Road Traffic Accident Severity Prediction using Deep Learning

Submitted in partial fulfillment of the requirements
of the degree of

BACHELOR OF ENGINEERING
In
COMPUTER ENGINEERING

By

Group No: 47

Roll No.	Name
2003079	Viren Kalra
2003085	Harsh Kasliwal
2003112	Zeel Mody
2003135	Vaibhavi Pore

Guide:

Prof. VAIBHAV AMBHIRE

(Assistant Professor, Department of Computer Engineering, TSEC)

Computer Engineering Department
Thadomal Shahani Engineering College
University of Mumbai
2023-2024

CERTIFICATE

This is to certify that the project entitled **"Road Traffic Accident Severity Prediction using Deep Learning"** is a bonafide work of

2003079	Viren Kalra
2003085	Harsh Kasliwal
2003112	Zeel Mody
2003135	Vaibhavi Pore

Submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"BACHELOR OF ENGINEERING"** in **"COMPUTER ENGINEERING"**.

Mr. Vaibhav Ambhire
Guide

Dr. Tanuja Sarode
Head of the Department

Dr. G.T. Thampi
Principal

Project Report Approval for B.E

Project report entitled *"**Road Traffic Accident Severity Prediction using Deep Learning**"* by

2003079	Viren Kalra
2003085	Harsh Kasliwal
2003112	Zeel Mody
2003135	Vaibhavi Pore

is approved for the degree of ***"BACHELOR OF ENGINEERING" in***

"COMPUTER ENGINEERING".

Examiners

1. --------------------------------

2. --------------------------------

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have not been properly cited or from whom proper permission has not been taken when needed.

1) ____________________

Viren Kalra, 2003079

2) ____________________

Harsh Kasliwal, 2003085

3) ____________________

Zeel Mody, 2003112

4) ____________________

Vaibhavi Pore, 2003135

Date:

Abstract

Road accidents represent a significant and persistent threat to public safety in India, demanding innovative solutions to mitigate their impact. The proposed system is devoted to addressing this critical issue by predicting accident severity, a vital step towards enhanced road safety. We draw upon diverse datasets gathered from various Indian states meticulously curated from Kaggle and other reputable sources. These datasets undergo careful merging and comprehensive preprocessing to ensure high-quality data inputs for our predictive model. The urgency of India's road safety situation cannot be overstated, making the proposed system a critical endeavor. By doing so, we aim to provide valuable insights and early warning systems capable of reducing the frequency of accidents and diminishing their impact on the public. The proposed system seeks to offer a comprehensive understanding of accident data through the lens of advanced machine learning techniques. The utilization of a CNN for predicting accident severity, complemented by selectKBest() for feature analysis, represents a groundbreaking shift in the approach to road safety across various Indian states. Thereafter, we have further implemented RNN and LSTM to indicate\ the accurate model. Furthermore, the proposed system introduces an innovative way to present its predictions: visually, through pie charts. This visualization method simplifies the communication of accident severity distribution, making it easily interpretable for a wide audience. Through this research initiative, we aspire to contribute significantly to ongoing endeavors aimed at enhancing road safety and accident prevention across multiple Indian states. The proposed system has the potential to assist authorities, policymakers, and individuals in making informed decisions that will result in a safer road environment, ultimately safeguarding lives and well-being. Road safety is an obligation shared by society, and the proposed system endeavors to play a pivotal role in the broader mission of reducing accidents and preserving lives within these diverse regions.

TABLE OF CONTENTS

List of Figures

Chapter 1

Introduction

1.1 Introduction

In a world increasingly defined by the dynamics of urbanization and expanding transportation networks, road safety has become an imperative concern. Accurate prediction of accident severity stands as a vital step towards minimizing the impact of road accidents on society. This project delves into the realm of accident severity prediction. Our methodology hinges on the Convolutional Neural Networks (CNN) for predictive modeling. By implementing these advanced techniques, we aim to unveil latent patterns within the accident data, empowering stakeholders with insights for effective decision-making and safety enhancement. The data foundation of this project comprises meticulously sourced datasets from Kaggle, a renowned platform for data science resources. These datasets underwent rigorous preprocessing, ensuring data cleanliness to maximize their relevance for our deep learning model. This project's core mission is to develop an innovative approach to accident severity prediction. By incorporating CNN, we aspire to provide a groundbreaking solution for road safety enhancement. The outcomes of our endeavor will be presented in a visually intuitive format through a map, offering a clear understanding of accident severity. As we navigate through the intricacies of our methodology, we anticipate shedding light on valuable insights that could significantly contribute to saving lives and resources by bolstering road safety practices in these states.

1.2Problem Statement & Objectives

1.2.1 Problem Statement

To design a predictive model that can accurately predict the severity of road traffic accidents based on various factors and conditions. The model will take into account various accident-related factors and conditions, such as weather conditions, road type, time of day, vehicle types involved, and other relevant attributes. The anticipated output will be a severity level classification, encompassing categories like low, moderate, or high. The dataset for this project will be sourced from reliable governmental sources, ensuring data authenticity and accuracy. It will comprise historical records of road traffic accidents, including attributes like weather conditions, road type, time of the day, vehicle types involved, and other relevant factors.

1.2.2 Objectives

The main objectives of this project are to design, develop, and deploy a comprehensive predictive model that can accurately assess and predict the severity of road traffic accidents based on a variety of factors and conditions. The project aims to leverage a robust dataset from reliable governmental sources, containing historical records of road traffic accidents with details such as weather conditions, road type, time of day, and the types of vehicles involved. By classifying accident severity levels into categories such as low, moderate, or high, the model will provide actionable insights for traffic management authorities, emergency services, and policy-makers to enhance road safety. A key objective is to extend the model's geographical coverage to include multiple regions or cities, offering predictions on a larger scale and benefiting diverse communities. To improve prediction accuracy and depth of insight, the project seeks to integrate more diverse data sources, such as video feeds from traffic cameras, social media data, and vehicle sensor data. This multi-source approach will enhance the model's predictive capabilities and contribute to more precise accident severity assessments. Collaboration with traffic management authorities is another primary objective, as integrating the prediction system into existing traffic control infrastructure can enable real-time traffic monitoring and control. Additionally, the project aims to evolve beyond predictions to detect traffic incidents as they occur and provide early alerts to relevant authorities, thus improving response times and potentially saving lives. The project will also encourage a community-driven approach to road safety by allowing users to report accidents and traffic conditions in real-time. By presenting the predictions on a map interface, users can access an enhanced view of predicted accident severity and associated traffic conditions, supporting safer travel choices. Exploring state-of-the-art deep learning architectures, such as neural networks, is another goal of the project, aiming to improve the accuracy and capabilities of the prediction model. Furthermore, the system's explainability will be enhanced to offer clearer insights into the decision-making process behind each prediction, providing valuable information for decision-makers and fostering trust in the model. Overall, the project's objectives revolve around creating a sophisticated, multi-source predictive model that provides comprehensive, real-time insights into traffic accident severity. By incorporating advanced technologies, real-time data, and a community-focused approach, the project aims to significantly improve road safety and traffic management in various regions.

1.3 Scope

Extend the project's scope to cover a larger geographical area, allowing it to provide traffic accident predictions for multiple regions or cities. Include more diverse data sources, such as video feeds from traffic cameras, social media data, and vehicle sensor data, to improve prediction accuracy. Collaborate with traffic management authorities to integrate the prediction system into their infrastructure for real-time traffic control. Expand the system to not only predict accident severity but also detect traffic incidents as they occur and provide early alerts to authorities. Implement features that allow users to report accidents and traffic conditions in real-time, fostering a community-driven approach to road safety. Integrate the output on the map to enable the users with the enhanced view about the accident predictions. Explore state-of-the-art deep learning architectures like convolutional neural networks for improved prediction capabilities. Enhance the system's explainability to provide clearer insights into why a particular prediction was made, making it more valuable for decision-makers.

