



Road Accident Prediction Using Machine Learning Models

Prepared By

Name: [Abdulrahman Mohammed Abdulrazeq]

IDNo.: [252001637]

Submitted in fulfillment of the requirements for the degree
Master of Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Ashraf El-Sayed

College of Informatics, Midocean University, Comoros

[2025]

1. Abstract

Road traffic accidents, a prevalent public health concern, are not only sources of human distress but also exert substantial economic and societal impacts. The consequences of these accidents extend beyond the immediate medical and rehabilitation costs, encompassing property damage, productivity losses, and increased insurance premiums. The long-term effects on victims, families, and communities further exacerbate this burden. The precise prediction of accident severity thus emerges as a pivotal factor in mitigating the societal impacts of road traffic incidents. Leveraging ML and data analysis techniques enables the exploration of diverse factors contributing to traffic accidents, such as weather conditions, traffic patterns, road design, and driver behavior. These factors are integral to the development of accurate models for predicting accident severity, which, in turn, aid emergency services in delivering prompt and effective responses. A notable benefit of such accurate predictions is the enhanced management of emergency resources, allowing for prioritization based on the severity of incidents. This leads to swifter response times and improved outcomes for those involved in accidents. Additionally, precise predictions can inform transportation management strategies and the design of safer road networks, ultimately reducing the likelihood and severity of future accidents.

2. Introduction

The alarming rise in worldwide road traffic accidents poses serious challenges to numerous aspects of human existence. However, there has been a lack of attention paid to the important aspects of traffic characteristics, causation analysis, accident severity analysis, and the relationships between diverse causal factors. The likelihood of accidents on different types of roads can be a significant analytical resource for traffic analysts and researchers aimed at identifying the major elements of road accident severity to improve road safety for automobiles as well as individuals. Using effective Machine Learning Models Like Random Forests and Decision Trees for the tabular data to predict the severity of road accidents, helps analyzing different cause factors and the direct importance of these factors (features)

3. Background:

Road traffic accidents represent one of the leading causes of death worldwide. According to the World Health Organization (WHO) [1], almost 1.3 million people die in road traffic accidents, while almost 50 million suffer non-fatal injuries. The leading cause of death of children and young people between 5 and 29 years is road traffic crashes. The serious problem is that 93% of fatalities come from the low-income and middle-income countries in development. Except for human suffering, road traffic accidents significantly influence the economy, costing countries up to 3% of their Gross Domestic Product (GDP). The occurrence of traffic accidents can be related to a variety of factors. Some of these factors are subjective: driver knowledge, training level, experience, the influence of alcohol, drugs, etc., while others are objective, such as: road volume (annual average daily traffic volume—AADT), road geometry (curvature, slopes, lane width, shoulder width), type of road (freeway or two-lane road), road conditions (pavement quality and potential damage of pavement surface), weather conditions (wind, ice, snow, rain, etc.), cars maintenance, speed limits, frequency of police controls, etc. An increase in a number of

road vehicles leads to heavier traffic volume on almost all roads, which increases the chances of collision between vehicles. On the other hand, the same traffic volume on the freeway and the two-lane roads have a different possibility of collision. This variety of factors makes complicated efforts to predict the number of traffic accidents for new roads or reconstruction of existing roads. In order to predict the number of traffic accidents in the near future on the existing or the newly planned/built road section, it is necessary to know what factors will mostly influence the occurrence of a road traffic accident. Objective factors that cannot be easily changed in the future are recognized as: the type of road, road geometry and AADT (which will likely to be increased in the future). Other factors can be changed (pavement can be repaired, weather conditions are subject to daily change, car maintenance can be improved by frequent checks, speed limits can also be adapted, etc.).

The classical machine learning models such as: artificial neural network (ANN), random forest regression (RFR), support vector machine (SVM), extreme learning machine (ELM), K-nearest neighbors (KNN) and decision tree (DT) are extensively used in modelling in various branches of science. The SVM is a widely used discriminant technique based on the statistical learning theory, well recognized for its strong generalization ability. The optimal network is obtained by exploring the balance between the complexity of the model and the training error. The ELM designs a single-layer feedforward network by randomly generating the input weights and biases of the hidden layers.

The vast variety of state-of-the-art machine learning techniques is suitable for sequence data like ensemble learning models, such as: XGBoost and LightGBM and CatBoost. XGBoost model exerts its advantages, especially for high prediction accuracy and interpretability. LightGBM model enables large amounts of data and GPU training. The LightGBM models are proven to be more accurate and faster than XGBoost. Furthermore, data fusion enables stronger forecasting accuracy, according to the integration of gradient boosting-based categorical attributes supported by the CatBoost algorithm. Applying artificial neural networks have proved its feasible feature in recent years by predicting and presenting desired results. Obtained results illustrate that the variables such as Road curvature, number of lanes, Weather and lighting conditions, Presence of road signs, Holiday or school season flags, Speed limits and time of day are the most influential aspects that can raise traffic accidents in urban roads.

4. Related works:

Dragan Gatarić & partners *1, found The developed optimal neural network models showed good generalization capabilities for the collected data foresee, and could be used to accurately predict the observed outputs, based on the input parameters. The highest values of r^2 for developed models ANN1 and ANN2 were 0.986, 0.988, and 0.977, and 0.990, 0.969, and 0.990, accordingly, for training, testing and validation cycles.

Koteswararao Kodepogu*2 & partner's findings advocate for the continuous development and refinement of sophisticated models, aiming to diminish the frequency and gravity of road accidents.

