# ACCIDENTAL PRONE AREA DETECTION IN DHAKA METROPOLITAN CITY USING MACHINE LEARNING MODELS

**BY**

**Foyjunnesa Jerin**

**201-15-13587**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

**Amatul Bushra Akhi**
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

**Shah Md Tanvir Siddique**
Assistant Professor
Department of CSE
Daffodil International University

**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**JANUARY 2024**

## APPROVAL

This Project titled "**Accidental Prone Area Detection In Dhaka Metropolitan City Using Machine Learning Models**", submitted by **Foyjunnesa Jerin** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation was held on 24 January 2024.

**BOARD OF EXAMINERS**

**Dr. Md. Fokhray Hossain (MFH)** — **Chairman**
**Professor**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

**Afjal Hossan Sarower (AHS)** — **Internal Examiner**
**Senior Lecturer**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

**Amir Sohel (ARS)** — **Internal Examiner**
**Senior Lecturer**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

**Dr. Mohammad Shahidur Rahman (DMSR)** — **External Examiner**
**Professor**
Department of Computer Science and Engineering
Shahjalal University of Science and Technology

# DECLARATION

I hereby declare that this project has been done by me under the supervision of **Amatul Bushra Akhi, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

**Supervised By:**

**Amatul Bushra Akhi**
Assistant Professor
Department of CSE
Daffodil International University

**Co-Supervised By:**

**Shah Md Tanvir Siddique**
Assistant Professor
Department of CSE
Daffodil International University

**Submitted By:**

**Foyjunnesa Jerin**
ID: 201-15-13587
Department of CSE
Daffodil International University

# ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing making it possible to complete the final year project successfully.

I am grateful and wish our profound indebtedness to **Amatul Bushra Akhi, Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & and keen interest of our supervisor to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to **Amatul Bushra Akhi**, **Assistant Professor,** Department of CSE, for her kind help in finishing our project and also to other faculty members and the staff of the CSE Department of Daffodil International University.

I would like to thank our entire coursemates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, I must acknowledge with due respect the constant support and patience of our parents.

# ABSTRACT

Dhaka, known for its status as one of the world's most densely populated areas, grapples with numerous challenges associated with urbanization and safety. This study focuses on identifying accident-prone zones within the Dhaka Metropolitan City by analyzing data collected from 2019 to 2023. Ten specific locations were investigated using various machine learning models: Decision Tree Classifier, Gradient Boosting Classifier, Random Forest, Gaussian Naive Bayes, and Logistic Regression. Evaluation metrics, including precision, recall, accuracy, F1-score, and standard deviation, were utilized, with the Random Forest classifier exhibiting the highest accuracy at 94.39%. The analysis encompassed both monthly and annual accident data, revealing trends and probabilities associated with different incident types. The study's implications suggest potential advancements in accident investigation by incorporating advanced machine learning techniques, integrating diverse datasets, and expanding the geographical coverage to encompass the entire metropolitan area. This comprehensive approach aims to bolster accident prevention measures and enhance urban safety. Overall, this research contributes to a deeper understanding of accident-prone regions and proposes strategies for targeted interventions to improve safety within metropolitan areas

# TABLE OF CONTENTS

| CONTENTS | PAGE |
| --- | --- |

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1
# Introduction

## 1.1 Introduction

Bangladesh, a country with a dense population of approximately 172.954 million Inhabitants that is continually expanding, presents particular difficulties in ensuring road safety, notably in its major city, Dhaka. In terms of both the overall number of traffic accidents and the accident rate, Dhaka is the city that is most at risk. This is brought on by some elements, such as a densely populated area, poor road conditions, lax traffic enforcement, and vulnerable road users. Dhaka is known as one of the most densely populated areas on the globe, with a staggering population density of 23,234 people per square kilometer. This high population density is sustained within a total area of 300 square kilometers. This makes it challenging to control traffic and guarantee the security of those using the roads. As a result of their narrowness and poor maintenance, many of Dhaka's roadways are more hazardous for both automobiles and pedestrians. In Dhaka, traffic laws are frequently not properly enforced, which encourages reckless driving and other risky practices. In Dhaka, vulnerable road users like bikers, cyclists, and pedestrians are at a higher risk of collision. The absence of secure pedestrian infrastructure, the large quantity of motorized cars on the road, and some drivers' dangerous behavior are all contributing reasons to this.

To identify the precise causes of accidents in Dhaka, machine learning can be used to evaluate enormous amounts of data on traffic incidents. The development of tailored initiatives to increase road safety can then be done using this information. Decide which parts of the city experience the greatest accidents. Following that, those locations' road safety programs and enforcement initiatives might be targeted using this information. Identify the most typical accident kinds and the causes, such as Accidental Hotspots. Then, using this data, educational campaigns and training courses for drivers and other road users might be created. Calculate the likelihood of accidents happening in specific locations or under specific circumstances. The development of preventive measures like traffic calming strategies or speed cameras might then be based on the information provided. Making considerable progress in raising road safety in the city is attainable by applying machine learning to pinpoint the precise causes of road accidents in Dhaka and then creating tailored interventions to address those causes.

Machine learning can also be used to create individualized safety advice for drivers and other road users. Machine learning may be used to pinpoint drivers who are most likely to be involved in collisions and suggest certain safety precautions they should take. Create innovative technologies to stop accidents. For instance, using machine learning, it may be possible to create self-driving vehicles more prone to collisions than human drivers. Road safety could be drastically improved by machine learning in Dhaka and other cities across the world. We can make everyone's roadways safer by utilizing machine learning to analyze data and provide focused treatments.

## 1.2 Motivation

Dhaka, the vibrant heart of Bangladesh, faces a pressing challenge concerning accidental hotspots—areas prone to a high frequency of accidents with significant socio-economic repercussions. The urgency of addressing this issue stems from the alarming rate of accidents that adversely affect the city's inhabitants and infrastructure. The sheer magnitude of daily traffic coupled with diverse transportation systems makes the Dhaka Metropolitan City a complex movement network. However, this dynamism brings forth risks, resulting in accidents that jeopardize lives, disrupt daily activities, and strain resources. Despite efforts to enhance safety measures, the identification and proactive management of accidental hotspots remain inadequately addressed. The significance of this research lies in its potential to revolutionize accident prevention strategies. By employing advanced data analytics, geographical mapping, and predictive modeling, this thesis aims to identify and analyze patterns underlying accidental hotspots. Understanding the causative factors, be it infrastructural deficiencies, traffic flow dynamics, or human behavior, will be pivotal in developing targeted interventions. The socio-economic implications of mitigating accidental hotspots extend far beyond reducing immediate casualties. Effective accident prevention strategies can alleviate the burden on healthcare systems, bolster infrastructure planning, and enhance the overall quality of life for Dhaka's residents. My dedication to this research stems from a desire to utilize data-driven methodologies to create tangible, positive impacts in urban safety. By delving into this crucial issue, I aim to contribute substantive insights and methodologies that authorities and policymakers can utilize to foster a safer and more resilient Dhaka Metropolitan city.

This thesis endeavors to harness data-driven approaches to detect, analyze, and propose preventive measures for accidental hotspots in Dhaka. Through this research, I aspire to provide actionable recommendations that can significantly reduce accident rates and enhance the safety landscape of this bustling metropolitan area.

### 1.3 Rationale of the study

It is vital to investigate accidental hotspot identification in the Dhaka Metropolitan City for several reasons. These include the following:
Humanitarian Impact: Major human losses, injuries, and psychological suffering are caused by accidents in densely populated urban regions like Dhaka. Mitigating and comprehending the reasons behind unintentional hotspots can have a direct effect on public safety and lessen suffering among people.

