ADDIS ABABA UNIVERSITY
COLLEGE OF TECHNOLOGY AND BUILT ENVIRONMENT
SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

Developing a Crash Severity Prediction Model Using Machine Learning Technique

By: Kidane Yifru
Advisor: Dr. Sosina Mengistu (PhD)

A thesis submitted to
The School of Graduate Studies
In partial fulfillment of the requirements for the degree of Master of Science in
Road Safety
(Road and Transport Engineering)

September, 2025

The undersigned have examined the thesis entitled "**Developing a Crash Severity Prediction Model Using Machine Learning Technique**" presented by Kidane Yifru, a candidate for the Degree of Master of Science, and hereby certify that it is worthy of acceptance.

Submitted By:

Kidane Yifru	____________________	____________________
Student	Signature	Date

Approved By:

Dr. Sosina Mengistu	____________________	____________________
Advisor	Signature	Date
____________________	____________________	____________________
External Evaluator	Signature	Date
____________________	____________________	____________________
Internal Examiner	Signature	Date
____________________	____________________	____________________
Chairperson	Signature	Date

UNDERTCKING

I certify that the research work titled **Developing Crash Severity Prediction Model Using Machine Learning Technique** is my work. The work has not been presented elsewhere for assessment. Where material has been used from other sources, it has been properly acknowledged or referred to.

Kidane Yifru

Name Signature Date

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS

AADT: Average annual daily Traffic
ADB: Ada boost Classifier
CNN: Convolutional Neural Networks
DNN: Dynamic neural network
DT: Decision Tree
ERA: Ethiopian Road Administration
ETRE: Ethiopian Toll Roads Enterprise
GDP: Gross Domestic Product
GBM: Gradient Boosting Machines
ITS: Intelligent Transport System
KNN: K-Nearest Neighbors
LR: Logistic Regression
ML: Machine Learning
NB: Nave baye
NN: Neural Networks
RF: Random Forest
RNN: Recurrent Neural Networks
RTC: Road Traffic Crash
SMOTE: Synthetic Minority Over-sampling Technique
SVM: Super Vector Machines
WHO: World Health Organization
XGB: Extreme Gradient Boost Classifier

ACKNOWLEDGMENTS

Firstly, I am extremely grateful to GOD the Almighty and our mother St. Mary the Virgin for giving me the strength, protection, and support.

Secondly, my deepest thankful to my supervisor, Dr. Sosina Mengistu for her invaluable advice, continuous support, and patience during my MSc study and her immense knowledge and plentiful experience have encouraged me in all the time of my academic research.

I would like to thank the Ethiopian Toll Roads Enterprise, Ethiopian Road Safety and Insurance Fund Service, and the School of Civil and Environmental Engineering, Addis Ababa University, Addis Ababa, for offering me an ideal environment in which I felt free and remained concentrated on my research.

Finally, I would like to express my gratitude to all members of my family. Without their understanding and encouragement in the past few years, it would have been impossible for me to complete my studies.

ABSTRACT

Road Traffic crashes (RTCs) deaths and injuries are a universal global public health problem and significantly affect a country's economy, up to 3% of annual GDP. *This paper examined the relationship between traffic flow features and expressway RTC severity. This study utilized a dataset containing detailed information on road traffic crashes and traffic flow. A comprehensive set of machine learning algorithms was involved, including LR, KNN, ADB, RF, XGB, NB, SVM, and DT, which were employed to develop the predictive model. The performance of the Model was assessed in terms of accuracy, precision, recall, and the F1 score. The best among the evaluated eight models for crash Severity prediction was the Random Forest model with an accuracy of 83%, a precision of 82.9 %, a recall of 83%, and an F1-score of 82.8 %. The study also found that chainage, months, crash types, and weather conditions are critical impacting features for crash severity prediction on the Addis-Adama Expressway. Moreover, traffic flow characteristics and Road traffic crash severity are weakly correlated.*

Thus, this study proves that the temporal and spatial factors are important for predicting RTC severity. The random forest model demonstrates to be an effective tool for forecasting RTC severity, which could be an effective model. Future work should focus on developing a prediction tool with an interactive map that could proactively monitor crash hot spots for potential crash severity in the Expressway, Ethiopia.

Keywords: RTC Analysis, Machine Learning Models, Crash Severity, Prediction Model, Confusion Matrix

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the study

Road traffic crash (RTC) deaths and injuries remain a major health and development concern in many countries. According to the World Bank Road Safety Status Report, 2023, the estimated road traffic deaths were 1.19 million, and injuries are the leading cause of death among people aged 5-29 years. Moreover, RTCs not only result in tragic loss of life but also incur significant financial burden on victims and their families, both the direct cost of treating injuries and the indirect cost of lost productivity from those disabled or killed. More broadly, road traffic injuries significantly depress national economies, costing countries 1-3% of their annual gross domestic product (WHO, 2023).

Severe RTC has devastating effects on low- and middle-income countries like Ethiopia due to its seriousness and the limited resources to develop feasible countermeasures for reducing this ever-growing challenge (Zenu, S., et al., 2023; Kidane, M., 2025; Deme, D., 2018; Abera, M. A., 2019; and Deme and Bari, 2018). In Ethiopia, RTC remains the main cause of deaths among the active working force ages 18–50 years, with fatality (66 % in 2018) and injuries (60 % in 2018), which are above the African prevalence. Buses and commercial vehicles are involved in crashes at a high rate, accounting for nearly 63% of fatal crashes (Federal Police Commission, 2019). This growing crisis has heightened awareness across society regarding the seriousness of road safety issues and the urgent need for improvement. More recently, road safety has received increased international attention, notably through objectives set by the World Health Organization (WHO), which aims to significantly diminish these adverse outcomes by 2030, emphasizing the critical need for more strong analytical methods. (Benfaress et al., 2024)

Although the crash injury severity analysis almost entirely depended on statistical techniques. Such as ordered probit models, and multinomial logit (Iranitalab & Khattak, A., 2017). These favored models for their interpretability, which clearly indicated relationships between intangible factors and crash results. Nevertheless, they frequently faced shortcomings when trying to represent the complicated, non-linear relationships and complex trends in real-world crash data. Their formative assumptions regularly fell short of the multidimensional nature of traffic crashes. Here, prediction of RTC with traditional statistical techniques tends to be inadequate due to built-in assumptions, but machine learning techniques have mighty analytical capabilities for large datasets and identifying complex trends that could be used for predicting, if not averting, road crashes (Benfares, I., et al, 2024). These methods have shown much potential for road safety when considering their capacity for recognizing non-evident relationships between crash information and predicting crash outcome severity.

Thus, utilizing machine learning algorithms to identify the intricate interplay factors that contribute to RTC severity and modelling is important for devising accurate and actionable RTC predictions and safety interventions. Consequently, this study aimed to develop the most suitable and effective RTC severity prediction model by rigorous testing and applying various ML algorithms.

1.2. Statement of the problem

The increasing prevalence of RTC is a major public health challenge faced by many countries, particularly in low- and middle-income countries, including Ethiopia. RTC is the main cause of death for children and young adults aged 5 to 29 years (WHO, 2018). In Ethiopia, RTC fatality (66 % in 2018) and injuries (60 % in 2018) persist among the active working force ages 18–50 years. Nearly 63% of fatal crashes involved buses and commercial vehicles (Federal Police Commission,2019). Thus, Ethiopia, as a nation, has a disproportionally high road crash fatality rate (28.2 deaths per 100,000 person) , compared to the global average (17.05 deaths per 100,000 person) (WHO 2019), but there is very little research dealing with this critical issue in the country. However, the severity of crashes occurring on the roads remains a critical area for attention. The available statistics also suggest that focusing on predicting the severity of RTC and identifying the risk factors that are related to the severity of crashes may be an effective way to prevent injury and death using traffic flow and crash data.

Despite advancements in road safety measures, understanding the nuanced dynamics of how traffic flow contributes to crashes has not been widely investigated. Besides, to the best of the authors' knowledge, only a few studies predict crash severity on expressways using machine learning worldwide, and these are rarely found in Ethiopia. As a result, this study aims to address this gap using a Machine Learning model and inform targeted interventions and improve road safety on the Addis Ababa -Adama Expressway.

1.3. Objectives

1.3.1. General Objective

The primary aim of this study is to develop a crash severity prediction model using machine learning techniques.

1.3.2. Specific Objectives

In line with the general objective, the specific objectives of the study are:

1. Investigate the interaction between road traffic crash severity and attributes of traffic flow.
2. Develop the road traffic crash severity prediction model and compare the performance of different machine learning algorithms in crash severity prediction.
3. Identify the significant contributing features (variables) that influence the severity level of RTAs.

1.4. Research questions/Hypothesis

This thesis aims to address the following research questions:

- ✓ Is there any relationship between traffic flow characteristics and crash severity?
- ✓ Which machine learning algorithms demonstrate the highest accuracy and reliability in predicting crash severity in this context?
- ✓ Which features have a significant contribution to crash severity?

1.5. Significance of the study

This study has significant importance in addressing the persistent problem of road safety in Ethiopia. The study introduces an innovative, data-driven approach to understanding and mitigating severe RTCs by predicting and assessing the RTC severity using an ML prediction model. The results will offer valuable insights about the recent methodology of RTC severity prediction, contributing factors for severe RTCs, enabling policymakers, road safety authorities, and transportation planners to implement effective interventions that can reduce deaths and injuries.

The study also shows the growing body of knowledge in traffic safety and artificial intelligence, particularly in the context of developing countries. While much of the existing research has focused on developed nations, this study emphasizes the unique challenges faced in Ethiopia, Addis-Adama Expressway, offering context-specific solutions that could serve as a blueprint for other similar expressways. Additionally, by improving the allocation of post-crash management resources and improving traffic management strategies, the study aligns with international and national road traffic safety goals and fosters economic growth through efficient transportation.

Finally, this study will not only address a vital public safety issue but also reveal the transformative potential of technology in resolving real-world challenges, making it a significant step forward for both academia and practice.

1.6. Scope of the study

This study focuses on predicting crash severity specifically on the Addis Ababa-Adama Expressway, recognizing its strategic importance as Ethiopia's busiest and most critical transportation route. It aims to analyze traffic dynamics and historical crash patterns to develop a machine learning model tailored to the unique conditions of this expressway.

The study encompasses various influencing factors such as traffic volume, spatial-temporal characteristics, weather conditions, and driver demography to identify key determinants of crash severity. Data will be collected from relevant sources such as traffic management systems and crash records. The research will utilize machine learning algorithms to develop predictive models, which will be tested and validated using historical data from the expressway.

The findings of this study are intended to provide insights that are practical for policymakers, traffic authorities, and road safety experts in Ethiopia. However, while the research will make use of comprehensive data from the Addis Ababa-Adama Expressway, its conclusions may not be universally applicable to other roads with different traffic and environmental conditions. The study also operates within the limitations of data availability, quality, and temporal scope, focusing on current and historical datasets rather than long-term trends.

By narrowing its focus to this expressway, the research aims to deliver actionable results while setting a foundation for further studies that could generalize its approach to other road networks.

1.7. Organization of the study

The study is structured into the following chapters to provide a clear and systematic exploration of the study topic:

Introduction This chapter offers an introduction to the study background, a summary of global and local road safety concerns, a motivation for reducing the severity of crashes and their socio-economic impacts, a rationale for developing a crash severity prediction model, a study purpose and questions, a study importance, and a study scope and limitation. It prepares for the thesis and identifies a study issue it wishes to solve.

Literature Review: Here, we discuss related studies and current literature regarding road traffic crashes, evaluation of RTC severity methodologies, machine learning approaches for road safety, and variables impacting RTC severity. We find gaps in current studies and position our work among the broader body of work in academia.

Methodology: This section of research describes the study's structure, explains procedures for data collection, and elaborates on the steps of data preprocessing. Additionally, it describes machine learning algorithms employed, model evaluation criteria, and model training, validation, and testing procedures.

Results and Discussion: It describes the results of the analysis of data related to traffic and RTCs, the prediction of RTC gravity by machine learning models, and ML model evaluation. It also provides a comparative analysis of data and model performance, identifies major factors affecting RTC gravity, and interprets the results of the study from a road safety perspective.

Conclusion and Recommendations: The last chapter will concisely highlight the notable achievements of the study, emphasize its contributions to road safety practice and research, and make policy recommendations for policymakers and road safety practitioners. It also includes potential for further studies.

Appendices: The appendices include the ETRE Vehicle classification table, sample raw data, data collection forms, and code snippets.

1.8. Limitations of the study

This study produces a crash Severity Prediction Model using machine learning techniques for the Addis-Adama Expressway, Ethiopia. Whilst generalizing results, the following limitations should be considered during generalization of results:

Also, the success of the predictions heavily depended on the quality and completeness of the historical crash data employed. Missing or incorrect data points might substantially influence model performance. Second, the study may not have accounted for all potential factors leading to RTC severity. Relevant variables like driver behavior, road condition, and weather condition were not adequately addressed, thereby tethering the scope of analysis in a sense of incompleteness. Third, models constructed during the study may not generalize across other Expressways nor across other conditions, and the study may not have investigated all possible algorithms nor all possible combinations of algorithms and hyperparameter optimization that are potentially beneficial for prediction improvement.

Finally, machine learning's complexities and opacities, as typified by the Random Forest, pose additional challenges. Such a method complicates interpretation, leading to impediments for understanding the basic constituents shaping the predictions themselves. Such limitations serve to emphasize the need for additional study and development in the field of predicting crashes by machine learning.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Overview of Road Traffic Crash Analysis

Road traffic crash analysis uses various statistical and ML techniques to determine the factors that cause crashes. Research frequently uses these methods to model the probability of crash occurrence and the severity of crash injuries (Mannering and Bhat, 2014). Some studies used conventional data analysis techniques such as binary logit and probit models, while others evolved to consider multiple discrete outcomes like multinomial logit model, nested logit model, and the random parameters logit model. Although statistical models can be mathematically well-interpreted and easy to understand the function of specific predictor variables, they have some limitations, such as poor prediction accuracy, producing biased model estimation, etc. Consequently, the analytical landscape for crash analysis has undergone a significant transformation over time from univariate and multivariate count models for instantaneous probability of an crash outcome to simple binary discrete and multiple discrete outcomes models to address crash-severity.

Henceforth, effective prediction of crash severity is essential for enabling the government to implement targeted interventions and allocate resources efficiently for emergency response and crash prevention (Obasi, I. C., & Benson, C., 2023).

However, all these models have limitations in capturing the complex, non-linear relationships in crash data (Yu & Abdel-Aty, 2020) and need more advanced computational methods. A growing number of studies now employ advanced machine learning algorithms to investigate factors associated with injury severity.

2.2. Machine Learning Techniques

Machine learning algorithms can discover intricate correlations between the many factors influencing injury severity and make highly accurate predictions. Moreover, the core of crash severity prediction lies in the selection and application of appropriate machine learning models. Thus, the prominent, frequently employed, and highly effective machine learning algorithms in crash severity prediction are grouped as follows.

2.2.1. Supervised Learning Models

Supervised learning models are popular for RTC prediction because of their ability to learn from labeled datasets. Among these, Decision Trees and Random Forests have been pretty commonly utilized. Chang and Chen (2005) used decision trees to figure out which factors matter most in RTC severity, such as weather conditions and road type. The results showed that decision trees could effectively classify crashes, but struggled a bit when the data was more complicated. Meanwhile, Random Forests tend to be better than a single decision tree because they combine multiple trees to make a more accurate prediction. Beshah and Hill (2010) used Random Forests for RTC severity prediction; they showed that it was more accurate than the usual classical statistical models. The random forest can handle many data types well and is less likely to overfit; for these reasons, many studies prefer this method.

Support Vector Machines (SVMs) have recently been introduced for transportation applications (Zhang and Xie, 2007), particularly in classification tasks, e.g., crash types classification at urban intersections (Abdel-Aty and Haleem, 2011) and prediction of crashes (Li et al., 2008). However, SVMs are complex to estimate, and it can be difficult to generalize to other data sets. Moreover, SVMs tend to behave as black-boxes (difficult to provide the interpretable parameters) (Lord, D., & Mannering, F., 2010).

2.2.2. Ensemble Learning Techniques

Ensemble learning techniques are a combination of multiple learning algorithms to improve predictive performance, have gained traction in RTC prediction. Gradient Boosting Machines (GBMs), an advanced ensemble method, have been employed by Liu et al. (2017) to predict Crash hotspots. The study highlighted that GBMs outperformed simpler models by effectively capturing complex interactions between variables. GBM's ability to reduce both bias and variance in predictions have made it particularly effective in RTC prediction tasks.

2.2.3. Neural Networks and Deep learning

With a large amount of data, neural networks (NNs) have been used to predict RTCs. Dong et al. (2018) employed a deep neural network to model traffic crash data, found patterns that simple methods could not see. Their result showed that deep learning could make accurate predictions, especially when trained on big, varied datasets. But these methods act as black boxes, which makes it hard to understand how they work. This makes it difficult to use in real-world applications. Different types of neural networks (NNs), like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been tested for RTC prediction over space and time. Zhou et al. (2020) created a model that combines a CNN and an RNN to analyze both location and time data. This model predicts crashes more accurately than standalone models. This approach shows the potential of deep learning or understanding the complex traffic pattern.

2.3. Factors Influencing Crash Severity

Many factors contribute to road traffic crashes (RTCs). These include traffic flow, driver actions, and weather or environmental conditions. Understanding these factors can help us create better safety measures. The causes of RTCs involve multiple elements, for example, traffic patterns (Garber et al, 2000), weather (Infante et al, 2023), driver behavior (Rolison et al, 2018), and visibility conditions, vehicle conditions (Baru et al, 2019). Recognizing these causes is vital to RTC prevention and reducing their effects.

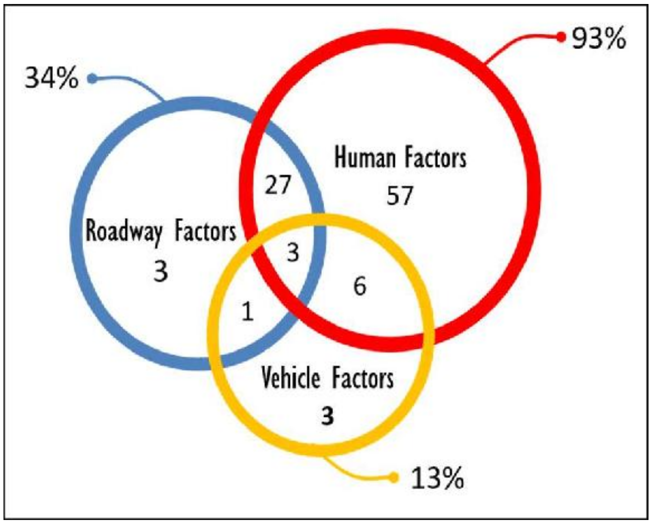


FIGURE 1: THE CONTRIBUTION FACTOR OF RTCS. (SOURCE: AASHTO, 2010)

2.3.1. Traffic flow characteristics

Traffic flow characteristics are a major cause for road traffic crashes, which are defined as the interactions between vehicles, drivers, and infrastructure, and include volume, speed, density, and road capacity. The term traffic volume refers to the number of vehicles passing a road section during a certain period of time (e.g., 5 minutes, one hour, or a day). Likewise, average annual daily traffic (AADT) is the number of vehicles passing through a given road section within a year, divided by 365. Studies have shown that low traffic volumes and higher percentages of heavy vehicles often correlate with increased severity of crashes (Casado-Sanz, 2020). Moreover, Yasmin et al. (2017) examined the impact of traffic flow on crash frequencies in urban areas, indicating that higher traffic volume correlates with increased crash rates due to reduced maneuverability and reaction times.

Speed is another vital factor in traffic flow that significantly affects RTC severity. According to Elvik et al. (2004), the relation between speed and RTCs is non-linear; as speed increases, the probability of severe injuries and fatalities escalates disproportionately. High-speed traffic flows can lead to more severe crashes, as noted by Anderson et al. (2018), who found that speed limits designed to optimize flow often overlook safety, resulting in higher crash rates in high-speed zones.

Furthermore, Noland and Quddus (2004) highlighted that differential speeds among vehicles can contribute to RTCs, particularly in mixed traffic environments where trucks and cars interact. This suggests that maintaining a uniform speed across vehicles in a flow could mitigate some risks associated with RTCs. Sometimes, speed variation in vehicle speeds within the same lane can lead to crashes. Moreover, higher speeds generally contribute to more severe crashes due to the increased momentum and reduced reaction times.

2.3.2. Human Behavior

Driving behavior encompasses the intentional and unintentional characteristics, actions, attitude, and decisions of a driver. It is a complex psychological and physical process influenced by driver age, sex, experience, and the surrounding environment (Alavi, et al, 2017; Lefevre et al., 2014; and Lian et al., 2017). Furthermore, Drivers may be on the phone, talking, and becoming fatigued while driving, and the responses of different drivers are not similar, and even the same driver possibly make different decisions (Yeon et al., 2019; Moser et al., 2015). Hence, driving behavior is widely considered the most critical factor that directly affects both the occurrence and the severity of RTCs.

2.3.3. Weather conditions

The weather conditions, such as rain, fog, sunny, and snow, have enormous impacts on RTC severity and occurrence. It mainly affects the visibility and the friction of the road, such that they may change the driving behaviors (Ahmed and Ghasemzadeh, 2018) and vehicle stability (Yang et al., 2020). For example, the low visibility caused by heavy fog or rain heavily impacts the perception and control of vehicles (Ahmed and Ghasemzadeh, 2018; Fridman et al., 2019). Thus, these factors may lead to many causes for RTCs at different levels.

2.3.4. Roadways

The condition of roads has a significant impact on road traffic crashes. Several factors related to road condition can contribute to crashes, such as Surface quality, wetness, potholes, and debris can contribute to crash rates. (Zhang, X., & Wang, Y., 2015). Studies have shown a clear correlation between road conditions and crash rates. (Elvik, R. 2009, O'Neill, B., & amp; T. S., 2011, and Kopacz & Garcia, L. 2010. Thus, road conditions are essential for predicting the occurrence of RTC. In conclusion to the influencing factors, road traffic systems involve dynamic situations that change rapidly and unexpectedly. However. Roadway condition effects are having a more consistent effect on safety than traffic, so a systematic, holistic, and balanced approach must be adopted to reliably identify the causal or contributing factors for RTCs.

2.4. Data Challenges in Crash Prediction

2.4.1. Data Quality and Feature Engineering

The quality of data and the selection of relevant features are the major issues for the success of machine learning models in RTC prediction. Hence, traffic and crash data should be collected and recorded with due attention. Moreover, Xu et al. (2014) underlined the importance of feature selection and engineering, noting that inappropriate or redundant features could degrade model performance.

2.4.2. Handling Imbalanced Data

RTCs often involve imbalanced datasets, where severe crashes, including fatalities, are less frequent than minor ones. Rahman and Hasan (2018) explored various techniques to address this issue, such as the synthesis of Minority Over-Sampling Techniques (SMOTE) and cost-sensitive learning. This method has been shown to improve models' ability to predict rare but severe crashes, which are of particular interest in public safety.

2.4.3. Model Interpretability

Model interpretability is the degree to which humans can understand the reasoning behind an ML model's prediction. Even though ML models offer high predictive accuracy, their lack of interpretability remains a significant challenge. Lipton (2016) discussed the need for interpretable machine learning models, especially in high-stakes applications like traffic crash prediction.

2.5. Summary of Literature Review and Research Gap

From the literature review above, the Methods of RTC analysis have been categorized into two, namely statistical and Machine Learning models. The first category of methods is a statistical model, which has grown from a simple probability count to multiple discrete outcomes. On the other hand, the contributing factors to RTCs include weather conditions, traffic flow characteristics such as traffic volume, human factors, and pavement conditions have been reviewed.