Such efforts contribute significantly to the enhancement of traffic control and public safety measures, it was observed that AdaBoost and Naive Bayes demonstrated the highest accuracy. Each of these models achieved an accuracy rate of 85.3%, marking them as the most precise among the five models examined

Jing Gan^{*3} and partners proposed a novel method for predicting the severity of traffic accidents based on the Deep Forests algorithm. To verify the superiority of our proposed method, several other machine learning algorithm-based prediction models were implemented to predict traffic accident severity with the same dataset, and the prediction results show that the Deep Forests propose a novel method for predicting the severity of traffic accidents based on the Deep Forests algorithm. To verify the superiority of our proposed method, several other machine learning algorithm-based prediction models were implemented to predict traffic accident severity with the same dataset, and the prediction results show that the Deep Forests algorithm present good stability, fewer hyper-parameters, and the highest accuracy under different level of training data volume. Recall score: 0.93 and F1: 0.88 It is expected that the findings from this study would be helpful for the establishment or improvement of effective traffic safety system within a sustainable transportation system

MUBARIZ MANZOOR and partners ^{*4} mentioned that the performance of the proposed approach Obtained results demonstrate that the RFCNN enhanced the decision-making process and outperformed other models with 0.991 accuracy, 0.974 precision, 0.986 recall, and 0.980 F-score using the 20 most Significant features in predicting the severity of accidents.

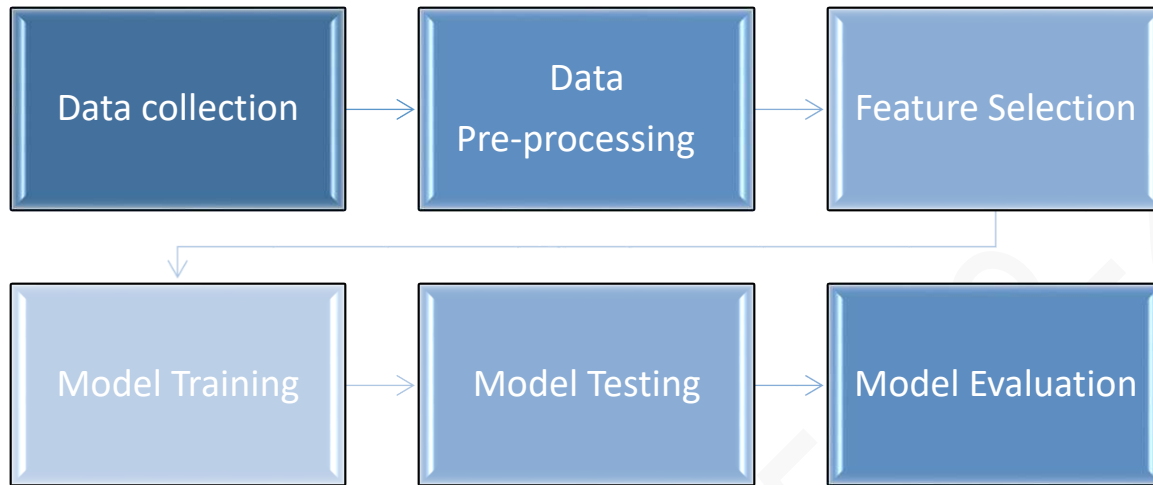
5. Research objectives:

The research delineates the foundational concepts, theoretical frameworks, and data sources prevalent in the field. The findings advocate for the continuous development and refinement of sophisticated models, aiming to diminish the frequency and gravity of road accidents. Such efforts contribute significantly to the enhancement of traffic control and public safety measures.

The overarching goal of road accident severity prediction is:

1. The early and precise determination of an accident's severity
2. Facilitating prompt and efficient deployment of emergency response teams.
3. Minimize fatalities, injury severity, and the economic burden of these incidents.
4. Aids policymakers in identifying high-risk areas and implementing appropriate safety measures.
5. Foster the development of preventive strategies, such as enhanced traffic management, road design improvements, and public awareness campaigns, to reduce accident likelihood.
6. The ultimate objective is to elevate emergency response efficiency, decrease the frequency of collisions, and enhance overall traffic safety

6. Research Methodology



7. Expected outcomes:

1. Accurate and advanced Machine Learning model can precisely predict the severity of road accidents accordance based on multiple combined factors falls under different categories such as: Road State, Driver features, Climate situation.
2. Feature extraction where we can find the most features can affect the accident accordance and severity, where government entities can take actions to prevent these accidents by applying the multiple safety measures and procedures.

8. Gantt chart

Tasks	January 2026	February 2026	March 2026	April 2026
Data Collection				
Data Pre-Processing				
Feature Selection				
Model Training				
Model Testing				
Thesis Write-up				
Model Evaluation				

9. References:

1. Predicting Road Traffic Accidents—Artificial Neural Network Approach Dragan Gatarić 1, Nenad Ruškić 2 , Branko Aleksić 1, Tihomir Đurić 1, Lato Pezo 3 and Milada Pezo 5
<https://www.mdpi.com/2298222>
2. Machine Learning for Road Accident Severity Prediction Koteswararao Kodepogu1* , Vijaya Bharathi Manjeti2 , Atchutha Bhavani Siriki3 <https://doi.org/10.56578/mits020403>
3. An Alternative Method for Traffic Accident Severity Prediction: Using Deep Forests Algorithm Jing Gan ,1Linheng Li ,1Dapeng Zhang ,2Ziwei Yi ,1andQiaojun Xiang
https://www.researchgate.net/publication/347689690_An_Alternative_Method_for_Traffic_Accident_Severity_Prediction_Using_Deep_Forests_Algorithm
4. RFCNN: Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model MUBARIZ MANZOR1, MUHAMMAD UMER SALEEM ULLAH 2,3, SAIMA SADIQ 2,ABID ISHAQ2, 2,HAMZAAHMADMADNI 4,ANDCARMENBISOONI
<https://ieeexplore.ieee.org/document/9536744>