Chapter 2

Review of Literature

2.1 Domain Explanation

Deep learning is the branch of machine learning which is based on artificial neural network architecture. A deep learning model uses layers of interconnected nodes called neurons that work together to process and learn from the input data. In a fully connected Deep neural network, there is an input layer and one or more hidden layers connected one after the other. Each neuron receives input from the previous layer neurons or the input layer. The output of one neuron becomes the input to other neurons in the next layer of the network, and this process continues until the final layer produces the output of the network. The layers of the neural network transform the input data through a series of nonlinear transformations, allowing the network to learn complex representations of the input data. In this project following deep learning concepts have implemented and compared:

1. **Convolutional Neural Networks (CNN):** In a one-dimensional CNN, the convolutional filters slide across the input data along one dimension, typically time or sequence length. Each filter learns to detect local patterns or features within a small window of the input. The output of these filters is then typically passed through an activation function and pooled to reduce dimensionality and extract the most relevant features. This process is

repeated through multiple layers, allowing the network to learn increasingly complex patterns. Finally, the learned features are flattened and fed into one or more fully connected layers for classification or regression.

2. **Recurrent Neural Networks (RNN):** RNNs are designed to handle sequential data by processing inputs in a sequential manner, where the output at each time step depends not only on the current input but also on previous inputs in the sequence. They have been widely used in tasks such as speech recognition, language modeling, and time series prediction. However, traditional RNNs suffer from the vanishing gradient problem, which limits their ability to capture long-range dependencies in sequences.
3. **Long Short-Term Memory (LSTM):** LSTMs are a type of RNN architecture specifically designed to address the limitations of traditional RNNs. LSTMs introduce a more sophisticated memory mechanism that allows them to selectively remember or forget information over long sequences, making them more effective at capturing long-term dependencies. LSTMs have become the preferred choice for many sequential data tasks due to their ability to maintain context over extended periods.

2.2 Review of Existing Systems

2.2.1 Deep Learning-based Prediction of Traffic Accident Risk in Vehicular Networks [1]

In this reference paper, a traffic accident risk forecasting algorithm based on deep learning is proposed. The forecasting model can either extract multi-dimensional features through convolution layers of CNN or select and classify the features by a decision tree again and again in the training process of the RF, and the prediction results can approach the real value infinitely. Simulations showed that the forecasting model has higher AUC and accuracy and lower loss, compared to the CNN based method. Feature extraction is composed of multiple layers - input layer, convolutional layers, ReLU layers, pooling layers, and flatten layers. In order to reduce parameters and avoid overfitting, the pooling layer performs dimensionality reduction. In the

Feature Classifier, RF is the combination of decision tree and Bagging. The forecasting model is a hybrid model, the CNN based method is used to obtain multi-dimensional features from distinct angles, and then the RF is used for feature classification and training.

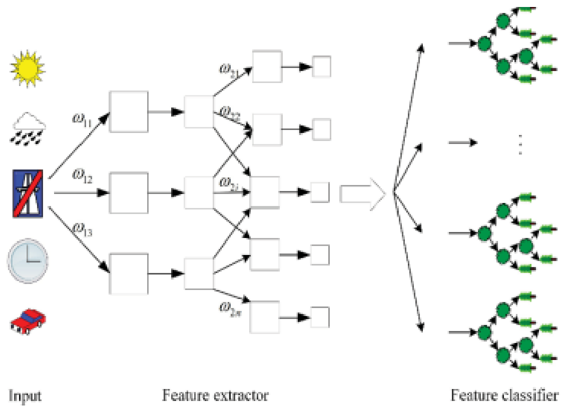

Fig. 2.1 The Architecture of the Forecasting Model

2.2.2 Accident Detection, Severity Prediction, Identification of Accident Prone Areas in India and Feasibility Study using Improved Image Segmentation, Machine Learning and Sensors [2]

The system provides a three-phase solution to analyze road accidents in India using machine learning and computer vision. The first phase includes the identification of accident prone areas in India based on the number of accidents occurring across various states and union territories. Here, linear regression was applied which handled the continuous data promptly giving a much appreciable accuracy of 95%. The second phase of the solution involves the classification of car accidents. In this phase, watershed segmentation was applied to the videos and images of car accidents. To further enhance the segmentation, they made use of masking on top of Fast RCNN. The third phase of the solution involves the designing of a mitigation alert system using the mask R-CNN model.

2.2.3 A Deep Learning based Accident Detection System [3]

The proposed system in this reference paper provides a mechanism that sends an alert to the control room to ensure timely action to help the victim of the accident. For the working of the system, deep learning technique was used that uses convolutional neural networks. The trained model was tested with external images and obtained an accuracy of eight-five percent. The sequential model mainly consists of the convolution, Batch Norm and the ReLu activation function. Video classification is done by performing the technique of averaging of frames. The probability of occurrence of accidents were checked frame by frame and if the probability indicates an accident, then emergency messages are sent to nearby control rooms using the GSM module.

2.2.4 Traffic Accident Prediction based on CNN model [4]

This proposed system uses Java high-level neural network API framework, India Accident 2016-2018 data set to simulate the suggested traffic accident prediction algorithm. The proposed Convolutional neural network (CNN) is compared with traditional neural network based back propagation prediction methods.

2.2.5 Severity Prediction of Traffic Accidents with Recurrent Neural Networks [5]

The proposed severity prediction model is based on LSTM-RNN. The RNN model comprises five main layers which include input, LSTM, two dense layers, and a Softmax layer. The network is trained with Backpropagation Through Time(BPTT) and Stochastic Gradient Descent(SGD) algorithm. To avoid over-fitting in the proposed RNN model, three standard techniques were used. These were Gaussian noise injection into the training data, using a ReLU activation function in the hidden layers, and subsequently applying the dropout technique. The hyper-parameters of the RNN model were selected by performing a systematic grid search implemented in scikit-learn using 100 epochs.

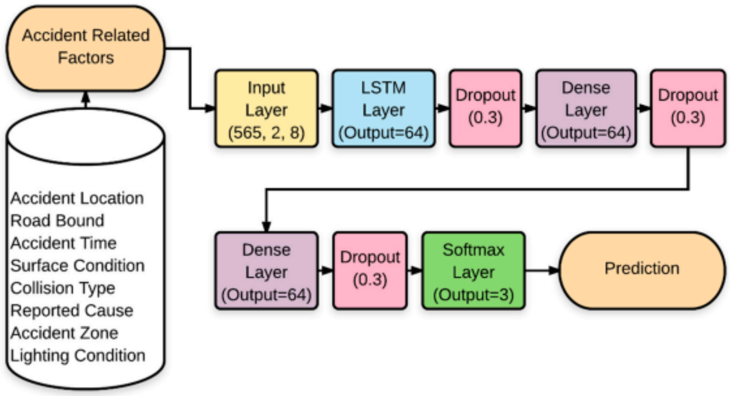

Fig. 2.2 The high-level architecture of the proposed RNN model

2.2.6 A Road Accident Prediction Model Using Data Mining Techniques [6]

The working of the proposed model is divided into four modules – Rule Mining, Risk Prediction, Graph Plot and New Data Entry. Rule mining is done using Apriori Algorithm and Risk Prediction is done using SVM (Support Vector Machine) Algorithm.

2.2.7 Deep learning based detection and localization of road accidents from traffic surveillance videos [7]

This paper uses a deep learning approach based on spatio-temporal autoencoder and sequence-to-sequence long short-term memory (LSTM) autoencoder. The research proposed in this paper addresses the task of accident detection by following an unusual activity detection approach based on deep learning and one-class classification paradigm The crux of the proposed approach is that it learns in an unsupervised manner alleviating the need to generate labeled anomalous road accident data.

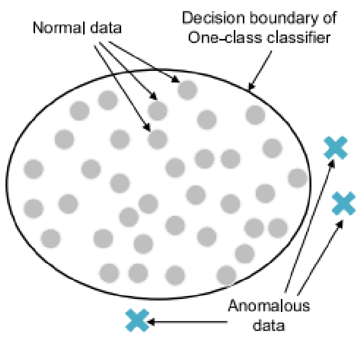

Fig. 2. One-class classification.

Fig. 2.3 One-class classification

2.3 Limitation of Existing Systems

Road accidents occur spontaneously. There are many different factors involved in the occurrence of an accident and the factors that contribute majorly in the accidents are variable and vague in nature. The main research gap in the existing systems is that the input parameters might not be sufficient in accurately predicting road accidents of varied types. Insufficient input parameters can lead to a lack of comprehensive understanding of accident risk. Some crucial variables, like road conditions, driver behavior, and vehicle maintenance, may not be adequately considered, limiting the models' ability to provide valuable insights. Unusual or outlier situations that can lead to accidents may not be effectively captured with a limited set of parameters. For instance, rare weather events or extraordinary driving behavior might be missed, reducing the models' robustness.