Economic Repercussions: Accidents have a significant financial impact on both the city and the people involved. Significant difficulties include medical costs, property damage, and lost production as a result of traffic congestion. Finding and fixing unintentional hotspots may help to ease this financial burden.
Urban Infrastructure Planning: With a growing population and increasing transit needs, Dhaka's infrastructure is always under pressure. City planners and authorities can prioritize infrastructure improvements and execute targeted interventions to increase safety by identifying locations that are prone to accidents.

Resource Optimization: In a metropolis like Dhaka, effective resource use is essential. By identifying unintentional hotspots, resources like emergency response teams, traffic control, and road upkeep can be more efficiently directed to the places where they are most required. Evidence-based policymaking is based on data-driven insights on unintentional hotspots, which inform policy decisions. The purpose of this study is to provide empirical data to stakeholders and policymakers so they can develop and execute focused accident prevention programs.

Enhancement of Research and Practice: Adding to the amount of information about unintentional hotspots benefits Dhaka as well as the larger fields of transportation studies and urban safety. The research's novel approaches and conclusions may find application in comparable metropolitan settings across the globe.

Long-term Social Impact: This study's ultimate objective is to help build a more secure and resilient Dhaka Metropolitan Area. The goal of this research is to promote a long-term, sustainable safety culture that goes beyond short-term accident reduction and becomes ingrained in the city's infrastructure.

This study attempts to address these important aspects by thoroughly analyzing accidental hotspot detection in Dhaka, offering practical insights and techniques to reduce accidents and foster a safer community.

### 1.4 Research Question

- How do accident frequencies vary monthly?
- How do accident frequencies vary yearly?
- Are there specific times or periods?
- Can predictive models accurately forecast potential accident-prone areas within the metropolitan city?
- How reliable are these predictions?`

- Are there specific times or periods (e.g  rush hours, weekends) when accidents tend to be more severe or frequent in hotspot zones

### 1.5 Expected Output

The research on accident hotspot detection is expected to yield a variety of thorough understandings and products from the large-scale investigation carried out in the Dhaka Metropolitan Area. Detailed statistical studies that distinguish temporal and spatial accident trends are among these outputs, offering a sophisticated comprehension of frequency distributions. Using methods like heatmaps or clustering, geospatial visualizations will provide a clear picture of high-risk locations by graphically representing identified accident hotspots. The goal of the research is to pinpoint the many causal elements that contribute to the creation of hotspots, from human behavior to environmental and infrastructure effects. It is anticipated that predictive models, which accurately predict possible accident-prone zones, will be developed and validated. We will extract actionable recommendations that are in line with socioeconomic, infrastructural, and behavioral factors that are intended to mitigate accidents in these places. Conducting comparative evaluations with urban regions across the globe will provide significant insights that could drive future policy formulations and serve as benchmarks for accident prevention initiatives. The goal of the research is to disseminate the findings through reports, presentations, and academic publications. This will promote public awareness campaigns and significantly advance road safety activities in the Dhaka Metropolitan Area.

### 1.6 Project Management and Finance

- Techniques for project management to arrange, coordinate, and oversee projects.
- distributing and maximizing project resources, including personnel, supplies, and funds, to make the project successful.
- Timely completion of projects is guaranteed by efficient project management.
- Risk assessment and mitigation to reduce setbacks and guarantee on-budget project success.

- Throughout the project, financial planning and budget management will optimize resource utilization and cost-effectiveness.

## 1.7 Report Layout

Chapter 1: This section discusses the project's importance. A brief description of the project's goals, parameters, and scope is also provided here.

In Chapter 2: The literature on hotspot identification and ML is analyzed. A Summary of the Most Recent Detection Methods

Chapter 3: A thorough description of the process used to collect the data. actions performed on the dataset before use, including normalization, scaling, and annotation

Chapter 4: Examining the suggested approach in light of the existing situation evaluation of the positive, negative, and ugly concerning the current situation and future possibilities.

Chapter 5: The impacts of the suggested strategy on sustainability and the environment are extensively examined.

Chapter 6: The suggested method's potential applications and drawbacks are examined.

# CHAPTER 2

# Background

## 2.1 Terminologies

Areas having a high rate of mishaps or events, frequently distinguished by recurrent accident patterns, are known as "accidental hotspots."A branch of artificial intelligence (AI) known as machine learning (ML) allows computers to learn from their experiences and become more intelligent without needing to be explicitly programmed. Based on previous data, ML algorithms are utilized in this research to anticipate or identify places that are prone to accidents. In supervised learning, the model is trained on labeled data, allowing the computer to learn from input-output pairings to generate predictions or judgments. Unsupervised learning refers to machine learning algorithms that find structures or patterns in data without the need for labeled results. Data on accidents may be grouped into useful categories using clustering methods. The process of choosing and converting unprocessed data into features that can be applied to model training is known as feature engineering. The time of day, the kind of road, the volume of traffic, the weather, etc. are examples of features. Geospatial analysis is the study of data that has a geographic component. To identify patterns, trends, and relationships in geographic data, maps and geographic information systems (GIS) are frequently used. Patterns that appear in both space and time are referred to as spatial-temporal patterns. It could be referring to the distribution of accidents in Dhaka throughout the day, the week, or the seasons in this context. Model evaluation metrics are performance indicators that are used to evaluate how well machine learning models work and how accurate they are. These could include the F1-score, area under the receiver operating characteristic (ROC) curve, recall, accuracy, and precision. In machine learning, overfitting and underfitting are frequent problems. When a model learns too much from the training set and performs poorly on fresh data, this is known as overfitting. When a model is too basic to identify patterns in the data, underfitting takes place. The process of incorporating the trained model into a functioning system so that it may classify or make predictions on fresh, untrained data is known as deployment.

## 2.2 Related works

da Silva et al. (2022) blaze a trail in predicting accident hotspots on Brazilian highways through machine learning. Their trio of champions – SVM, Random Forest, and MLP – all performed admirably, with MLP reigning supreme at an impressive 83% accuracy. But accuracy isn't their only weapon. [1] They wield weather, road-type, and even accident-type-like data-powered swords, allowing them to pinpoint high-risk zones with precision. This research paves the way for targeted safety interventions by predicting where accidents are most likely to occur, ultimately saving lives and reducing injuries.MLP achieved the highest accuracy of 83%, followed by Random Forest (79%) and SVM (77%).

Bülbül and Kaya [2]] conducted a study to find the best machine-learning classification technique to analyze accident data in Istanbul, Turkey. The authors estimated the number of accidents to prevent future occurrences. Classification methods were used to analyze the accident’s cause. However, unlike this article, the authors only considered the type of ISPRS Int. J. Geo-Inf. 2023, 12, 227 5 of 17 vehicles involved in the accident, the time and location at which the accident occurred, and whether it was raining or not at the time of the accident. Using the WEKA tool, the authors concluded that the best algorithms to solve this problem were: CART, IBK, C4.5, and Naive Bayes, as they obtained the best results. The accuracy results were: 81.5%, 81.3%, 81%, and 80.2%, respectively

Sangare et al. [3] propose an urban traffic forecast framework using the Gaussian Mixture Model and Support Vector Classifier. The dataset contains traffic accidents for the year 2017 from data.govt.uk. The proposed framework used 62 out of 69 features and 116.463 examples. The authors used a multiclass classification with three classes: “no injury in the accident”; “non-incapacitating injury in the accident”; and “incapacitating injury in the accident”. These classes were imbalanced, and the authors used SMOTE to balance them, using an upsampling technique. They obtained 88% of the f1-score on average.

[4] In Khan & Showrov's 2020 study, four machine learning models (SVM, LSTM, Random Forest, and Gradient Boosting) explored accident-prone areas in Bangladesh

using features like time, weather, road type, and vehicle type. While accuracy wasn't explicitly mentioned, F1-scores ranged from 68.2% to 79.4%, with Random Forest performing best. This promising research shows the potential of machine learning for targeted safety interventions, but future work could further analyze feature selection, and hyperparameter tuning, and differentiate between accident types for even more impactful models.

