better detection and accuracy in classification.

Chapter 3: Proposed Methodology
Further elaboration on how each model contributes to spatiotemporal feature extraction from video data, besides detailing how a full system design makes sure it's robust and efficient for a classification, is discussed.

Chapter 4: Simulation and Results
This chapter addresses the implementation of deep learning models on a traffic video dataset, explaining its experimental setup, procedures, and performance evaluation. The chapter performs further comparative analysis to highlight the accuracy and efficiency of models in accident detection and classification of severity levels.

Chapter 5: Conclusion & Future Scope
This final chapter will summarize the contributions and findings of this thesis and stress the improvements brought by the deep learning models.

CHAPTER 2: LITERATURE REVIEW

This chapter presents a comprehensive review of the literature related to Machine learning and Deep learning models used in past for Accident Detection and severity. The purpose is to explore the existing knowledge and research efforts in the field of Severity Classification. The literature review aims to identify the gaps, challenges, and opportunities for further research in this area.

2.1 Introduction

Traffic accidents still rank among the highest threats concerning road safety. Ranking among the causes of death globally is road traffic injuries. It is important to have proper detection and classification of accidents, especially in real-time, to help increase the response time and improve traffic management. The traditional methods for this, which include sensor-based monitoring or human monitoring, are inefficient in complex urban settings. With advancements in deep learning and video surveillance technologies, researchers are now focusing on automated methods for the detection of accidents and severity classification.

Accident Frames | **Severity & Non severity frames**

Figure 3 Accident and severity Frames

2.2 Deep Learning Models for Accident Detection

CNNs have been used extensively to detect traffic accidents in real time based on video data deep learning techniques. CNNs are effective for extract features from images or videos in the Spatial domain we can say that CNNs are suitable for this scenario.

In future work – Yu et al. (2024) implemented a finer accident detection model using transformer learning. Their model was able to classify accidents and estimate their scale based on video input, with the features captured in traffic videos being spatio-temporal. In the paper [1] the transformer-based architecture received a high accuracy in multi-class accident detection, so it can be potentially used in ‘‘urban traffic surveillance’’. Diagram is attached for reference showing the mode of transportations in which accident occurred

Lakshmy et al. (2022) proposed to study the IoT implementation along with the CNNs to detect the car accidents in real time. When integrating the data from the sensors with the video analysis, their model improved the rate of accident detection and prompt reporting to the officials. Indeed, the accuracy of traffic accident identification was identified as an aspect that requires prompt video and IoT monitoring in the study [5].

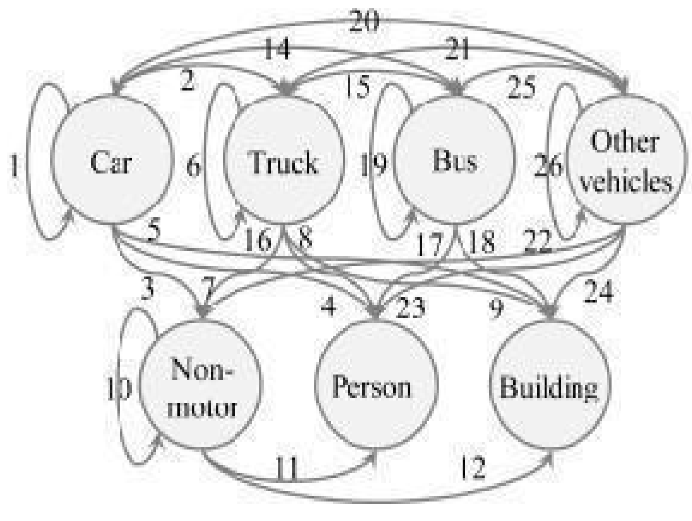


Figure 4 The accident category graph. The connection represents the participants in each accident category [1]

2.3 Severity Classification in Traffic Accidents

The identification of the degree of accident severity cannot be considered as a simple task of accident detection. Correct partitioning of severity necessitates models to assess the level of harm that a given accident caused, which may be minute and hard to discern. In the severity classification task, the task is usually to decide whether the accident falls under low, medium or high severity.

Sattar et al. (2024) used Graph Neural Networks (GNNs) to predict the severity of an accident. Compared to traditional methods, GNNs are capable of modeling the dependence between different factors of an accident, including the speed of the vehicle, the storm events, and the type of road. As a result, this approach showed better performance than the standard models of machine learning for various traffic injuries' severity prediction rate of complex accidents. [3]

Aboulola et al. (2024) applied transformer learning for demarcating the level of severity in road traffic accidents. When coupled with traffic flow data, environmental conditions such as weather and road conditions, then their model was able to predict the severity of accidents accurately. The study provided evidence of transformer-based models' capability of processing large data to correctly estimate the severity of an accident [2].

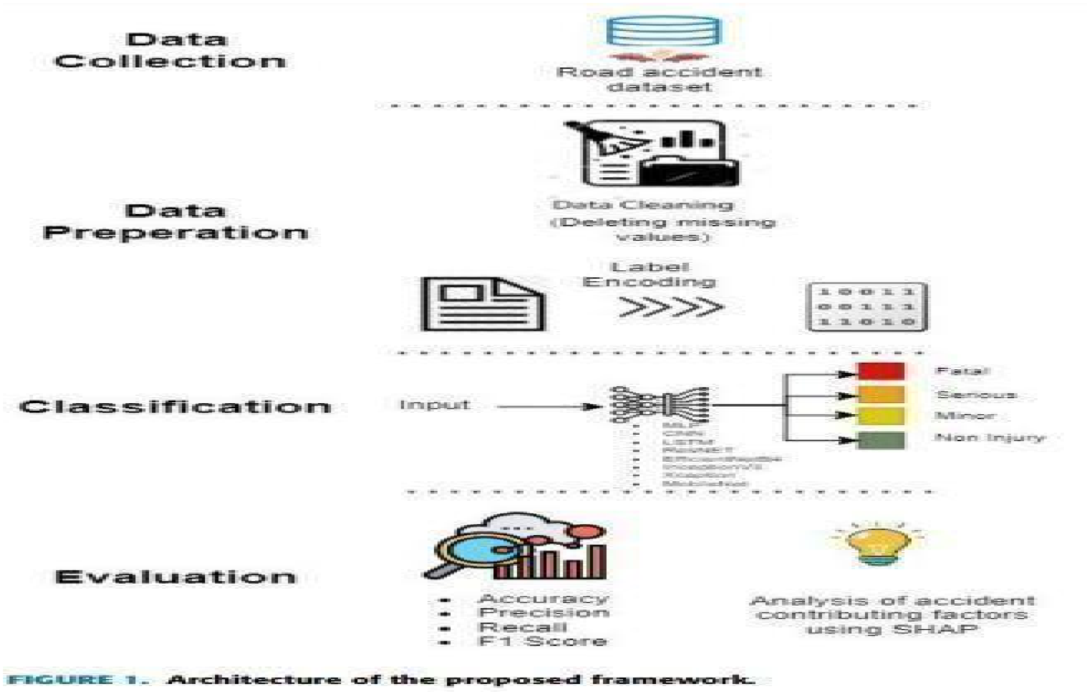


Figure 5 Architecture of the proposed framework by Aboulola et al [2]

2.4 Challenges in Accident Detection and Severity Classification

The identification of and classification of accidents and their severity is not without some difficulties. Kyi et al., (2022) noted that traffic accident data has a data imbalance problem due to the fact that accident occurrence is less common than non-accidents. This creates an imbalance within the model performance because the models lean towards the majority class, which cuts the effectiveness of accident detection and severity estimation in half [4].