Further, there should be integration of the model with the systems of the vehicle for real-time prediction. Not integrating the model with these systems means missing out on valuable real-time data that could be used for more accurate and timely accident predictions. Also, many modern vehicles are equipped with Advanced Driver Assistance Systems (ADAS), including features like lane departure warnings, adaptive cruise control, and collision avoidance systems. Integrating the accident severity prediction model with these systems could enable proactive safety measures, such as automatically applying the brakes or steering to avoid accidents.

The study's limitations include potential biases from the dataset's self-selection, limited exploration of advanced neural architectures, and exclusion of crucial variables like ethnicity and weather conditions, which could affect the analysis of road accident severity [1]. The proposed alert system's limitations include potential challenges in real-time data accuracy, integration issues, lack of discussion on privacy and security concerns, and uncertainties about the practicality and effectiveness of the mentioned tracking system for accident-prone areas [2]. Rigorous data verification and robust security measures are crucial before practical implementation. The limitations of the proposed system include reliance on fixed camera locations, potential gaps in coverage, uncertain performance in various conditions, and lack of privacy considerations, indicating the need for extensive testing and ethical considerations before widespread implementation [3]. The proposed model's limitations include its modest impact on prediction rate, lack of exploration into additional features, and limited comparison with diverse machine learning algorithms beyond Back Propagation networks [4]. These factors raise concerns about its generalizability and effectiveness in real-world highway scenarios, necessitating further research and validation. The proposed RNN model's limitations include fixed hyper-parameters sensitivity, dependency on sequence length, and a focus on temporal-contextual factors, potentially overlooking complex spatial relationships [5]. The study lacks real-time adaptability and dynamic network adjustments, hindering its practical application in diverse traffic scenarios. Further research is needed to enhance the model's flexibility and real-world usability. The model's limitations include the omission of crucial factors like driver emotions, lack of comprehensive accuracy validation, limited regional applicability, absence of real-time data integration, and theoretical proposals for mobile app integration, highlighting the need for more diverse data, psychological factors inclusion, rigorous validation, and real-world testing before practical implementation [6]. The model's limitations include potential challenges in handling diverse accident scenarios not well-represented in training data, lack of discussion on real-world implementation challenges like varying conditions and camera angles, limited exploration of false positive/negative rates, and uncertainties about efficiency and scalability when deployed on specific hardware like PYNQ [7]. Computational resource requirements for real-time processing are not addressed, underscoring the need for comprehensive testing and evaluation before practical deployment.

Chapter 3

Proposed System

3.1 Framework

The framework for Accident Prediction includes the dataset that is sourced from governmental bodies, encompassing factors pertinent to road accidents such as weather conditions, road conditions, and accident timing. Subsequently, the data undergoes preprocessing, entailing handling missing values, outliers, and formatting for compatibility with deep learning models. This involves structuring various convolutional layers, pooling layers, and fully-connected layers within a CNN architecture, followed by fine-tuning of parameters. Furthermore, activation functions, loss functions, and optimizer algorithms are specified. The refined CNN model ingests input, processes it, and yields predictions regarding accident occurrences based on parameters like location, time, and weather conditions. For Graphical User Interface, the project implements various routing APIs and libraries. OpenLayers library is used to display various types of geographic data like routes and coordinates. On the other hand, OpenRouteService API serves as an open-source routing solution, facilitating route planning and optimization. Through OpenRouteService, the project computes routes between destinations, tailoring criteria such as speed, distance, and avoidance of certain road features like tolls or ferries. The coordinates defining a route are fed into our deep learning model. This model then computes the likelihood of accidents occurring along this route. Subsequently, areas with heightened probabilities of accidents are identified and highlighted, serving as a proactive warning mechanism for users navigating through these regions.

3.2 Design Details

3.2.1 GUI Design

1. Form interface -

   The form takes input from the user. The input fields include ‘Age of driver’, ‘Age of vehicle’, ‘Engine Capacity’, ‘Gender’ and ‘Vehicle Type’.

2. Map interface -

   Map interface is implemented using OpenLayers which is an open-source JavaScript library that enables developers to create interactive maps on web pages. It provides a powerful set of tools for displaying, interacting with, and styling geographic data on the web. OpenLayers supports a wide range of map data sources, including popular ones like OpenStreetMap, Google Maps, Bing Maps, and many others.

3. Route interface -

   The route between two coordinates is displayed using OpenRouteService(ORS) API. ORS provides a RESTful HTTP API that provides routing capabilities. It obtains optimal routes between locations considering factors like distance, travel time, traffic conditions, and elevation.

3.2.2 Deployment Diagram

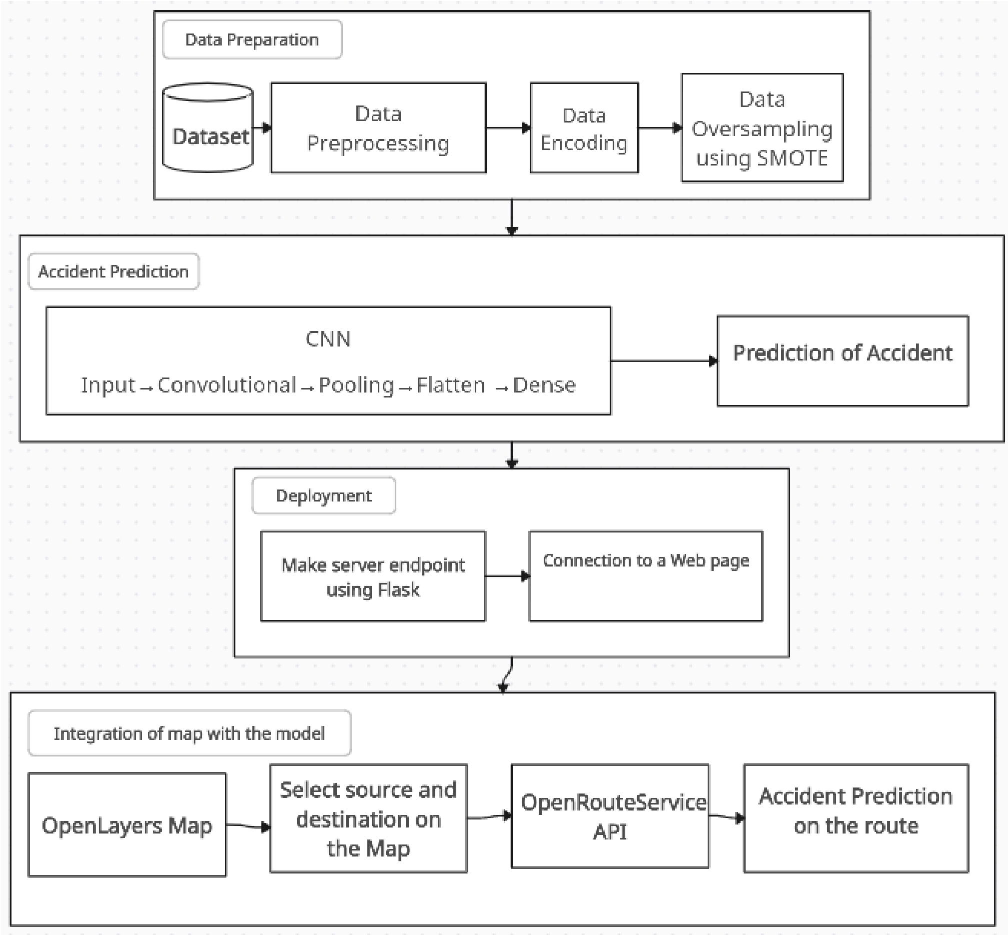

Fig. 3.1 Prediction Model Development

In an effort to enhance road safety, the project harnesses the power of artificial intelligence. This system leverages a convolutional neural network (CNN), a type of artificial neural network particularly adept at analyzing image data. By feeding the CNN data on traffic flow patterns and weather conditions, the system can identify patterns and relationships that influence accident likelihood. However, this data needs some preparation before it can be used by CNN. Raw data from various sources undergoes preprocessing, a cleansing process that ensures its compatibility with the model. This might involve cleaning inconsistencies, converting non-numerical data into a format the CNN understands, and even balancing the dataset using a technique called SMOTE (Synthetic Minority Oversampling Technique) to account for potential biases. Once the data is prepped, the CNN analyzes it and generates a prediction for the likelihood of an accident. To make this information readily available, the CNN model is then deployed as a web service, essentially making it accessible to other applications and websites. This deployment phase involves creating a user-friendly interface using Flask, a Python web framework, and integrating it with a mapping service like OpenLayers Map. This allows users to select a route on the map. The system then retrieves detailed route information using the OpenRouteService API and combines this with the accident probability predicted by the CNN model. This combined analysis provides users with an estimate of the likelihood of accidents along their chosen route, empowering them to make informed decisions about their travels.