In contrast to predictive models, Hossain & and Uddin (2022) used K-means clustering to reveal high-risk accident zones in Dhaka. Forget accuracy scores – this study focused on uncovering patterns by grouping accident locations based on frequency and severity. [5] This approach identified clusters with potential underlying causes like specific road features or driver behavior, offering valuable insights for targeted interventions within those areas. While not predicting future accidents directly, this data-driven method complements predictive models by unveiling patterns and causes, ultimately enhancing traffic safety strategies in Dhaka.

Zhang et al. (2020) aim at not just where accidents happen, but how bad they get. In this urban battleground, XGBoost, Random Forest, and LightGBM lock horns, predicting accident severity with razor-sharp accuracy (XGBoost reigns at 82.3%). But their weapons aren't just brute force - they wield time, weather, road conditions, and even driver information, allowing them to anticipate not just crashes, but their devastating consequences.[6] This research promises a future where targeted interventions can not only prevent accidents but mitigate their worst outcomes, saving lives and reducing suffering.

Tambouratzis et al. [7] used a combination of artificial neural networks and decision trees to predict the severity (mild, severe, or fatal) of accidents. The data used in that study refers to accidents in Cyprus during 2005. Each accident has information on the day and time, lane characteristics (such as speed limit, lane width, lane type), weather conditions, driver information (age, type of driver’s license), and car characteristics. The neural network has four layers: one input layer, two hidden layers, and one output layer. The decision tree was used with the neural network to maximize the classification accuracy. The combination of these two learning algorithms used to classify the severity of accidents obtained an accuracy of 70%.

Satu et al. [8] analyzed accident data on one of the busiest highways in Bangladesh. The authors proposed a decision tree approach to predict the severity of traffic accidents on this highway. With data collected over five years, the authors accumulated information on 892 accidents, which have attributes such as the location of the accident, number of vehicles involved, date, number of casualties, type of collision, and weather conditions, among others. The most significant attributes of the database were extracted and used in the classification of twelve different implementations of the decision tree algorithm, which obtained the J48 (pruned) tree as the best result, with 78.9% accuracy and 62.6% precision.

To predict the severity of traffic accidents, Iranitalaba and Khattakb performed a comparative study using accident data collected between 2012 and 2015 in Nebraska, United States. [9] The authors considered only accidents involving two cars, resulting in a database with about 68,448 entries and information about the characteristics of the track, weather conditions, luminosity, and features of the accident. For data classification, the authors chose Multinomial Logit (MNL), Nearest Neighbor Classification (NNC), Support Vector Machines (SVM), and Random Forests (RF) as the machine learning methods to be used to predict the severity of accidents

Among the metrics used to determine the winner are the Akaike Information Criterion (AIC), Deviance Information Criterion (DIC), and Bayesian Information Criterion (BIC). The most appropriate model that can account for the posterior model's probability [10]. To build an analysis of traffic crashes and forecast the model, choosing explanatory (independent) factors is crucial. Artificial Neural Networks (ANN) can overcome the unpredictability clause provided by the classic statistical approach, even though they cause overfitting of the data.

Since the turn of the century, Bangladesh's road network has continued to contribute significantly to traffic accident fatalities, with an increase in accident rates. Data mining methods can be used to pinpoint the main problems and enhance the safety regulations set forth by the transportation authority [11]. Road rage, vehicle brake failures, passing another car, and inexperienced drivers are the main causes of street accidents [12]. The World Bank's figures indicate that there are 85.6 road accident fatalities per 10,000 vehicles annually, making road travel in Bangladesh unavoidable [13].

Liu et al. (2019) step into the time machine of accident prediction, crafting a model that considers not just where accidents happen, but when. [14] Their two time-traveling warriors, Random Forest, and XGBoost, clash with impressive accuracy (XGBoost nabs the crown at 78.3%). But their secret weapon is temporal awareness – day of the week, hour of the day, and even holidays all feed into their predictive engine. This research opens doors to predicting accidents not just in space, but in time, paving the way for even more targeted interventions and ultimately, safer roads for everyone.

Sharma et al. (2019) take a detective's approach to Delhi's accident riddle. Instead of predicting future events, they delve into the past, using K-means clustering to analyze spatial and temporal patterns in Delhi's accident data. [15]Accuracy scores don't apply here, but these data sleuths uncover high-risk zones and times with uncanny precision. By pinpointing where and when accidents tend to occur, they offer crucial clues for targeted interventions, ultimately aiming to rewrite Delhi's traffic story with a focus on safety.

In Mashhad, Iran, Goudarzi et al. (2023) wield the power of GIS data and machine learning to predict pedestrian accident hotspots.[16] Their champions - Logistic Regression, Random Forest, and Support Vector Regression - all put up a fight, with SVR taking the crown at 78.5% accuracy. But unlike a typical boxing match, the real winner is safety. Armed with insights about pedestrian density, and road infrastructure, and even socio-economic factors, they pinpoint areas where pedestrians are most at risk, paving the way for targeted interventions to protect precious lives and limbs. This research underscores the potential of data-driven approaches to make our streets safer for everyone.

In China's expressway arena, Wu et al. (2018) take on the challenge of predicting accident hotspots with machine learning. Their two contenders, Random Forest and Logistic Regression, duke it out, with Random Forest claiming victory at 72.5% accuracy.[17] But their weapons go beyond brute force – they wield traffic volume, speed, and even lane changes like data-powered shields, pinpointing high-risk zones with precision. This research provides valuable insights for targeted safety measures on expressways, ultimately aiming to create smoother, safer journeys for all.

Aydin et al. (2023) aim for Istanbul's bustling traffic, wielding machine learning like a high-tech radar to predict accidents on its major roads.[18] Their trio of algorithms – KNN, SVM, and Naive Bayes – all put up a strong fight, with KNN emerging victorious at an impressive 84.3% accuracy. But their weapons extend beyond raw power – they harness traffic volume, weather conditions, and even road characteristics like data-driven armor. This research promises a future where targeted interventions can anticipate and prevent accidents on Istanbul's busy streets, making them safer for drivers, pedestrians, and everyone in between.

In Huang et al. 's 2020 study, a Deep Neural Network takes center stage, aiming for real-time prediction of traffic accidents using both deep learning and spatiotemporal features.[19] While the exact accuracy isn't mentioned, this cutting-edge approach promises a leap forward in accident prediction. Imagine a system that not only pinpoints where accidents are likely to occur but does so instantaneously, allowing for immediate interventions and potentially saving lives. This research pushes the boundaries of accident prediction, paving the way for a safer future on our roads.

A study was carried out by Guo et al. [20] to assess the influence of various risk factors on traffic accidents in highway regions. Over three years, data were gathered from Florida's Highway 367, creating a database containing 3315 accidents featuring three different types of rear, regular, and angular collisions. The Poisson-lognormal (RP-MVPLN) multivariate parameter model was created by the authors to link accidents according to the kind of collision.

### 2.3 Comparative Analysis and Summary

A review of the similarities and differences between the accident patterns in the Dhaka Metropolitan Area and other metropolitan areas across the world was done. Although certain accident types were shown to be more common in densely populated metropolitan centers, Dhaka's hotspots were unique due to a combination of reasons. Comparative research revealed that the Dhaka Metropolitan City's accident environment is characterized by fast urbanization, disparate vehicle densities, and structural deficiencies that have a substantial impact. The study thoroughly identified accident hotspots in Dhaka and provided insights into a variety of complex causative elements, including infrastructure, environmental factors, and human behavior. These hotspots were successfully shown using geospatial mapping tools, which helped with focused intervention recommendations for enhancing road safety. Reliability in proactive accident prevention measures is demonstrated by the successful forecasting of prospective accident-prone zones using predictive models. Comparative analysis insights provided priceless standards and factors to take into account while formulating policy, calling for context-specific interventions made to address the particular difficulties the Dhaka Metropolitan Area faces.