The focus of many studies in the past has been on the impact factors of RTCs and their relationship with the spaciotemporal characteristics of the regional or national level. However, not many studies investigated the impacts of these factors at the corridor level for a short-term period or in real time. In addition, most studies have used a clustering method or traditional statistical models, such as regression models. However, advanced machine learning models, such as the random forest model and the dynamic neural network (DNN) model, have not been used extensively in the field of RTC analysis.

Therefore, this study focuses on developing a Road Traffic crash Severity Model, using advanced analytical techniques and ML algorithms. By examining patterns of vehicular crashes, we aim to unravel the intricacies of how traffic flow influences the Severity of crashes. Our analysis will incorporate data from historical crash reports and traffic flow to create a comprehensive picture of road safety in varying contexts.

The model will facilitate proactive measures, allowing traffic engineers and policymakers to devise targeted interventions, such as road design improvements and enhanced law enforcement strategies. By integrating safety considerations into traffic management practices, we can cultivate safer road environments that not only reduce the incidence of crashes but also promote sustainable mobility on an expressway.

In summary, this study aims to bridge the knowledge gap between traffic flow and safety analysis through sophisticated modeling techniques, finally contributing to the growth of safer transportation systems. The implications of our findings are expected to extend beyond academic inquiry, offering valuable insights for practitioners and stakeholders committed to developing safer roads for all users.

CHAPTER THREE

3. METHODOLOGY OF THE STUDY

3.1. Introduction

This Chapter outlines a quantitative research design to determine the significance and nature of relationships between variables (Traffic flow characteristics and RTC features), identify the influence of various features on crash severity and also use predictive modeling approach to develop a robust prediction model for crashes severity, and identify the most effective machine learning approach to provide actionable insights for enhancing road safety on the Addis Ababa Expressway. The study focused on traffic crash severity level as the target variable, which has represented four possible outcomes: Fatal, Serious Injury, Slight Injury, and Property Damage only

To achieve our objectives, we used an ML technique by implementing eight algorithms: Logistic Regression, Naive Bayes Classifier, Decision Tree Classifier, K-nearest neighbor, Support Vector Machine, AdaBoost, Random Forest Classifier, and XG Boost Classifier. Model performance was evaluated using previously established metrics.

3.2. Study Area

Addis Ababa- Adama Expressway is located in the eastern part of Addis Ababa City, Oromia Regional State. It has approximately a length of 78+000km. The express road starts from about 2km away from the University of South Africa, square ends at Welenchitye, the western part of Adama city. It is a toll road having six lanes and 31 total width. The road has 5 ramps that connect towns found between Addis Ababa to Adama City.

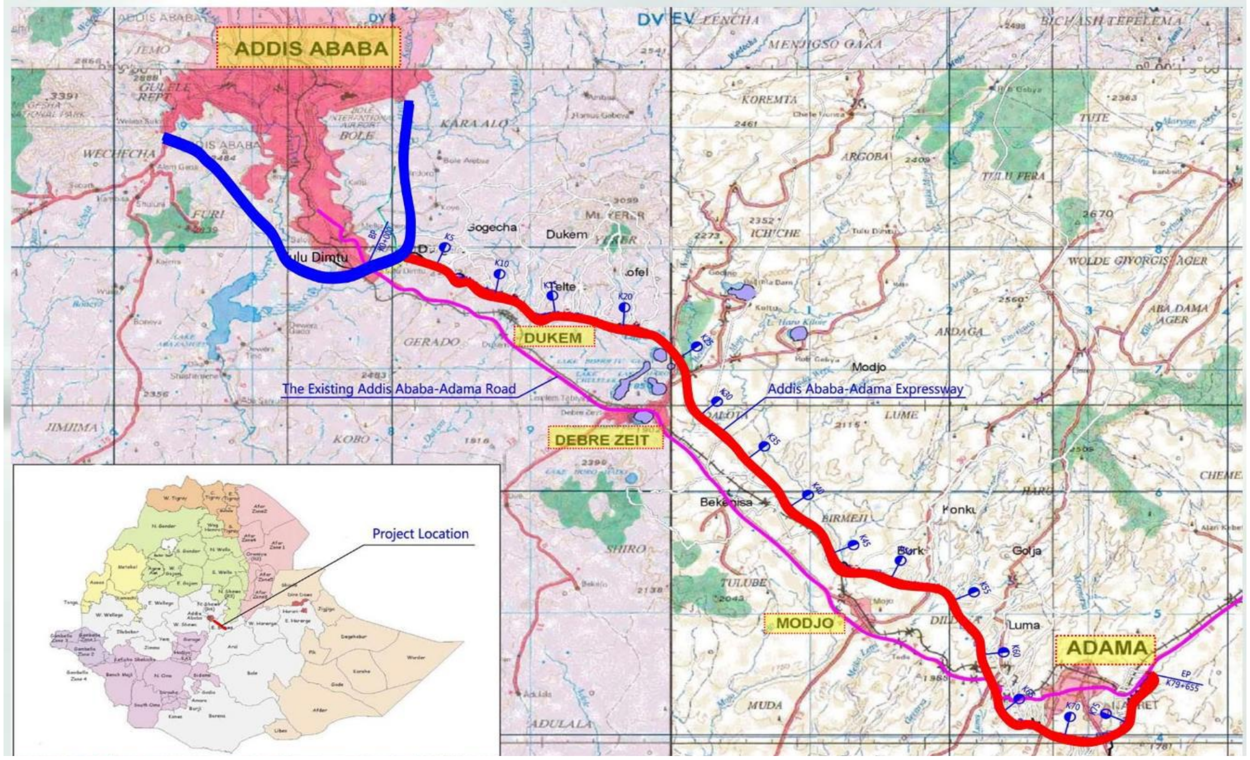


FIGURE 2: LOCATION OF THE STUDY AREA (ERA, 2017)

3.3. Methods of Data Collection

3.3.1. Data Collection

The study utilizes multiple dataset sources from ETRE: road traffic crash records and toll transaction dataset for traffic flow spans from January 1, 2015, to December 30, 2024, representing the ten years of data collection by ETRE. The RTC records provides detailed road crash data regarding the factors contributing to crash severity in the Addis Adama Expressway between 2015 and 2024, such information were the date and time of the crash, the crash location (in kilometers), the type of crash, the cause of the crash, the types of vehicles involved, crash severity, the gender, age of the drivers, road surface condition data, weather condition and visibility levels. The traffic dataset contains timestamps for entry and exit, the distance between interchanges, and the vehicle. The kilometer readings for crash locations were taken from kilometer markers on the median guardrail, positioned every 100 meters, with intermediate stations at 20-meter intervals marked on the median curbstone. Furthermore, road surface condition data, weather conditions, and visibility levels play a key role in crash severity assessment (Zhou et al., 2019). The integration of these diverse data allows for an all-inclusive analysis to develop the best crash severity prediction model. The data was cleaned, sorted, filtered, and organized into different variables and groups for this study.

3.3.2. Data Description

TABLE 1: DESCRIPTION OF VARIABLES

Variable Name	Description
Hour	Time of the crash (In 24-hour format)
Date	Day on which the crash occurred
Day of week	A day when a crash occurred (Monday - Sunday)
Age	The age of the driver
Age group of the driver	The age group of the driver
Sex of driver	Gender of driver
License level	The driving license of the driver
Educational level	Drivers' highest education level
Vehicle driver relation	What's the relation of a driver to the vehicle
Driving Experience	How many years of driving experience does the driver have
Types of Vehicles	What is the type of vehicle?
Owner of the Vehicle	Who's the owner of the vehicle
Service Year of the Vehicle	The last service year of the vehicle
Chainage	Locality of the crash site
Road Section	Are there any lanes or medians at the crash site?
Geometrical Condition	Road alignment with the terrain of the land
Direction	The direction of the road toward Addis OR Adama
Driving license issued Region.	The regional government office that issued a license to the driver
Road Surface Conditions	What was the condition of the road surface?
Light conditions	lighting conditions on the site
Weather Conditions	weather situation at the site of the crashes
Crash Type	What types of collision?
Number of Vehicles involved	Total number of vehicles involved in the crash
Number of casualties	Total number of casualties in a crash
Cause of the Crash	What was the cause of the crash?
Crash Type	Types of crashes that occurred
Driver at Fault	An crash due to the driver
Fatal	The number of fatalities in the crashes
Serious Injury	The number of Serious injuries in the crashes
Slight in jury	The number of Slight In jury from the crashes
Crash severity	How severe was the crash? (Target variable)
Entrance and Exit time stamp	The time in hrs. and minutes of a vehicle entry and exit to the interchange
Distance	The distance between interchanges

3.4. Data Preprocessing and Feature Engineering

3.4.1. Data Cleaning

Cleaning data is a critical step in the ML pipeline, during which the removal or fixing of missing features is done. The dataset may contain many unnecessary entries that directly impact the performance and reliability of the analysis. Before model utilization, data cleaning and preprocessing steps were conducted to ensure data quality and consistency. Missing data is a common problem and can arise for various reasons (e.g., data entry errors, system failures, non-response). The solution of such a problem required the application of parameters that directly factor for RTC. Missing data were addressed through imputation techniques, and duplicate or erroneous records were removed to prevent biases in model estimation. Transformation techniques, such as normalization and categorical encoding, are applied to standardize variables for machine learning analysis. These steps were crucial in preparing the data for subsequent model utilization in Python. Python was the primary programming tool used for data processing, model utilization, model calibration, and validation.

3.4.2. Data Merging and Preprocessing

Data Merging and preprocessing steps were crucial in preparing the data for subsequent model utilization in Python. The Pandas library in Python was used to process, combine, and align the traffic volume and RTC data, which are both time series, having different time intervals. The process of merging the different datasets involved using pandas. Merge function in Python's Pandas library. This function aligned traffic data and RTC data horizontally, combining them into a single, comprehensive dataset suitable for analysis. The merging process used the time variables of years, months, days, and hours to align it with the RTC data. The resulting dataset thus integrates traffic conditions with RTC information. This ensured that all relevant traffic data within the same hour of RTC occurrence was included.

3.4.3. Feature Selection

Feature engineering is performed to create meaningful explanatory variables such as time of day, day of the week, travel time, age group, experience range, and speed of the vehicle, which are incorporated to account for variations in crash occurrence over time. These enhancements improve the model's ability to detect important patterns and trends in crash data. These results show the complex relationship of various factors contributing to RTC Severity and provide valuable insights for further analysis and targeted interventions.

Categorical encoding is an important part of the data preprocessing pipeline. Label Encoding was applied across the entire data frame to convert categorical variables into numerical representations that ML algorithms can understand and process. As a result, the data (numerical inputs) were ready for model algorithms.

3.4.4. Handling Class Imbalance

Given the potential for an imbalanced distribution of crash severity classes, techniques such as oversampling (SMOTE) were employed to ensure the model gets sufficient exposure to the minority class to learn its patterns. SMOTE is applied before the initial train-test split to avoid the risk of data leakage, where the synthetic samples generated in the training set might be based on information present in the test set, leading to an overly optimistic evaluation. And also used a strict train-test separation for an unbiased final evaluation of the model's generalization.

3.5. Machine Learning Models

The selection of these ML algorithms for RTC severity prediction is a clear, comprehensive, and structured choice. We choose different types of ML models, including linear models such as Logistic Regression for baseline understanding, non-linear methods such as DT, KNN, SVM to capture complex relationships, and advanced ensemble techniques like AdaB, RF, XGB, known for their high predictive power. This diverse range allows for effective exploration of how different model architectures learn patterns, addressing various aspects of the bias-variance trade-off by including models prone to higher bias but lower variance, those with lower bias but higher variance, and vigorous ensemble methods that strike an optimal balance.

Moreover, this selection carefully considers the trade-off between ease of model understanding versus how well it predicts. It integrates highly transparent algorithms like Decision Trees and Logistic Regression, which are important for gaining actionable insights into the factors driving Crash severity, alongside more complex black-box yet powerful models such as XGBoost and Random Forest, which are geared towards maximizing prediction accuracy. This broad range of algorithms also provides varying degrees of robustness to different data characteristics, such as outliers, scaling issues, and high dimensionality, ensuring a resilient analysis across diverse data conditions and a comprehensive evaluation based on various metrics, finally maximizing the likelihood of identifying the most suitable and insightful model for the task.

Thus, based on our review of the previous studies, we selected eight machine learning models, i.e., LR, NB, DT, KNN, SVM, AdaB, RF, and XGB, as candidates for RTC severity prediction and to understand the factors influencing it.

3.6. Model Training and Validation

3.6.1. Data Splitting

The model training and validation were carried out by splitting the integrated dataset into training, validation, and testing subsets. primarily, the total data set was split into **80%** for training and **20%** for testing. Secondly, the split training data was further divided for training and validation. i.e., **20%** of the already-split training data was set aside for validation, and the remaining **80%** already-split training data was used for the actual training of the model. This data split was reproducible, which is crucial for consistent results.

3.6.2. Performance Metrics

Crash severity prediction is a classification task. It sorts the outcomes into different severity levels. We use standard metrics to measure the performance of the model which includes Accuracy, Precision, Recall and F1 score (Chen et al.,2015 and Tang et al.,2019).

Accuracy: shows the percentage of correctly predicted outcomes out of all predictions.

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

Precision: Precision measures the proportion of genuinely correct identifications, focusing on the quality of positive predictions.

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

Recall (Sensitivity): Recall, also known as sensitivity, quantifies the proportion of actual positive instances that were correctly identified by the model, indicating its ability to capture all relevant cases.

$$R = \frac{TP}{TP+FN}$$

F1-score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure that is particularly useful when dealing with imbalanced datasets. It is often the primary evaluation metric in such scenarios.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Where: TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative
The term is interpreted as follows: true positive means a correct positive result, true negative means a correct negative result, false positive means an incorrect positive result, and false negative means an incorrect negative result.

3.6.3. Model Evaluation

We evaluated the trained models using several common metrics. These include accuracy, precision, recall, and F1-score. We used these metrics to get an overall understanding of how well the models predicted. To understand what influenced the models' predictions, we used SHAP (Shapley Additive Explanations). This helped us identify the main features that affected the prediction of RTCs. We calculated these metrics using true positives, true negatives, false negatives false positives. The model's performance was measured based on accuracy, precision, recall, and F1-score (Chen et al.,2015 and Tang et al.,2019).

3.7. Software and Tools

Python was the primary programming tool used for data processing, model utilization, model calibration, and validation. The libraries/frameworks Scikit-learn, Pandas, NumPy, itertools, aborn, Sklearn, tracemalloc, and Matplotlib were used.

CHAPTER FOUR

4. RESULT AND DISCUSSION

4.1. Traffic flow and Road Traffic Crash analysis

The Ethiopian Toll Roads Enterprise (ETRE) collects 24-hour real-time toll traffic flow data year-round. Traffic flow and RTC analysis were visualized by advanced visualization techniques using Python, Matplotlib, and Seaborn modules to ensure a clear and meaningful depiction of the data.

4.1.1. Traffic flow pattern analysis and correlation with RTC Severity

The traffic pattern shown in Figure 2 below presents the number of vehicles using the Addis Ababa -Adama Expressway over 24 hours for the last ten year. The aggregate traffic flow for all vehicle types combined indicates, with low activity during late night/early morning (0-5 AM), a rise in the morning, and a significant surge in the late afternoon/evening, and drops off sharply after 9 PM. Note that V-1(Small automobiles cars, Jeep, Land rover, taxi and pick up), V-2(Minibus), V-3(Medium Bus and ISUZU), V-4(Big size Bus and Dump Trucks),V-5(Heavy Trucks and Trailers),V-6(Heavy Truck and trailers) and V-7(Heavy Truck and trailers)

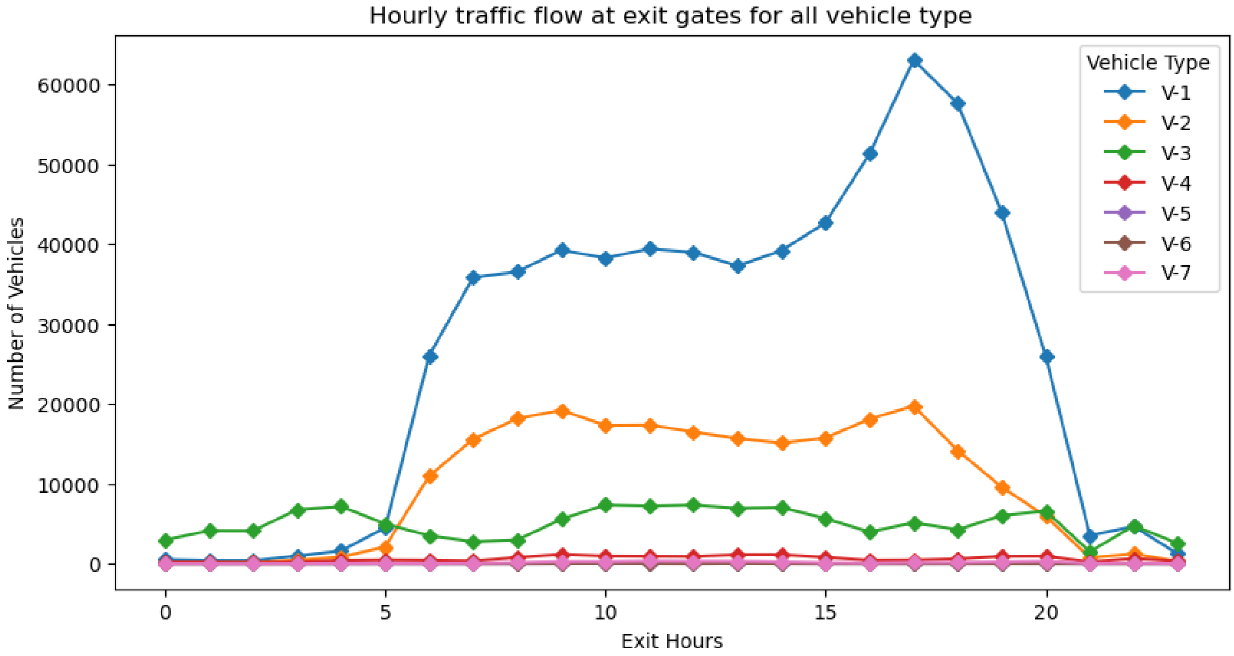


FIGURE 3: HOURLY TRAFFIC FLOW FOR ALL VEHICLE TYPES

The average speeds shown in Figure 4 are lowest in the very early morning and late night. They experience a sharp increase to reach their highest levels during the late morning and early afternoon 6 to 16 hours, after which they gradually decline through the evening.

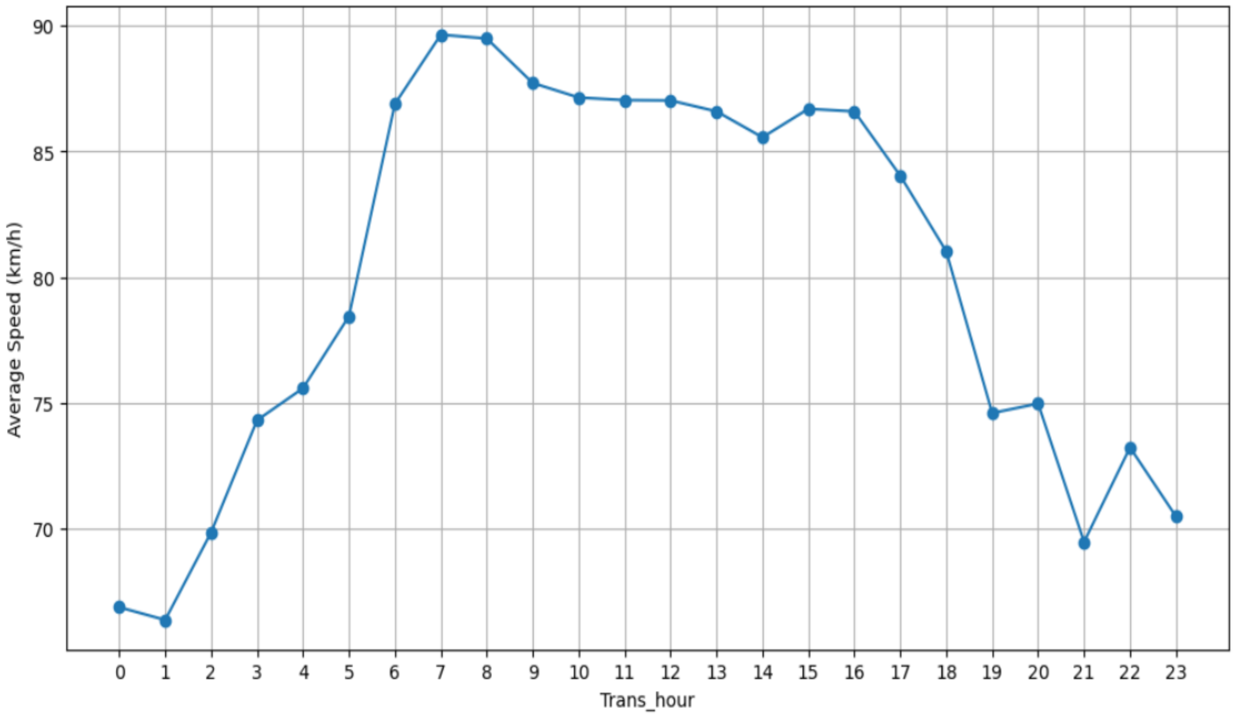


FIGURE 4: AVERAGE SPEED(KM/H) ACROSS THE DAY

The distribution of vehicle types across the expressway speed ranges depicted in Figure 5 indicates vehicle types 1.0, 2.0, and 3.0 account for the vast majority of observed vehicles, predominantly traveling within the 75-150 mph speed ranges. Conversely, Vehicle Types 4.0 through 7.0 exhibit significantly lower vehicle counts, with differing dominant speed ranges and overall, less presence, suggesting they might be specialized or less common vehicle types. Generally, extreme low (20-25 mph) and high (175-250 mph) expressway speed ranges show minimal vehicle activity across all types, highlighting that most vehicles operate within a moderate to high-speed spectrum typical of expressways.