Furthermore, Yu et al. (2024) also observed that categorizing severity level as low, medium and high poses a challenge owing to the small variation in sheep's visual features of the accidents. Similarly, Aboulola et al (2024) emphasized the importance of capturing more context data about the circumstances of the accident; such as road and weather conditions.

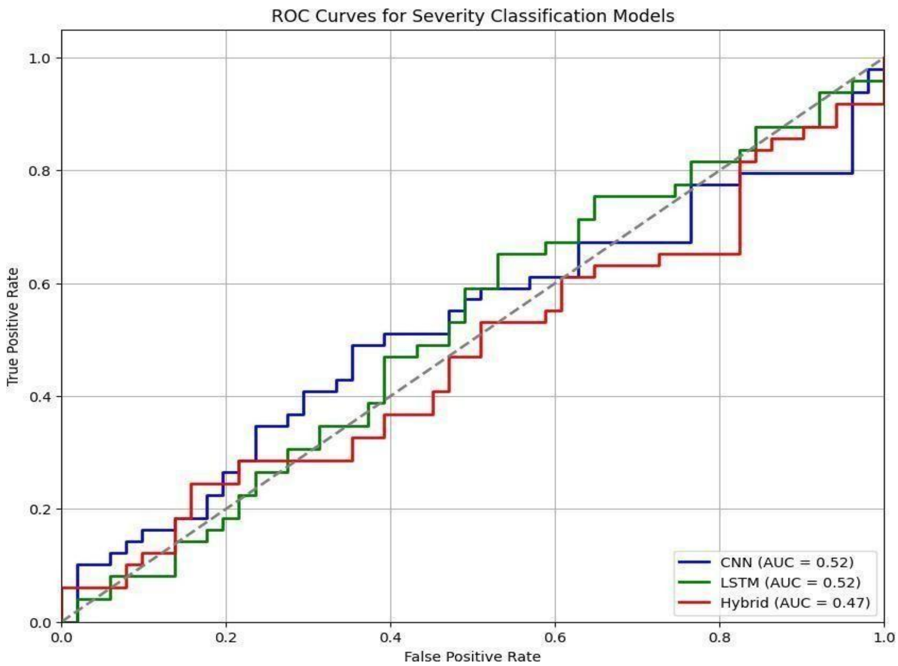


Figure 6 ROC curves comparing the performance of CNN, LSTM, and hybrid models on severity classification tasks [4]

2.5 Conclusions

Research using Deep Learning has shown remarkable improvement in traffic accident identification and severity prediction. Densely connected networks like CNN-LSTM and Movinet have shown remarkable performances in terms of unearthing both spatial and temporal characteristics and as a result, they are perfect when it comes to real-time traffic surveillance systems. But difficulties persist with data skewness and identifying the proper classification of accident severity level. Future work needs to work to improve the robustness of the models used that will incorporate more than one data source as well as address how to deal with imbalanced datasets.

CHAPTER 3: PROPOSED DESIGN

This chapter describes a methodology for the proposed real-time detection and severity classification of traffic accidents through deep learning approaches within a video surveillance system. This study aims at efficiently designing an accurate detection system that can identify in real time any kind of traffic accidents on roadways and classify their severity. In this approach, four deep learning models will be trained and compared. They include LRCN, LSTM, 3D CNN, and Movinet. The main model will be LRCN because it can process both spatial and temporal features of video data. This chapter outlines the research design, data preprocessing methods, model architectures, and how it will be trained. The proposed models will be experimented and validated on a big-sized dataset of traffic videos with performance metrics tested to validate the methodology.

3.1 Introduction to Proposed Design

3.1.1 Overview of The Problem

With the ever-increasing vehicles on the roads that effectively cause significant damage in terms of loss of lives and property, accidents in the roads have emerged to be one of the most significant safety concerns. Traditional methods of accident detection and severity assessment are slow and inefficient, thus causing delays in emergency responses. Therefore, deep learning approaches for the accurate detection of accidents and classification of their severity using video surveillance have become the need of the time.

3.1.2 Chosen Approach

First, we collected the Dataset and then we trained it for Detecting the Accident from the videos, Once Accident is detected we Trained our dataset for Severity classification. For this research, four deep learning models were selected that are

- LRCN
- LSTM
- 3D CNN
- Movinet

In order to attack the problem of traffic accident detection and severity classification. The Long-term Recurrent Convolutional Network, LRCN, was mainly used due to its ability to interpret video data that have spatial features extracted by a CNN and temporal sequences processed by an LSTM.

From the above options, LSTM is chosen as it captures long dependencies. Now, from the view of processing time series data, LSTM efficiency is to be seen. Then, 3D CNN will be chosen because there is a simultaneous processing of both spatial and temporal information that exists in video frames. Last but not the least, as it processes real-time video, Movinet is chosen for testing the efficiency of accident detection scenarios. The performances of all the models will then be compared to conclude which is the most effective approach for achieving accurate and real-time traffic accident detection and severity classification.