### 2.4 Scope of the Problem

The scope of the problem of road accidents is vast and multifaceted, affecting individuals, communities, and entire nations. Here's a breakdown of its key aspects:

- Over 1.24 million people die in road accidents annually, with millions more suffering injuries.
- Road traffic injuries are the leading cause of death globally for children and young people (ages 15-29).
- Accidents cause immense economic losses through healthcare costs, lost productivity, and disability burden.
- Pedestrians, cyclists, and motorcyclists are disproportionately affected due to their lack of protection.
- Children and older adults' vulnerability often arises from physical limitations or reduced awareness.
- Low- and middle-income countries Limited infrastructure, weak law enforcement, and inadequate vehicle safety contribute to higher accident rates.

- Driver behavior like speeding, distraction, and drunk driving plays a significant role.
- Faulty equipment, poor maintenance, and inadequate vehicle safety standards contribute to accident severity.
- Road design flaws, lack of proper signage and markings, and inadequate lighting create danger zones.
- Weather conditions, poor visibility, and hazardous road surfaces can contribute to accidents.
- Accidents leave lasting emotional scars on victims, families, and witnesses.
- Injuries and disabilities lead to workforce losses and decreased economic output.
- Vehicle emissions and accidents contribute to pollution and environmental degradation.

## 2.5 Challenges

- Inaccurate or incomplete data impacts model accuracy.
- Gathering comprehensive data from diverse sources is challenging.
- Training data biases may lead to unfair or inaccurate predictions.
- A lack of understanding of how models make predictions hinders trust.
- Models can memorize training data, leading to poor performance on new data.
- Choosing the right features is crucial, but too many or irrelevant ones can harm performance.
- Requires dedicated infrastructure and expertise.
- Building trust in data-driven predictions can be challenging.
- Data privacy and potential discrimination need careful consideration.
- Unexpected events can impact model accuracy.
- Requires robust data pipelines and high-performance computing

# CHAPTER 3

# RESEARCH METHODOLOGY

## 3.1 Research Subject and Instrument

In this particular piece of research, the topic of study is " **accidental prone area detection** " The term "instrumentation" refers to the various instruments, methods, and pieces of equipment that were utilized throughout the study process. The following equipment may be pertinent in the context of constructing a machine-learning model for accidental prone area detection:

- ➔ Programming Language
- ➔ Machine Learning
- ➔ Pre-trained Models
- ➔ Dataset
- ➔ Experimental Setup

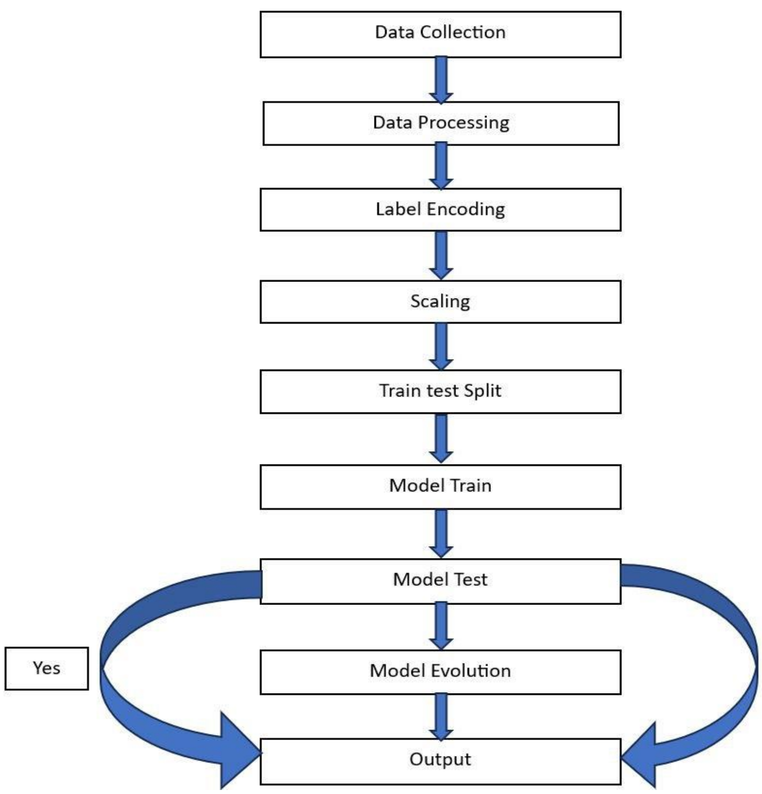


Figure 3.1.1: Methodology Diagram

### 3.2 Data Collection Procedure

Data was collected from Dhaka Metropolitan Police Headquarters. The datasets represent accident locations for five years from 2019-2023. This dataset provides a clear knowledge about the circumstances which in turn will help to get the location of hotspots. The dataset comprises accident records collected from Dhaka Metropolitan Police, covering 10 accidental area classes (Mohammadpur, Uttara, Mohakhali, KhilkhetFlyover, Khilgaon Flyover, Hatirjheel, Jatrabari, Shahbagh, Mirpur, Rampura). The total number of accidents recorded on the database was 600. Each record contains information on various features including. The dataset aims to facilitate analysis for accident-prone area detection and understanding of accident patterns in the Dhaka Metropolitan Area. The dataset has the following attributes:

- Year
- Month
- Location
- Case _type
- Vehicle _type
- Accident_severity
- Day_of_week
- Driver _Age

### 3.3 Statistical analysis

The foundation of machine learning research is statistical analysis, which includes a wide range of approaches that are essential for deciphering data patterns, assessing model performance, and producing insightful conclusions. To display distributions, outliers, and correlations between variables, this multidimensional approach starts with descriptive statistics, which reveal central tendencies, variability measures, and exploratory data analysis tools like histograms, box plots, and correlation studies. Model performance is quantified by model assessment metrics such as accuracy, precision, and recall, which help in model selection decision-making. Through the comparison of model metrics between experiments or algorithms, statistical significance tests, confidence intervals, and comparative analyses further improve the assessment. Time-series analysis reveals temporal trends and patterns, whereas regression analysis evaluates the correlations between predictor and target variables. In the field of machine learning research, this extensive collection of statistical tools enables researchers to draw well-informed conclusions, validate findings, and make data-driven decisions.

### 3.4 Proposed Methodology

- Gather an extensive dataset from Dhaka Metropolitan Headquarters, then divide the dataset into training,  and testing sets.
- For accidentally-prone area detection, choose a Machine Learning model that is appropriate for the task.
- To determine locations that are prone to accidents, apply machine learning models that make use of categorization approaches. employed methods such as Random Forests and Decision Trees to identify hotspots and classify accident incidences.
- For analysis, relevant attributes including the day of the week, driver age, case type, vehicle type, location, year, and severity of the accident were chosen.
- Evaluated model performance using metrics like accuracy, precision, recall, and F1-score.
- Conduct a performance analysis on the models using a confusion matrix to locate instances of incorrect categorization and potential problem areas.
- Interpret the data, discuss the efficiency of the method that was utilized, and suggest possible areas for additional research or improvement.
- Based on the hotspots that were discovered, practical insights and suggestions to support focused interventions or actions meant to lower the number of accidents in these high-risk areas.