3.3 Methodology

3.3.1 Prediction Model Development

1. **Data Preparation**

   The collection of dataset is done from Kaggle which includes road accident-related factors like weather conditions, road conditions, time of the accident and other such relevant factors. The dataset was in two different parts - Accident data and Vehicle data. The Accident dataset contained features related to the accident such as Latitude, Longitude, Time, Junction Detail, Light Conditions. The Vehicle dataset contained features related to the vehicle involved in the accident such Age of Vehicle, Engine Capacity, Age of Driver, Vehicle Manoeuvre.

2. **Data Preprocessing**

   In the data preparation phase, various transformations are applied to each feature to ensure data quality and suitability for subsequent modeling tasks. It involved removal of irrelevant features which do not contribute to predictive modeling. The features with greater amounts of missing values were removed. Values of the features which had a minor percentage of missing values, were imputed. Two types of imputation methods are involved - Imputation by mean values and Imputation by mode values. Further, the two different datasets were merged into one single dataset.

3. **Data Encoding**

   Data encoding is performed on the preprocessed dataset using pd.get_dummies() method. It is used for converting categorical variables into indicator variables. It is a crucial preprocessing step for deep learning tasks, where algorithms typically require numerical input.

4. **Data Oversampling**

   The distribution of classes in the dataset was imbalanced, with one class having far fewer instances than the others. The 'Severe' accident prediction class had very few instances than the 'Slight' and 'Serious' accident prediction class. This can lead to a biased deep learning model that favors the majority class and performs poorly on the minority class. Hence SMOTE is implemented, which stands for Synthetic Minority Over-sampling Technique, is a popular method used in machine learning and particularly in the field of imbalanced class distribution. It is designed to address the issue of imbalanced datasets, where one class (typically the minority class) is significantly underrepresented compared to the other classes.

5. **Model Configuration**

   The preprocessed dataset is trained on three deep learning models - CNN, RNN and LSTM. CNN model involves two convolutional layers, two pooling layers, one flatten layer and two dense layers. RNN model involves RNN layer, a dropout layer and two dense layers. The LSTM model involves an LSTM layer, two dropout layers and two dense layers. The activation functions used in all the three models are relu function and softmax function.

6. **Model API**

   Flask API is implemented for serving predictions using the pre-trained Deep Learning model. The ‘predict’ endpoint is defined to accept POST requests containing input data in JSON format. It preprocesses the input data using standard scaling, then passes it to the pre-trained model to make predictions. Finally, it returns the predictions as JSON to the client.

3.3.2 Map Representation

1. **OpenLayers**

   OpenLayers is an open-source JavaScript library that is implemented for displaying and interacting with maps on web pages and to create maps with various layers, markers, vector shapes, and controls.

2. **Obtain the route directions**

   OpenRouteService is an open-source routing service that provides route planning and optimization. It is implemented for routing profiles tailored to different types of transportation modes, including car and bike. OpenRouteService calculates routes between locations based on different criteria such as fastest route, shortest route, and route avoiding toll roads or ferries.

3. **Alert the users based on their travel routes**

   The route’s certain latitudes and longitudes are given input into our deep learning model which computes the probability of accident occurrence. The coordinates with high probability of accident occurrence are highlighted to inform and alert the users.

Chapter 4
Implementation Details

4.1 Experimental Setup

4.1.1 Dataset Description

Dataset: Indian Road Accident Dataset

Format: Excel (.csv)

Rows: 86,000

Columns of the dataset-

a. Date
b. Time
c. Junction_Detail
d. Latitude
e. Longitude
f. Light_Conditions
g. Number_of_Casualties
h. Number of Vehicles
i. Engine Capacity
j. Vehicle manoeuvre
k. Vehicle type
l. Road_Surface_Conditions
m. Road_Type
n. Speed_limit
o. Weather_Conditions
p. Age_Band_of_Driver
q. Age_of_Vehicle
r. Accident_Severity
s. Junction Control
t. Urban_or_rural
u. Sex_of_driver
v. Point_of_impact

Metadata Highlights:

1. Data Source: Kaggle
2. Collection Period: 2010

3. Collection Method: The dataset was obtained from Kaggle, a popular data science platform, where it was made available for public use. If there were any specific data sources or preprocessing steps, they are documented on the Kaggle dataset page.
4. Target Variable: Accident_Severity

The primary purpose of this dataset is to support the development and evaluation of deep learning algorithms for the prediction of accident severity in India. It can be used for research, analysis, and modeling to improve road safety and enhance emergency response systems. Researchers, data scientists, and policymakers may utilize this dataset to gain insights into accident patterns, contributing factors, and severity outcomes. The dataset is intended for educational purposes, model development, and fostering discussions related to road safety and accident prevention strategies.

4.2 Software Requirements

1. **Keras -** Keras is designed with a focus on simplicity and ease of use, which makes it an excellent choice for both beginners and experienced deep learning practitioners. It offers a modular and user-friendly API for defining and training neural networks. Users can quickly build models by stacking layers, and it supports various types of layers like convolutional, recurrent, and densely connected layers. Keras provides a wide range of pre-processing and data augmentation tools for working with image and text data. It is closely integrated with TensorFlow and can run on both CPU and GPU, allowing for efficient neural network training and deployment.

2. **scikit-learn -** Scikit-learn is built on NumPy and SciPy, and it is designed for machine learning, data mining, and data analysis. It is open-source and easy to install via Python's package manager, pip. The library includes a variety of machine learning algorithms for classification, regression, clustering, and dimensionality reduction. It also provides tools for model selection, hyperparameter tuning, and evaluation. Scikit-learn supports various feature extraction and selection methods, making it valuable for data preprocessing tasks. It has a consistent and user-friendly API, which simplifies the process of building and evaluating machine learning models.

3. **OSRM -** OSRM is particularly useful in geographic information systems (GIS) and location-based services. It is open-source and can be customized to fit specific routing needs. OSRM can compute routes for different transportation modes, including car, bicycle, and pedestrian, considering factors like road types, speed limits, and traffic conditions. It provides detailed route instructions, estimated travel times, and distance information, making it valuable for applications like GPS navigation and logistics.

4. **Numpy -** NumPy is the foundation of many scientific and data-related Python libraries. It provides powerful array objects that are efficient for numerical computations. NumPy arrays are homogeneous and support vectorized operations, which make it possible to perform element-wise

calculations on large datasets without the need for explicit loops. It includes mathematical functions for operations like linear algebra, statistics, and Fourier analysis, which are essential for scientific computing.

5. **Pandas -** Pandas is built on top of NumPy and provides high-level data structures like DataFrames and Series, which are particularly well-suited for working with structured data. DataFrames allow for easy indexing, slicing, and reshaping of data, making data manipulation and analysis more straightforward. Pandas supports data reading and writing in various formats, including CSV, Excel, SQL databases, and more. It offers powerful tools for data cleaning, transformation, aggregation, and visualization, which are crucial in the field of data analysis.

6. **Flask -** Flask is a lightweight WSGI web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around Werkzeug and Jinja, and has become one of the most popular Python web application frameworks. Flask offers suggestions, but doesn't enforce any dependencies or project layout. It is up to the developer to choose the tools and libraries they want to use. There are many extensions provided by the community that make adding new functionality easy.