Note that V-1(Small automobiles cars, Jeep, Land rover, taxi and pick up), V-2 (Minibus), V-3 (Medium Bus and ISUZU), V-4 (Big size Bus and Dump Trucks), V-5(Heavy Trucks and Trailers), V-6 (Heavy Truck and trailers) and V-7 (Heavy Truck and trailers)

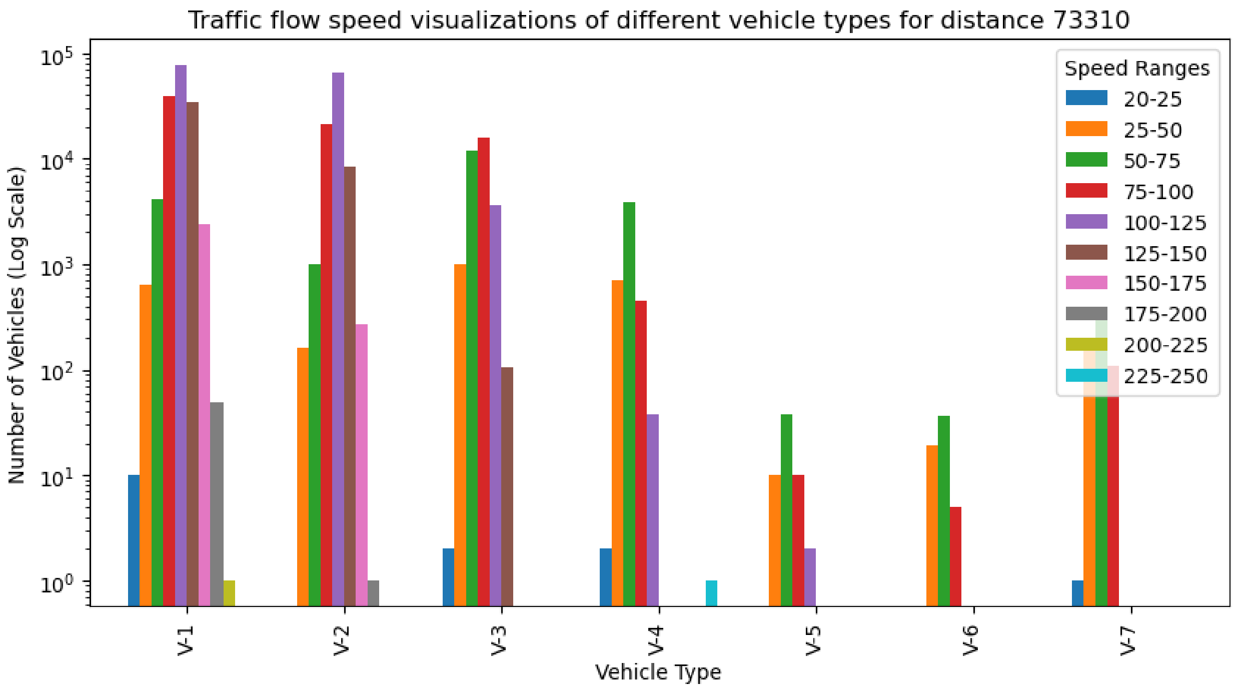
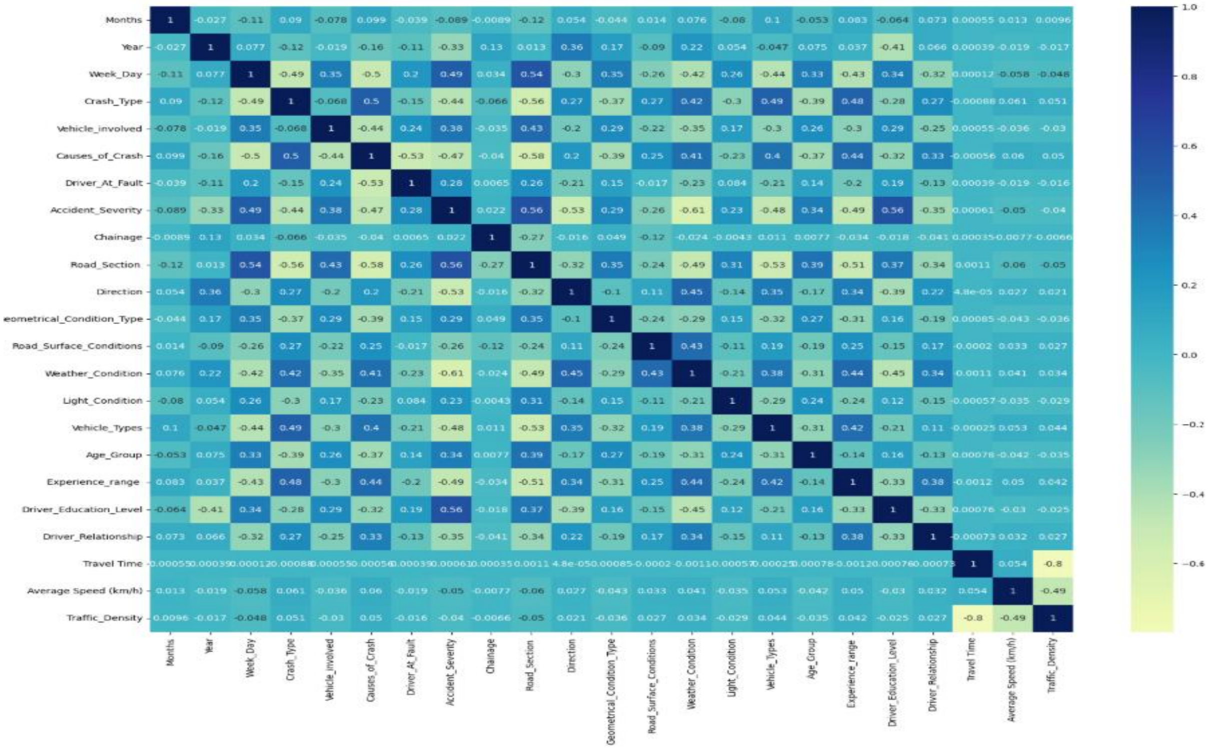


FIGURE 5 : DISTRIBUTION OF SPEED RANGES AND VEHICLE TYPE

The correlation between crash severity and various traffic flow characteristics, such as average speed (0.05), traffic density (0.0007), and travel time (0.057), was found to be generally weak, as shown in Figure 6, which means their variations have a negligible effect on the severity of an Crash.



	Months	Year	Week_Day	Crash_Type	Vehicle_involved	Causes_of_Crash	Driver_At_Fault	Accident_Severity	Chainage	Road_Section	Direction	Geometrical_Condition_Type	Road_Surface_Conditions	Weather_Condition	Light_Condition	Vehicle_Types	Age_Group	Experience_range	Driver_Education_Level	Driver_Relationship	Travel Time	Average Speed (km/h)	Traffic_Density
Months	1	-0.027	-0.11	0.09	-0.078	0.099	-0.039	-0.089	-0.0089	-0.12	0.054	-0.044	0.014	0.076	-0.08	0.1	-0.053	0.083	-0.064	0.073	0.00055	0.013	0.0096
Year	-0.027	1	0.077	-0.12	-0.019	-0.16	-0.11	-0.33	0.13	0.013	0.36	0.17	-0.09	0.22	0.054	-0.047	0.075	0.037	-0.41	0.066	0.00039	-0.019	-0.017
Week_Day	-0.11	0.077	1	-0.49	0.35	-0.5	0.2	0.49	0.034	0.54	-0.3	0.35	-0.26	-0.42	0.26	-0.44	0.33	-0.43	0.34	-0.32	0.00012	-0.058	-0.048
Crash_Type	0.09	-0.12	-0.49	1	-0.068	0.5	-0.15	-0.44	-0.066	-0.56	0.27	-0.37	0.27	0.42	-0.3	0.49	-0.39	0.48	-0.28	0.27	-0.00088	0.061	0.051
Vehicle_involved	-0.078	-0.019	0.35	-0.068	1	-0.44	0.24	0.38	-0.035	0.43	-0.2	0.29	-0.22	-0.35	0.17	-0.3	0.26	-0.3	0.29	-0.25	0.00055	-0.036	-0.03
Causes_of_Crash	0.099	-0.16	-0.5	0.5	-0.44	1	-0.53	-0.47	-0.04	-0.58	0.2	-0.39	0.25	0.41	-0.23	0.4	-0.37	0.44	-0.32	0.33	-0.00056	0.06	0.05
Driver_At_Fault	-0.039	-0.11	0.2	-0.15	0.24	-0.53	1	0.28	0.0065	0.26	-0.21	0.15	-0.017	-0.23	0.084	-0.21	0.14	-0.2	0.19	-0.13	0.00039	-0.019	-0.016
Accident_Severity	-0.089	-0.33	0.49	-0.44	0.38	-0.47	0.28	1	0.022	0.56	-0.53	0.29	-0.26	-0.61	0.23	-0.48	0.34	-0.49	0.56	-0.35	0.00061	-0.05	-0.04
Chainage	-0.0089	0.13	0.034	-0.066	-0.035	-0.04	0.0065	0.022	1	-0.27	-0.016	0.049	-0.12	-0.024	-0.0043	0.011	0.0077	-0.034	-0.018	-0.041	0.00035	-0.0077	-0.0066
Road_Section	-0.12	0.013	0.54	-0.56	0.43	-0.58	0.26	0.56	-0.27	1	-0.32	0.35	-0.24	-0.49	0.31	-0.53	0.39	-0.51	0.37	-0.34	0.0011	-0.06	-0.05
Direction	0.054	0.36	-0.3	0.27	-0.2	0.2	-0.21	-0.53	-0.016	-0.32	1	-0.1	0.11	0.45	-0.14	0.35	-0.17	0.34	-0.39	0.22	4.8e-05	0.027	0.021
eometrical_Condition_Type	-0.044	0.17	0.35	-0.37	0.29	-0.39	0.15	0.29	0.049	0.35	-0.1	1	-0.24	-0.29	0.15	-0.32	0.27	-0.31	0.16	-0.19	0.00085	-0.043	-0.036
Road_Surface_Conditions	0.014	-0.09	-0.26	0.27	-0.22	0.25	-0.017	-0.26	-0.12	-0.24	0.11	-0.24	1	0.43	-0.11	0.19	-0.19	0.25	-0.15	0.17	-0.0002	0.033	0.027
Weather_Condition	0.076	0.22	-0.42	0.42	-0.35	0.41	-0.23	-0.61	-0.024	-0.49	0.45	-0.29	0.43	1	-0.21	0.38	-0.31	0.44	-0.45	0.34	0.0011	0.041	0.034
Light_Condition	-0.08	0.054	0.26	-0.3	0.17	-0.23	0.084	0.23	-0.0043	0.31	-0.14	0.15	-0.11	-0.21	1	-0.29	0.24	-0.24	0.12	-0.15	-0.00057	-0.035	-0.029
Vehicle_Types	0.1	-0.047	-0.44	0.49	-0.3	0.4	-0.21	-0.48	0.011	-0.53	0.35	-0.32	0.19	0.38	-0.29	1	-0.31	0.42	-0.21	0.11	-0.00025	0.053	0.044
Age_Group	-0.053	0.075	0.33	-0.39	0.26	-0.37	0.14	0.34	0.0077	0.39	-0.17	0.27	-0.19	-0.31	0.24	-0.31	1	-0.14	0.16	-0.13	0.00078	-0.042	-0.035
Experience_range	0.083	0.037	-0.43	0.48	-0.3	0.44	-0.2	-0.49	-0.034	-0.51	0.34	-0.31	0.25	0.44	-0.24	0.42	-0.14	1	-0.33	0.38	-0.0012	0.05	0.042
Driver_Education_Level	-0.064	-0.41	0.34	-0.28	0.29	-0.32	0.19	0.56	-0.018	0.37	-0.39	0.16	-0.15	-0.45	0.12	-0.21	0.16	-0.33	1	-0.33	0.00076	-0.03	-0.025
Driver_Relationship	0.073	0.066	-0.32	0.27	-0.25	0.33	-0.13	-0.35	-0.041	-0.34	0.22	-0.19	0.17	0.34	-0.15	0.11	-0.13	0.38	-0.33	1	-0.00073	0.032	0.027
Travel Time	0.00055	0.00039	0.00012	0.00088	0.00055	0.00056	0.00039	0.00061	0.00035	0.0011	4.8e-05	0.00085	-0.0002	-0.0011	0.00057	0.00025	0.00078	-0.0012	0.00076	0.00073	1	0.054	-0.8
Average Speed (km/h)	0.013	-0.019	-0.058	0.061	-0.036	0.06	-0.019	-0.05	-0.0077	-0.06	0.027	-0.043	0.033	0.041	-0.035	0.053	-0.042	0.05	-0.03	0.032	0.054	1	-0.49
Traffic_Density	0.0096	-0.017	-0.048	0.051	-0.03	0.05	-0.016	-0.04	-0.0066	-0.05	0.021	-0.036	0.027	0.034	-0.029	0.044	-0.035	0.042	-0.025	0.027	-0.8	-0.49	1

FIGURE 6 : CORRELATION HEAT MAP FOR CRASH SEVERITY AND TRAFFIC CHARACTERISTICS

4.1.2. Road Traffic Crash Severity Analysis

Figure 7 illustrates a clear diurnal pattern in the frequency of crashes. The first peak occurs during the morning rush hour, between 6 a.m. and 7 a.m., and the second, more significant peak, occurs around lunchtime, specifically between 12 p.m. and 1 p.m. After this daytime peak, the number of crashes drops dramatically and remains low throughout the afternoon and evening, with the fewest crashes occurring in the late-night hours. Thus, crashes are most frequent during typical daytime and commuting hours, with a prominent peak around lunchtime. Conversely, the number of crashes is at its lowest during the late night and very early morning hours.

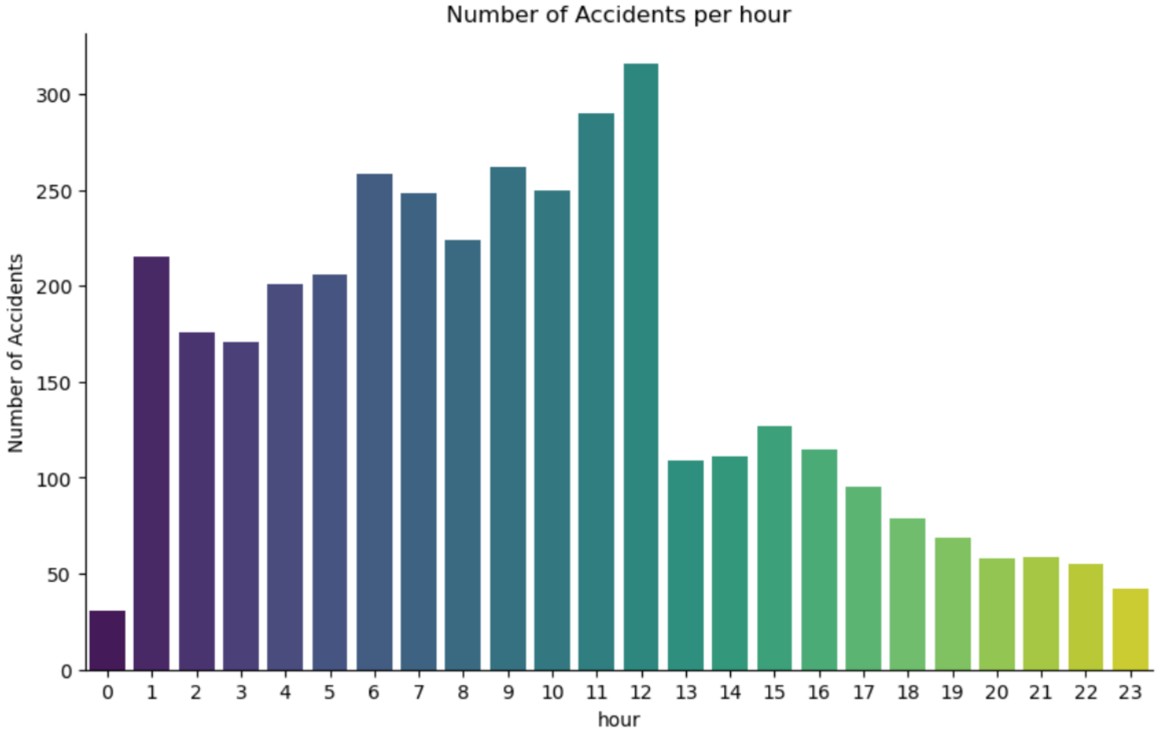


FIGURE 7: CRASH FREQUENCY BY HOURS OF THE DAY

Figure 8 depicts the distribution of crashes throughout the week. The number of crashes is generally high throughout the week, particularly very high on Tuesday and Saturday. There is a slight decrease on Friday and a more noticeable dip on Monday. Weekend crash numbers (Saturday and Sunday) are moderate to high, with Sunday seeing a considerable decrease compared to Saturday, approaching the levels of mid-week high-crash days. Thus, the pattern could suggest different traffic patterns, driver behavior, or other factors influencing crashes on specific days.

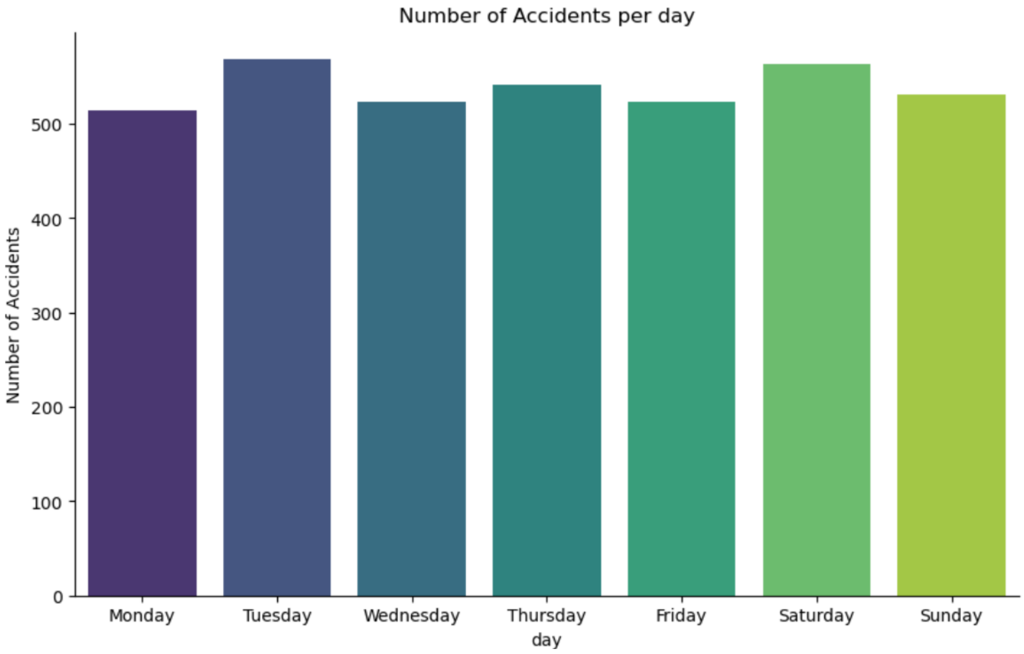


FIGURE 8: CRASH DISTRIBUTION BY DAY OF THE WEEK

Figure 9 clearly depicts the seasonal trend of the number of crashes, that is, the number of crashes is relatively low during the winter and spring months, with the lowest point in February at slightly above 250 crashes. The number of crashes gradually increases from March to May and shows a significant rise in June and July. The peak in crashes occurs during the summer months of July and August, with July having the highest number of crashes, approaching 400. The number of crashes begins to decline in September and continues to decrease through the fall, before slightly leveling off in November and December. Thus, the number of crashes during the summer season is significantly higher, with a peak in July and August, conversely, lowest in the winter and early spring.

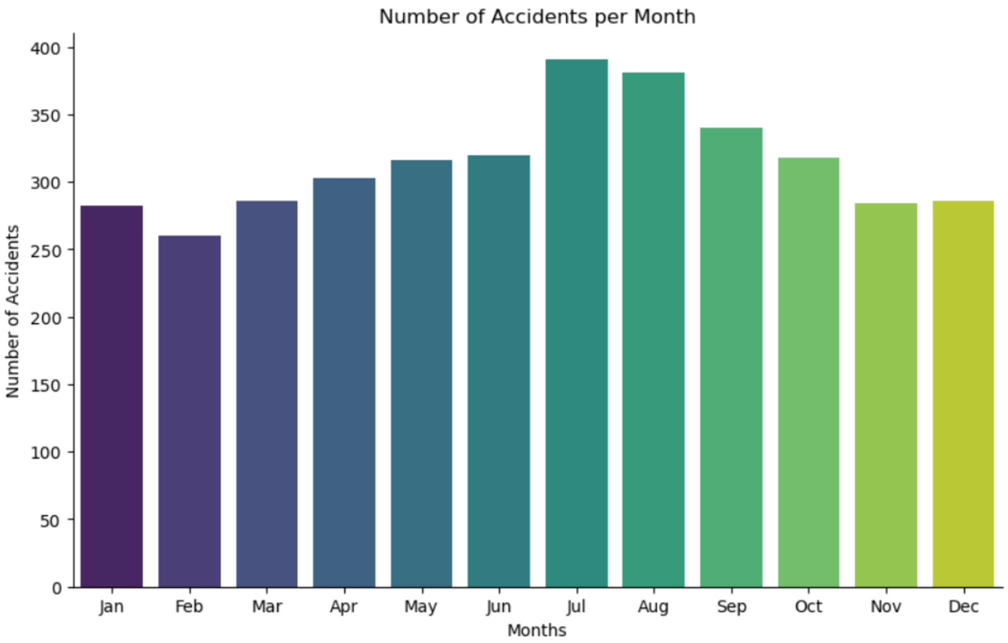


FIGURE 9: CRASH DISTRIBUTION BY MONTH OF THE YEAR

Figure 10 shows that the most RTCs happened in a good weather. Bad weather conditions like heavy wind, severe winds, snow, and fog are considered as dangerous. However, they make up a small part all crashes in the dataset. This may because drivers are more careful in bad.

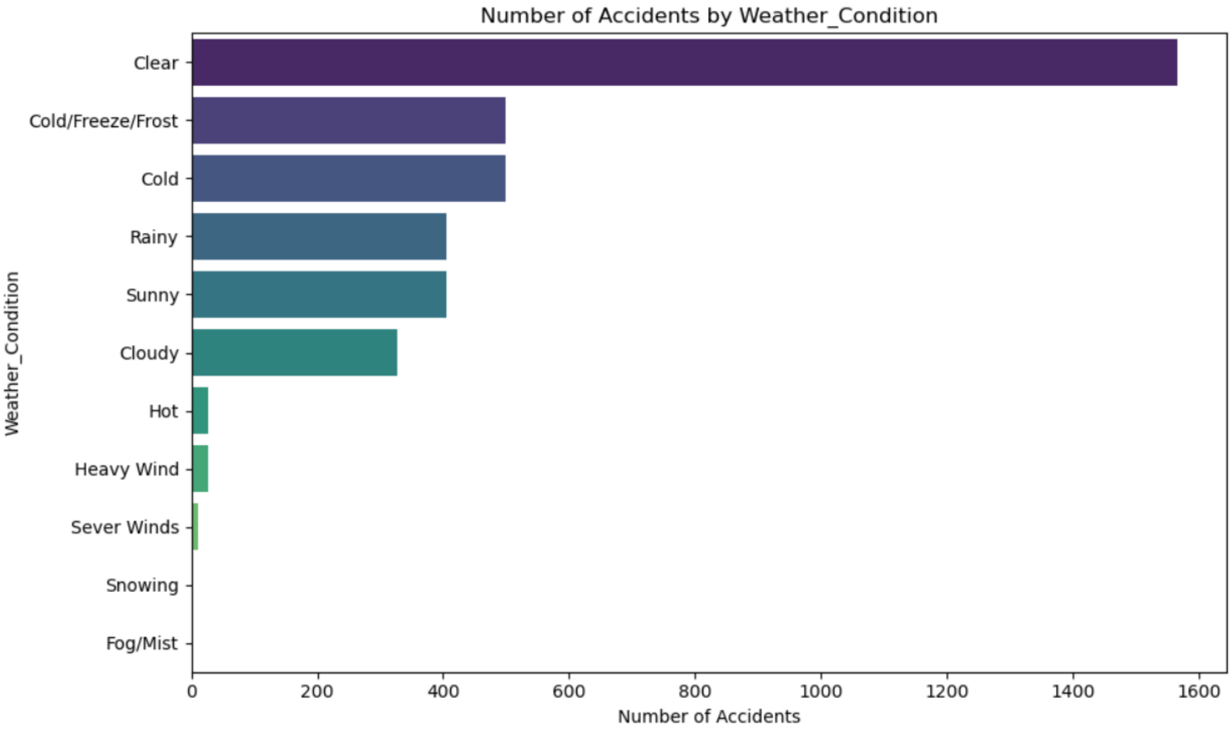


FIGURE 10: CRASH FREQUENCY BY WEATHER CONDITION

This distribution likely reflects traffic patterns with more vehicles on the road during the day, and potentially varying levels of driver caution or visibility challenges in different light conditions. The vast majority of crashes happen during Daylight hours, as shown in Figure 10, while darkness is the second highest contributor; its count is significantly lower than that of daylight crashes. Conditions around sunrise and sunset contribute very little to the total number of crashes. Despite the lower numbers, crashes in darkness, especially without streetlights, can often be more severe due to reduced visibility.