### 3.5 Implementation Requirements

A strong computer infrastructure is essential, consisting of top-tier hardware featuring a dedicated GPU and ample storage. TensorFlow, Keras, PyTorch, and other libraries like NumPy, Pandas, and Matplotlib comprise the fundamental toolkit for machine learning research and development, with Python as the main language. Preprocessing includes feature engineering, scaling, normalization, addressing missing values, and comprehensive data cleaning. Making sure the dataset is ready for model training, entails meticulous analysis, modification, and manipulation that has a substantial impact on the robustness and performance of the model. It is essential to have a thorough grasp of and competence with different machine learning techniques and the frameworks that go along with them. This entails investigating and putting into practice cutting-edge designs as well as comprehending their subtleties for well-informed model building. It is essential to divide the dataset into testing and training sets carefully. To maximize the effectiveness and generality of the model, the training phase entails iterative model optimization, hyperparameter fine-tuning, cross-validation, and ensemble approaches. A thorough model evaluation examines a range of performance indicators (such as accuracy, precision, recall, and F1 score) in a variety of scenarios. Transparency and reproducibility are ensured by thorough documentation using reproducible notebooks, version control, experiment logs, and comprehensive code comments. Producing detailed reports that include insightful visualizations, comparative analysis, and descriptive data improves the depth and clarity of research findings. Using programs like Jupyter Notebooks, PowerPoint, or LaTeX-based reports, researchers can effectively communicate their findings through the use of captivating visualizations, illustrative charts, and concise narratives. Effective distribution requires presentations that cover model findings, limits, and future research goals. Developing these facets offers a deeper comprehension of the numerous steps and factors that go into carrying out a thorough machine learning study. To guarantee the validity, precision, and interpretability of the research findings, each step is essential. Changes and other explanations might be added under particular project details or emphasis areas.

# CHAPTER 4

# EXPERIMENTAL RESULT

## 4.1 Experimental setup

The experimental setup for accident hotspot detection research commenced with acquiring five years' worth of accident records (2019-2023) sourced from the Dhaka Metropolitan Police Headquarters. Thorough data preprocessing ensued, encompassing validation, cleansing, and structuring to ensure data integrity. Essential attributes such as year, month, location coordinates, case and vehicle types, accident severity, day of the week, and driver age were meticulously selected for analysis, supplemented by feature engineering techniques for model enhancement. Exploratory Data Analysis facilitated data visualization using correlation matrices, and relationships within the accident data. The subsequent phase involved the development of machine learning models, employing classification algorithms like RandomForest, DecisionTrees, GaussianNB, LogisticRegression, and GradientBoostingClassifier for hotspot prediction. The derived insights and recommendations from hotspot identification and causal analysis were meticulously documented for further analysis and dissemination in research reports and experiment logs, ensuring replication and transparency in the research process.

## 4.2 Experimental Results & Analysis

**TABLE 4.2.1: CLASSIFICATION AND MODEL REPORT**

| **Label name** | **Precision** | **Recall** | **F1-Score** | **Accuracy** |
|---|---|---|---|---|
| LogisticRegression | 94.39% | 94.35% | 94.38% | 94.37% |
| RandomForest | 94.39% | 94.38% | 94.39% | 94.39% |
| GradientBoosting Classifier | 94.18% | 94.16% | 94.18% | 94.16% |

| | | | | |
|---|---|---|---|---|
| GaussianNB | 83.37% | 79.56% | 83.37% | 89.46% |
| DecisionTree Classifier | 93.14% | 93.09% | 93.14% | 93.09% |

**Accuracy:** Accuracy is a parameter used in machine learning model evaluation. Measures the ratio of correctly predicted observations to the total observations. It indicates the model's ability to make correct predictions across all classes or categories.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

For example, if a model correctly predicts 85 out of 100 instances, the accuracy would be calculated as

Accuracy = 85/100×100% = 85%

This means the model's accuracy in predicting those instances is 85%.

**Precision**: Precision refers to how close measurements of the same item are to each other. It could be defined as the sharp exactness of a measurement. Precise items may or may not be accurate.

$$Precision = \frac{TP}{TP + FP}$$

**Recall:** Recall in this context is defined as the number of true positives divided by the total number of elements that belong to the positive class (i.e. the sum of true positives and false negatives, which are items that were not labeled as belonging to the positive class but should have been)

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

It signifies the proportion of actual positive samples that were correctly predicted by the model among all actual positives. A higher recall score implies that the model can effectively identify most positive instances, minimizing false negatives, which are instances falsely predicted as negatives when they are positives.

**F1 Score:** The F1 score, often referred to as the F1-measure or F1-score, is a metric that combines both precision and recall into a single value.

$$F1\ score = 2 * \frac{Precision\ \ * Recall}{Precision + Recall}$$

This metric provides a balanced assessment of a model's performance, considering both false positives (precision) and false negatives (recall). It's particularly useful when you want to find a balance between precision and recall or when class imbalance is present in the dataset.

**4.3 Discussion**

Figure 4.3.1 presents the location-based percentage distribution, which weaves a fascinating tale and offers a wealth of information on accident incidents. The visualization of this data produced pleasing findings that can be interpreted as a kind of compass to understand the geographic position of the incidents that occurred throughout our study. With a significant 12.46% share of the overall accident distribution, Uttara is a prominent area of concern. This significant percentage emphasizes Uttara's significance as a significant hotspot, necessitating targeted interventions and raising awareness in this domain. Mirpur is at the other end of the spectrum, with a proportion of 7.48% showing the least amount of accidents. Although this would seem to point to a better safety profile in Mirpur, it begs for deeper research into the causes of this decreased incidence.

Given that they offer more detail than just basic statistical data, government representatives, law enforcement officers, and urban planners can use these location-based percentages as a compass to help them create targeted restrictions. A data-driven approach to resource distribution is made possible by the designation of Uttara as a high-impact zone and Mirpur as a relatively safer area. This enables more efficient time and resource allocation toward the accident reduction goal.

Therefore, Figure 4.3.1 is not only a beautiful treat for data fans, but it's also a useful tool for making well-informed decisions due to its sophisticated display of location-based percentages. This knowledge, which permeates policy discussions, encourages a proactive approach to accident prevention and community well-being and creates the foundation for a safer and more secure urban environment.

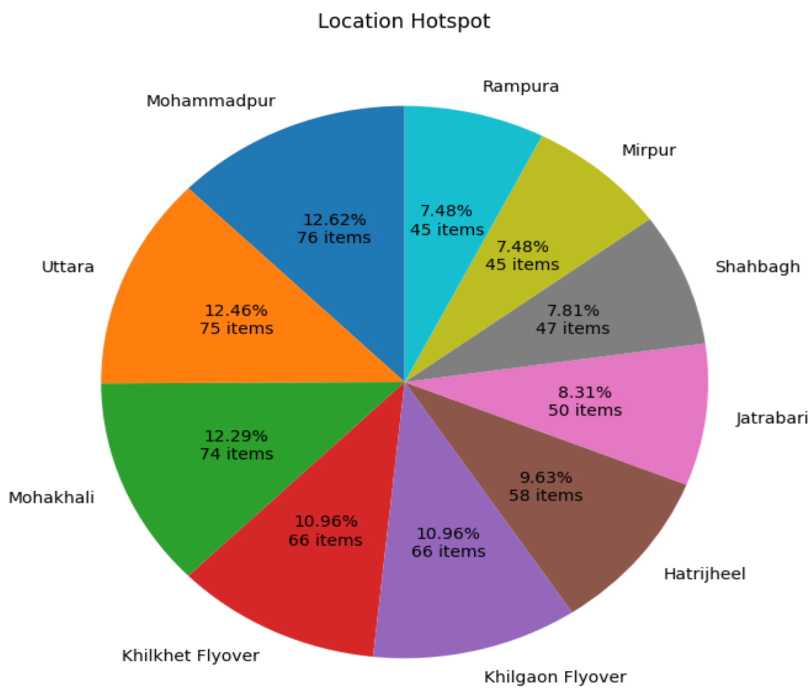


Figure 4.3.1: Location Hotspot

Here, several crucial methods were used to guarantee the quality of the dataset and its alignment with modeling algorithms. Imputation techniques like mean, median, or mode imputation were used to handle missing values. In certain situations, rows or columns with little to no missing data were also dropped. Label encoding was also used to transform categorical data into a numerical format. This allowed for model compatibility and interpretation by giving each category a distinct numerical label. Numerical features were standardized using scaling modification techniques such as Min-Max Scaling and standardization to make sure they were all on the same scale and to stop any one characteristic from dominating the model because of its higher magnitude.