4.3 Performance Evaluation Parameters

Evaluating the performance of the accident prediction model is done using accuracy which provides a general overview of correct predictions. Validation loss and training loss are also the key metrics used in the training process of deep learning models. Training loss refers to the error or discrepancy between the predicted values of the model and the actual values in the training dataset. During the training phase, the model iteratively adjusts its parameters to minimize this training loss. Validation loss, on the other hand, measures the error between the model's predictions and the actual values in a separate dataset called the validation dataset. This dataset is not used in the training process but is instead held out for evaluation purposes. The validation dataset helps assess the generalization ability of the model. F1 Score is a harmonic mean between recall and precision. Its range is [0,1]. This metric usually tells us how precise (It correctly classifies how many instances) and robust (does not miss any significant number of instances) our classifier is. Precision is a measure of a model's performance that tells you how many of the positive predictions made by the model are actually correct. It is calculated as the number of true positive predictions divided by the number of true positive and false positive predictions. Lower recall and higher precision give you great accuracy but then it misses a large number of instances. The more the F1 score the better will be performance. Confusion Matrix creates a $N \ X \ N$ matrix, where N is the number of classes or categories that are to be predicted.

Chapter 5

Result and Discussion

5.1 Implementation Results

```
940/940 [==============================] - 4s 4ms/step
  Modelname   Accuracy  Precision  F1 score    Recall
0       CNN  78.433591   0.779811  0.776198  0.784336
```

Fig 5.1 Evaluation parameters of CNN model

```
940/940 [==============================] - 5s 5ms/step
  Modelname   Accuracy  Precision  F1 score    Recall
0       RNN  71.885813   0.706235   0.70667  0.718858
```

Fig 5.2 Evaluation parameters of RNN model

```
940/940 [==============================] - 16s 16ms/step
  Modelname   Accuracy  Precision  F1 score    Recall
0      LSTM  71.107266   0.700183  0.691469  0.711073
```

Fig 5.3 Evaluation parameters of LSTM model

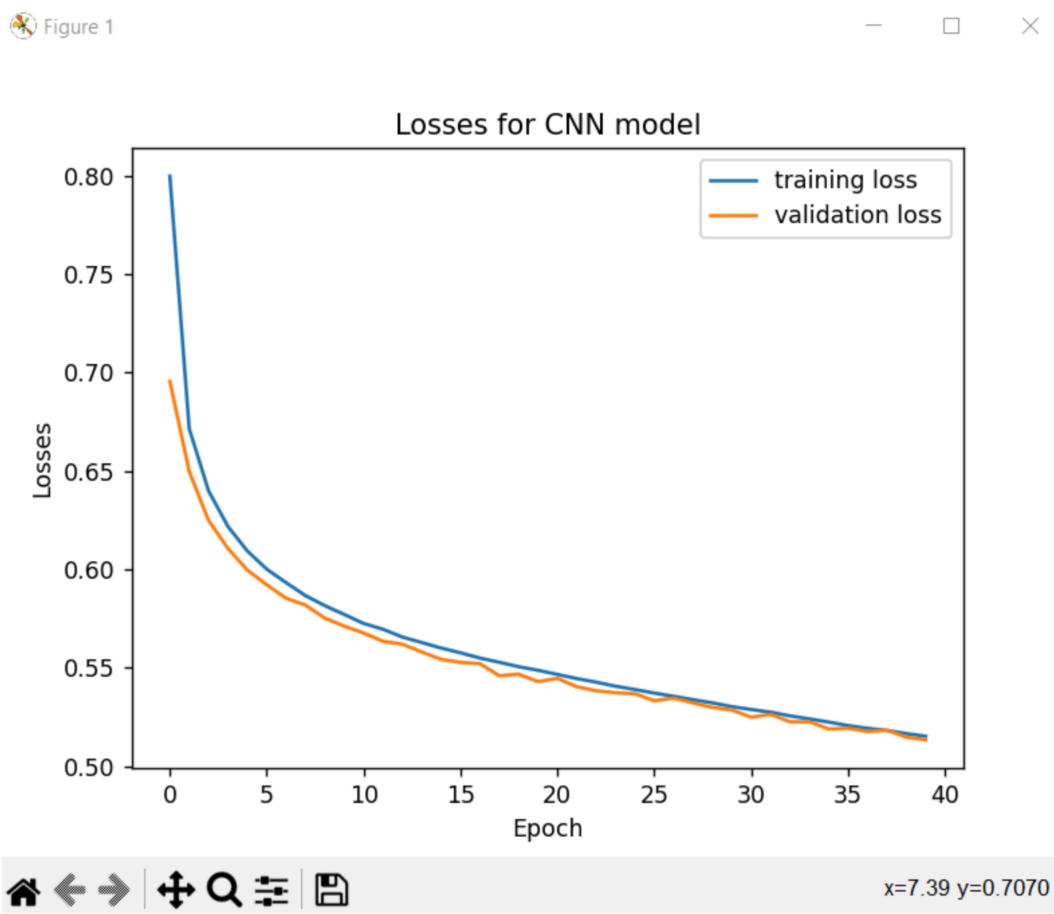

Fig 5.4 Loss Function of CNN model

The above figure shows the loss function of CNN Model. The validation loss is 50%.

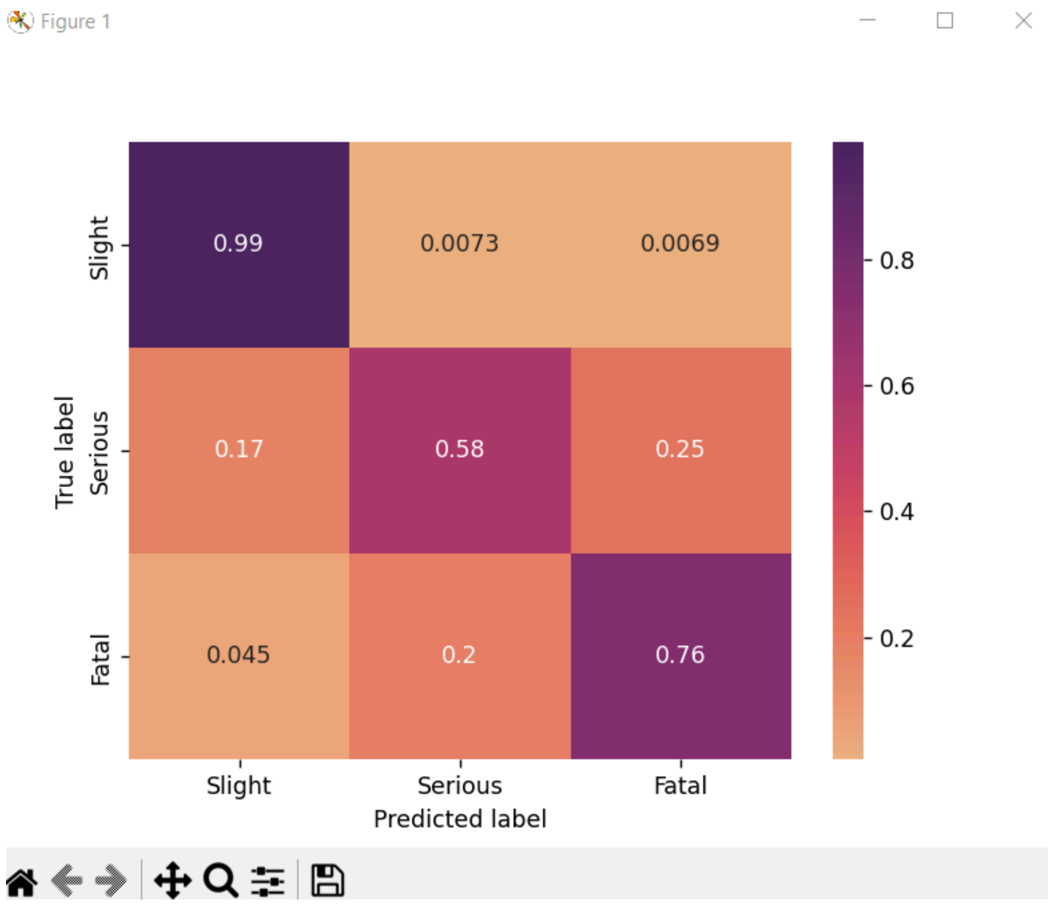

Fig 5.5 Confusion Matrix of CNN model

The above figure shows the confusion matrix for CNN model. The confusion matrix shows the correctly predicted outputs and the incorrectly predicted outputs.

