

# Analysis and Prediction of Road Accident Severity Using ML

M.Rajalakshmi\*

\*(Department of Computer Science, Sri Kaliswari College of Arts and Science, Sivakasi  
Email: [kaliswaricollege@gmail.com](mailto:kaliswaricollege@gmail.com))

\*\*\*\*\*

## Abstract:

Road accidents are a major cause of injuries and fatalities worldwide. Early prediction of accident severity can help traffic authorities implement preventive measures and improve road safety. This study focuses on the analysis of historical road accident data and the development of machine learning models to predict accident severity levels. The dataset includes factors such as weather conditions, time of accident, road type, traffic density, and vehicle type. Data preprocessing techniques including data cleaning, encoding, and feature selection are applied to prepare the dataset for model training. Machine learning algorithms such as Decision Tree, RandomForest are used to classify accident severity into three categories: Low, Medium, and High. The performance of the models is evaluated using Accuracy, Precision, Recall, and F1-score. A comparative analysis is conducted to determine the best-performing algorithm. Experimental results demonstrate that the proposed approach effectively predicts accident severity and supports data-driven decision-making. This system can assist traffic management authorities in identifying high-risk conditions and taking proactive measures to reduce accident severity and enhance road safety.

*Keywords* —Road Accident, Accident Severity, Machine Learning, Prediction, Random Forest, Classification..

\*\*\*\*\*

## I. INTRODUCTION

Road accidents are a major public safety concern worldwide, leading to serious injuries and fatalities. Predicting the severity of road accidents in advance can help authorities take preventive measures and improve road safety. Various factors such as weather conditions, time of accident, road type, and traffic density influence accident severity.

Machine learning techniques provide an effective way to analyze accident data and identify hidden patterns. In this study, Decision Tree and Random Forest algorithms are used to classify accident severity into three categories: Low, Medium, and High. These algorithms are selected due to their efficiency, interpretability, and strong performance in classification problems. The proposed system analyzes

historical accident data, applies preprocessing techniques, and evaluates model performance using metrics such as Accuracy, Precision, Recall, and F1-score. The best-performing model is identified based on comparative analysis.

## II. Motivation:

Road traffic accidents continue to pose significant challenges for public safety, leading to injuries, fatalities, and economic loss. Despite advances in road infrastructure and traffic management, emergency response systems are often delayed due to congestion and lack of prior information. This research is motivated by the need to leverage large-scale accident data to predict accident severity, identify accident hotspots, and optimize rescue pathways. By integrating machine learning techniques with spatial analysis and shortest path algorithms, the study aims to support authorities in proactive

planning, improve emergency response efficiency, and ultimately reduce fatalities and injuries on roads.

### III Objectives:

**\*Predict the severity of road accidents** (Minor, Serious, or Fatal) using Machine Learning algorithms applied to historical accident data.

**•Analyze the key factors influencing accident severity:** including weather conditions, time of day, road conditions, traffic density, and vehicle type.

**•Identify high-risk locations and conditions:** where severe accidents are more likely to occur

**•Provide actionable insights for authorities:** to enhance road safety, optimize emergency response, and reduce accident-related injuries and fatalities.

#### **Problem Statement:**

Road accidents are a leading cause of injuries, fatalities, and property damage worldwide. The severity of an accident is influenced by multiple factors, including weather conditions, time of day, road type, traffic density, and driver behavior. Conventional methods are often unable to accurately predict accident severity in advance, limiting the effectiveness of preventive measures.

This study aims to **analyze historical accident data** and apply **Machine Learning algorithms** to predict the severity of road accidents as Minor, Serious, or Fatal. By identifying high-risk conditions and locations, the proposed system can contribute to enhanced road safety, optimized emergency responses, and reduced accident-related losses.

### **Machine Learning Using Python:**

#### **3.1 Data Preprocessing:**

\*The historical accident dataset was cleaned to handle missing values and inconsistencies.

\*Categorical variables such as weather, road type, and vehicle type were encoded using **Label Encoding** and **One-Hot Encoding**.

\*Features were normalized or scaled where necessary to improve model performance.

\*The dataset was split into **training** and **testing** sets to evaluate model performance.

#### **3.2 Machine Learning Algorithms**

Several supervised classification algorithms were implemented using Python:

**1. Decision Tree** – A tree-based model that splits data based on feature values to predict accident severity.

**2. Random Forest** – An ensemble of decision trees that improves prediction accuracy and reduces overfitting.

\*The models were evaluated using **accuracy, precision, recall, and F1-score** metrics.