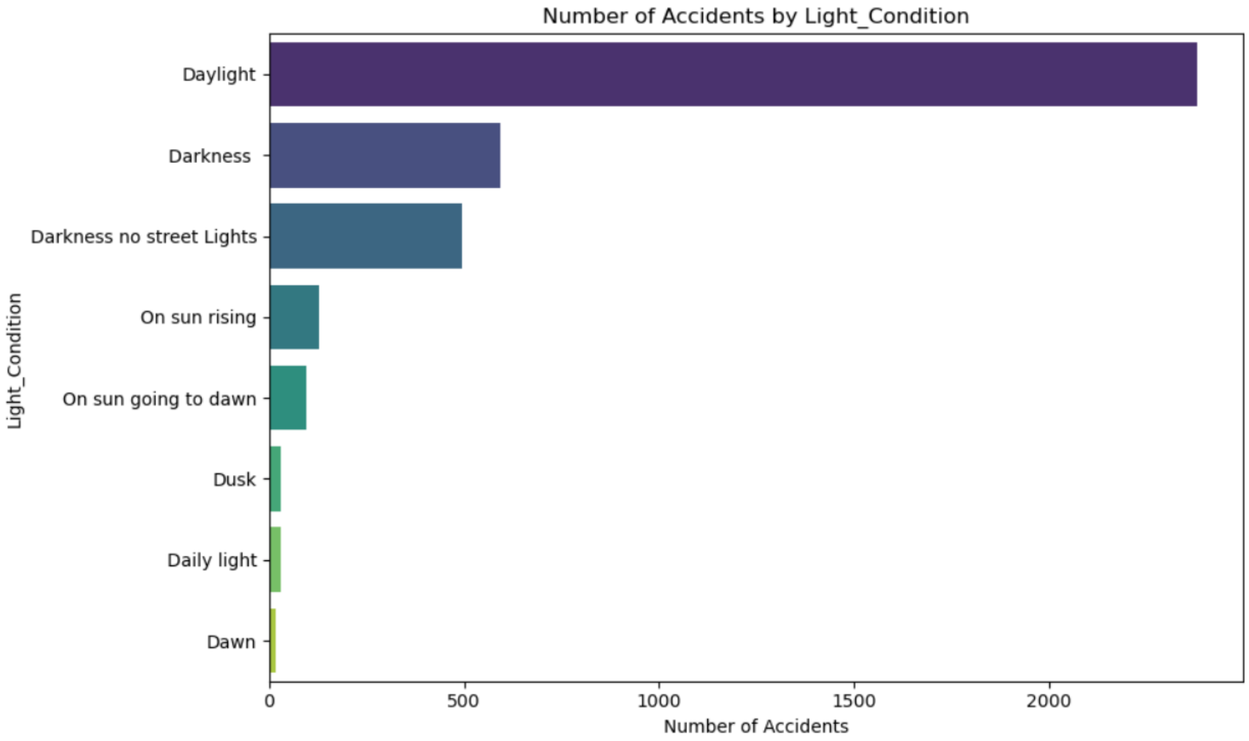


FIGURE 11: CRASH DISTRIBUTION BY LIGHTING CONDITION

Figure 12 shows that most crashes, over 1600, occurs on straight roads. This result is way higher than any of the other road conditions. The next highest number of crashes occurred on sloppy roads, with over 600. crashes on Sloppy or Downhill, Uphill, and curve roads occurred low in number, with the lowest number of crashes, around 400, happening on curve roads. Thus, contrary to what might be expected, the majority of crashes in this dataset occurred on straight roads. This suggests that factors other than road geometry, such as inattention, fatigue, or speeding, are likely the primary contributors to crashes.

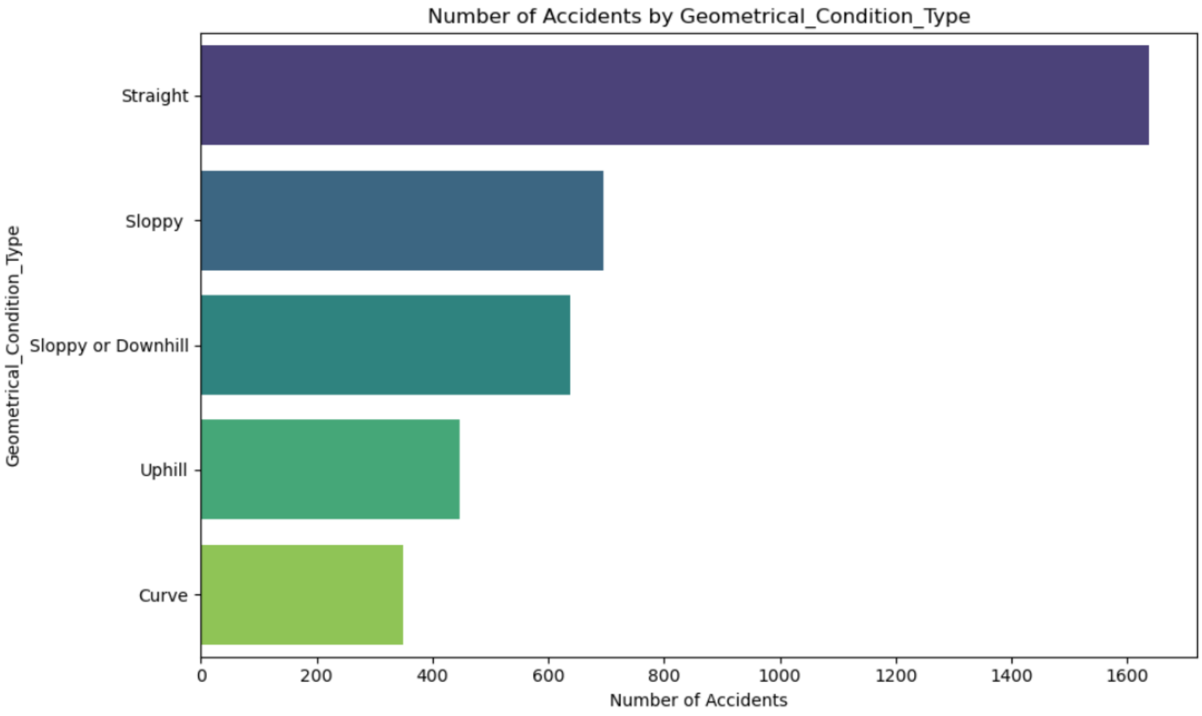


FIGURE 12: CRASH DISTRIBUTION BY GEOMETRIC CONDITION

The data in Figure 13 shows that the most common crash type is a crash with fixed object. Several crashes of this type exceed 1600. Next rear-end collisions are common. They happen more than 800 times. Rollover crashes are the third most frequent. There are slightly over 500 crashes. Close behind is overturning. Other crash types such as side impact collisions, head-on collisions, and crashes with animals, all have a very lower frequent, with lower than 200 crashes each. The least common crashes are crash with parked vehicle, vehicle on fire, and out of road, each with less than 100 crashes. Thus, crashes with a fixed objects are the most frequent crash type in the dataset. This indicate that most percentage of crashes are single-vehicle collision with fixed objects rather than another vehicle on expressway.

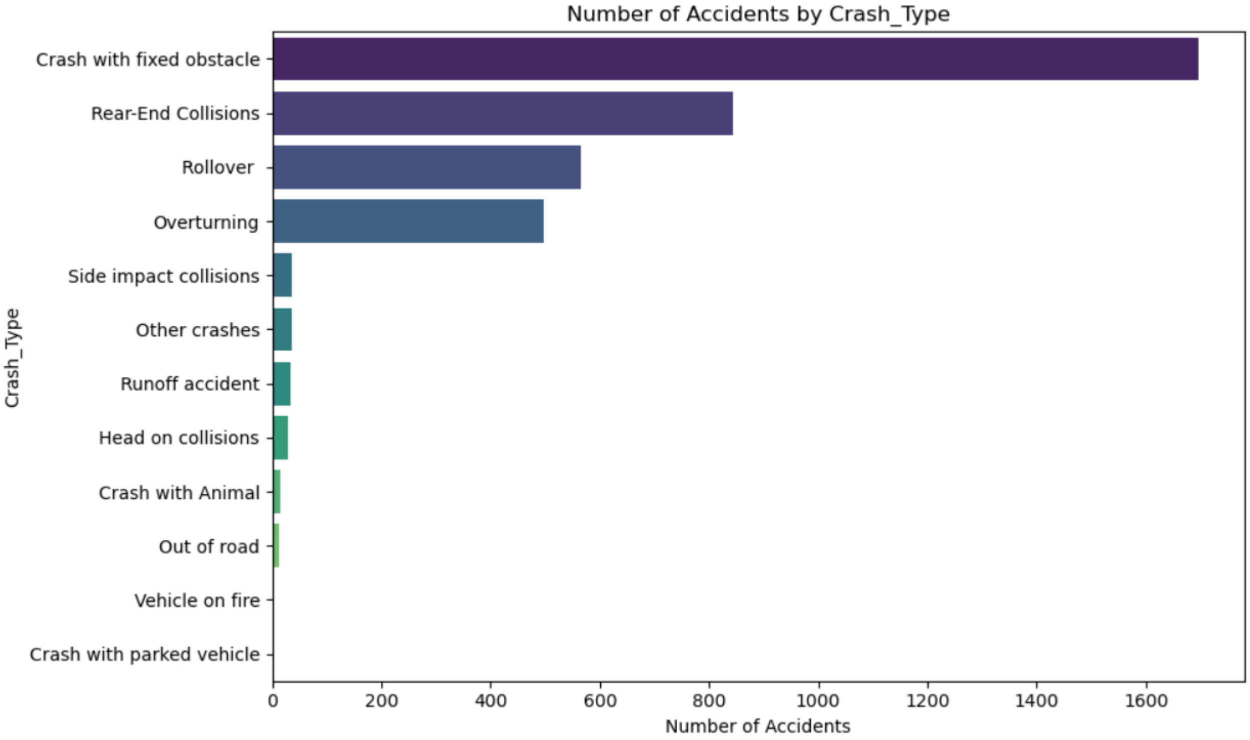


FIGURE 13: CRASH DISTRIBUTION BY CRASH TYPE

Figure 14 presents the causes of crashes in descending order of their frequency, allowing for quick identification of the most prevalent factors. Driver Error is overwhelmingly the leading cause of crashes. This highlights human error as the most significant contributor to crashes.

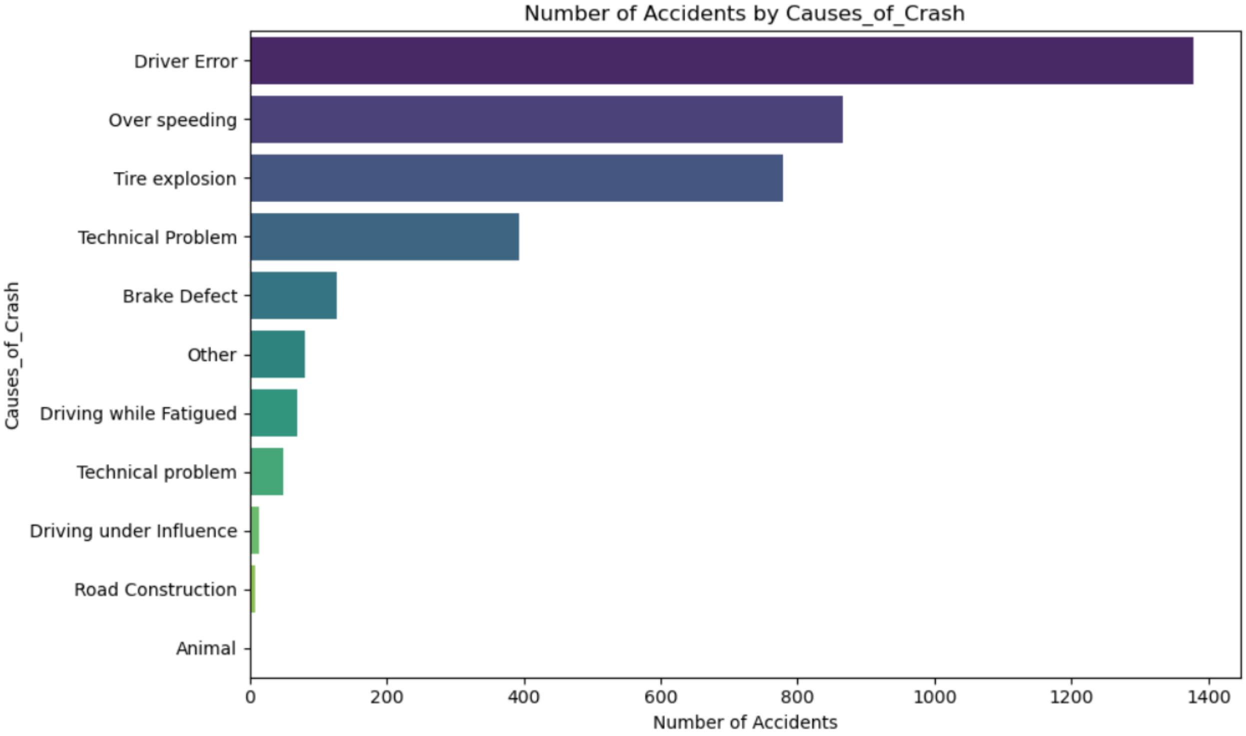


FIGURE 14: FREQUENCY OF CRASHES BY CAUSE OF CRASHES

Figure 15 shows that V-1(Small automobiles cars, Jeep, Land rover, taxi and pick up) is the dominant vehicle type involved in crashes with 43.2% of the total. The next largest group is V3(Medium Bus and ISUZU) with 23.2%, followed by V4 (Big size Bus and Dump Trucks) at 13.1% and V7(Heavy Truck and trailers) and above at 9.8%. The smallest shares are V2 (Minibus) and V6 (Heavy Truck and Trailers) with 9.7% and 0.4% respectively. The labels V5(Heavy Trucks and Trailers) and V6 appear on very small portions. Vehicle type V1 is highly involved in Crashes. It is almost nearly half of all crashes in the dataset. This shows that vehicle type V1 is a major contributor to RTC as compared to all others, which are involved in significantly fewer crashes.

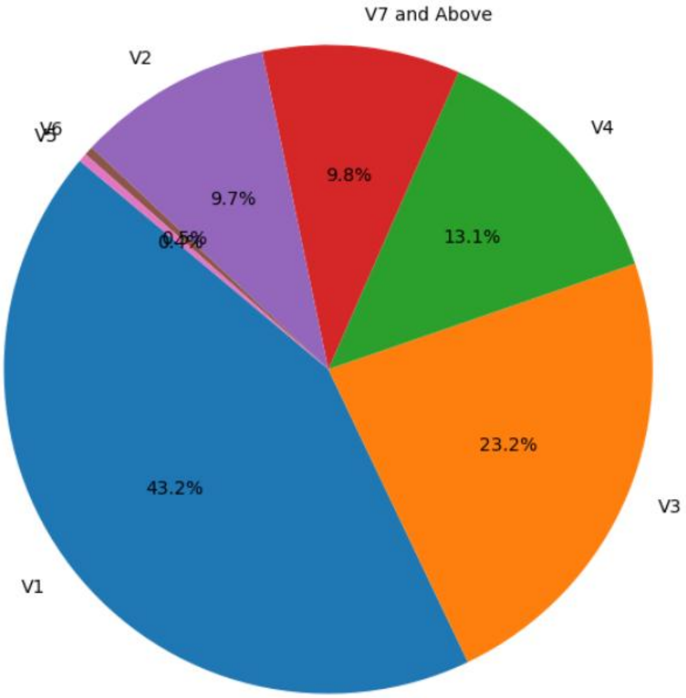


Figure 15: Frequency of Crashes by Vehicle type

Figure 16 shows the number of crashes for each age group. This describes the different age group involvement on RTCs. The middle-aged drivers (31-50) are involved in the highest number of crashes. Whereas, young adult drivers (18-30) are the next age groups. Drivers over 50 appear to be involved in fewer crashes. The distribution could be influenced by many factors such as driving experience, exposure, risk-taking behavior, or physical capabilities. The data points towards targeting crash prevention efforts more seriously towards the 31-50 and 18-30 age groups.

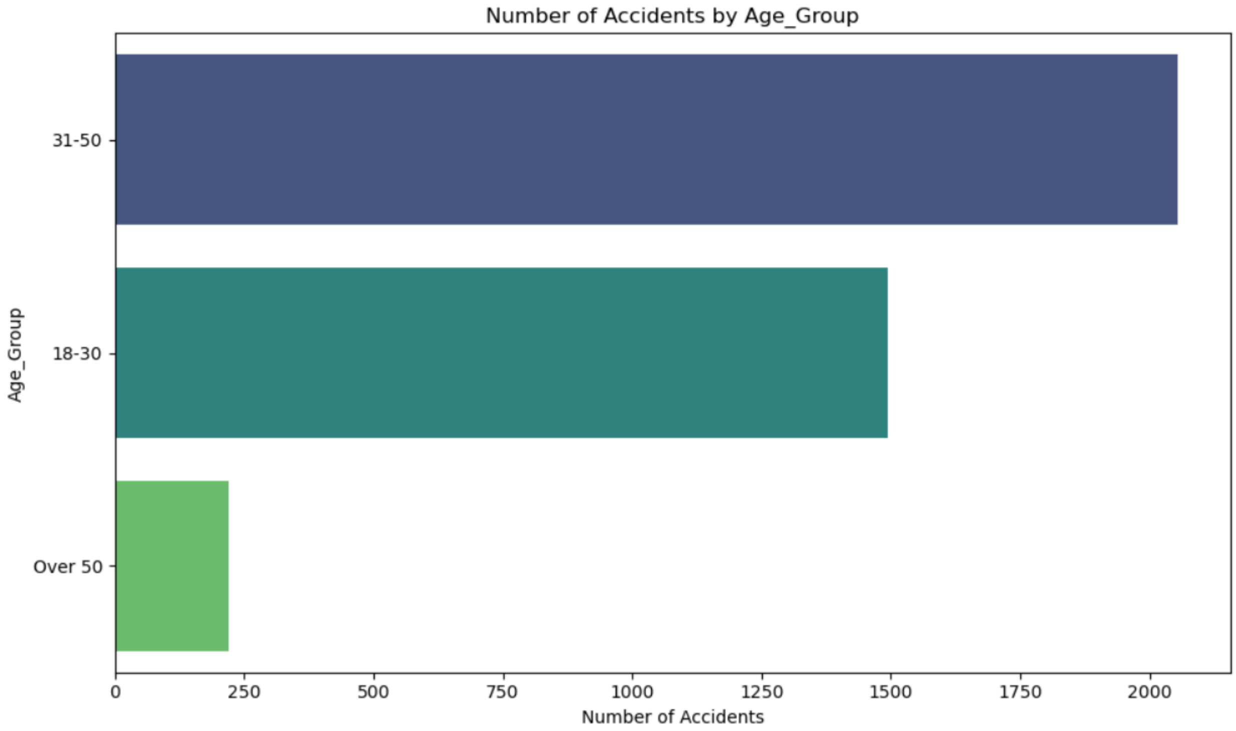


FIGURE 16: FREQUENCY OF CRASHES BY AGE GROUP

Figure 17 shows the crash severities in a comparative way, ordered from the most frequent to the least frequent. By far, the most common outcome of crashes is Suspected Slight Injury. This is followed by Property Damage only, indicating that a substantial portion of crashes result in material damage rather than injuries. Suspected Serious Injury cases are considerably less frequent. Fatal injury is the rarest outcome among the listed severities. This is a critical metric, despite its lower frequency. The figure intensely illustrates that crashes primarily result in slight injuries and property damage. Serious injuries and fatal outcomes are considerably less frequent. This distribution suggests that while efforts to prevent all types of crashes are important, a large proportion of incidents might not involve direct harm to individuals. However, the presence of fatal injuries, even in small numbers, indicates the most safety aspects of crash prevention.

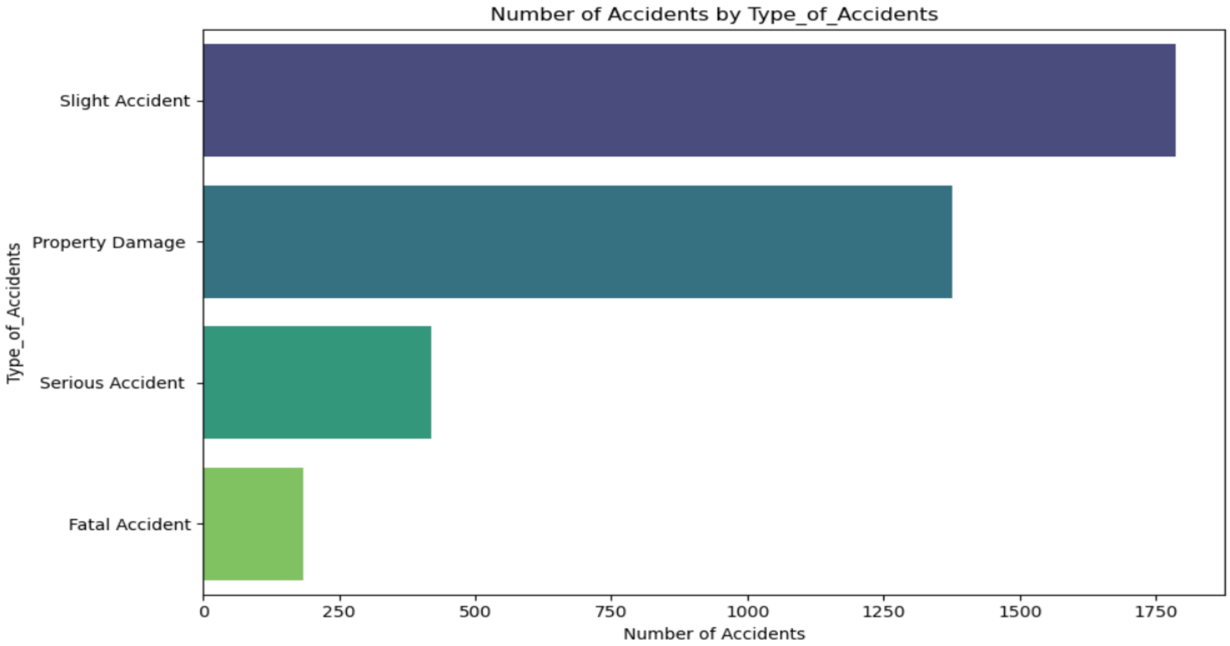


Figure 17 : Frequency of Crashes by Crash severity

4.2. Model Performance Evaluation

A suite of machine learning models, including Logistic Regression, Naive Bayes Classifier, Decision Tree Classifier, K-Nearest Neighbor, Support Vector Machine, AdaBoost, Random Forest Classifier, and XGBoost classifier, was trained and evaluated. The performance of each model was assessed using a range of metrics relevant to multi-class classification, including accuracy, precision, recall, and F1-score for each severity class, as well as a macro-averaged or weighted average for overall performance.