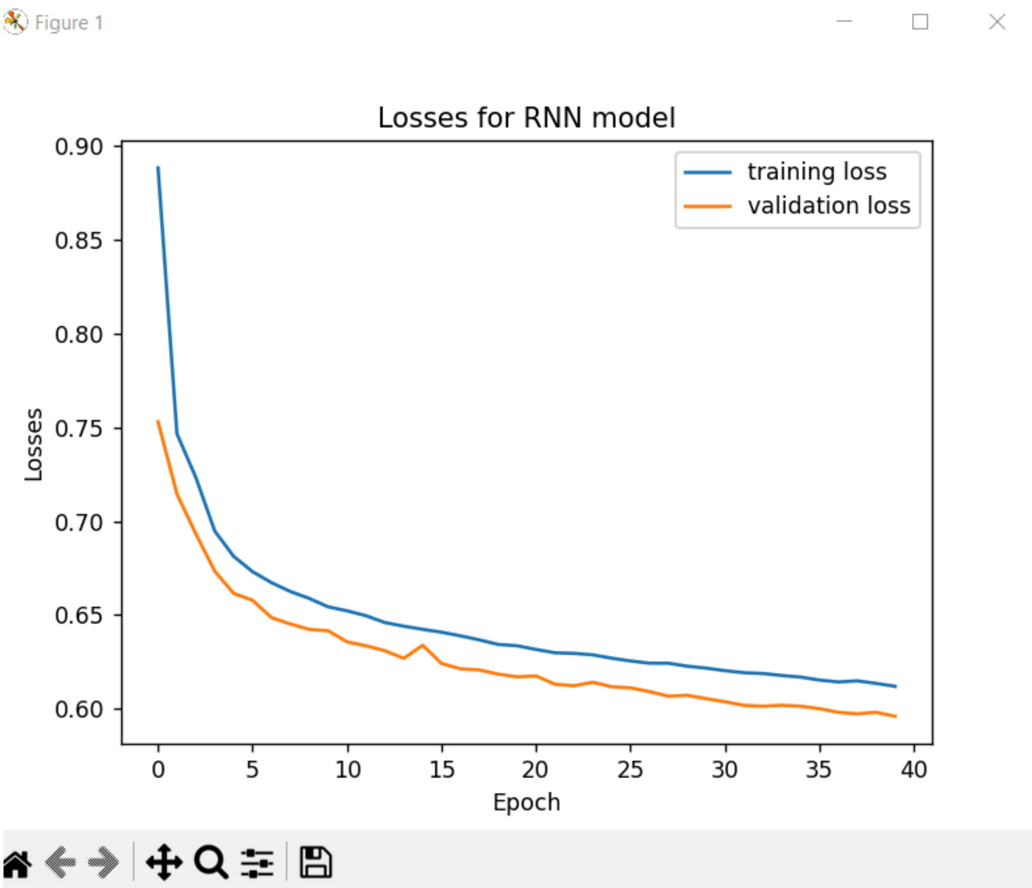

Fig 5.6 Loss function of RNN model

The above figure shows the loss function of RNN Model. The validation loss is 60%.

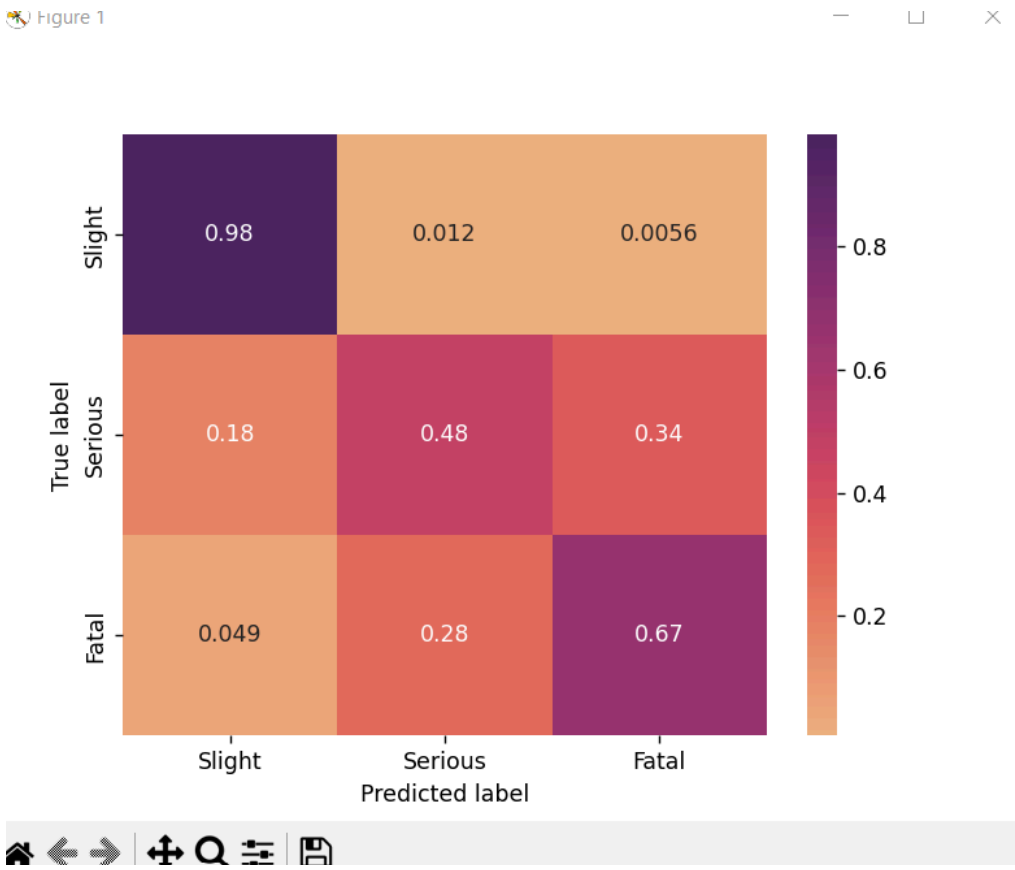

Fig 5.7 Confusion Matrix of RNN model

The above figure shows the confusion matrix for the RNN model. The confusion matrix shows the correctly predicted outputs and the incorrectly predicted outputs.