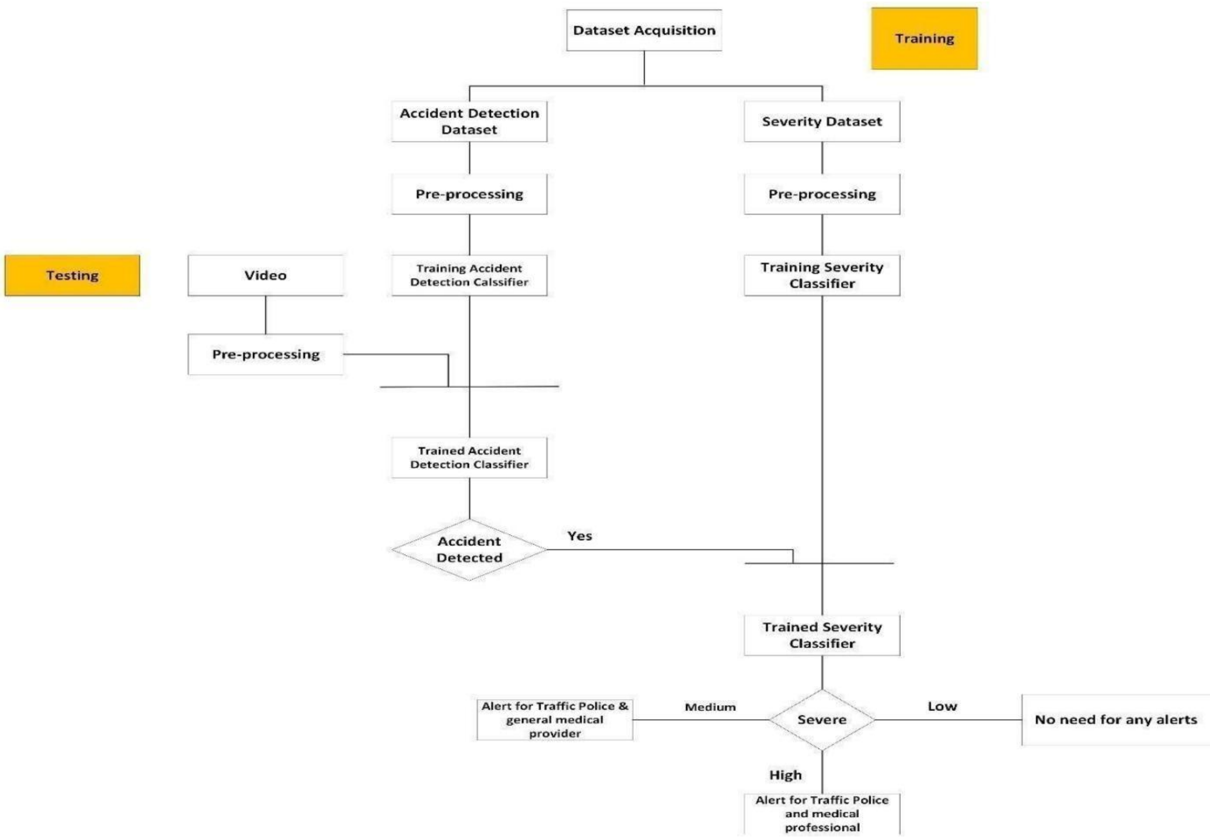


Figure 7 Block Diagram of Proposed Design

3.1.3 Objectives of the Design

- Design a Real-Time Accurate Detection System Developing a Deep Learning-Based Model for Real-Time Traffic Accident Detection Using Video Surveillance Data.
- Enforce models classifying accidents by severity so that the emergency responses could be given as minor versus major.
- Compare the results of four models, namely, LRCN, LSTM, 3D CNN, and Movinet for accident detection along with severity classification.
- Leverage the spatial features from individual video frames and temporal sequences for better detection.
- Models have to support real-time processing of video feeds, especially in live conditions

3.2 Dataset Collection and Preprocessing

3.2.1 Sources of Data & Type of Videos

The source of data for this study is a collection from repositories that feeds the research with accident and non-accident video clips.

- CADP Dataset comprises of several video scenes, in fact, body parts of accident scenes in which segmented scenes are allowed to elapse up to 5 seconds with frames of 30 fps. It is for accident identification and designing prototype response to the identified accidents.
- HEV-I (Honda Egocentric View-Intersection) is the video captured from first person view in the intersections in San Francisco and its neighboring areas. It encompasses many driving scenarios and is confined to the contact between vehicles and pedestrians, especially at junctions. The dataset has videos up to two minutes in length with clips extracted at ten frames per second and are in 20 second intervals.
- Although the JAAD Dataset (Joint Attention in Autonomous Driving) is set up for analyzing drivers' behaviors and pedestrians' actions, it does not contain any accident videos. It mainly addresses behavior of pedestrians, environment stimuli, and driving in dealing with non-accident situations.

- Accident Detection from CCTV Footage deals largely with accident videos extracted from CCTV footage that have been preprocessed into short video clips with a frame rate of 10 fps.
- Deep Accident (mini) concerns itself with accident scenes involving 57,000 frames that are made into videos of up to 5 seconds displaying diverse accident surroundings at 10 frames per second.

Table 1 Sources of Data

Dataset	Download Source	Accident	Non Accident	Total videos	Duration	Frame Rate
CADP (Car Accident Detection and Prediction)	CADP Dataset Release ReadMe - Google Docs	Yes	Yes	1416 (210 folders)	Videos scenes are broken into segments and segments have duration up to 5 s	segments(30 fps)
HEV-I (Honda Egocentric View-Intersection)	Dataset Download Page - Honda Research Institute USA (honda-ri.com)	Yes	Yes	230 video clips	Up to 2 min video clips which we split to shorten the duration up to 20 s)	10 fps
JAAD (Joint Attention in autonomous driving)	JAAD dataset (yorku.ca)	No	Yes	346 clips	5 to 10s	30 fps
Accident detection from CCTV footage	Accident Detection From CCTV Footage (kaggle.com)	Yes	No	989 jpg images(joined to form videos)	Resultant videos of up to 5 s	10 fps
DeepAccident(mini Dataset)	DeepAccident - Download	No	Yes	Total of 57000 frames (joined to form videos)	Resultant videos of up to 5 s	10 fps

3.2.2 Preprocessing Steps

Following were the steps that were followed,

1. **Data mounting:** Google Drive is mounted to access the dataset and load video files from specified directories.

2. **Dataset Preparation:** Datasets are extracted and arranged in specific directories (accident, nonaccidental, severe, and non-severe) based on classifications. Self-annotation was used to label each video as accident, non-accident, severe, or non-severe. Some 2,000 videos were put to use (1,000 for accident and non-accident) and some 1,500 videos were self-annotated (750 severe, 750 non-severe).

3. **Frame Extraction:** Frames were collected for every video in the dataset, using OpenCV to extract every given number of frames from the videos. Setting the frame count per video at fixed length means that the model will take in the same number of frames per video. Frames are resized to standard sizes namely 64x64 or 80x80 and normalized by dividing the pixel values by 255 to bring the pixel intensity range between 0 and 1.

```
video_reader = cv2.VideoCapture(video_path)
success, frame = video_reader.read()
resized_frame = cv2.resize(frame, (IMAGE_HEIGHT, IMAGE_WIDTH))
normalized_frame = resized_frame / 255
```

Figure 8 Code for Frame Extraction from videos

4. **Data Augmentation:** Methods applied to data augmentation included resizing and normalization of video frames to ensure consistency in data.

```
resized_frame = cv2.resize(frame, (IMAGE_HEIGHT, IMAGE_WIDTH))
normalized_frame = resized_frame / 255
```

Figure 9 Code for Data Augmentation

5. **Labeling:** Video labels are self-annotated since the videos were classified into two classes: accident and non-accident. For severity category, either "severe" or "non-severe" were used for labels. These then get translated into one-hot encoding for the model via the to_categorical () function by Keras.

```
one_hot_encoded_labels = to_categorical(labels)
```

Figure 10 Code used for Setting up labels.