\***Confusion matrices** were plotted to visualize prediction performance across Minor, Serious, and Fatal classes.

\*The model with the highest overall performance was selected for predicting accident severity.

#### **2. RELATED WORKS**

Road accident severity prediction has attracted significant research attention due to the increasing number of traffic-related injuries and fatalities worldwide. Researchers have applied various statistical and Machine Learning techniques to analyze accident datasets and identify factors influencing accident severity levels.

Quinlan (1986) introduced the Decision Tree algorithm, which has been widely used for classification tasks because of its interpretability and simplicity. Decision Trees have been applied in accident severity analysis to evaluate the impact of factors such as weather conditions, road type, and vehicle characteristics.

Breiman (2001) proposed the Random Forest algorithm, an ensemble learning approach that enhances prediction accuracy by combining multiple decision trees. Random Forest has demonstrated strong performance in traffic accident severity prediction due to its ability to handle large datasets and complex feature interactions.

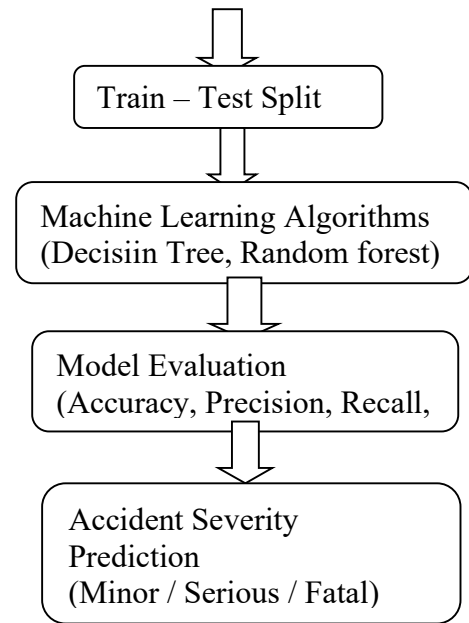
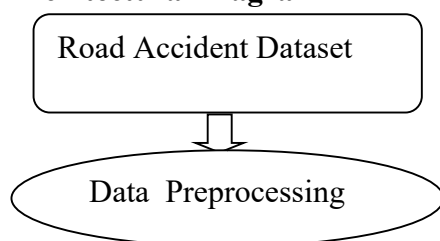
Recent studies have explored advanced Machine Learning models such as Support Vector Machines (SVM), Logistic Regression, and XGBoost for multi-class accident severity classification. These models have been used to analyze temporal, environmental, and traffic-related factors, showing that ensemble and boosting techniques often outperform traditional classification models.

In addition, spatial clustering methods such as DBSCAN and K-Means have been utilized to identify accident hotspots and high-risk zones. These approaches help authorities implement targeted safety interventions and improve emergency response strategies.

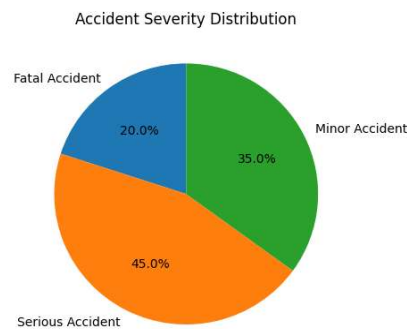
However, many existing studies focus on single-model approaches or limited feature analysis. The present study aims to compare multiple Machine Learning algorithms and provide a comprehensive analysis of accident severity factors to enhance road safety planning and decision-making.

### 3. IMPLEMENTATION AND WORKING

#### 3.1 Architectural Diagram



#### 3.2 Data Collection and Data Cleaning



The pie chart illustrates the distribution of accident severity in the dataset used for the study. Accident severity is categorized into three classes: Fatal Accident, Serious Accident, and Minor Accident. From the visualization, it can be observed that **Serious Accidents account for the largest proportion with 45%**, indicating that a significant number of accidents result in major injuries or considerable damage. **Minor Accidents represent 35% of the total cases**, which generally involve limited damage or minor injuries. **Fatal Accidents**

**constitute 20% of the dataset**, representing the most critical cases where loss of life may occur.

This distribution analysis helps in understanding the overall pattern of accident severity within the dataset and provides valuable insights for developing machine learning models to predict accident severity. By analyzing these categories, the proposed model can effectively classify accidents into different severity levels and assist in improving road safety analysis.