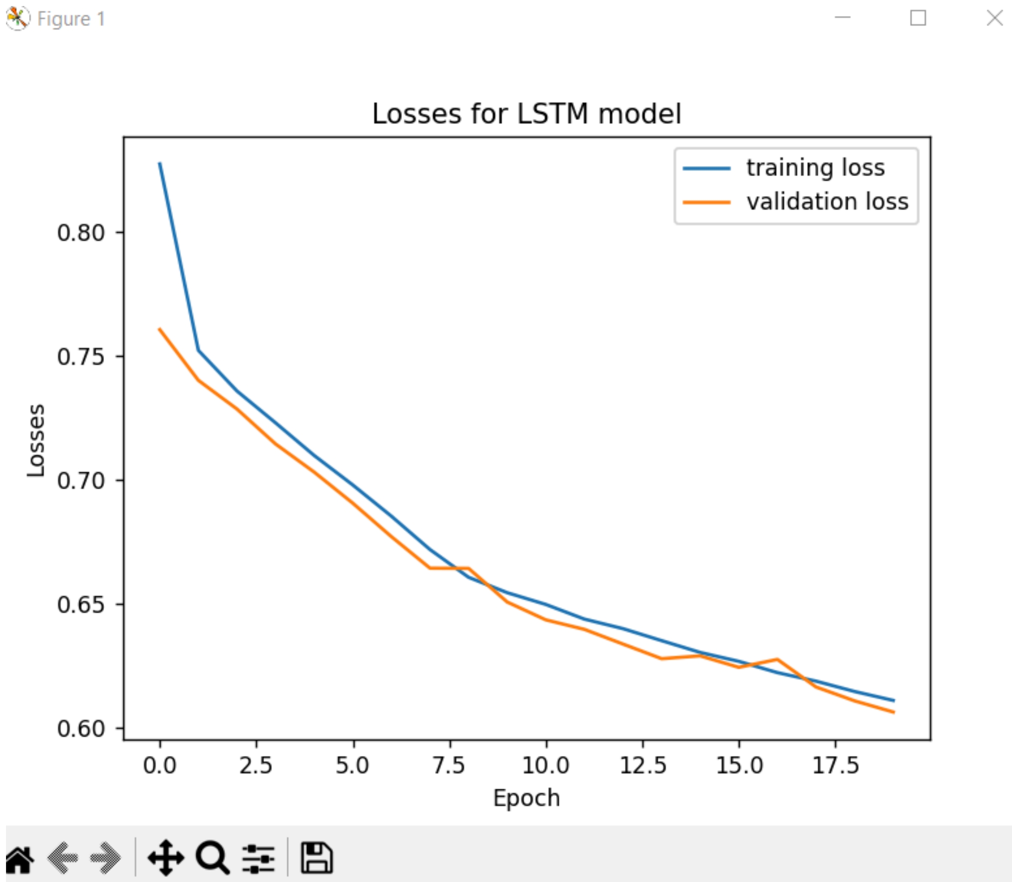

Fig 5.8 Loss function of LSTM model

The above figure shows the loss function of the LSTM Model. The validation loss is above 60%.

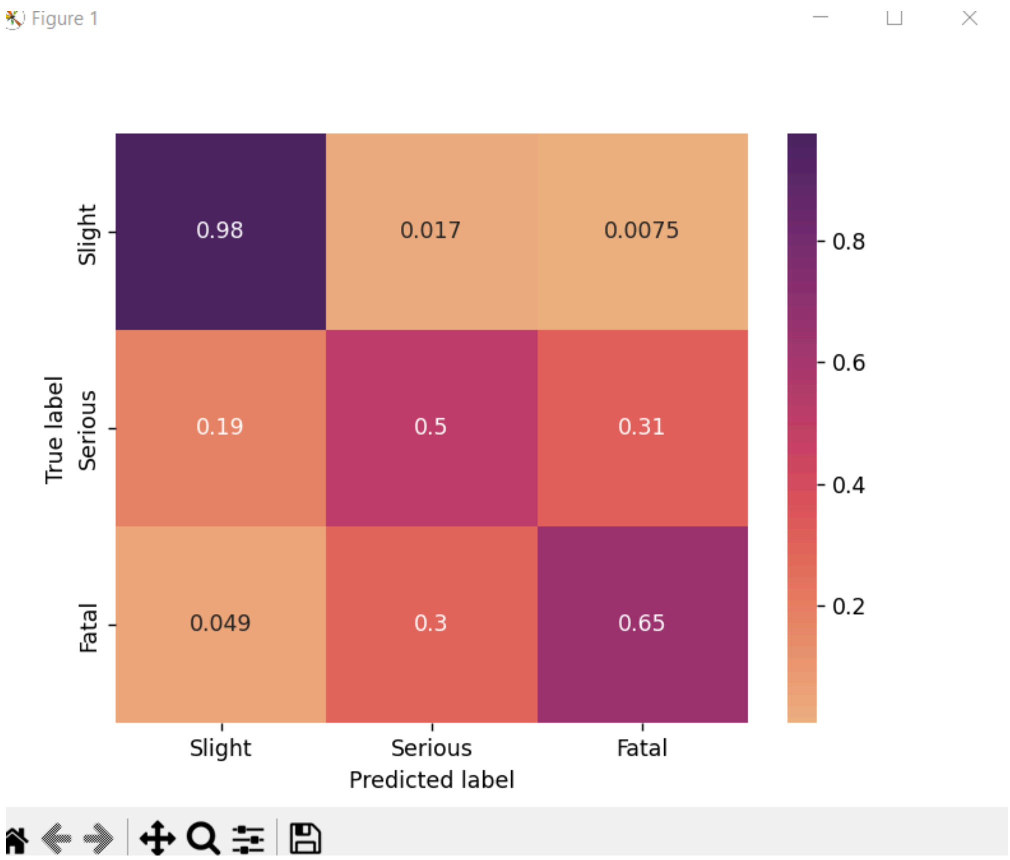

Fig 5.9 Confusion Matrix of LSTM model

The above figure shows the confusion matrix for the LSTM model. The confusion matrix shows the correctly predicted outputs and the incorrectly predicted outputs.

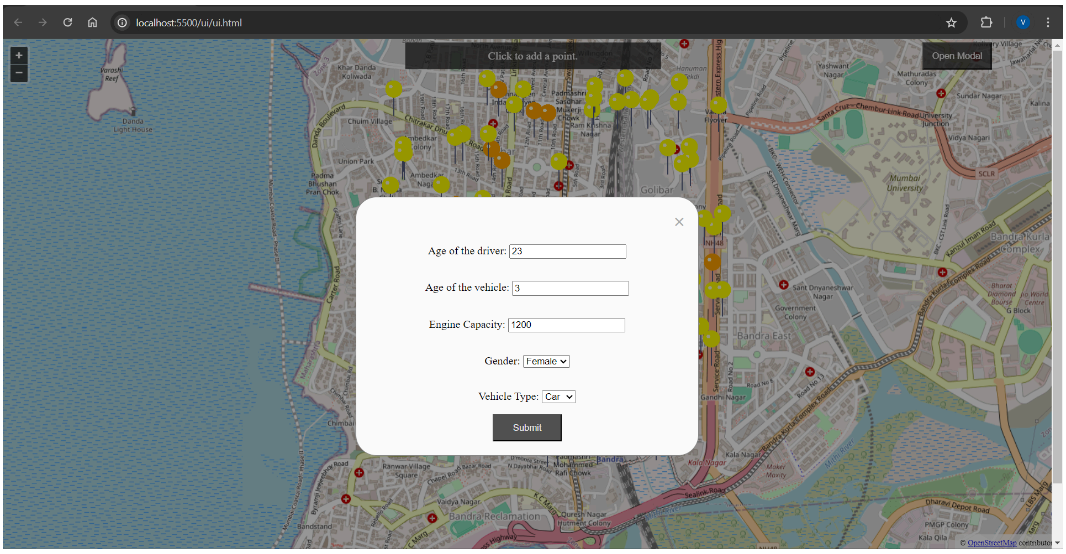

Fig 5.10 Input form

The above figure shows the very first page of our UI. The user is supposed to give an input regarding the Age of the driver, Age of vehicle, Engine Capacity, gender and vehicle.

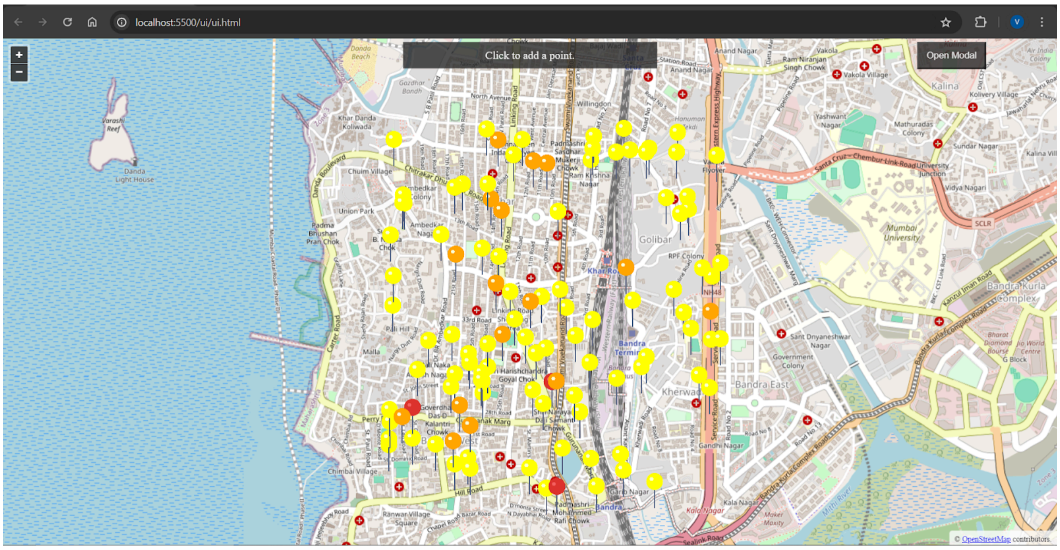

Fig 5.11 The plotted points on the map

The above figure shows the plotted data points from the dataset depicting all the three cases of incidents. Yellow indicates slight, Orange indicates serious and Red indicates fatal. This picture is relevant for Bandra(W).