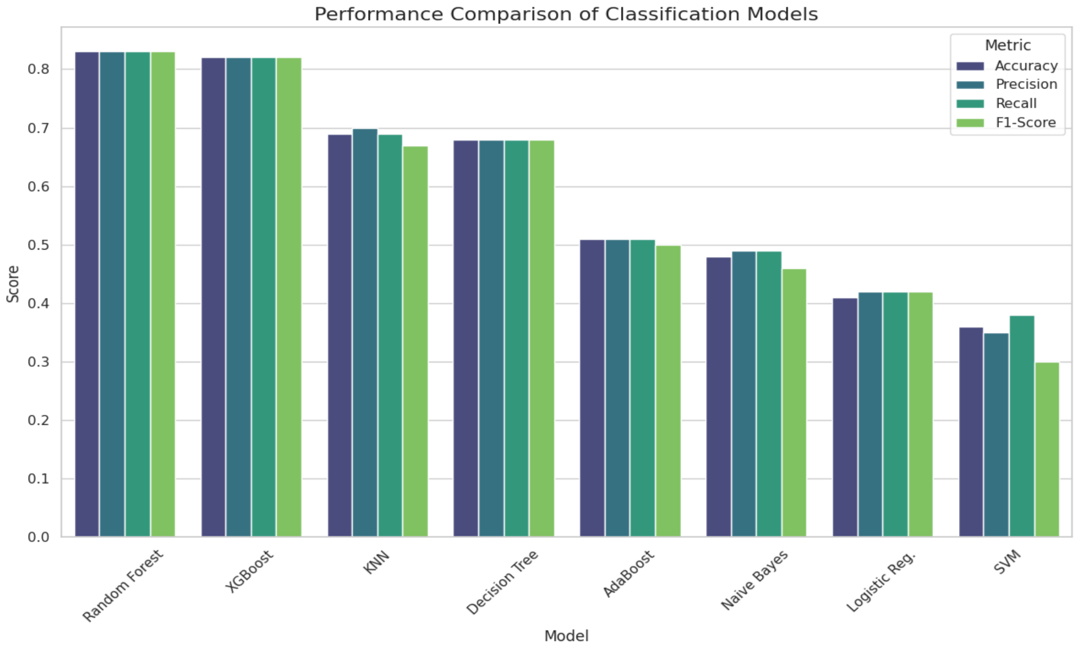


Figure 18: The comparison of ML for Crash Severity using all metrics.

From the analysis, Random Forest and XG Boost performed better than other models. According to Table 3, the Random Forest model had the highest accuracy at 83.0%, this shows that RF are good at correctly classify crash severity levels across different classes. At one end, the Support Vector Machine (SVM) model showed poor performance, with accuracy of 36.1%. this indicates it is not effective without major tuning or other methods. K-Nearest Neighbor (KNN), with an accuracy of 69.3%, and decision tree (DT), with an accuracy of 67.6%, proved themselves as moderate contender. Then comes Adaboost (AdaB), Naive Bayes (NB), and Logistic Regression (LR) models, having an accuracy of (51.2%), (48.4%), and (41.6%) respectively, suggesting a lowly accurate and unwell-balanced classifier. These models demonstrated fault classification across all severity levels, predicting every case with less precision and recall.

Table 2 : Performance Metrics of Different Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.83	0.829	0.831	0.828
XGBoost	0.821	0.821	0.821	0.821
KNN	0.693	0.704	0.694	0.666
Decision Tree	0.676	0.676	0.677	0.675
AdaBoost	0.512	0.505	0.512	0.5
Naive Bayes	0.484	0.491	0.488	0.463
Logistic Reg.	0.416	0.416	0.416	0.416
SVM	0.361	0.349	0.376	0.299

4.3. Feature Importance Analysis

Figure 18 shows the importance of different features in the model. It helps us see which variables matter most. The feature with the highest importance is chainage, with a score around 0.145. Next are the months and the crash type. These features have the biggest influence on the predictions. The location along a road, the time of year, and the kind of crash are the strongest and most important predictors. Similarly, weather conditions, causes of crash, and weekday have a moderate level of importance with scores ranging from about 0.06 to 0.07. The driver relationship, sex, and road surface conditions are less important, having scores below 0.02. This means these factors have an impact on the model's result.

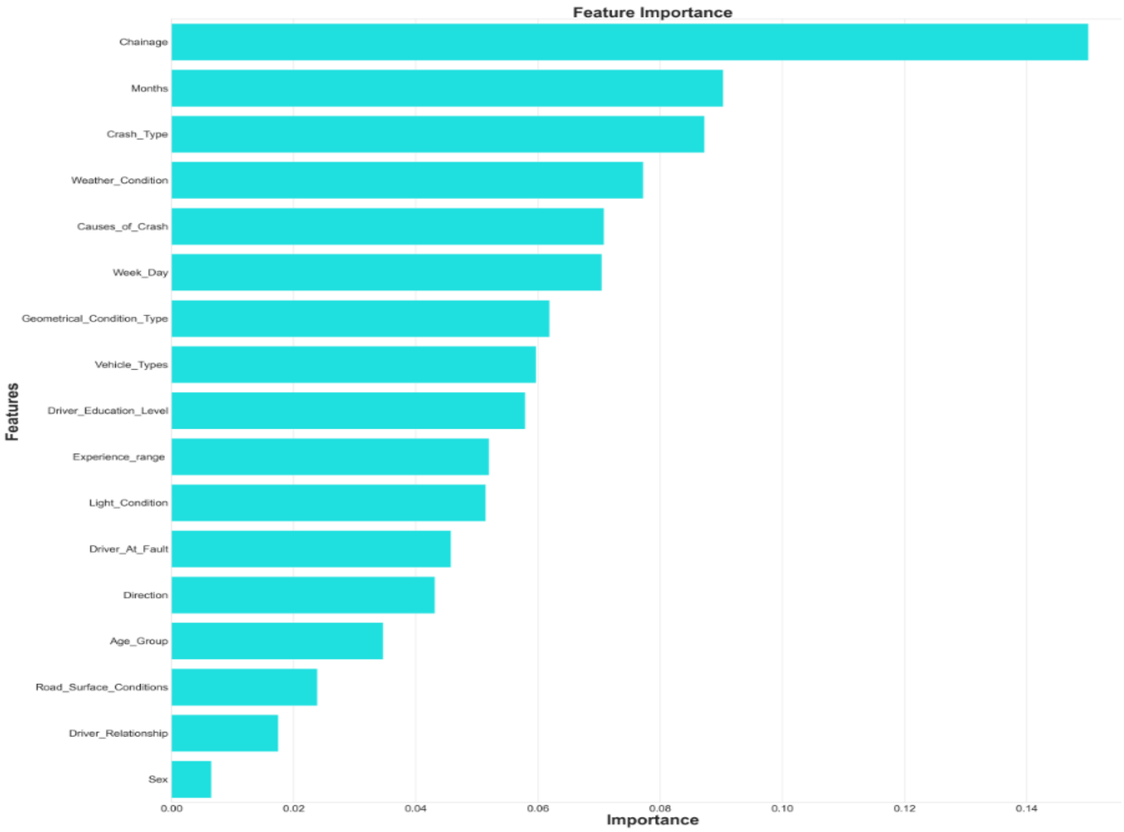


FIGURE 19: FEATURES IMPORTANCE OF RTC PREDICTION FOR RANDOM FOREST

This feature importance table 3 provides critical insights into which factors are most influential in determining the outcome that the model is predicting. It clearly shows that the Spatial factor chainage, the Temporal factor months, and the crash type are overwhelmingly the most significant predictors. Environmental factors like weather conditions and Demographic factors like driver education and Experience range have much less predictive power among the first ten predictive factors in this particular model.

TABLE 3: TOP 10 MOST IMPORTANT FEATURES FOR CRASH SEVERITY PREDICTION.

Rank	Features	Importance Score
1	Chainage	0.1497
2	Months	0.0898
3	Crash Type	0.0879
4	Weather conditions	0.0770
5	Causes of Crash	0.0707
6	Weekday	0.0701
7	Geometrical Condition Type	0.0615
8	Vehicle Types	0.0607

9	Driver_Education_Level	0.0591
10	Experience range	0.0523

4.4. Prominent Contributing Factors of RTC severity (SHAP Analysis)

ML models have a limitation in their interpretation to understand their predictions. Thus, a new analysis approach called Shapley Additive Explanation (SHAP) has been introduced by Lundberg and Lee (2017), which aims to determine how much each feature contributes to the prediction for individual instances and reveals each feature's overall impact. This study performed the SHAP analysis on the Random Forest classifier to interpret the model prediction. Features are ordered by importance from top to bottom, showing that Chainage and Weather condition are the most influential, where higher values of red dots generally lead to higher model outputs (positive SHAP values). Conversely, features lower on the list, like sex and direction, have minimal impact, with their SHAP values clustered near zero, indicating they contribute little to the model's predictions.

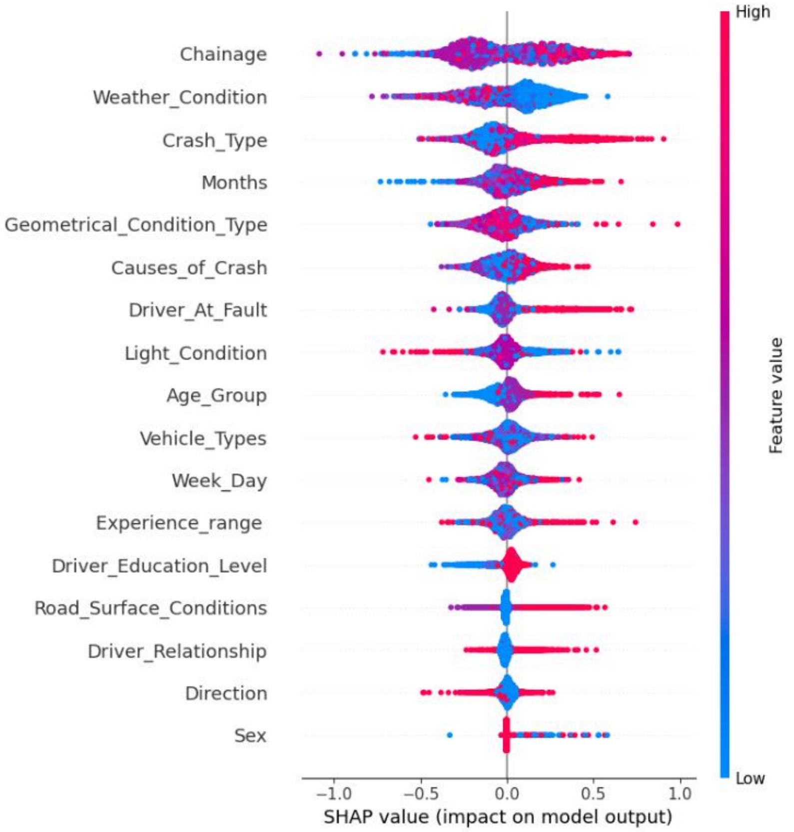


Figure 20: Global explanation obtained from SHAP for Random Forest.

4.5. Confusion Matrix Analysis

4.5.1. Random Forest

The Random Forest model performed the best model based on the provided data; it had the highest scores on all evaluation metrics. Its accuracy was 0.830, it has a precision of 0.829, it has a recall of 0.831, and it has an F1-score of 0.828. The confusion matrix shows strong model performance. It correctly spotted a lot of crashes across all categories, especially Fatal crashes (322), with only a few mistakes. The ability to accurately predict RTC severity outcome makes the RF model the most reliable choice for this specific classification task. Its metrics are better than other models like XGBoost and KNN, indicating their superior ability to correctly classify the data.

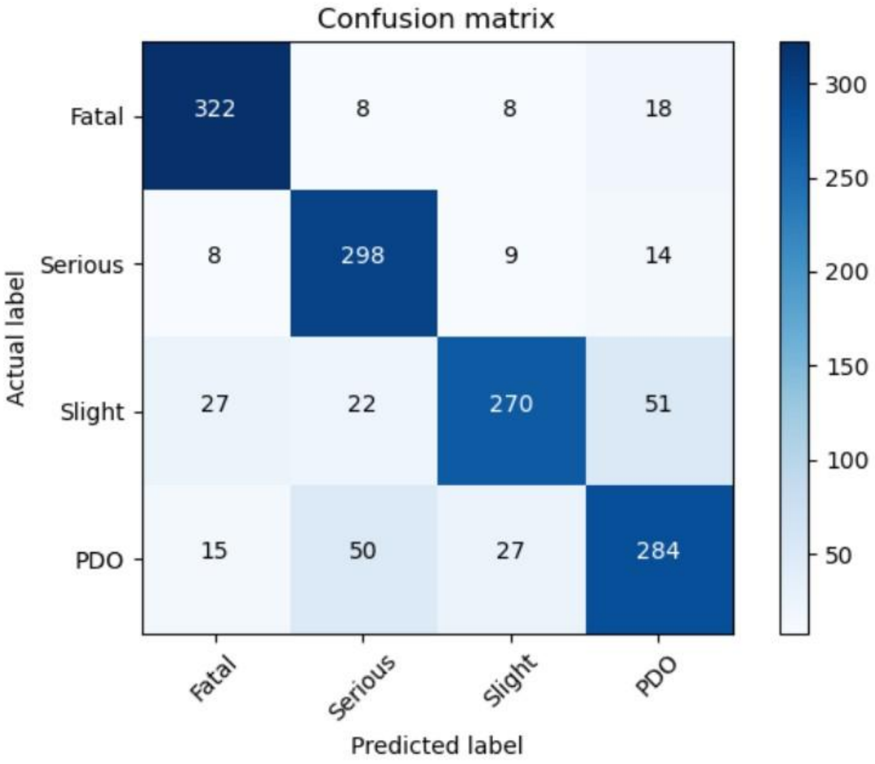


FIGURE 21: CONFUSION MATRIX FOR RANDOM FOREST MODEL

4.5.2. Extreme Gradient Boost

The confusion matrix for the XGBoost model shows good performance results. It is most likely to predict RTC correctly. The number of the diagonal, which indicates correct predictions, are high. The model identified 325 fatal crashes, 284 serious crashes, 272 slight crashes, and 290 property damage only (PDO) crashes. This shows that the model can tell the different severity levels well. The model makes a few mistakes. For instance, out of all actual Slight crashes, it predicted 17 as Fatal, 24 as Serious, and 57 as Property Damage only. Some Serious crashes were misclassified as Fatal (12), and some Fatal crashes were incorrectly labeled as less severe. However, these off-diagonal numbers are relatively low compared to the number of correct predictions, suggesting that the model is generally reliable and accurate.

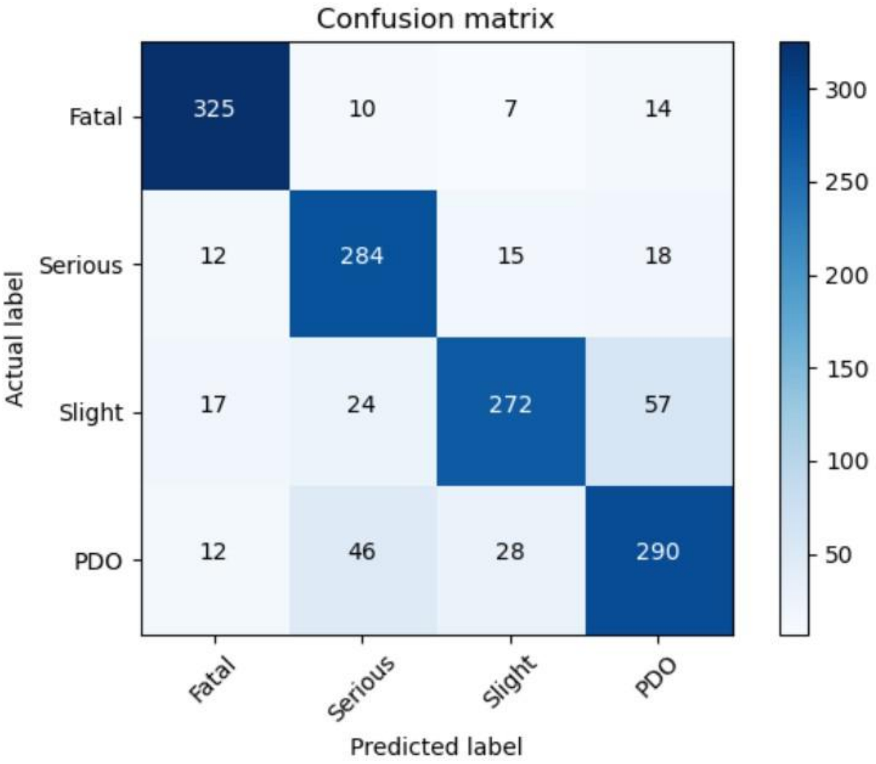


FIGURE 22 : CONFUSION MATRIX FOR XG BOOST CLASSIFIER

4.5.3. K-Nearest Neighbors

The K-Nearest Neighbors (KNN) model performs moderately well, with an overall accuracy of 0.693, a precision of 0.704, a recall of 0.694, and an F1-score of 0.666. These scores place it in the middle tier of the models tested, outperforming simpler models like Decision Tree and Naive Bayes but falling significantly behind the top performers, Random Forest and XGBoost. The confusion matrix for the KNN model reveals its specific strengths and weaknesses. The model correctly identified a high number of Fatal (333) and Slight (318) crashes, showing proficiency in classifying these two categories. However, its performance is weaker for the Serious and Property Damage Only (PDO) categories, correctly identifying only 227 and 109 cases, respectively. Furthermore, it shows a notable tendency to misclassify PDO Crashes as Slight (128 times) and Fatal (93 times), indicating a challenge in distinguishing between these severity levels.

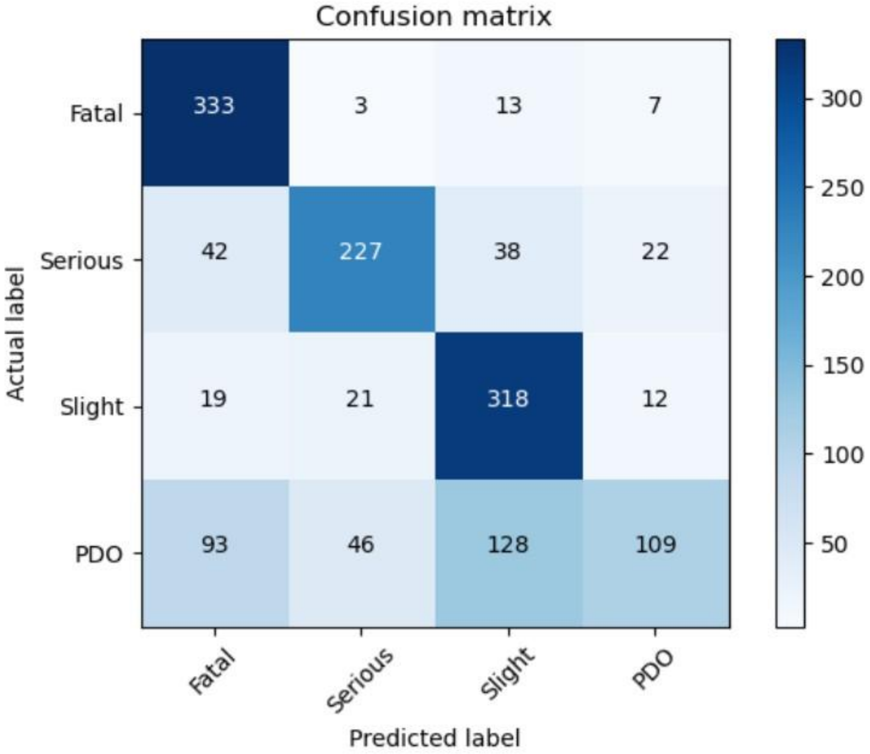


FIGURE 23: CONFUSION MATRIX FOR KNN

4.5.4. Decision Tree

The Decision Tree model is a lower-tier performer among the models tested, with an accuracy of 0.676. Its precision, recall, and F1-score are all similar, at 0.676, 0.677, and 0.675, respectively. These metrics indicate that it performs slightly better than simpler models like AdaBoost and Naive Bayes, but is significantly outperformed by more advanced models such as Random Forest, XGBoost, and KNN. The confusion matrix for the Decision Tree model reveals its performance in detail. It correctly identifies a decent number of Crashes across the categories, specifically 288 fatal, 215 serious, 242 slight, and 207 property damage only (PDO) incidents. However, the model struggles with misclassification, particularly for the Slight and PDO categories. For instance, 68 actual Slight crashes were incorrectly predicted as PDO, and 89 actual PDO crashes were incorrectly classified as Slight. This highlights the model's difficulty in accurately distinguishing between these severity levels.

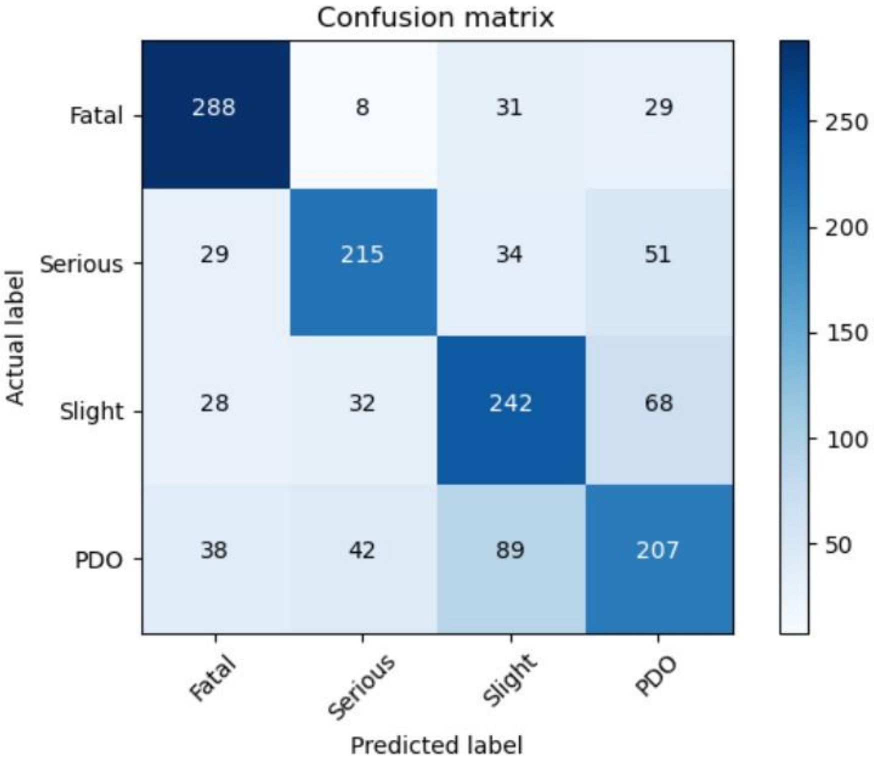


FIGURE 24: CONFUSION MATRIX FOR DECISION TREE

4.5.5. AdaBoost Classifier

The AdaBoost model is one of the lowest-performing models tested, with an accuracy of 0.512. Its precision is 0.505, recall is 0.512, and F1-score is 0.500. These low scores indicate that the model is only slightly better than a random guess and is significantly less effective than models like Random Forest, XGBoost, and KNN.

The confusion matrix for the AdaBoost model reveals significant misclassification issues. While it correctly identifies a moderate number of Fatal (189), Serious (217), and Property Damage Only (PDO) (234) Crashes, it performs poorly on the Slight category, correctly identifying only 87 cases. The model frequently misclassifies Slight Crashes as other categories, particularly PDO (110 times). This high rate of error, especially in a specific category, suggests the model struggles to accurately differentiate between all four levels of Crash severity.