.

3.3 Model Architecture:

The model LRCN is utilized for traffic accident detection and classification of severity in this thesis. It is mainly because it combines the power of Convolutional Neural Networks (CNNs) for spatial feature extraction with Long Short-Term Memory (LSTM) networks for temporal sequence learning. Therefore, this model is the best for video surveillance accident detection because of the reason that it can analyze video streams efficiently by catching both spatial and temporal patterns.

3.3.1 Overview of the LRCN Model:

The LRCN architecture manages video sequences by first extracting meaningful features from individual video frames (spatial features) using CNNs and then process the features over time using an LSTM network (temporal sequence learning). This two-stage approach makes LRCN effective at detecting traffic accidents in dynamic, real-world environments where the spatial arrangement of objects (cars, pedestrians, road signs) and their movement over time are critical.

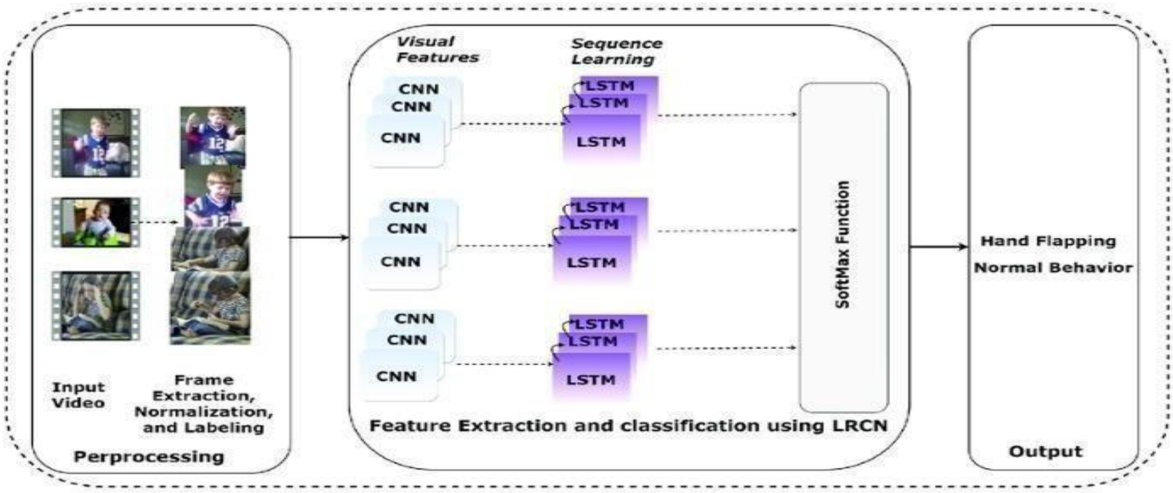


Figure 11 LRCN Model

3.3.2 CNN for Spatial Feature Extraction:

The CNN component is used for extracting spatial features from the video frame in which every individual frame is processed. This presents a configuration of several convolutional layers followed by some pooling layers in such a way that all the significant patterns such as edges, corners, and the basic structures of objects like vehicle shapes and pedestrian shapes could be captured.

1. **CNN Layer Architecture:**
    - Input frame size: 64x64 or 80x80 pixels.
    - A few numbers of convolutional layers are applied with ReLU activation.
    - Stride sizes are also used in max pooling layers applied after each convolutional block in order to decrease the spatial dimension.

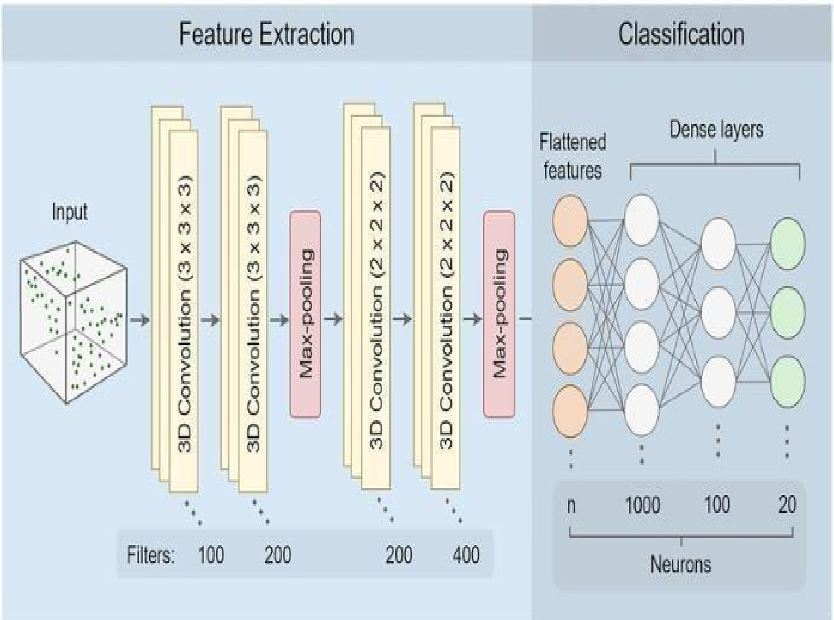


Figure 12 CNN architecture

2. **Convolutional Layers:**

Convolutional layers apply filters to the input frames to identify edges, textures and objects. It produces feature maps by sliding dot products of filters with receptive fields on the input frame.

- **Mathematical Formulation:**

$$Y(i,j) = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} X(i+h, j+w) \cdot K(h,w)$$

Where:

- $Y(i,j)$ is the output of the convolution.
- X represents the input frame.
- K is the convolutional kernel (filter).
- H and W represent the height and width of the kernel.

3. **Pooling Layers:**

The pooling layers were used to decrease the spatial dimensions of the feature maps, thereby preserving information and reducing the level of computational complexity. The max pooling down samples the feature maps by picking the maximum value in each patch of the feature map.

- **Mathematical Formulation:**

$$(\ ) = \qquad (\ ,\ )$$

3.3.3 LSTM for Temporal Sequence Learning

In this architecture, 40 frames per video fed into the LSTM layer are processed sequentially, which lets the model learn the temporal dynamics of traffic events. The LSTM network processes the sequential data from video frames; it learns the temporal dynamics involved in accident scenarios. Long short-term memory is particularly suited for sequence data as it captures long term dependencies very well and can handle temporal information efficaciously.

1. LSTM Layer Details:

The LSTM layer, therefore, takes as input the feature vectors that CNN produces from all the frames. It works through those in sequence. This allows the LSTM to retain and learn temporal patterns arising from the sequence of frames, which is pretty much fundamental to determining whether there is an accident taking place, about to occur, or has just occurred.