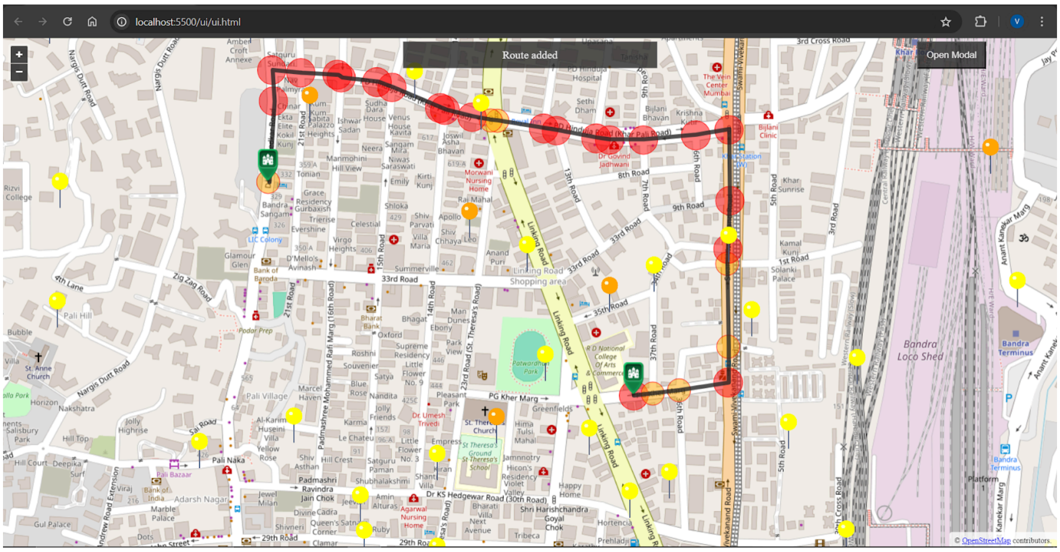

Fig 5.12 The final results

The above figure depicts our actual prediction between the two selected paths.

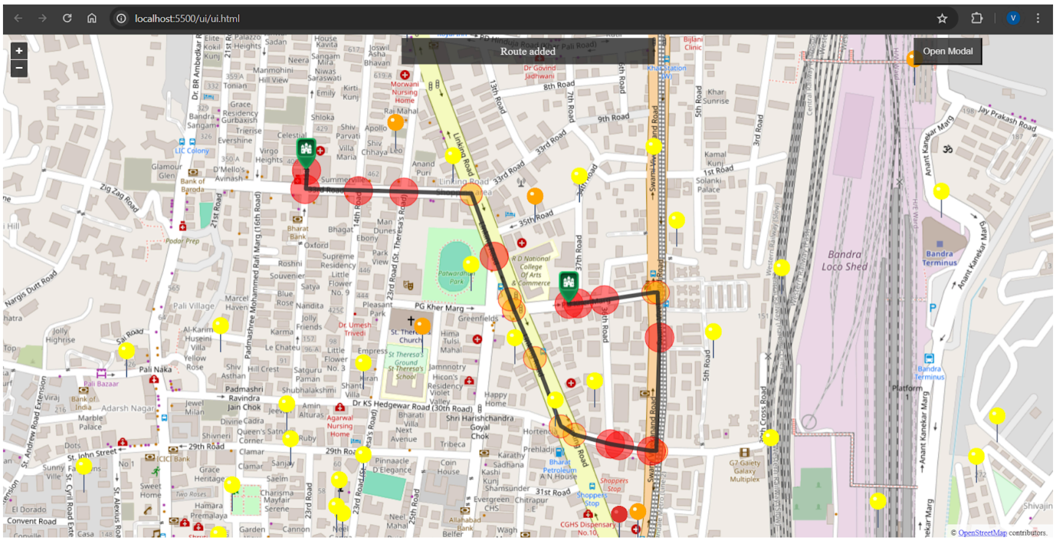

Fig 5.13 The final results

The above figure depicts our actual prediction between the two selected paths.

Chapter 6
Conclusion and Future Work

6.1 Conclusion

The proposed system represents a significant step towards addressing the critical issue of road safety in India. The persistently high rate of road accidents poses a formidable challenge, and our dedication to predicting accident severity is a testament to our commitment to mitigating these risks. Drawing upon datasets from various Indian states and harnessing the power of advanced data analysis and deep neural networks, we have strived to offer a practical solution to enhance public safety. The multi-state approach adopted in the proposed system is a key strength, as it allows us to gain a comprehensive understanding of road safety issues in various regions across India. This adaptability is critical, as road safety challenges can vary significantly from one state to another. By merging and preprocessing data from these regions, we have ensured that our predictive model is rooted in reliable and high-quality data, serving as a solid foundation for our research. Our methodology, incorporating Convolutional Neural Network (CNN) for accident severity categorization, is innovative and promising. The CNN's pattern recognition capabilities, enables us to make precise predictions and categorize accident severity into slight, serious, and severe classes. By doing so, we provide stakeholders with actionable insights that can inform decisions aimed at preventing accidents and minimizing their consequences. One of the hallmarks of the proposed system is the visual representation of predictions using pie charts. This feature enhances the accessibility of our findings and simplifies data interpretation for a broader audience, including authorities, policymakers, and the general public. We believe that clear and intuitive communication of accident severity distribution is crucial for effective decision-making and the implementation of proactive safety measures. In the larger context, our research endeavors to contribute to a safer and more secure road environment for the people of India. It reflects the collective responsibility of society to reduce accidents and protect lives. While road safety is a complex and ongoing challenge, the proposed system is a significant step towards addressing this issue, and we hope that our findings and methodologies will be utilized to enhance road safety initiatives across the diverse landscape of India.

6.2 Future Work

Future work on the project could explore the integration of advanced route planning algorithms with predictive models. By incorporating deep learning techniques, such as convolutional neural networks and recurrent neural networks, into route planning systems, it becomes possible to not only predict the likelihood of accidents along different paths but also to dynamically adjust routes in real-time to minimize the risk of accidents. This could involve developing algorithms that analyze historical accident data, real-time traffic conditions, weather patterns, and other relevant factors to identify the safest path from source to destination. Additionally, the system could visualize these paths and highlight the route with the lowest probability of accident occurrence, providing users with valuable information to make safer travel decisions. This approach not only enhances road safety but also contributes to more efficient and reliable transportation systems.

References

[1] Zhao, H., Zhang, J., Li, X., Wang, Q., & Zhu, H. (2020). "Deep Learning-based Prediction of Traffic Accident Risk in Vehicular Networks" IEEE.

[2] Dhanya Viswanath ,Preethi K, Nandini R, Bhuvaneshwari R ,"A Road Accident Prediction Model Using Data Mining Techniques" in *2021*,pp. 02-05

[3] Gokul Rajesh, Amitha Rossy Benny, Harikrishnan A, James Jacob Abraham and Nithin Prince John, "A Deep Learning based Accident Detection System," in *2020 International Conference on Communication and Signal Processing, July 28 - 30, 2020, India*

[4] Amani Thaduri, Vijayakumar Polepally, Swathi Vodithala, "TRAFFIC ACCIDENT PREDICTION BASED ON CNN MODEL"; 2021,pp.2-3

[5] Maher Ibrahim Sameen, Biswajeet Pradhan,"Severity Prediction of Traffic Accidents with Recurrent Neural Networks ",2017

[6] Vipul Gaurav, Sanyam Kumar Singh, Avikant Srivastava,"Accident Detection, Severity Prediction, Identification of Accident Prone Areas in India and Feasibility Study using Improved Image Segmentation, Machine Learning and Sensors" in *2019 International Journal of Engineering Research & Technology* , 2019

[7] Karishma Pawar, Vahida Attar," Deep learning based detection and localization of road accidents from traffic surveillance videos",in *2021*