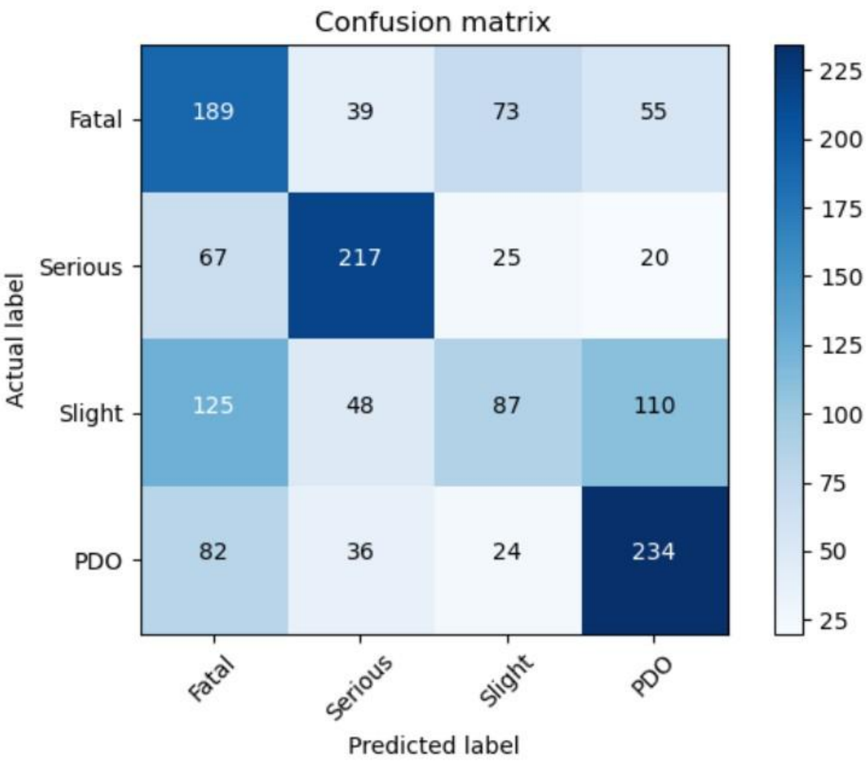


FIGURE 25 : CONFUSION MATRIX FOR AD BOOST

4.5.6. Naïve Bayes Classifier

The Figure displays a confusion matrix for a machine learning model, a Naive Bayes (NB) classifier, given the file name. The matrix evaluates the model's performance on a multi-class classification problem involving four categories: Fatal, Serious, Slight, and PDO (which likely stands for property damage only). The Actual label represents the true class of the data points, while the Predicted label shows what the model predicted. The diagonal cells, represented by darker blue squares, show the number of correctly classified instances for each class. For example, the model correctly classified 218 instances as Fatal, 206 as Serious, 52 as Slight, and 216 as PDO. These are the true positives for each class. The off-diagonal cells indicate misclassifications. For instance, 51 Fatal instances were incorrectly predicted as Serious.

The matrix highlights the model's strengths and weaknesses in distinguishing between the different crash severity classes. While the model shows a relatively high number of correct predictions on the diagonal, there are also significant misclassifications. For example, the model struggled to correctly identify Slight crashes, as many were misclassified as Fatal (170) and PDO (102). Similarly, many Fatal crashes were misclassified as PDO (55), which could have serious implications depending on the application. The model seems to perform best on accuracy in the

Fatal and PDO classes. Still, the high number of false negatives for Slight and Serious crashes suggests a potential area for improvement. The overall performance can be further analyzed by calculating metrics such as precision, recall, and F1-score from the values presented in this matrix.

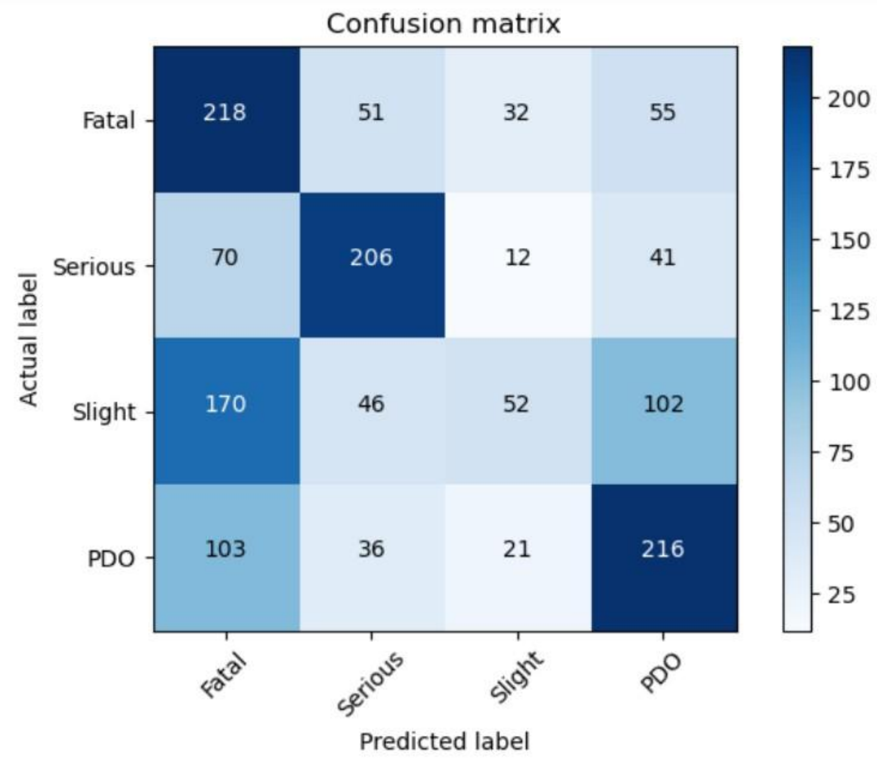


FIGURE 26: CONFUSION MATRIX FOR NAÏVE BAYES

4.5.7. Logistic regression

The provided image, a confusion matrix for Logistic Regression, shows that this model is one of the worst-performing models in the set. With an accuracy, precision, and recall of **0.416**, its performance is only slightly better than a random guess. The matrix reveals that the model struggles with classifying all categories, particularly identifying fatal crashes.

The confusion matrix highlights the model's key weakness: misclassification. While it correctly identified 209 serious and 193 property damage only (PDO) crashes, it performed poorly on the Fatal category, correctly identifying only 79 cases. A large number of fatal crashes were incorrectly classified as property damage only (147). This high misclassification rate for a critical category indicates that the Logistic Regression model is unreliable for this task.

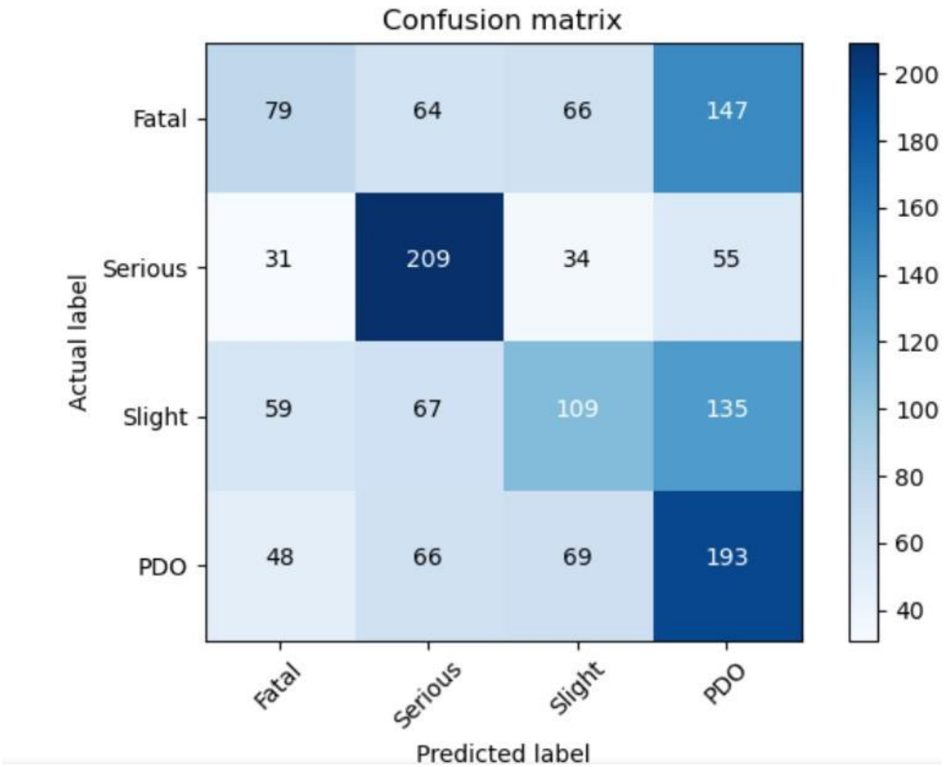


FIGURE 27: CONFUSION MATRIX FOR LOGISTICS REGRESSION

4.5.8. Support Vector Machine

The Support Vector Machine (SVM) model is the worst-performing model of all those tested. Its accuracy is 0.361, with a precision of 0.349, a recall of 0.376, and an F1-score of 0.299. These metrics are the lowest among all models, indicating that SVM is highly ineffective for this classification task.

The confusion matrix for the SVM model confirms its poor performance, revealing severe misclassification issues. While it correctly identified a moderate number of Serious crashes (274), its performance for other categories is poor. It struggled significantly with the slight and property damage only (PDO) categories, correctly identifying only 2 and 90 cases, respectively. The model also shows a strong bias towards classifying crashes as Fatal or Serious, incorrectly labeling many slight and PDO Crashes as Fatal (139 and 150 times, respectively) and Serious (130 and 132 times). This bias and the high misclassification rate make the SVM model unsuitable for this dataset.

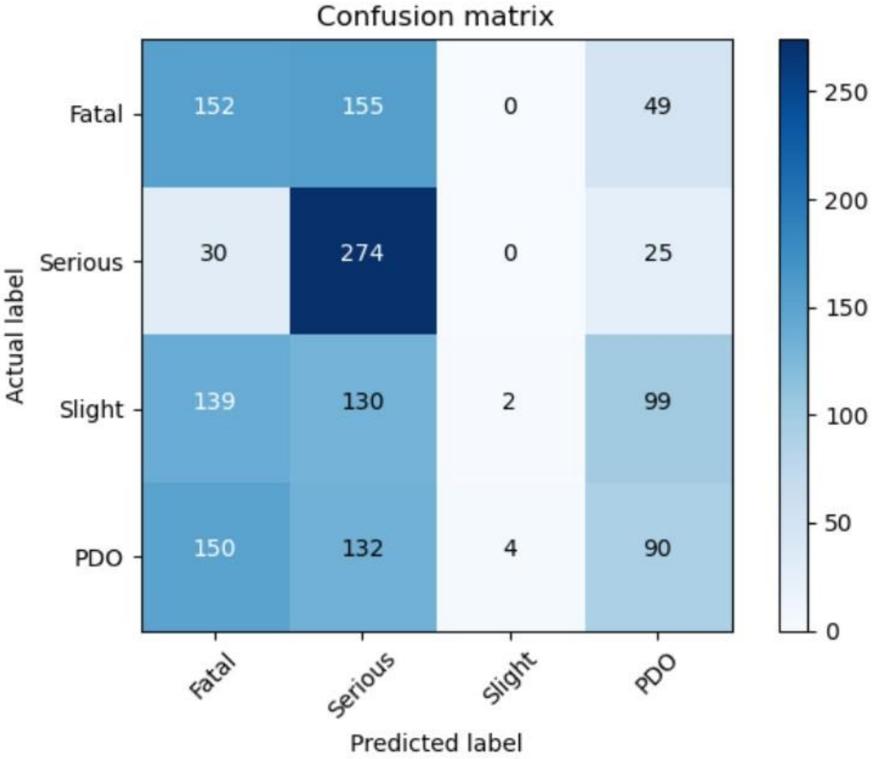


FIGURE 28: CONFUSION MATRIX FOR SVM

Thus, these models' superior performance, evidenced by their high accuracy and F1-scores, indicates their ability to capture complex non-linear relationships and interactions between diverse factors influencing crash severity outcomes, which traditional statistical methods might struggle to identify.

4.6. Sensitivity Analysis / Robustness Testing

Provides a clear overview of the model's robustness and sensitivity to different data splits (test sizes). Random Forest is the most robust model, as shown in Table 4, with the highest and most stable performance across all metrics by its narrow range of scores. This suggests it is highly reliable and less sensitive to training and testing data changes. Conversely, the Decision Tree is the least robust model. It has the lowest scores and the widest ranges for its metrics, particularly its F1 score. This indicates that its performance is susceptible to the specific data it is trained on and is the least reliable model for this task. The other models—XGBoost, SVM, AdaBoost, Naive Bayes, and Linear Model—fall between these extremes, with varying degrees of stability and performance.

TABLE 4: SENSITIVITY TESTING FOR MODEL STABILITY SUMMARY

Model	Accuracy Range	Precision Range	Recall Range	F1 Range
Random Forest	0.7410–0.7489	0.6570–0.6699	0.7410–0.7489	0.6833–0.6909
XGBoost	0.7128–0.7250	0.6400–0.6556	0.7128–0.7250	0.6710–0.6825
SVM	0.7007–0.7141	0.5949–0.6067	0.7007–0.7141	0.6401–0.6535
AdaBoost	0.6929–0.7005	0.5888–0.5936	0.6929–0.7005	0.6313–0.6397
Naive Bayes	0.6706–0.6736	0.6030–0.6178	0.6706–0.6736	0.6267–0.6330
Linear Model	0.6826–0.6863	0.5915–0.6026	0.6826–0.6863	0.6223–0.6276
KNN	0.6385–0.6504	0.5884–0.5963	0.6385–0.6504	0.6101–0.6167
Decision Tree	0.5905–0.6012	0.5979–0.6150	0.5905–0.6012	0.5937–0.6053

The result of the study regarding to the relationship between traffic flow characteristics and Road traffic Crash severity has weak, due to data sourced from toll transaction at the interchanges, however, most studies on this topic use real-time data from loop detectors both upstream and downstream of crash sites (Theofilatos, A., et al., 2017; Mehrannia, P., et al., 2023; and Zhang, H., et al., 2020). Consequently, those features have a limited influence on the model Random Forest. Thus, the author recommends that future research use real-time traffic data to more accurately evaluate the link between traffic flow characteristics and the severity of road traffic crashes.

On the other hand, the study develops a crash severity prediction based on the historical crash data using a random forest model, which is highly reliable and stable. It was consistent with similar studies in other contexts (Alotaibi, J., 2025; Masud, S. S. B., et al, 2024; Khanum et al, 2024; Iranitalab, A., & Khattak, A.,2017). Moreover, the study identified the crash severity prediction influential features, such as Spatial characteristics like chainage, geometric condition, Temporal characteristics like months, weekdays, and weather conditions, which were also commonly reported as significant predictors of crash severity globally (Nagesh U B1, et al., 2021).

This research provided a unique contribution to the local understanding of Ethiopian road safety by developing a specific model for prediction and analyzing factors influencing RTC severity on the Addis-Adama Expressway. Despite its promising result, the study acknowledges several limitations that point toward future research directions. To enhance model accuracy, future research should aim to incorporate the real-time traffic flow data (e.g., Individual vehicle speeds, lane-specific flow, etc.), and comprehensive driver behavior data (such as Fatigue, phone usage, etc.), possibly by integrating in-vehicle sensor technology.

This study developed a model to predict RTC severity using data from the expressway. In the Future, a real-time prediction model could be created. This model would use live traffic data and crash reports. it allows for quicker interventions. This would involve considerations of computational efficiency and installation of sensors on the infrastructure. While it is good for the Addis Ababa-Adama Expressway, the RF model's generalizability to other roads might require recalibration or retraining with local data due to variations in road infrastructure and traffic conditions.

Therefore, this study presents a machine learning model for RTC severity prediction for the Addis Ababa-Adama Expressway, indicating the serious role of crash patterns. The results provide valuable understandings for stakeholders to develop evidence-based road safety strategies, finally contributing to reducing crash severity and improving expressway safety.

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

Modeling RTC likelihood (Frequency or Probability) and severity is a crucial area in Safety Engineering, enabling proactive risk management and design of effective countermeasures. For likelihood prediction, classic statistical models like Poisson and negative binomial regression are commonly used, while to model severity, models for discrete outcomes are employed, such as Logistic and multinomial logistic regression. Recent advancements have been a significant shift toward ML techniques, including RF, SVM, XGB, and deep learning, which often provide superior predictive accuracy by capturing complex and nonlinear relationships between factors and the outcomes. Henceforth, the study analyzed the factors associated with the Traffic flow and the RTC Severity, explored their interrelationships, and developed an RTC severity prediction model using various machine learning algorithms.

In this study, before investigating the different models of crash severity, traffic flow, and RTC crash analysis, the association between traffic flow and road traffic crash severity was assessed. Predictors were derived from a wide range of features, some of which are temporal and some others are spatial. Therefore, predictors of RTC severity in this study were selected with reference to the literature, based on availability.

Despite the availability of RTC data and transaction traffic data gathered by the expressway, there was a lack of real-time traffic flow data, which affects the correlation of RTC severity, which can lead to weak correlation.

Eight models, such as LR, NB Classifier, DT Classifier, KNN, SVM, AdaB, RF Classifier, and XGB classifier, were chosen, compared, and evaluated based on various criteria, and several modelling trials were conducted to retain models with the best overall accuracy, recall, and F-score. Thus, from the above eight chosen algorithms Random Forest classifier yields superior predictive capabilities with an accuracy of 83%, a precision of 82.9 %, a recall of 83%, and an F1-score of 82.8 %.

The resulting factors associated with RTC based on the Random Forest model were the Chainage of the roadway, months of the year, crash type, weather condition, cause of crash, weak days, Geometric condition, and Vehicle type. The chainage of the roadway, months of the year, crash type, and weather conditions promoted higher RTC severity. In contrast, sex, driver relationship, age group, and direction contributed to lower RTC severity.

Overall, the RTC severity prediction model of this work contributes to improving the knowledge of factors associated with RTC severity in the Addis-Adama Expressway. Implementing such models, in practice, adds a quantifiable performance dimension to the decision-making process, from planning to design and operations. To employ these RTC prediction models on other expressways, calibration of the parameters might be necessary to account for conditional differences.

The outcome of the random forest model appears highly promising, and, thus, more combinations of variables and hyperparameter tuning could be explored to achieve a better understanding of the effects of different traffic-related variables and other independent variables on RTC severity. Although this study aimed to be as comprehensive as possible in the inclusion of different independent variables, there is room to explore different modelling structures and different unobserved heterogeneity levels.

Therefore, the selected random forest model can be further developed and applied in an actual real-time scenario by integrating the existing Addis-Adama Expressway Traffic Management Systems. The model application is not only solving current problems but also propelling the system toward becoming a dynamic, responsive, and sustainable system.

5.2. Recommendation

From the above list of findings, the following recommendations can be made for RTC severity prediction on the Addis-Adama Expressway in other roads with similar contexts.

- To fully investigate the correlation of traffic flow attributes and RTC severity, real-time data from loop detectors both upstream and downstream of crash sites is needed.
- To raise the prediction accuracy, the model's hybrid hyperparameter tuning is needed.
- Targeted safety intervention should be done by focusing on the high-risk chainages and specific months identified by the model.
- Use the RF model to strategically allocate post-crash management and emergency response resources to monitor hot spots proactively.

- Promote the collection of more granular data on driver behavior and specific road conditions to further refine prediction models.

5.3. Further work

1. Creating a real-time operational crash severity prediction model that continuously updates predictions based on live traffic and RTC feeds, allowing for truly proactive interventions.
2. Develop an interactive web application that integrates a random forest model (for prediction).

REFERENCE

Abdel-Aty, M. A., & Haleem, K. (2011). Forecasting freeway traffic speeds using support vector machines. UCF, Department of Civil, Environmental, and Construction Engineering, Orlando, Florida.

Abera, M. A. (2019). Evaluation of road traffic accident in Addis Ababa–Adama Expressway. *International Journal of Advanced Research, Ideas and Innovations in Technology*, *5*(5), 450-456.

Ahmed, M., & Ghasemzadeh, A. (2018). Impact of weather conditions on driver behavior: A review. *Transportation Research Part F: Traffic Psychology and Behaviour*, *57*, 1-13.

Alavi, S. S., Mohammadi, M. R., Souri, H., Kalhori, S. M., Jannatifard, F., & Sepahbodi, G. (2017). Personality, driving behavior and mental disorders factors as predictors of road traffic accidents based on logistic regression. *Iranian journal of medical sciences*, *42*(1), 24.

Alotaibi, J. (2025). Enhancing Traffic Accident Severity Prediction: Feature Identification Using Explainable AI. *Vehicles*, *7*(2), 38.

Baru, A., Azazh, A., & Beza, L. (2019). Injury severity levels and associated factors among road traffic collision victims referred to emergency departments of selected public hospitals in Addis Ababa, Ethiopia: the study based on the Haddon matrix. *BMC emergency medicine*, *19*(1), 2.

Beshah, T., & Hill, S. (2010). Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in Ethiopia. In *Proc. AAAI Spring Symp., Artif. Intell. Develop*, Volume 24, Princeton, NJ, USA: Citeseer, Pages 1173–1181.

Benfaress, I., Bouhoute, A., & Zinedine, A. (2024). Enhancing Traffic Accident Severity Prediction Using ResNet and SHAP for Interpretability. *AI*, *5*(4), 2568-2585.

Bi, N., & Sadia, H. (2023). Accident severity detection using machine learning. *International Journal Of Engineering And Management Research*, *13*(3), 203-208.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge university press.

Casado-Sanz, N., Guirao, B., & Attard, M. (2020). Analysis of the risk factors affecting the severity of traffic accidents on Spanish crosstown roads: The driver's perspective. *Sustainability*, *12*(6), 2237..

Chang, L. Y., & Chen, C. Y. (2005). Analysis of traffic accident severity using a decision tree model. *Accident Analysis & Prevention*, *37*(6), 1045-1052.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., & Guan, H. (2015). A multinomial logit model-

Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes.

Accident Analysis & Prevention, 80, 76–88.

Creswell, J. W., & Creswell, J. D. (2005). *Mixed methods research: Developments, debates, and dilemma* (pp. 315-26). Oakland, CA: Berrett-Koehler Publishers.

Deme, D. (2018). Traffic Accident Causes and Its Countermeasures on the Addis Ababa-Adama Expressway. *Journal of Equity in Sciences and Sustainable Development (JESSD)*, *2*(2), 13-23.

Deme, D., & Bari, M. B. (2018). Traffic Accident Causes and Its Countermeasures on Addis Ababa-Adama Expressway. *Madda Walabu University Journal of Equity in Sciences and Sustainable Development*, *2*(2), 13-23.

Dong, C., Shao, C., Li, J., & Xiong, Z. (2018). An improved deep learning model for traffic crash prediction. *Journal of Advanced Transportation*, *2018*(1), 3869106.

El-Basyouny, K., & Sayed, T. (2009a). A multivariate Poisson-lognormal model for crash counts. *Accident Analysis & Prevention*, *41*(4), 856-865.

Elvik, R. (2004). Road safety effects of speed limits. *Accident Analysis & Prevention*, *36*(5), 901-912.

Elvik, R. (2009). The safety effect of road surface condition: A literature review. *Accident Analysis & Prevention*, *41*(1), 160-167.

Federal Police Commission. (2019). *Annual crash data report*. Addis Ababa.

Fitzpatrick, K., et al. (2003). Impact of roadway conditions on crash rates: A review. *Transportation Research Record, 1840*(1), 18-26.

Fridman, L., et al. (2019). Driving in adverse weather conditions: Impact on driver behavior and safety. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 700-715.

Garber, N. J., & Ehrhart, A. A. (2000). Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. *Transportation Research Record*, *1717*(1), 76-83.

Geedipally, V. K., & Lord, D. (2009). Bivariate Poisson-lognormal model for crash counts. *Journal of Transportation Engineering*, *135*(11), 875-885.

Hauer, E. (2004). Safety in numbers: A common explanation for the effect of volume on accidents. *Accident Analysis & Prevention*, *36*(1), 814-817.

Hossain, S., et al. (2019). Spatio-temporal analysis of road traffic accidents using machine learning. *Accident Analysis & Prevention*, *131*, 15-28.

Infante, P., Jacinto, G., Afonso, A., Rego, L., Nogueira, P., Silva, M., ... & Manuel, P. R. (2023). Factors that influence the type of road traffic accidents: A case study in a district of portugal. *Sustainability*, *15*(3), 2352.

Iranitalab, A., & Khattak, A. (2017). Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis & Prevention*, *108*, 27-36.