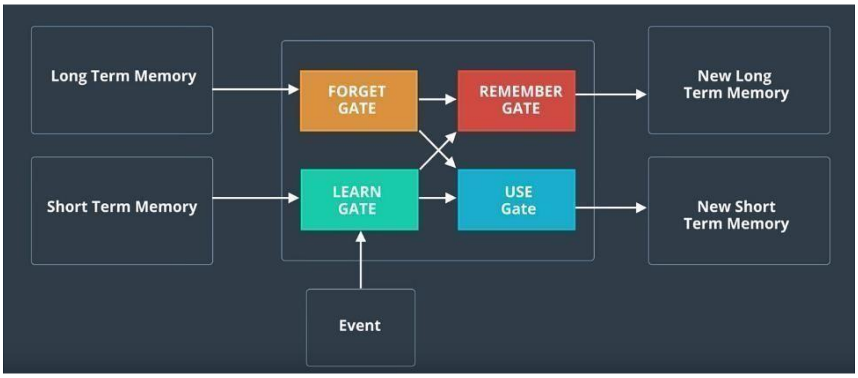


Figure 13 LSTM architecture

2. LSTM Mathematical Formulation:

An LSTM has three important gates: the input gate, the forget gate, and the output gate. These gates have control of how information flows within the network.

- Forget Gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f)$$

- Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)$$

- Output Gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o)$$

- Cell State Update:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

Where:

- σ is the sigmoid activation function.
- W_f, W_i, W_o are weight matrices.

3.3.4 Fully Connected Layer and Output

Finally, it is passed to a fully connected (dense) layer that takes the final hidden state of the LSTM and the annotation of the event it predicts either whether an accident occurred or not or the severity for an accident.

- **SoftMax Output for Classification**

The final dense layer uses SoftMax activation function to output the probability of belonging to one of the different classes:

- Accident/non-accident
- Severe/non-severe

- **Mathematical Formulation:**

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

Where:

- z_i represents the score for class i.
- The softmax function converts the raw output scores into probabilities.

3.4 Training Strategy

3.4.1 Data Splitting

For fair evaluation of the model, the dataset was divided into three sets, **75%** used for training, **15%** used to validate the model, and **10%** for testing. This will ensure that most of the data would be available for training, middle quantities would be used to hyperparameter tune-up, and unseen data to test up the performance of the model.

3.4.2 Hyperparameters

- **Learning Rate:**

The learning rate used was **0.0001**. This small learning rate was chosen to ensure slow and steady convergence of the deep learning models. A lower learning rate helps avoid large weight updates, which could destabilize the learning process, especially for complex models like LRCN and 3D CNN.

- **Batch Size:**

Training the models was done using a batch size of **4**. This small batch size is useful in cases where we want to deal with large datasets such as videos and use minimal GPU memory. Furthermore, using a small batch size will also allow for the weight updates to occur more frequently and a better generalization to take place for video-based model.

3.4.3 Optimizer

Taking into account that we deal with large volumes of data, and therefore it is necessary to have an optimizer with an adaptive learning rate, I noted the **Adam** optimizer. Adam is popular for deep learning models since it has the benefits of both **RMS Prop** and **SGD** (Stochastic Gradient Descent), making it possible to reach a result faster and being less likely to get the algorithm stuck at a local minimum.

3.5 Summary:

This system design focuses on deep learning models which include LRCN, LSTM, 3D CNN, and Movinet for the traffic accident detection and applications of severity classification in video surveillance. To ensure that the input data was as accurate as possible, the dataset was preprocessed by frame extraction, normalization, and labeling by hand. They trained the training strategy using **75/15/10** split for training, validation, and test data respectively using learning rate of **0.0001** with the batch size of **4** along with Adam optimizer. These models were trained using back propagation to enhance performance in real time traffic incident detection and classification accurately.

CHAPTER 4: RESULTS & DISCUSSION

4.1 Overview

This chapter provides the evaluation and analysis of the deep learning frameworks LRCN, LSTM, 3D CNN, and Movinet for traffic accident identification and intensity assessment. The findings are computed using the evaluation criterion of accuracy, precision, recall and F1-score with the purpose of comparing the performance of every model in accident recognition and severity forecasting. Of all the models developed, the LRCN model which acted as the main model had higher accuracy to the other models that were used to compare and contrast the effectiveness of the various kinds of models. This chapter also presents the issues faced and future enhancement, and evaluative results of the proposed approach for real-time accident detection systems.

4.2 Accident Detection Results

We are going to see that how each Deep learning model performed while detecting that Accident Occurred or not. As discussed, we used 2000 videos for training each Deep learning model.

4.2.1 LRCN Model Performance

The highest accuracy of **90.63%** for accident detection was achieved using the LRCN model, which captures features both spatially and temporally from video data. Its sequential processing through CNN to extract spatial features and LSTM to capture temporal dynamics makes it very suitable for analyzing the flow of a traffic accident.

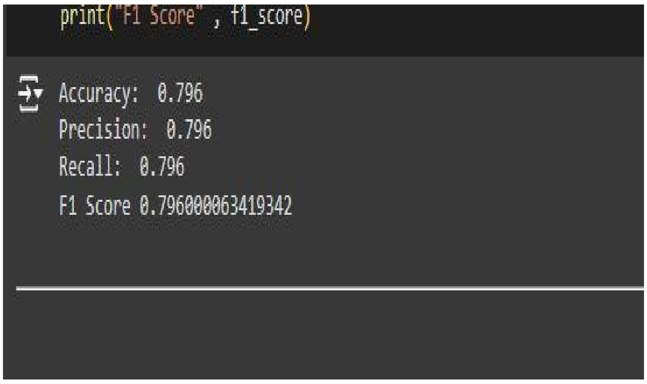
	precision	recall	f1-score	support
0	0.86	0.91	0.88	235
1	0.91	0.87	0.89	265
micro avg	0.89	0.89	0.89	500
macro avg	0.89	0.89	0.89	500
weighted avg	0.89	0.89	0.89	500
samples avg	0.89	0.89	0.89	500

Class accuracy: [0.9063829787234042, 0.8679245283018868]

Figure 14 Accident Detection Results for LRCN model

4.2.2 LSTM Model Performance

For accident detection and considering temporal variations of the sequences, the LSTM model stands at an accuracy of **79.6 percent**. Long term dependencies were effectively implemented in it, which made it possible to analyze continuous frames over time. On the other hand, the absence of spatial feature extraction during LSTM processing may be a reason why it ranked slightly lower than LRCN's accident detection performance in terms of percentage.