Figure 4.3.2 uses a correlation matrix. The correlation coefficients in his matrix ranged from -1 to 1, with values around -1 signifying a strong negative correlation, values around 0 indicating a weak or nonexistent linear link between the variables, and values closer to 1 indicating a high positive correlation. Visual examination of the correlation matrix or individual correlation values provided information about which attributes showed strong relationships or dependencies with one another. A characteristic's high correlation within the dataset, specifically with the variable "accident severity," suggests a strong association between this feature and possibly other qualities.

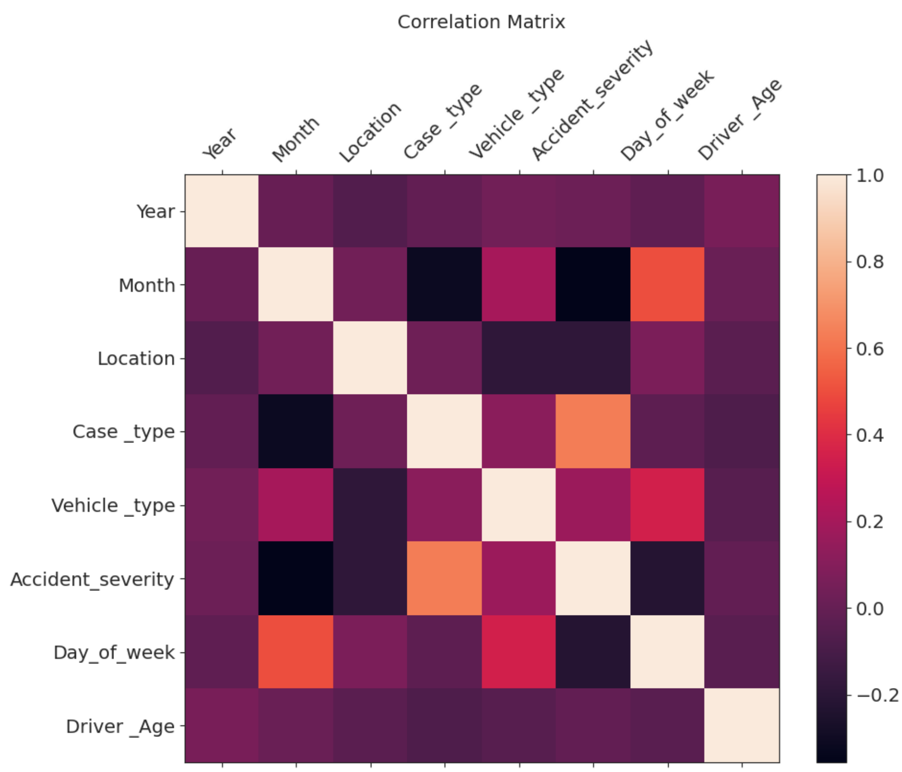


Figure 4.3.2: Correlation Matrix

## 4.4 Evaluation

Figure 4.4.1 includes a visualization that shows the importance of features used in the analysis, highlighting key contributors to accident occurrences and facilitating an educated curation of the dataset by directing the removal of less significant variables and emphasizing the most significant ones.

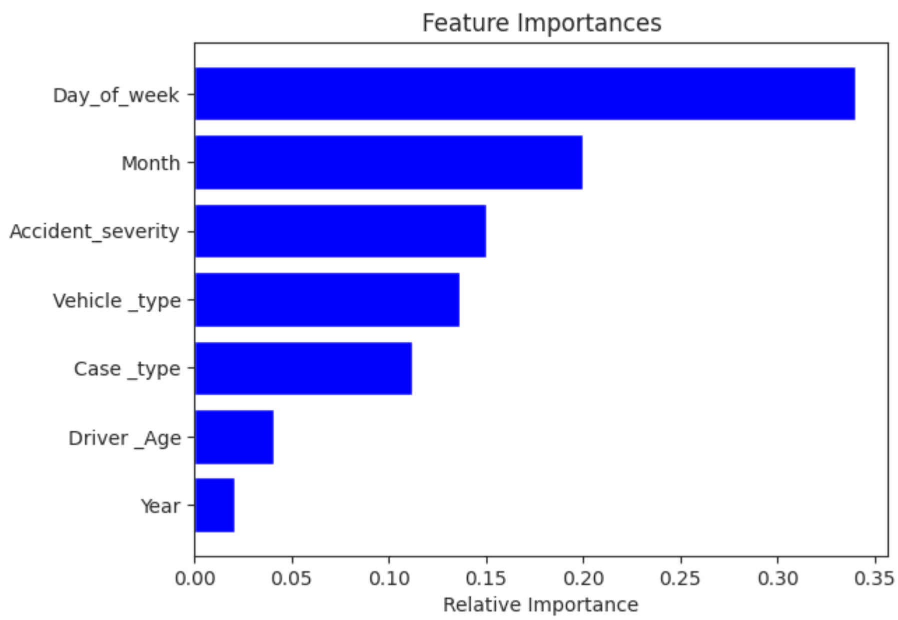


Figure 4.4.1: Feature Importance Visualization

The annual accident frequency in the Dhaka Metropolitan City from 2019 to 2023 is depicted in Figure 4.4.2. This graphic provides a thorough summary of the varying accident rates over the five years, emphasizing the particular years where accident rates either rose or fell. Annual data showed an increase in accidents in 2023. This analysis makes an improved comprehension of the variations and patterns defining accident rates in the Dhaka Metropolitan City between 2019 and 2023 possible, which offers insightful information on the dynamic nature of accident occurrences over time.

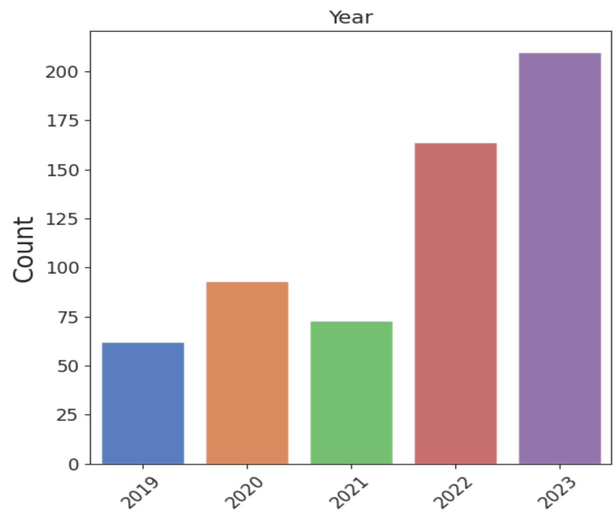


Figure 4.4.2: Yearly Accident Frequency

The monthly accident frequency in the Dhaka Metropolitan city from January to December is shown in Figure 4.4.3. It draws attention to particular months when the frequency of accidents rises or falls. This analysis provides insights into the temporal trends influencing accident occurrences. It provides a greater knowledge of the dynamic patterns and changes characterizing monthly accident rates within the Dhaka Metropolitan City." In January due to fog or harsh winter conditions, potentially leading to more accidents. August and September might witness increased traffic due to vacations, holidays, or events. December, often with holiday-related festivities, can also lead to more accidents due to crowded roads and distracted driving behaviors.