Jagannathan, R., Petrovic, S., Powell, G., & Roberts, M. (2013). Predicting road accidents based on current and historical spatio-temporal traffic flow data. In *Computational Logistics: 4th International Conference, ICCL 2013, Copenhagen, Denmark, September 25-27, 2013. Proceedings 4* (pp. 83-97). Springer Berlin Heidelberg.

Kahane, C. J. (2002). Effects of Roadway Condition on Traffic Safety. Transportation Research Board.

Khanum, H., Garg, A., & Faheem, M. I. (2023). Accident severity prediction modeling for road safety using random forest algorithm: an analysis of Indian highways. *F1000Research*, *12*, 494.

Kidane, M., Sultan, M., Azazh, A., Beza, L., & Waganew, W. (2025). Rising epidemic of road traffic injuries in Ethiopia: A systematic review of available literature. *Pan African Journal of Emergency Medicine and Critical Care*, *3*(1).

Kopacz, K., & Garcia, L. (2010). The role of road conditions in traffic accidents: A data analysis approach. *Journal of Traffic and Transportation Engineering*.

Lambodar Gourish Hebbar1, Raghavendra Bhardwaj D G 2, Vardhan Kumar G K3, Varun S4, Roopashree C S 5,. (2025). Modeling the Accident Severity Level: A Survey. *International Journal of Innovative Research in Technology*, Volume 11 Issue 11 | ISSN: 2349-6002).

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, *44*(5), 291-305

Leedy, P. D., Ormrod, J. E., & Johnson, L. R. (2014). *Practical research: Planning and design* (p. 360). Pearson Education.

Lefevre, S., et al. (2014). Driving behavior modeling based on perception and decision. *IEEE Transactions on Intelligent Transportation Systems*, *15*(3), 1334-1345.

Li, X., Lord, D., Zhang, Y., & Xie, Y. (2008). Predicting motor vehicle crashes using support vector machine models. *Accident Analysis & Prevention*, *40*(4), 1611-1618.

Lian, Y., et al. (2017). Human driving behavior modeling and prediction based on deep learning. *Neurocomputing*, *267*, 239-247.

Lipton, Z. C. (2016). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *14*(3), 31-40.

Liu, J., et al. (2017). Predicting traffic accident hotspots using gradient boosting machines. *Accident Analysis & Prevention*, *106*, 157-167.

Lord, D., et al. (2007). The Conway-Maxwell-Poisson regression model for crash counts: An application to a road safety problem. *Accident Analysis & Prevention*, *39*(5), 903-911.

Lord, D., et al. (2009). The Poisson-lognormal model for crash counts: An empirical Bayes application. *Accident Analysis & Prevention*, *41*(1), 168-175.

Ma, J., & Kockelman, K. M. (2006). A multivariate Poisson-lognormal model for crash counts. *Transportation Research Record*, *1950*(1), 89-97.

Madushani, J. S., Sandamal, R. K., Meddage, D. P. P., Pasindu, H. R., & Gomes, P. A. (2023). Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers. *Transportation Engineering*, *13*, 100190.

Maher, M. J. (1990). A bivariate negative binomial regression model for traffic accident counts. *Traffic Engineering & Control*, *31*(6), 332-335.

Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic methods in accident research*, *1*, 1-22.

Maria, A. (1997, December). Introduction to modeling and simulation. In *Proceedings of the 29th conference on Winter simulation* (pp. 7-13).

Masud, S. S. B., Hossain, A., Akter, N., & Mohammed, H. (2024). Fatal Crash Occurrence Prediction and Pattern Evaluation by Applying Machine Learning Techniques. *The Open Transportation Journal*, *18*(1).

Mehrannia, P., Bagi, S. S. G., Moshiri, B., & Al-Basir, O. A. (2023). Deep representation of imbalanced spatio-temporal traffic flow data for traffic accident detection. *IET Intelligent Transport Systems*, *17*(3), 606-619.

Mekonen, E. K. (2016). The Economic Effect of Road Traffic Accidents In Ethiopia: Evidences from Addis Ababa City. *ITIHAS-The Journal of Indian Management*, *6*(2).

Miaou, S. P., & Song, J. J. (2005). Bayesian multinomial logit models for injury severity in traffic accidents. *Accident Analysis & Prevention*, *37*(1), 1-15.

Moser, S., et al. (2015). Inter-driver variability in naturalistic driving: A comparison of two data sets. *Traffic Injury Prevention*, *16*(sup2), S105-S110.

Nagesh U B1 (2021). *Analysis and Prediction of Road Accidents using machine learning techniques: A mixed logit approach*. (PhD dissertation, University of Central Florida).

NGuessan, A. K. (2010). *Modelling injury severity in traffic accidents: A mixed logit approach*. (PhD dissertation, University of Central Florida).

Noland, R. B., & Quddus, M. A. (2004). A spatially disaggregate analysis of road casualties in London. *Accident Analysis & Prevention*, *36*(6), 947-954.

Obasi, I. C., & Benson, C. (2023). Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents. *Heliyon*, *9*(8).

ONeill, B., & T. S. (2011). The effects of road surface management on safety. *Accident Analysis & Prevention*, *43*(1), 293-298.

Oh, J., et al. (2006). The gamma distribution for traffic accident counts: A new approach. *Accident Analysis & Prevention*, *38*(5), 987-994.

Park, B. J., & Lord, D. (2007). A multivariate Poisson-lognormal model for accident counts. *Accident Analysis & Prevention*, *39*(6), 1167-1175.

Peden, M. M. (Ed.). (2004). *World report on road traffic injury prevention*. World Health Organization.

Rahman, A., & Hasan, M. (2018). Addressing class imbalance in road accident severity prediction using machine learning techniques. *Accident Analysis & Prevention*, *118*, 223-233.

Rolison, J. J., Regev, S., Moutari, S., & Feeney, A. (2018). What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers opinions, and road accident records. *Accident Analysis & Prevention*, *115*, 11–24.

Rosen, H. E., Bari, I., Paichadze, N., Peden, M., Khayesi, M., Monclús, J., & Hyder, A. A. (2022). Global road safety 2010–18: an analysis of global status reports. *Injury*.

Segui Gomez, M., Addo-Ashong, T., Raffo, V. I., & Venter, P. (2021). *Road Safety Data in Africa: A Proposed Minimum Set of Road Safety Indicators for Data Collection, Analysis, and Reporting*. World Bank.

Shah, S. A. R., Ahmad, N., Shen, Y., Kamal, M. A., Basheer, M. A., & Brijs, T. (2019). Relationship between road traffic features and accidents: An application of two-stage decision-making approach for transportation engineers. *Journal of Safety Research*, *69*, 201-215.

Shmueli, G., et al. (2005). A useful distribution for modeling count data. *Journal of Quality Technology*, *37*(3), 322-334.

Subrahmainiam, K., & Subrahmainiam, K. (1973). On the bivariate Poisson distribution. *Journal of the American Statistical Association*, *68*(344), 955-958.

Tang, J., Liang, J., Han, C., Li, Z., & Huang, H. (2019). Crash injury severity analysis using a two-layer stacking framework. Accident Analysis & Prevention, 122, 226–238.

Theofilatos, A., Yannis, G., Vlahogianni, E. I., & Golias, J. C. (2017). Modeling the effect of traffic regimes on safety of urban arterials: The case study of Athens. *Journal of traffic and transportation engineering (English edition)*, *4*(3), 240-251.

Wambulwa, W. M., & Job, S. (2020). *Guide for Road Safety Opportunities and Challenges: Low-and Middle-Income Country Profiles*.

WHO. (2015). *Global status report on road safety 2015*. World Health Organization.

WHO. (2018). *Global status report on road safety*. World Health Organization.

WHO. (2023). *Global status report on road safety 2023*. World Health Organization.

Winkelmann, R. (2003). *Econometric analysis of count data*. Springer-Verlag.

World Health Organization. (2008). *Global Road Safety Partnership: Speed management: a road safety manual for decision-makers and practitioners*.

World Health Organization. (2017). *Save lives: a road safety technical package*.

Xu, J., et al. (2014). Feature selection for traffic accident prediction. *Transportation Research Part C: Emerging Technologies*, *40*, 17-29.

Yang, J., et al. (2020). Vehicle stability control in adverse weather conditions. *Journal of Dynamic Systems, Measurement, and Control*, *142*(6), 061001.

Yasmin, S., et al. (2017). Impact of traffic flow on accident frequencies in urban areas. *Journal of Traffic and Transportation Engineering (English Edition)*, *4*(4), 312-321.

Yeon, Y., et al. (2019). Modeling of driving behavior based on deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, *20*(11), 4165-4176.

Yu, R., Zheng, Y., Qin, Y., & Abdel-Aty, M. (2020). Crash injury severity analyses with multilevel thresholds of change modelling approach for at-fault out-of-state drivers. *Journal of Transportation Safety & Security*, *12*(9), 1164-1181.

Zenu, S., Tesfaye, A., Roba, G., & Negera, E. (2023). Factors in road traffic crashes among drivers in Buno Bedelle Zone, Southwest Ethiopia: a case control study. *Journal of road safety*, *34*(4), 1-10.

Zhang, H., Li, S., Wu, C., Zhang, Q., & Wang, Y. (2020). Predicting Crash Frequency for Urban Expressway considering Collision Types Using Real-Time Traffic Data. *Journal of advanced transportation*, *2020*(1), 8523818.

Zhang, X., & Wang, Y. (2015). Analysis of the relationship between highway surface conditions and traffic accidents. *International Journal of Transportation Science and Technology*, *4*(4), 355-362.

Zhang, H., & Xie, Y. (2007). Application of support vector machines in intelligent transportation systems. *Transportation Research Record*, *1993*(1), 10-18.

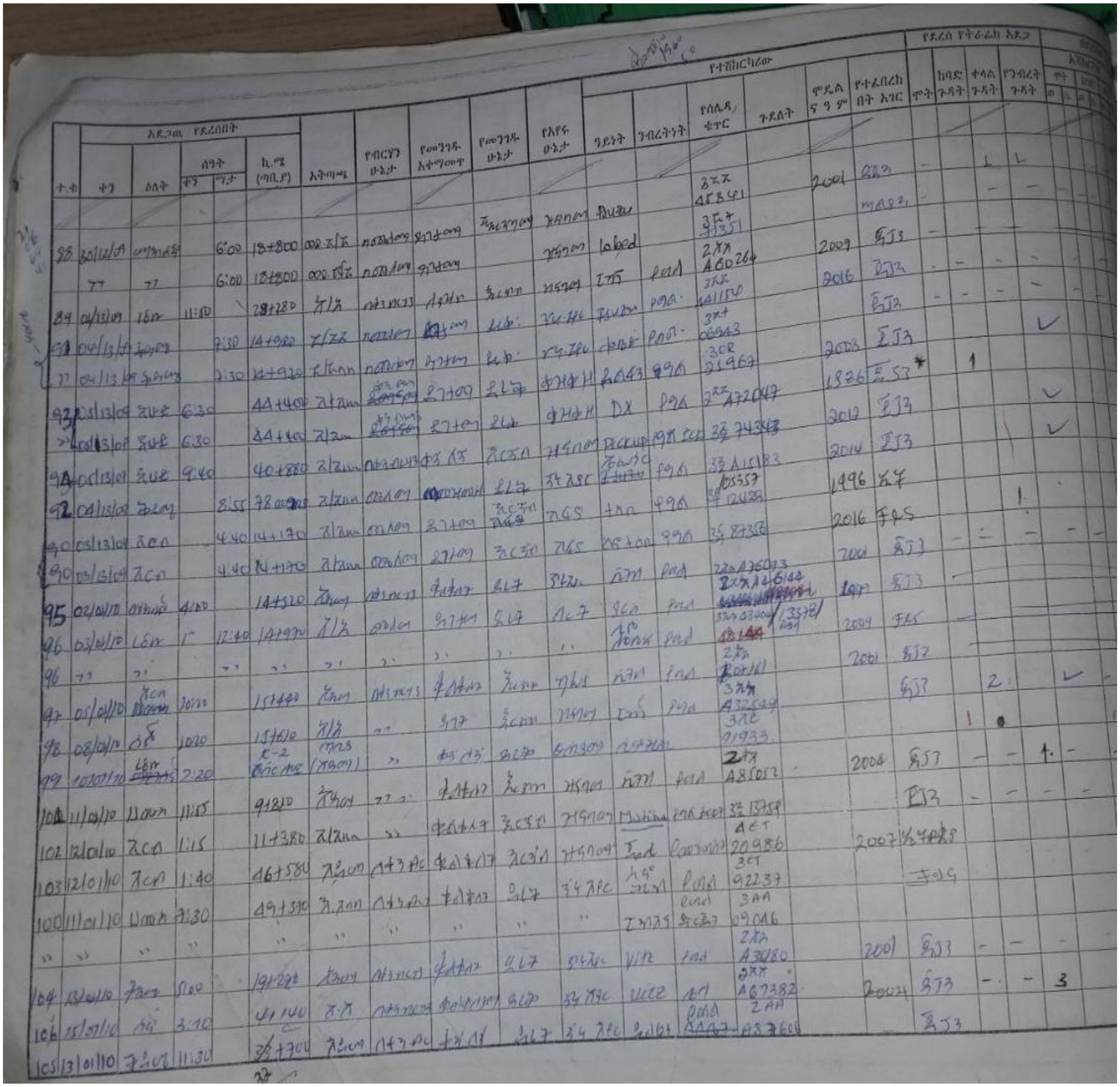
Zhou, H., et al. (2020). A CNN-RNN hybrid model for spatio-temporal prediction of traffic accidents. *Accident Analysis & Prevention*, *140*, 105527.

APPENDICES

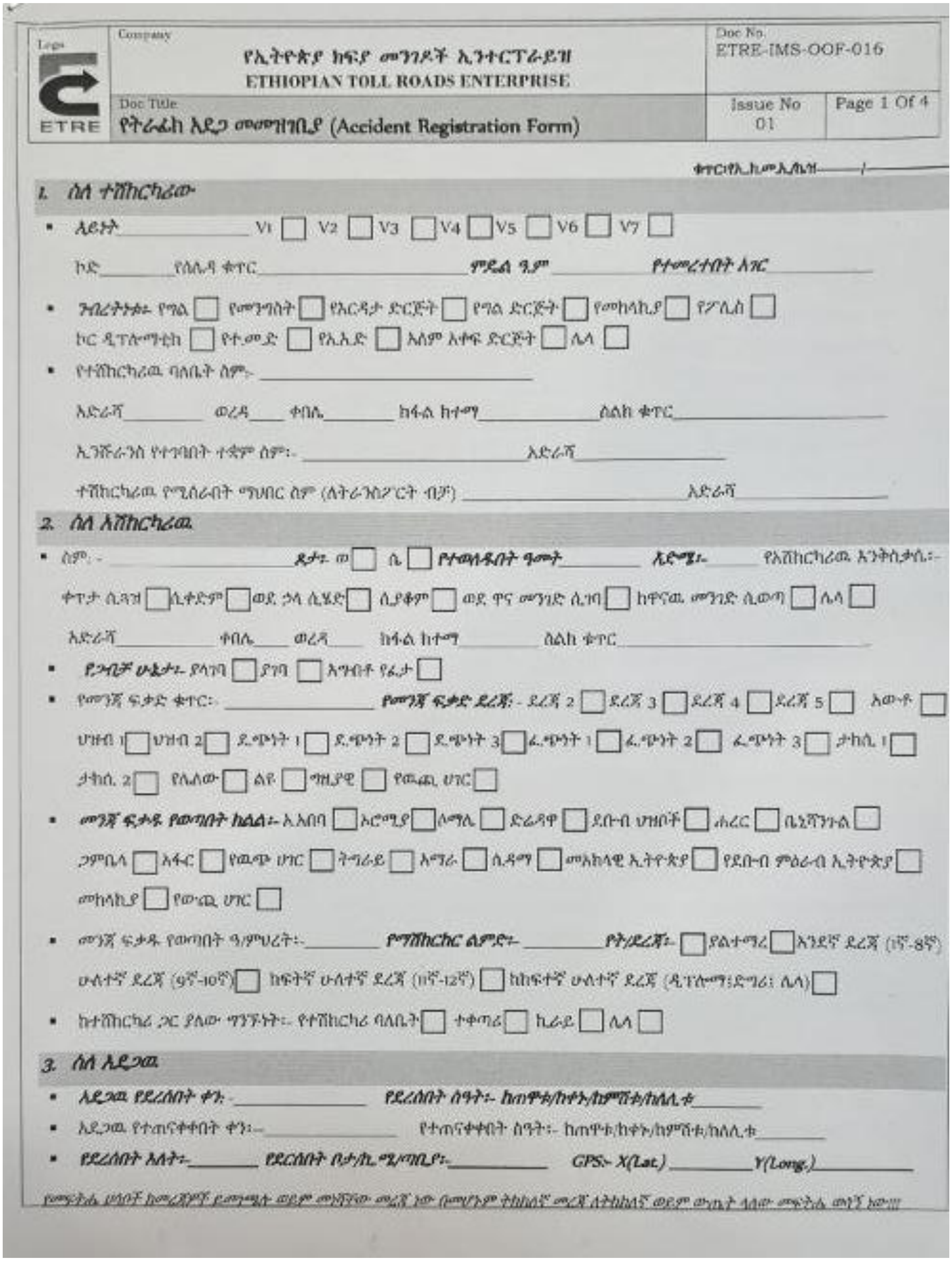
Appendix 1: ETRE Vehicles Classification Parameters

Vehicles Class	**Classification Parameters**					
	i-Axle number	**ii**-Tire number	iii-Height of vehicle head(m)	iV-Axle distance (m)	Description of vehicles or	Remarks
V1	2	4	<1.3	<2.4	Small automobiles cars, Jeep, Land rover, taxi, pick up	At the same time to satisfy the parameters I, II, and III **or** IV in any conditions
V2	2	4	≥1.3	≥2.4	Minibus	At the same time to satisfy the parameters I, II, III **and** IV conditions
V3	2	6	—	—	Medium Bus、ISUZU	
V4	3	—	—	—	Big size Bus Dump Trucks	
V5	4	—	—	—	Heavy Trucks, Trailers	
V6	5	—	—	—	Heavy Truck Trailers	
V7	**≥6**	—	—	—	Heavy Truck Big trailers	

Appendix 2: Sample Raw data



APPENDIX 3: DATA COLLECTION FORMS.



Logo ETRE	Company የኢትዮጵያ ክፍያ መንገዶች ኢንተርፕራይዝ ETHIOPIAN TOLL ROADS ENTERPRISE	Doc No. ETRE-IMS-OOF-016	
	Doc Title የትራፊክ አደጋ መመዝገቢያ (Accident Registration Form)	Issue No 01	Page 1 Of 4

ቁጥር፡የኢ.ክ.መ.ኢ/ሲ/ዝ———/———

1. ስለ ተሽከርካሪው

- **ዓይነት**________ V1 ☐ V2 ☐ V3 ☐ V4 ☐ V5 ☐ V6 ☐ V7 ☐

ኮድ______ የሰሌዳ ቁጥር__________ **ሞዴል ዓ.ም** _______ **የተመረተበት አገር**_______

- **ንብረትነቱ፦** የግል ☐ የመንግስት ☐ የእርዳታ ድርጅት ☐ የግል ድርጅት ☐ የመከላከያ ☐ የፖሊስ ☐
ኮር ዲፕሎማቲክ ☐ የተ.መ.ድ ☐ የአ.አ.ድ ☐ አለም አቀፍ ድርጅት ☐ ሌላ ☐
- የተሽከርካሪዉ ባለቤት ስም፦ ________________

አድራሻ______ ወረዳ____ ቀበሌ______ ክፋል ከተማ________ ስልክ ቁጥር__________

ኢንሹራንስ የተገባበት ተቋም ስም፡- ______________ አድራሻ_________

ተሽከርካሪዉ የሚሰራበት ማህበር ስም (ለትራንስፖርት ብቻ) ______________ አድራሻ_________

2. ስለ አሽከርካሪዉ

- ስም፡- ______________ **ጾታ፦** ወ ☐ ሴ ☐ **የተወለዱበት ዓመት**_______ **ዕድሜ፦**_____ የአሽከርካሪዉ እንቅስቃሴ፡-

ቀጥታ ሲጓዝ ☐ ሲቀድም ☐ ወደ ኋላ ሲሄድ ☐ ሲያቆም ☐ ወደ ዋና መንገድ ሲገባ ☐ ከዋናዉ መንገድ ሲወጣ ☐ ሌላ ☐

አድራሻ________ ቀበሌ____ ወረዳ____ ክፋል ከተማ_______ ስልክ ቁጥር____________

- **የጋብቻ ሁኔታ፦** ያላገባ ☐ ያገባ ☐ አግብቶ የፈታ ☐
- የመንጃ ፍቃድ ቁጥር፡ __________ **የመንጃ ፍቃድ ደረጃ፡** - ደረጃ 2 ☐ ደረጃ 3 ☐ ደረጃ 4 ☐ ደረጃ 5 ☐ አውቶ ☐

ህዝብ 1 ☐ ህዝብ 2 ☐ ደ.ጭነት 1 ☐ ደ.ጭነት 2 ☐ ደ.ጭነት 3 ☐ ፈ.ጭነት 1 ☐ ፈ.ጭነት 2 ☐ ፈ.ጭነት 3 ☐ ታክሲ 1 ☐

ታክሲ 2 ☐ የሌለው ☐ ልዩ ☐ ግዚያዊ ☐ የዉጪ ሀገር ☐

- **መንጃ ፍቃዱ የወጣበት ክልል፦** አ.አበባ ☐ ኦሮሚያ ☐ ሶማሌ ☐ ድሬዳዋ ☐ ደቡብ ህዝቦች ☐ ሐረር ☐ ቤኒሻንጉል ☐

ጋምቤላ ☐ አፋር ☐ የዉጭ ሀገር ☐ ትግራይ ☐ አማራ ☐ ሲዳማ ☐ መአከላዊ ኢትዮጵያ ☐ የደቡብ ምዕራብ ኢትዮጵያ ☐

መከላከያ ☐ የውጪ ሀገር ☐

- መንጃ ፍቃዱ የወጣበት ዓ/ምህረት፡-_______ **የማሽከርከር ልምድ፦** _______**የት/ደረጃ፦** ☐ ያልተማረ ☐ አንደኛ ደረጃ (1ኛ-8ኛ)