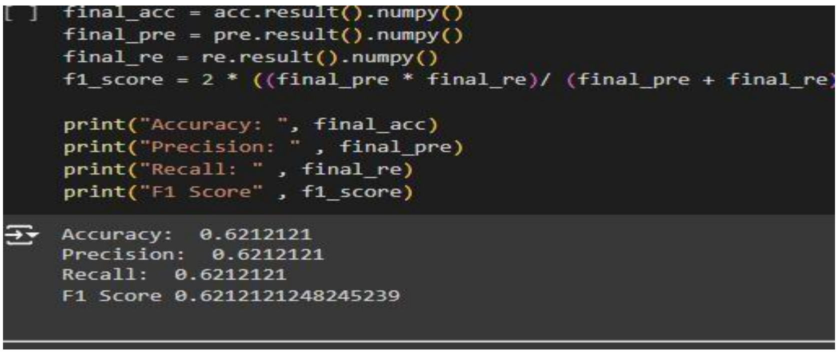
```
print("F1 Score" , f1_score)

Accuracy:  0.796
Precision:  0.796
Recall:  0.796
F1 Score 0.796000063419342
```

Figure 15 Accident Detection Results for LSTM model

4.2.3 3D-CNN Model Performance

For accident detection the 3D CNN model gave **62.1%** as accuracy as it can process spatial and temporal data at one time from the frames of the video. One of the major advantages of this approach is its ability to capture motion redundancies between frames, but it can get complicated, therefore computationally intensive. Although being capable of processing the video data as a whole, the 3D CNN could not surpass the results of models such as LRCN, probably because of the fewer efficiency of handling the long-term dynamics.



```
final_acc = acc.result().numpy()
final_pre = pre.result().numpy()
final_re = re.result().numpy()
f1_score = 2 * ((final_pre * final_re)/ (final_pre + final_re

print("Accuracy: ", final_acc)
print("Precision: " , final_pre)
print("Recall: " , final_re)
print("F1 Score" , f1_score)

Accuracy:  0.6212121
Precision:  0.6212121
Recall:  0.6212121
F1 Score 0.6212121248245239
```

Figure 16 Accident Detection Results for 3D-CNN model

4.2.4 MoviNet Model Performance

Movinet model trained on the accident data set had the accuracy of detecting accidents at **76.6%** using the real time video processing of the proposed model. Moreover, Movinet has been pre-trained to work on video streams which are ideal for real-time analysis. Nevertheless, it was slightly faster and more efficient than the method, but it was not as accurate as LRCN, which may be attributed to the fact that it is more generalized to a larger database than focusing specifically on traffic accident videos.

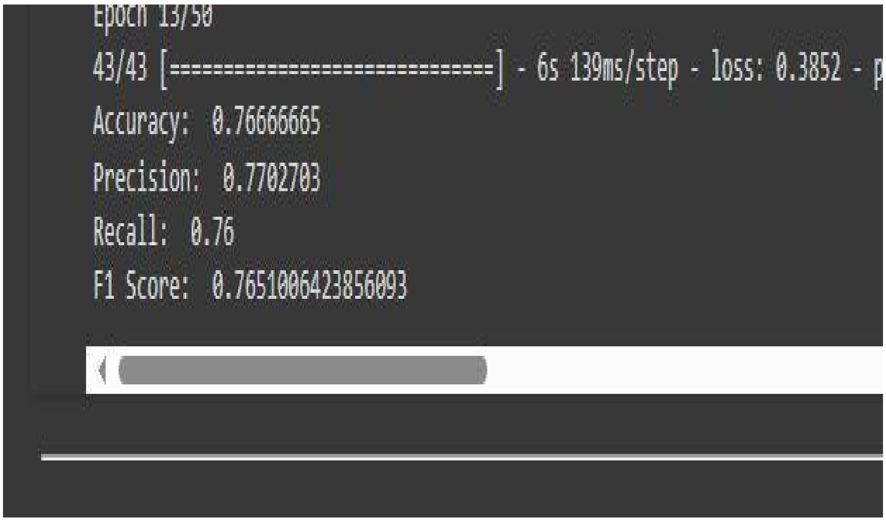


Figure 17 Accident Detection Results for MoviNet model

4.3 Severity Classification Results

We are going to see that how each Deep learning model performed while classifying the Severity of the Accident. As discussed, we used 1500 videos for training each Deep learning model.

4.3.1 LRCN Model Performance

The LRCN model showed a high level of accuracy in classification of severity with **83.33%** using a score of 1 for severe accidents and 0 for non-severe one. Its capacity to analyze features of space in frames of the video and temporal sequences over time lets it provide a better perception of accidents and their evolution. Due to this, makes LRCN most appropriate for localizing the extent of an accident within dynamic traffic scenes efficiently.

	precision	recall	f1-score	support
0	0.89	0.75	0.81	32
1	0.65	0.83	0.73	18
micro avg	0.78	0.78	0.78	50
macro avg	0.77	0.79	0.77	50
weighted avg	0.80	0.78	0.78	50
samples avg	0.78	0.78	0.78	50

Class accuracy: [0.75, 0.8333333333333334]

Figure 18 Severity Classification Results for LRCN model

4.3.2 LSTM Model Performance

A clear limitation of the LSTM approach, which was great at sequence processing, was the ability of correctly classify the severity of an accident; in this case, the LSTM model attained an accuracy of **55.7%.** Although LSTM has shown high efficacy for capturing temporal structure of the video sequences, it is quest of extracting spatial features capable of describing the severity level accurately, which has resulted in lower accuracy of LSTM when compared with models like LRCN.

```
final_accuracy = acc.result().numpy(
final_precision = pre.result().numpy
final_recall = re.result().numpy()
f1_score = 2 * ((final_precision * f

print("Accuracy: ", final_accuracy)
print("Precision: " , final_precisio
print("Recall: " , final_recall)
print("F1 Score" , f1_score)
```

```
Accuracy:  0.557971
Precision:  0.557971
Recall:  0.557971
F1 Score 0.5579710006713867
```

Figure 19 Severity Classification Results for LSTM model

4.3.3 3D-CNN Model Performance

Hence, the 3D CNN model for the severity classification had a classification accuracy of **52 %** with equal processing of both spatial and temporal data. Although utilized in the most effective manner for detecting general motion and transformations from one frame to another of a video, its discriminative capability in terms of accident severity was mediocre. It was noted that the model in question had a higher number of parameters compared to other networks, such as LSTM or LRCN, and it had a less efficient way of accounting for temporal dependencies – this could well explain the lower accuracy of estimations related to the severity of the accident.