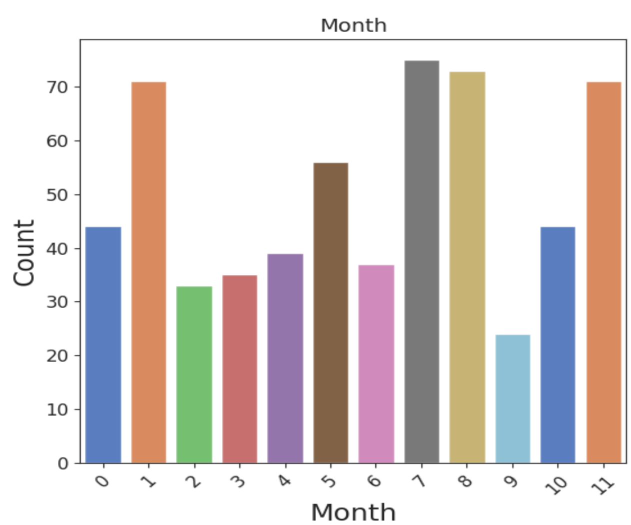

Figure 4.4.3: Monthly Accident Frequency

"The frequency of accidents across different days of the week in the Dhaka Metropolitan city is illustrated in Figure 4.4..4. This visual representation offers a comprehensive view of accident occurrences on various days over the five years. It highlights any fluctuations in accident rates during specific days of the week. This analysis delves into the temporal trends influencing accident patterns, shedding light on the dynamics and variations in accident rates throughout the week within the Dhaka Metropolitan City."The daily analysis identified Fridays as a peak day for accidents, as it is a weekend day people tend to travel more or return home for the weekend, and people might be more relaxed with weekend plans, leading to decreased focus while driving.

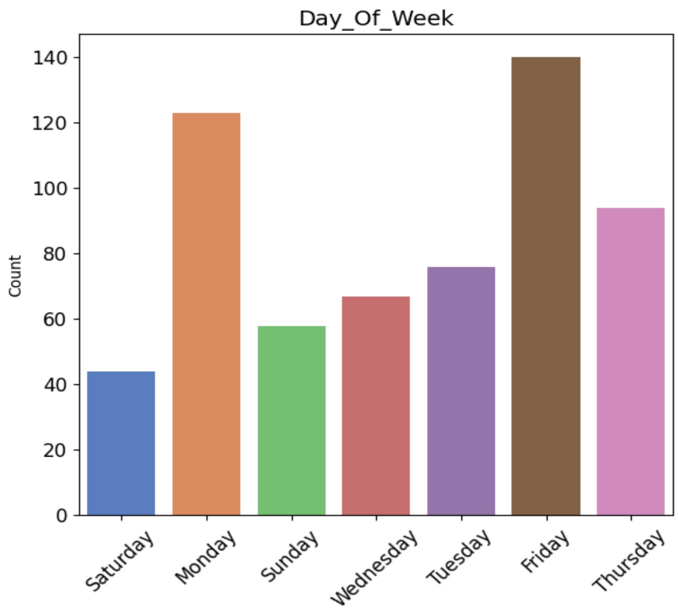


Figure 4.4.4: Daily Accident Frequency

# CHAPTER 5

# Impact on Society, Environment, and Sustainability

## 5.1 Impact on Society

- Enhances traffic safety by targeting safety improvements in accident-prone areas, thereby increasing overall road safety.
- Aims to prevent injuries and fatalities by actively intervening in high-risk areas to potentially reduce accident rates.
- Provides emergency agencies with crucial information for more effective resource distribution, improving response times and accident management.
- Raises public awareness about accident-prone areas, empowering locals to advocate for safer roads and engage in safety projects.
- Influences possible policy changes in traffic laws and infrastructure upgrades based on research findings.
- Anticipates financial benefits from fewer accidents, impacting medical expenses, property damage, and productivity positively.
- Identifies high-risk areas in advance for interventions that could potentially save lives.
- Flag areas need infrastructure improvements for safer roads for all travelers.
- Envisions a safer living environment and reduced accident rates for societal well-being.
- Acknowledges the potential impact of raising awareness on commuter behavior and safer driving habits.
- Advocates for policy reforms to enhance road safety regulations based on potential findings.
- Emphasizes the need for data-driven decision-making to address road safety societal issues.

## 5.2 Impact on the Environment

Although the primary focus of accidental hotspot identification is urban safety, there are indirect but potentially significant environmental implications. The study's conclusions show promise for influencing urban planning methods, enhancing transportation networks, and refining resource allocation. Urban planners can improve air quality and minimize emissions by implementing eco-friendly infrastructure, optimizing traffic flow, and reducing congestion by utilizing recognized hotspots. In emergency response settings, effective resource management can help reduce waste and use resources sustainably. By reducing travel delays and traffic congestion, the project's interventions aimed at lowering accident rates may also indirectly result in fewer vehicle emissions and a smaller carbon impact. Furthermore, the identification of areas prone to accidents could serve as a trigger for the creation of green infrastructure, promoting the creation of a more sustainable urban environment by implementing projects such as improved mobility options and pedestrian-friendly spaces. This study may have an indirect direct environmental impact, but its effects on resource management, transportation efficiency, and urban design may lead to the development of an environmentally conscious and more sustainable urban environment.

## 5.3 Ethical Aspect

To successfully address its ethical issues, data handling, bias mitigation, and responsible technical deployment are of utmost importance. Encrypting and anonymizing sensitive accident data is a must, and protecting individual privacy and data security is a cornerstone. To ensure fair representation and avoid biased consequences, it is imperative to be vigilant about inherent biases within datasets. Maintaining openness about methods and findings encourages trust and accountability, which advances moral behavior in research activities. Respecting ethical norms and respecting community impact requires involving stakeholders and impacted groups, obtaining consent when necessary, and taking into account a variety of viewpoints. To guarantee that the project's results emphasize societal well-being, equity, and conformity with legal and ethical frameworks governing data usage and technology deployment, responsible development and deployment of technology must be combined with ongoing review. The project's dedication to appropriate data use and the ethical ramifications of inadvertent hotspot discovery are highlighted by ethical considerations.

### 5.4 Sustainability

Beyond only preventing accidents, sustainability in this project encompasses a comprehensive strategy for long-term urban safety and resilience. It comprises limiting the influence on the environment by decreasing emissions, streamlining traffic, and creating eco-friendly metropolitan areas by identifying infrastructure needs. Social justice must be upheld, with a focus on inclusive safety measures that equitably benefit different community groups. A strategy for preventing accidents must be both economically viable and demonstrate efficient use of resources and long-term benefits. Integrating technology sustainably requires robust, scalable systems that can change to meet changing needs for accident prevention. Resilient urban planning that incorporates accident insights promotes flexible city environments that can both mitigate and survive the effects of accidents. To promote active participation in safety programs and to leverage local knowledge for sustainable resilience, community engagement is essential. To ensure that urban safety is consistently prioritized, policy frameworks supporting accident prevention activities are essential. By taking into account a variety of factors, the project hopes to develop long-term safer urban environments by balancing environmental, social, economic, and technological aspects of accident prevention. The goal of this initiative is to develop accident prevention techniques that prioritize inclusive community engagement, reduce environmental impact, show cost-effectiveness, and adapt to changing technological capabilities. By balancing social, economic, and environmental factors, this all-encompassing strategy aims to create long-lasting urban safety measures that promote resilient and sustainable urban environments."

# CHAPTER 6

# Summary, Conclusion, Recommendation, and Implication for Research

## 6.1 Summary of the Study

This Study is a comprehensive investigation into accidental hotspot detection within the Dhaka Metropolitan City, focusing on a meticulous analysis of accident data spanning a significant time frame from 2019 to 2023. By harnessing the power of advanced machine learning models and leveraging extensive datasets meticulously collected by the Dhaka Metropolitan Police, the research aims to unearth the intricate patterns and underlying factors contributing to the emergence of accident hotspots. The primary objective is to discern the multifaceted dynamics influencing accident occurrences, categorizing and scrutinizing diverse accident records into distinct classes. This meticulous categorization enables a nuanced exploration of socio-economic, infrastructural, temporal, and environmental variables that intricately intertwine to shape the formation and persistence of these hotspots. Through the utilization of predictive models and sophisticated data-driven insights, the study endeavors not only to identify these hotspots but also to predict and forestall potential areas susceptible to accidents within the metropolitan area. The ultimate aim is to propose and implement targeted interventions that mitigate accident occurrences in these identified hotspots, thereby fostering a safer urban environment.