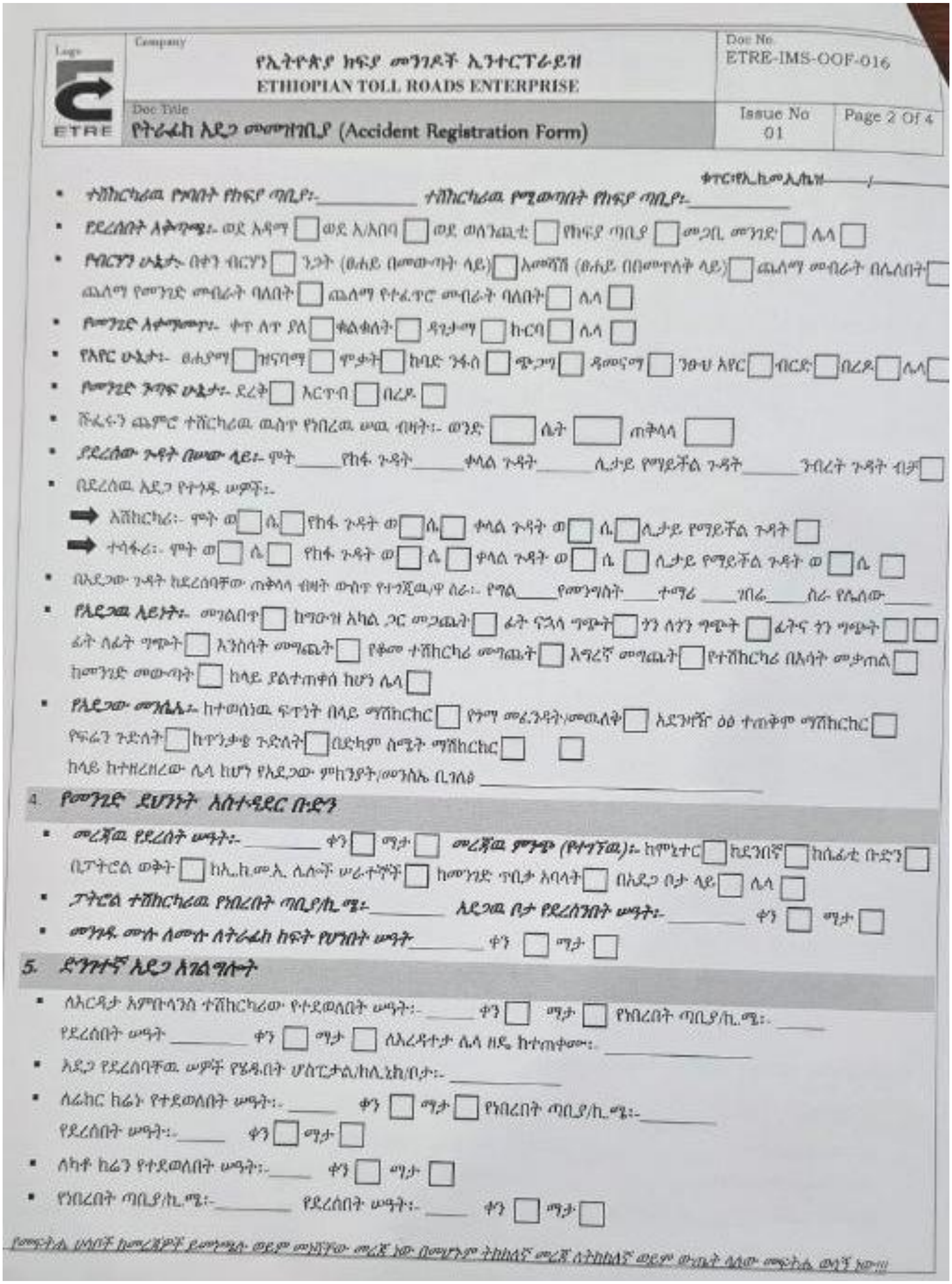
ሁለተኛ ደረጃ (9ኛ-10ኛ) ☐ ከፍተኛ ሁለተኛ ደረጃ (11ኛ-12ኛ) ☐ ከከፍተኛ ሁለተኛ ደረጃ (ዲፕሎማ፣ድግሪ፣ ሌላ) ☐

- ከተሽከርካሪ ጋር ያለው ግንኙነት፡- የተሽከርካሪ ባለቤት ☐ ተቀጣሪ ☐ ኪራይ ☐ ሌላ ☐

3. ስለ አደጋዉ

- **አደጋዉ የደረሰበት ቀን፡** -__________ **የደረሰበት ሰዓት፦ ከጠዋቱ/ከቀኑ/ከምሽቱ/ከሌሊቱ**______
- አደጋዉ የተጠናቀቀበት ቀን፡-__________ የተጠናቀቀበት ሰዓት፡- ከጠዋቱ/ከቀኑ/ከምሽቱ/ከሌሊቱ______
- **የደረሰበት እለት፦**______ **የደረሰበት ቦታ/ኪ.ሜ/ማቢያ፦**________ **GPS፦ X(Lat.)**_______**Y(Long.)**________

የሚሞሉት ሀሳቦች ከመረጃዎች ይመጣሉ ወይም መዝገብ መረጃ ነው በማንኛውም ትክክለኛ መረጃ ለትክክለኛ ወይም ውጤት ላለው መፍትሔ መሆኑ ነው!!!



Logo	Company የኢትዮጵያ ክፍያ መንገዶች ኢንተርፕራይዝ ETHIOPIAN TOLL ROADS ENTERPRISE	Doc No ETRE-IMS-OOF-016	
ETRE	Doc Title የትራፊክ አደጋ መመዝገቢያ (Accident Registration Form)	Issue No 01	Page 2 Of 4

ቁጥር፡የኢ.ኪ.መ.ኢ.ሲ.ዝ______/______

- ተሽከርካሪዉ የገባበት የክፍያ ጣቢያ፡-__________ ተሽከርካሪዉ የሚወጣበት የክፍያ ጣቢያ፡-__________
- የደረሰበት አቅጣጫ፡- ወደ አዳማ ☐ ወደ አ/አበባ ☐ ወደ ወሰንጨቲ ☐ የክፍያ ጣቢያ ☐ መጋቢ መንገድ ☐ ሌላ ☐
- የብርሃን ሁኔታ፡- በቀን ብርሃን ☐ ንጋት (ፀሐይ በመውጣት ላይ) ☐ አመሻሽ (ፀሐይ በበመጥለቅ ላይ) ☐ ጨለማ መብራት በሌለበት ☐
ጨለማ የመንገድ መብራት ባለበት ☐ ጨለማ የተፈጥሮ መብራት ባለበት ☐ ሌላ ☐
- የመንገድ አቀማመጥ፡- ቀጥ ለጥ ያለ ☐ ቁልቁለት ☐ ዳገታማ ☐ ኩርባ ☐ ሌላ ☐
- የአየር ሁኔታ፡- ፀሐያማ ☐ ዝናባማ ☐ ሞቃት ☐ ከባድ ንፋስ ☐ ጭጋግ ☐ ዳመናማ ☐ ንፁህ አየር ☐ ብርድ ☐ በረዶ ☐ ሌላ ☐
- የመንገድ ንጣፍ ሁኔታ፡- ደረቅ ☐ እርጥብ ☐ በረዶ ☐
- ሹፌሩን ጨምሮ ተሽከርካሪዉ ዉስጥ የነበረዉ ሠዉ ብዛት፡- ወንድ ☐ ሴት ☐ ጠቅላላ ☐
- ያደረሰው ጉዳት በሠው ላይ፡- ሞት_____ የከፋ ጉዳት_____ ቀላል ጉዳት_____ ሊታይ የማይችል ጉዳት_____ ንብረት ጉዳት ብቻ ☐
- በደረሰዉ አደጋ የተጎዱ ሠዎች፡-

➡ አሽከርካሪ፡- ሞት ወ ☐ ሴ ☐ የከፋ ጉዳት ወ ☐ ሴ ☐ ቀላል ጉዳት ወ ☐ ሴ ☐ ሊታይ የማይችል ጉዳት ☐

➡ ተሳፋሪ፡- ሞት ወ ☐ ሴ ☐ የከፋ ጉዳት ወ ☐ ሴ ☐ ቀላል ጉዳት ወ ☐ ሴ ☐ ሊታይ የማይችል ጉዳት ወ ☐ ሴ ☐

- በአደጋው ጉዳት ከደረሰባቸው ጠቅላላ ብዛት ውስጥ የተገኟዉ/ዋ ስራ፡- የግል_____ የመንግስት_____ ተማሪ _____ ገበሬ_____ ስራ የሌለው_____
- የአደጋዉ አይነት፡- መገልበጥ ☐ ከግዑዝ አካል ጋር መጋጨት ☐ ፊት ኋላ ግጭት ☐ ጎን ለጎን ግጭት ☐ ፊትና ጎን ግጭት ☐ ☐
ፊት ለፊት ግጭት ☐ እንስሳት መግጨት ☐ የቆመ ተሽከርካሪ መግጨት ☐ እግረኛ መግጨት ☐ የተሽከርካሪ በእሳት መቃጠል ☐
ከመንገድ መውጣት ☐ ከላይ ያልተጠቀሰ ከሆነ ሌላ ☐
- የአደጋው መንስኤ፡- ከተወሰነዉ ፍጥነት በላይ ማሽከርከር ☐ የጎማ መፈንዳት/መዉለቅ ☐ አደንዛዥ ዕፅ ተጠቅሞ ማሽከርከር ☐
የፍሬን ጉድለት ☐ ከጥንቃቄ ጉድለት ☐ በድካም ስሜት ማሽከርከር ☐ ☐
ከላይ ከተዘረዘረው ሌላ ከሆነ የአደጋው ምክንያት/መንስኤ ቢገለፅ __________

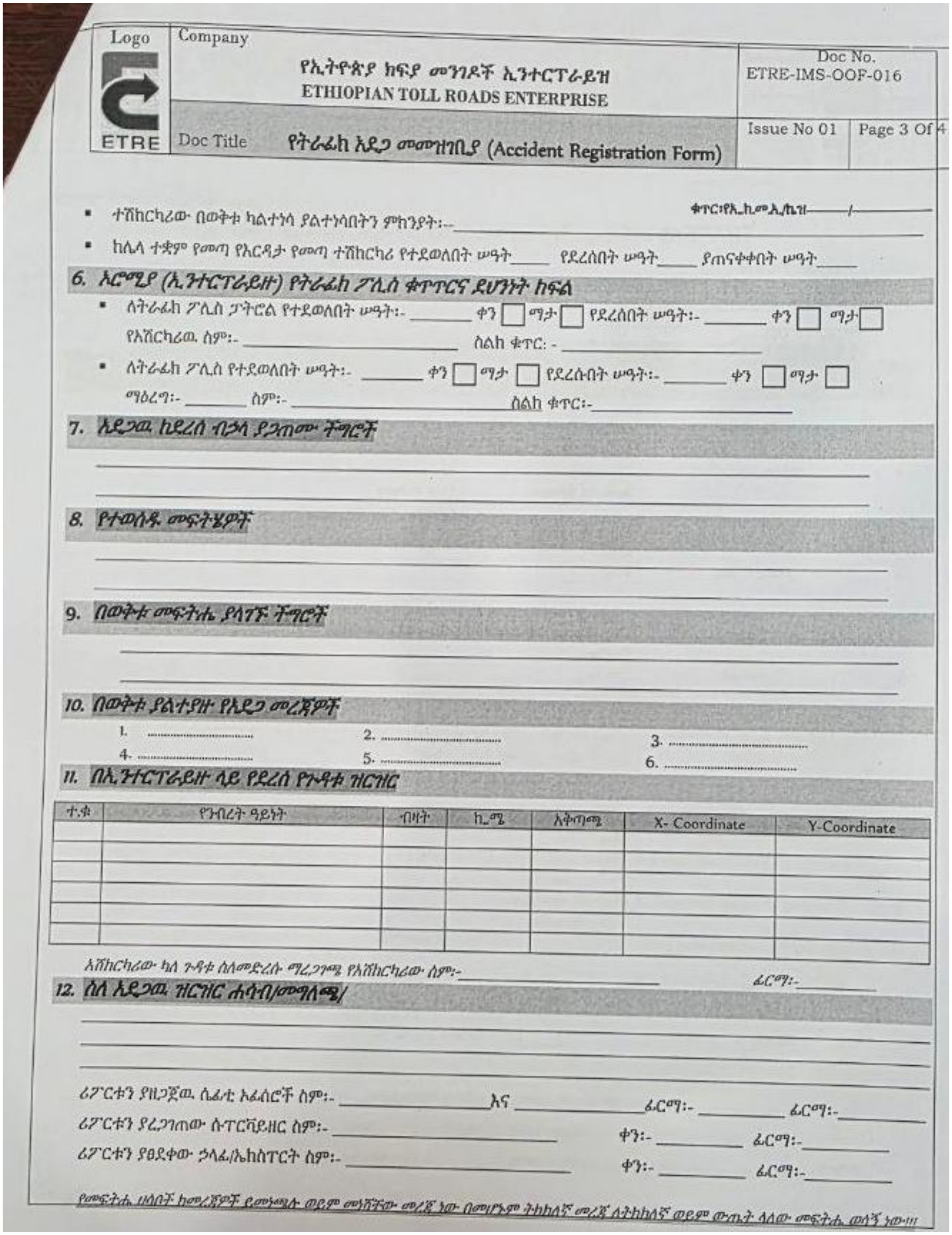
4. የመንገድ ደህንነት አስተዳደር ቡድን

- መረጃዉ የደረሰት ሠዓት፡-________ ቀን ☐ ማታ ☐ መረጃዉ ምንጭ (የተገኘዉ)፡- ከሞኒተር ☐ ከደንበኛ ☐ ከሴፍቲ ቡድን ☐
በፓትሮል ወቅት ☐ ከኢ.ኪ.መ.ኢ. ሌሎች ሠራተኞች ☐ ከመንገድ ጥበቃ አባላት ☐ በአደጋ ቦታ ላይ ☐ ሌላ ☐
- ፓትሮል ተሽከርካሪዉ የነበረበት ጣቢያ/ኪ.ሜ፡-________ አደጋዉ ቦታ የደረሰንበት ሠዓት፡-________ ቀን ☐ ማታ ☐
- መንገዱ መሉ ለመሉ ለትራፊክ ክፍት የሆነበት ሠዓት________ ቀን ☐ ማታ ☐

5. ድንገተኛ አደጋ አገልግሎት

- ለእርዳታ አምቡላንስ ተሽከርካሪው የተደወለበት ሠዓት፡-______ ቀን ☐ ማታ ☐ የነበረበት ጣቢያ/ኪ.ሜ፡-______
የደረሰበት ሠዓት ________ ቀን ☐ ማታ ☐ ለእረዳትታ ሌላ ዘዴ ከተጠቀሙ፡ __________
- አደጋ የደረሰባቸዉ ሠዎች የሄዱበት ሆስፒታል/ክሊኒክ/ቦታ፡- __________
- ለሬከር ክሬን የተደወለበት ሠዓት፡-______ ቀን ☐ ማታ ☐ የነበረበት ጣቢያ/ኪ.ሜ፡-________
የደረሰበት ሠዓት፡-______ ቀን ☐ ማታ ☐
- ለካቶ ክሬን የተደወለበት ሠዓት፡-______ ቀን ☐ ማታ ☐
- የነበረበት ጣቢያ/ኪ.ሜ፡-________ የደረሰበት ሠዓት፡-______ ቀን ☐ ማታ ☐

የመፍትሔ ሀሳቦች ከመረጃዎች ይመነጫሉ ወይም መነሻቸው መረጃ ነው በማንኛውም ትክክለኛ መረጃ ለትክክለኛ ወይም ውጤት ላለው መፍትሔ ወሳኝ ነው!!!



Logo	Company	Doc No.
ETRE	የኢትዮጵያ ክፍያ መንገዶች ኢንተርፕራይዝ ETHIOPIAN TOLL ROADS ENTERPRISE	ETRE-IMS-OOF-016
	Doc Title **የትራፊክ አደጋ መመዝገቢያ (Accident Registration Form)**	Issue No 01 | Page 3 Of 4

ቁጥር፡የኢ_ኪመኢ/ኪዝ______/________

- ተሽከርካሪው በወቅቱ ካልተነሳ ያልተነሳበትን ምክንያት:-________________________
- ከሌላ ተቋም የመጣ የእርዳታ የመጣ ተሽከርካሪ የተደወለበት ሠዓት_____ የደረሰበት ሠዓት_____ ያጠናቀቀበት ሠዓት_____

6. ኦሮሚያ (ኢንተርፕራይዞ) የትራፊክ ፖሊስ ቁጥጥርና ደህንነት ክፍል

- ለትራፊክ ፖሊስ ፓትሮል የተደወለበት ሠዓት:- _______ ቀን ☐ ማታ ☐ የደረሰበት ሠዓት:- _______ ቀን ☐ ማታ ☐
  የአሽርካሪዉ ስም:- ________________________ ስልክ ቁጥር: - ____________________
- ለትራፊክ ፖሊስ የተደወለበት ሠዓት:- _______ ቀን ☐ ማታ ☐ የደረሱበት ሠዓት:- _______ ቀን ☐ ማታ ☐
  ማዕረግ:- _______ ስም:- ________________________ ስልክ ቁጥር:-____________________

7. አደጋዉ ከደረሰ ብኃላ ያጋጠሙ ችግሮች

____

____

8. የተወሰዱ መፍትሄዎች

____

____

9. በወቅቱ መፍትሔ ያላገኙ ችግሮች

____

____

10. በወቅቱ ያልተያዙ የአደጋ መረጃዎች

1. .... 2. .... 3. ....
4. .... 5. .... 6. ....

11. በኢንተርፕራይዙ ላይ የደረሰ የጉዳቱ ዝርዝር

ተ.ቁ	የንብረት ዓይነት	ብዛት	ኪ.ሜ	አቅጣጫ	X- Coordinate	Y-Coordinate

አሽከርካሪው ካለ ጉዳቱ ስለመድረሱ ማረጋገጫ የአሽከርካሪው ስም:-____________________ ፊርማ:-__________

12. ስለ አደጋዉ ዝርዝር ሐሳብ/መግለጫ/

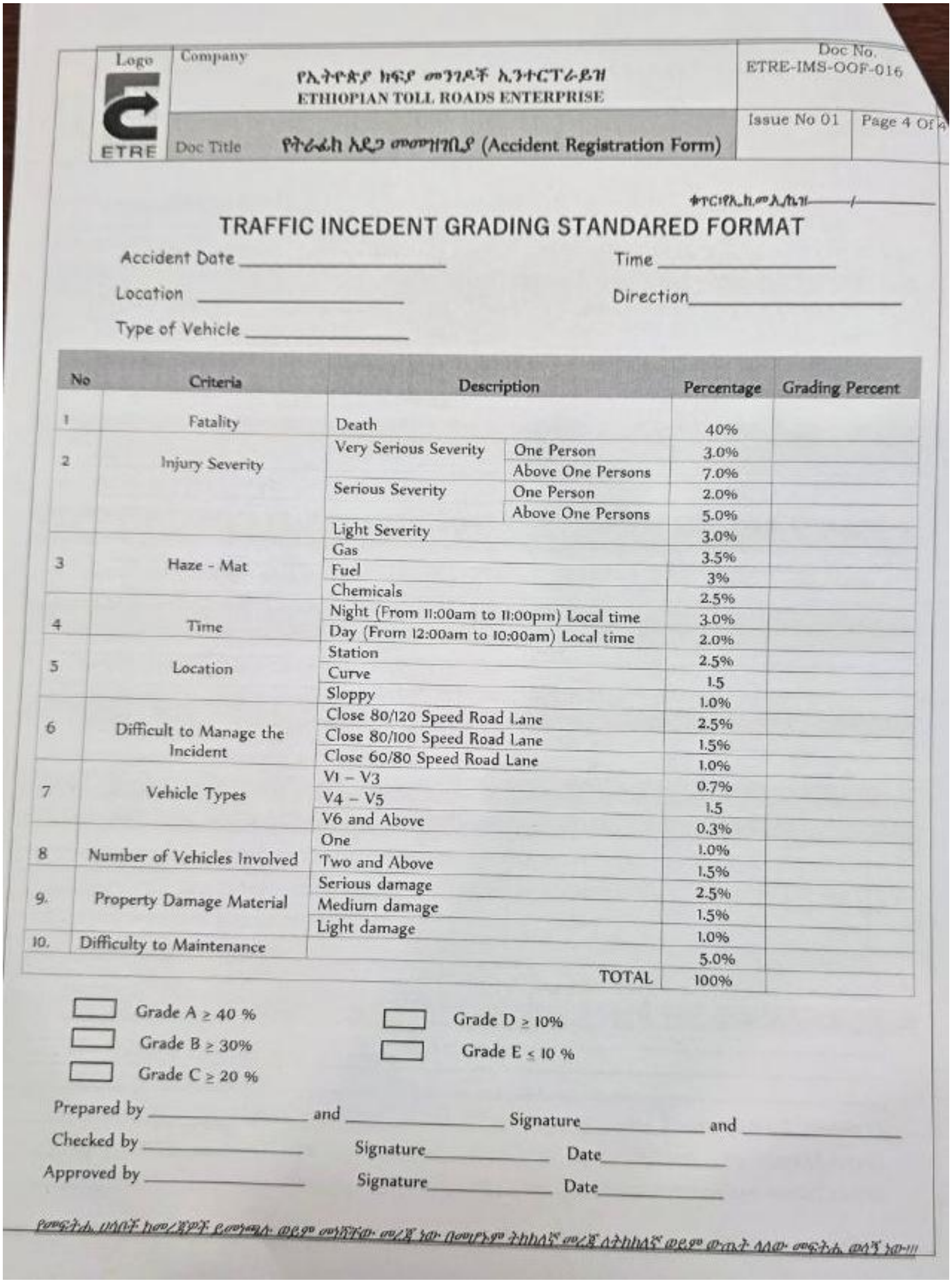
____

____

____

ሪፖርቱን ያዘጋጀዉ ሴፊቲ ኦፊሰሮች ስም:- ________________ እና ________________ ፊርማ:- __________ ፊርማ:-__________
ሪፖርቱን ያረጋገጠው ሱፐርቫይዘር ስም:- ________________________ ቀን:- __________ ፊርማ:-__________
ሪፖርቱን ያፀደቀው ኃላፊ/ኤክስፐርት ስም:- ________________________ ቀን:- __________ ፊርማ:-__________

የመፍትሔ ሀሳቦች ከመረጃዎች ይመነጫሉ ወይም መነሻችን መረጃ ነው በማሆኑም ትክክለኛ መረጃ ለትክክለኛ ወይም ውጤት ላለው መፍትሔ ወሳኝ ነው!!!



Logo	Company	Doc No.
ETRE	የኢትዮጵያ ክፍያ መንገዶች ኢንተርፕራይዝ ETHIOPIAN TOLL ROADS ENTERPRISE	ETRE-IMS-OOF-016
	Doc Title የትራፊክ አደጋ መመዝገቢያ (Accident Registration Form)	Issue No 01 · Page 4 Of 4

ቁጥር፡የኢ.ክ.መ.ኢ/ቢዝ ______/______

TRAFFIC INCEDENT GRADING STANDARED FORMAT

Accident Date ____________________ Time ____________________

Location ____________________ Direction ____________________

Type of Vehicle ____________________

No	Criteria	Description		Percentage	Grading Percent
1	Fatality	Death		40%	
2	Injury Severity	Very Serious Severity	One Person	3.0%	
			Above One Persons	7.0%	
		Serious Severity	One Person	2.0%	
			Above One Persons	5.0%	
		Light Severity		3.0%	
3	Haze - Mat	Gas		3.5%	
		Fuel		3%	
		Chemicals		2.5%	
4	Time	Night (From 11:00am to 11:00pm) Local time		3.0%	
		Day (From 12:00am to 10:00am) Local time		2.0%	
5	Location	Station		2.5%	
		Curve		1.5	
		Sloppy		1.0%	
6	Difficult to Manage the Incident	Close 80/120 Speed Road Lane		2.5%	
		Close 80/100 Speed Road Lane		1.5%	
		Close 60/80 Speed Road Lane		1.0%	
7	Vehicle Types	V1 – V3		0.7%	
		V4 – V5		1.5	
		V6 and Above		0.3%	
8	Number of Vehicles Involved	One		1.0%	
		Two and Above		1.5%	
9.	Property Damage Material	Serious damage		2.5%	
		Medium damage		1.5%	
		Light damage		1.0%	
10.	Difficulty to Maintenance			5.0%	
		TOTAL		100%	

☐ Grade A ≥ 40 %

☐ Grade B ≥ 30%

☐ Grade C ≥ 20 %

☐ Grade D ≥ 10%

☐ Grade E ≤ 10 %

Prepared by ______________ and ______________ Signature ______________ and ______________

Checked by ______________ Signature ______________ Date ______________

Approved by ______________ Signature ______________ Date ______________

የመፍትሔ ሀሳቦች ከመረጃዎች ይመነጫሉ ወይም መነሻቸው መረጃ ነው በመሆኑም ትክክለኛ መረጃ ለትክክለኛ ወይም ውጤት ላለው መፍትሔ ወሳኝ ነው!!!