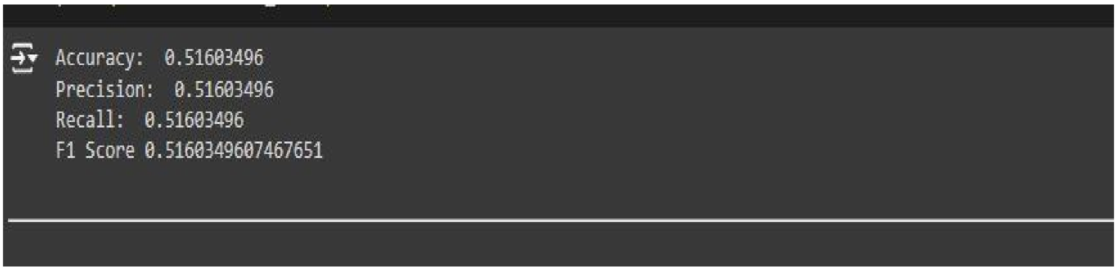


Figure 20 Severity Classification Results for 3D-CNN model

4.3.4 MoviNet Model Performance

The Movinet model attained an accuracy of **58%** in the classification of severity level, through using its aspects on real-time video processing. Although it is effective in terms of being fast and lightweight, its switch-off might be its pretrained form that restrained it from adapting best to the task of classifying severity of accidents. While exploring Movinet, it did not present as favorable results compared with another model – LRCN which examples performed worse only in recognizing between severe and non-severe accident cases.

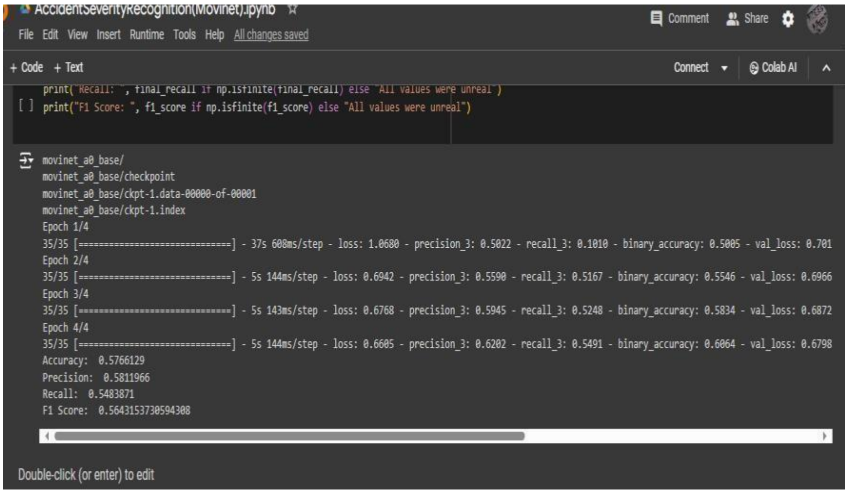


Figure 21 Severity Classification Results for MoviNet model

4.4 Comparison of Models

It also needs to be stated that a detailed comparison of the four models in terms of results of accidents detection and their severity revealed several important considerations. The LRCN model was relatively the best among all as it was able to achieve the highest accuracy in both activities because it so integrated CNN which recognizes spatial features and LSTM which recognizes temporal patterns. LSTM, though greatly capable to process sequential data, was less accurate as of LRCN, especially in accident detection since it lacks the ability to extract spatial features from the input data. 3D CNN, on the other hand, has both a spatial and temporal amelioration of data, but it failed to capture long term dependency, hence was less accurate. Finally, Movinet, which has primarily been discussed as real-time processing network gave efficient results but slightly less accurate as compared to LRCN; probable reason being this is generally trained for general video applications not specifically for traffic problems.

Comparison with other models shows that for tasks where spatial and temporal data are needed, including high-accuracy scenarios like traffic accident identification and severity determination, a hybrid approach like LRCN is appropriate.

Table 2 Table Showing the overall performance of models

Model	Batch Size	Frame Size	Accident Detection Accuracy	Precision (Accident)	Recall (Accident)	F1-Score (Accident)	Severity Classification Accuracy	Precision (Severity)	Recall (Severity)	F1-Score (Severity)
LRCN	4	64x64	90.63%	0.91	0.87	0.89	83.33%	0.65	0.83	0.73
LSTM	4	64x64	79.60%	0.79	0.79	0.79	55.70%	0.55	0.55	0.55
3D CNN	4	64x64	62.10%	0.62	0.62	0.62	52%	0.51	0.51	0.51
Movinet	8	64x64	76.60%	0.77	0.76	0.76	58%	0.56	0.54	0.57

4.5 Challenges and Limitations

4.5.1 Computational Limitations

Applying deep learning models for learning large videos especially LRCN and 3D CNN, was a resource-intensive process. Over usage of GPU was observed and the memory resources were in limited use which hampered the training and compare the result of different architectures.

4.5.2 Limited Dataset

Although the diversity in the dataset is good, further augmentation and variety in video data such as differences in lighting conditions, types of roads, and angles could possibly ensure a better generalization of the models. A much broader and more mixed dataset can help in avoiding overfitting, hence making these models better regarding robustness.

4.5.3 Difficulty in Differentiating Severity Levels

With the 3-class severity classification (low, medium, high), there is often not much visual difference between any of these classes. Classifying the accidents into three severities caused the models to fail to differentiate the severity of the accident while LSTM and 3D CNN were less accurate with the medium and high severity level predictions. Previous models such as Movinet and 3D CNN as well as models that can handle video data failed on this fine temporal analysis resulting in confusion between the medium and high severity accidents.

The model LRCN Achieved accuracies of **48% ,52% and 35%** for High, Medium and Low Category for Severity classification Respectively with training dataset of 1500 videos (499 for High and 508 for Medium and 532 for low)

```
WARNING:py.warnings:/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:13
  _warn_prf(average, modifier, msg_start, len(result))

              precision    recall  f1-score   support

           0       0.62      0.45      0.52       114
           1       0.38      0.52      0.44        92
           2       0.39      0.35      0.37       102

   micro avg       0.45      0.44      0.44       308
   macro avg       0.46      0.44      0.44       308
weighted avg       0.47      0.44      0.45       308
 samples avg       0.44      0.44      0.44       308

Class accuracy: [0.4824561403508772, 0.5217391304347826, 0.35294117647058826]
```

Figure 22 Severity Classification Results for 3 classes for LRCN model

CHAPTER 5: CONCLUSION AND FUTURE WORK

5.1 Conclusion

For this thesis, a deep learning-based approach for traffic accident detection and severity classification based on video surveillance was proposed and tested. Among the four models: LRCN, LSTM, 3D CNN, and Movinet, all four were trained to test their capability in traffic monitoring in real time. Among all the proposed models, the LRCN, CNN for the spatial feature extraction and LSTM for temporal sequence analysis showed the best result in terms of accuracy for both accident detection and severity classification.

According to the results, deep learning models possess promising ability to enhance road safety through automatic accident identification Although these outcomes illustrate the work of deep learning models in enhancing road safety through automatic accident identification, this study also described some of the major obstacles such as the imbalance of the dataset, the feasibility issue in terms of computation and explanation, and the difficulty in differentiating the severity of the accidents. Severities were classified into low medium, and high, and these were hard to differentiate based on visual differences with multiple classes leading to class imbalance.