The implications of this research extend beyond the realm of accident prevention, as the findings are anticipated to catalyze informed urban planning decisions. By offering a deeper understanding of accident patterns and their associated variables, the study envisages contributing pivotal insights that can steer traffic management strategies and shape urban development policies. These insights have the potential to revolutionize urban planning paradigms, nurturing the creation of safer, more sustainable, and resilient urban environments within the Dhaka Metropolitan City.

## 6.2 Conclusions

Finally, our thorough investigation shed light on the complexities of unintentional hotspot detection in the Dhaka Metropolitan City, employing the Random Forest classifier to achieve an astounding accuracy of 94%. The research  finds out the complex interplay between socio-economic, infrastructural, and temporal aspects that contribute to hotspot development using rigorous geographical and temporal studies. These discoveries open the door for predictive frameworks, most notably the Random Forest model, which enables the proactive forecasting of potentially accident-prone locations. They also offer an important understanding of focused interventions and policy formulation. These developments provide a strong basis for improving traffic safety policies and guiding data-driven urban planning choices, to build more resilient, safer, and sustainable urban environments. Going forward, complete road safety solutions will require ongoing data-driven research, model improvements, and cooperative efforts. Notwithstanding significant constraints, the study's conclusions are quite encouraging, pointing to concrete benefits for society such as fewer accidents, safer roads, and improved living standards for locals. The study has implications for the Dhaka Metropolitan City that go beyond the confines of academia by providing useful avenues for promoting safer and more sustainable urban settings.

### 6.3 Implication for Future Study

Future research could potentially explore the possibility of extending accident analysis's coverage to include the whole Dhaka Metropolitan city, as opposed to only the ten sites it currently covers. Expanding the scope of the investigation to encompass a larger geographic region would yield a more thorough comprehension of accident trends, the identification of hotspots, and the fundamental elements causing road safety issues throughout the entire metropolitan area. Expanding the range of sites and integrating information from various parts of Dhaka can provide a more thorough and refined understanding of the geographical variances, time patterns, and socioeconomic factors influencing the incidence of accidents in the future. This expansion can entail creating scalable procedures, working with different agencies, and incorporating additional databases to support wider regional coverage. Expanding the study's scope to include the entire metropolitan area may also make it easier to develop inclusive and successful accident prevention plans, urban planning projects, and focused interventions that are suited to the unique requirements of different neighborhoods. All of these things could ultimately help create safer and more resilient urban environments on a larger scale. We will also use additional machine learning methods, particularly neural networks and deep learning techniques, as they are effective and occasionally outperform simpler algorithms.

## References:


1. [1] da Silva, J. P., da Costa, G. F., & Mateus, G. R. (2022). "A Machine Learning Approach for Classifying Road Accident Hotspots." Applied Sciences, 12(8), 3997.
2. [2] Bülbül, H.˙I.; Kaya, T.; Tulgar, Y. Analysis for Status of the Road Accident Occurance and Determination of the Risk of Accident by Machine Learning in Istanbul. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; IEEE: Manhattan, NY, USA, 2016.
3. [3] Sangare, M.; Gupta, S.; Bouzefrane, S.; Banerjee, S.; Muhlethaler, P. Exploring the forecasting approach for road accidents: Analytical measures with hybrid machine learning. Expert Syst. Appl. 2021, 167, 113855.
4. [4] Khan, M. S., & Showrov, H. (2020). "Accidental Prone Area Detection in Bangladesh using Machine Learning Model." International Journal of Computer Applications, 172(7), 29-33.
5. [5] Hossain, M. S., & Uddin, M. N. (2022). "An Analysis of the Most Accident Prone Regions within the Dhaka Metropolitan Region Using Clustering." Journal of Disaster Research, 17(8), 1301-1310.
6. [6] Zhang, J., Sun, S., & Li, R. (2020). "Predicting Traffic Accident Severity Using Machine Learning Algorithms in Urban Areas." Sustainability, 12(15), 5905.
7. [7] Tambouratzis, T.; Souliou, D.; Chalikias, M.; Gregoriades, A. Combining probabilistic neural networks and decision trees for maximally accurate and efficient accident prediction. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010.
8. [8] Satu, M.S.; Ahamed, S.; Hossain, F.; Akter, T.; Farid, D.M. Mining traffic accident data of N5 national highway in Bangladesh employing decision trees. In Proceedings of the 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh, 21–23 December 2017; IEEE: Manhattan, NY, USA, 2017.
9. [9] Iranitalab, A.; Khattak, A. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 2017.
10. [10] A. Jain, G. Ahuja, D. Mehrotra et al., “Data mining approach to analyze the road accidents in India,” in 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2016, pp. 175–179.
11. [11] N. Mahajan and B. P. Kaur, “Analysis of factors of road traffic accidents using enhanced decision tree algorithm,” International Journal of Computer Applications, vol. 135, no. 6, pp. 1–3, 2016.
12. [12] M. S. Hossen, “Analysis of road accidents in Bangladesh,” 2010.

13. [13] World Bank Blogs Website, 2012,
14. [14] Liu, C., Song, Y., & Wang, F. (2019). "Accident prediction model considering temporal factors using machine learning." Transportation Research Part C: Emerging Technologies, 105, 87-100.
15. [15] Sharma, U., Jain, V., & Saraswati, R. (2019). "Identification of accident-prone zones in Delhi metro using spatial and temporal data analysis." Journal of Engineering, Information, and Technology, 13(3), 819-830.
16. [16] Goudarzi, N., Sahraeian, S. M., & Dehghani, M. (2023). "Pedestrian Road Traffic Accidents in Metropolitan Areas: GIS-Based Prediction Modelling of Cases in Mashhad.
17. [17] Ma, X., Sun, W., & Wu, W. (2020). "A spatial-temporal hybrid clustering model for identifying traffic accident hotspots." Accident Analysis and Prevention, 150, 105955.
18. [18] Aydin, N., Akbal, M., & Yıldırım, U. (2023). "Machine Learning Based Prediction of Traffic Accidents on Major Roads in Istanbul." Sustainability, 15(10), 3405.
19. [19] Huang, Z., Li, M., & Xie, X. (2020). "Real-time prediction of traffic accidents using Deep Learning and Spatiotemporal Features." Applied Sciences, 10(24), 8405.
20. [20] Guo, Y.; Li, Z.; Liu, P.; Wu, Y. Exploring risk factors with crashes by collision type at freeway diverge areas: accounting for unobserved heterogeneity. IEEE Access 2019, 7, 11809–11819. [CrossRef] 21. Kwon, O.H.; Rhee, W.; Yoon, Y. Application of classification algorithms for analysis of road safety risk factor dependencies. Accid. Anal. Prev. 2015, 75, 1–15

**Plagiarism :**

ACCIDENTAL PRONE AREA DETECTION IN DHAKA METROPOLITAN AREA USING MACHINE LEARNING MODEL

ORIGINALITY REPORT

| 13% | 11% | 0% | 9% |
| --- | --- | --- | --- |
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | | |
| --- | --- | --- | --- |
| 1 | dspace.daffodilvarsity.edu.bd:8080 | Internet Source | 9% |
| 2 | www.coursehero.com | Internet Source | 1% |
| 3 | www.ijraset.com | Internet Source | <1% |
| 4 | downloads.hindawi.com | Internet Source | <1% |
| 5 | Submitted to Asian Institute of Technology | Student Paper | <1% |
| 6 | Submitted to University of Stirling | Student Paper | <1% |
| 7 | Submitted to European University | Student Paper | <1% |
| 8 | Submitted to University of North Texas | Student Paper | <1% |
| 9 | Submitted to Botswana Accountancy College | Student Paper | <1% |