However, the results of the research indicate that deep learning models such as LRCN offers a lot of potential for improving traffic flow and response time to road occurrences. The future of this work could be to extend the data set used, optimization of the models to better differentiate between severity and problems stemming from constrained computational resources. The studies laid down in this research form a good background on the implementation of AI surveillance systems in smart city hence leading to safer roads besides enhancing quick response to any eventuality.

5.2 Future Work

Future work may provide significant improvements in the accuracy of severity classification, especially in the three-class system, due to improvements in the deep learning models and the additional use of data, for example, post-accident reports, or vehicle sensors. It is also important to address the problem of unbalanced data and the creation of many more various and large datasets should help to increase the quality of the constructed models. Furthermore, research on novel architectures such as transformers or a combination of models with some novel attention-based mechanism can give additional fine-grained information for accidents in the models. Improving computational capabilities will also extend training on higher resolution videos which improves the real time performance of the system. At last, using these models in actual traffic monitoring systems and updating and enhancing the models or using the refined models in practical traffic monitoring systems in smart city will make the models more usable in practice for intelligent response for traffic incidents.

REFERENCES

[1] H. Yu, X. Zhang, Y. Wang, Q. Huang and B. Yin, "Fine-Grained Accident Detection: Database and Algorithm," in IEEE Transactions on Image Processing, vol. 33, pp. 1059-1069, 2024, doi: 10.1109/TIP.2024.3355812

[2] O. I. Aboulola, E. A. Alabdulqader, A. A. AlArfaj, S. Alsubai and T. -H. Kim, "An Automated Approach for Predicting Road Traffic Accident Severity Using Transformer Learning and Explainable AI Technique," in IEEE Access, vol. 12, pp. 61062-61072, 2024, doi: 10.1109/ACCESS.2024.3380895.

[3] K. A. Sattar, I. Ishak, L. S. Affendey and S. N. B. Mohd Rum, "Road Crash Injury Severity Prediction Using a Graph Neural Network Framework," in IEEE Access, vol. 12, pp. 37540-37556, 2024, doi: 10.1109/ACCESS.2024.3373885.

[4] K. P. Hlaing, N. T. T. Aung, S. Z. Hlaing and K. Ochimizu, "Analysis of accident severity factor in Road Accident of Yangon using FRAM and Classification Technique," 2019 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 2019, pp. 256-261, doi: 10.1109/AITC.2019.8921119.

[5] L. S, R. Gopan, M. M L, A. V and M. R. Elizabeth, "Vehicle Accident Detection and Prevention using IoT and Deep Learning," 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), THIRUVANANTHAPURAM, India, 2022, pp. 22-27, doi: 10.1109/SPICES52834.2022.9774089.

[6] G. A. Senthil, R. V. Lakshmi Priya, S. Geerthik, G. Karthick and R. Lavanya, "Safe Road AI: Real-Time Smart Accident Detection for Multi-Angle Crash Videos using Deep Learning Techniques and Computer Vision," *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2024, pp. 617-622, doi: 10.1109/ICAAIC60222.2024.10575074.

[7] G. A. Senthil, R. V. Lakshmi Priya, S. Geerthik, G. Karthick and R. Lavanya, "Safe Road AI: Real-Time Smart Accident Detection for Multi-Angle Crash Videos using Deep Learning Techniques and Computer Vision," *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2024, pp. 617-622, doi: 10.1109/ICAAIC60222.2024.10575074.

[8] J. Fang, J. Qiao, J. Xue and Z. Li, "Vision-Based Traffic Accident Detection and Anticipation: A Survey," in IEEE Transactions on Circuits and Systems for Video Technology, doi: 10.1109/TCSVT.2023.3307655.

[9] Z. U. Arifeen, J. -E. Hong, B. -S. Seo and J. -W. Suh, "Traffic Accident Detection and Classification in Videos based on Deep Network Features," 2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN), Paris, France, 2023, pp. 491-493, doi: 10.1109/ICUFN57995.2023.10199977

[10] K. Shaaban and M. S. Ghanim, "Modeling of Severity in Red-Light-Running Crashes Using Deep Learning Recognition," 2023 Intermountain Engineering, Technology and Computing (IETC), Provo, UT, USA, 2023, pp. 181-186, doi: 10.1109/IETC57902.2023.10152094

[11] J. A. Sowdagur, B. T. B. Rozbully-Sowdagur and G. Suddul, "An Artificial Neural Network Approach for Road Accident Severity Prediction," 2022 IEEE Zooming Innovation in Consumer Technologies Conference (ZINC), Novi Sad, Serbia, 2022, pp. 267-270, doi: 10.1109/ZINC55034.2022.9840576

[12] Y. Yao, M. Xu, Y. Wang, D. J. Crandall and E. M. Atkins, "Unsupervised Traffic Accident Detection in First-Person Videos," 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2019, pp. 273-280, doi: 10.1109/IROS40897.2019.8967556.

[13] Basheer Ahmed, M.I.; Zaghdoud, R.; Ahmed, M.S.; Sendi, R.; Alsharif, S.; Alabdulkarim, J.; Albin Saad, B.A.; Alsabt, R.; Rahman, A.; Krishnasamy, G. A Real-Time Computer Vision Based Approach to Detection and Classification of Traffic Incidents. Big Data Cogn. Comput. 2023, 7, 22. https://doi.org/10.3390/bdcc7010022

[14] M. Essam, N. M. Ghanem, and M. A. Ismail, “Detection of road traffic crashes based on collision estimation,” arXiv:2207.12886, 2022

[15] M. I. B. Ahmed et al., “A real-time computer vision-based approach to detection and classification of traffic incidents,” Big Data Cogn. Comput., vol. 7, no. 1, p. 22, 2023.

[16] T. K. Vijay, D. P. Dogra, H. Choi, G. P. Nam, and I. Kim, “Detection of road accidents using synthetically generated multi-perspective accident videos,” IEEE Trans. Intell. Transp. Syst., vol. 24, no. 2, pp. 1926– 1935, 2023

[17] D. Singh and C. K. Mohan, “Deep spatio-temporal representation for detection of road accidents using stacked autoencoder,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 3, pp. 879–887, 2019.

[18] R. E. Almamlook, K. M. Kwayu, M. R. Alkasisbeh, and A. A.Frefer, “Comparison of machine learning algorithms for predicting traffic accident severity,” 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2019 - Proceedings, pp. 272–276, May 2019, doi: 10.1109/JEEIT.2019.8717393.

[19] Road Safety: Data Show Improvements in 2017 But Renewed Efforts are Needed for Further Substantial Progress, Eur. Commission, Paris, France, Apr. 2018.

[20] A. S. Al-Ghamdi, “Using logistic regression to estimate the influence of accident factors on accident severity,” Accident Anal. Prevention, vol. 34, no. 6, pp. 729–741, 2002.

.