### 3.3 Data Preprocessing and Analysis of Data

Data preprocessing is an important step in machine learning. In this stage, the raw accident dataset is cleaned and prepared for analysis and model training. Missing values and duplicate records in the dataset are identified and removed to improve data quality. Categorical data such as weather conditions, road type, and accident severity are converted into numerical form using encoding techniques. This process helps the machine learning algorithms to understand the data effectively. After preprocessing, the dataset becomes structured and suitable for further analysis and prediction.

Data analysis is performed to understand the characteristics and patterns of road accidents in the dataset. Various visualization techniques such as bar charts, pie charts, and accuracy comparison graphs are used to analyze the data. For example, the distribution of accident severity (minor, serious, and fatal) is visualized using a pie chart to understand the proportion of each category. This analysis helps in identifying important factors influencing accident severity and supports the development of an accurate machine learning model for predicting accident severity.

### 3.4 Data Visualization

Data visualization is used to represent the dataset in graphical form for better understanding and analysis. It helps in identifying patterns, trends, and relationships within the data. In this project, different visualization techniques such as pie charts

and bar graphs are used to analyze the distribution of road accident severity.

The visualization helps to understand how accidents are distributed among different categories such as **Fatal, Serious, and Minor accidents**. For example, a pie chart is used to display the percentage distribution of accident severity in the dataset. These visual representations make it easier to interpret the data and identify important patterns.

Data visualization also supports the evaluation of machine learning models by comparing their performance using graphical representations such as accuracy comparison graphs. Overall, visualization plays an important role in analyzing the dataset and presenting the results clearly.

### 3.5 Cross Validation and Training the Model

In this study, machine learning models are trained to predict the severity of road accidents. The dataset is divided into training and testing sets to build and evaluate the models. The training data is used to train the algorithms so that they can learn patterns from the dataset.

To improve the reliability and accuracy of the model, cross validation is applied. Cross validation divides the dataset into multiple subsets and the model is trained and tested several times using different data splits. This helps to reduce overfitting and improves the model's generalization ability.

In this project, **Decision Tree and Random Forest algorithms** are used for training the model. The Decision Tree algorithm classifies accident severity based on different conditions such as weather, road type, and traffic conditions. Random Forest improves prediction accuracy by combining multiple decision trees.

The performance of the models is evaluated using metrics such as **Accuracy, Precision, Recall, and F1-score**. These evaluation metrics help in

identifying the best model for predicting road accident severity.

2	0.95	0.96	0.95	99
---	------	------	------	----

### 3.6 Testing and Integration with UI

After the training phase, the developed machine learning models are evaluated using the testing dataset to verify their prediction capability. The testing process ensures that the trained models can accurately classify accident severity levels based on the given input features. The performance of the models is assessed using evaluation metrics such as **accuracy, precision, recall, and F1-score**.

Once the models are tested and validated, the system is integrated with a simple **User Interface (UI)** to make the prediction system easy to use. The user interface allows users to input accident-related parameters such as road conditions, weather conditions, traffic information, and time of accident. Based on these inputs, the trained machine learning model processes the data and predicts the **severity level of the accident (Minor, Serious, or Fatal)**.

The integration of the machine learning model with a user interface improves usability and enables real-time prediction of accident severity. This approach helps users and authorities to analyze accident risk and take preventive safety measures effectively.

### 4.RESULT

Model	Accuracy
Decision Tree	94.8%
Random Forest	92.0%

Class	Precision	Recall	F1-Score	Support
0	0.94	0.94	0.94	107
1	0.91	0.89	0.90	44

### 5.CONCLUSION

In this study, machine learning techniques were applied to analyze and predict road accident severity using historical accident data. Data preprocessing techniques such as data cleaning and feature selection were performed to prepare the dataset for model training. The machine learning algorithms **Decision Tree and Random Forest** were used to classify accident severity into three categories: **Low, Medium, and High**.

The performance of the models was evaluated using metrics such as **Accuracy, Precision, Recall, and F1-score**. The experimental results showed that both models performed well in predicting accident severity. However, the **Decision Tree model achieved slightly higher accuracy than the Random Forest model**.

The results demonstrate that machine learning models can effectively predict accident severity based on various accident-related factors. This system can help traffic management authorities analyze accident patterns and take preventive measures to improve road safety. In future work, the system can be enhanced by using larger datasets and advanced machine learning techniques to further improve prediction accuracy.

### REFERENCES

- [1] Noushin Behboudia, Sobhan Moosavib, Rajiv Ramnathb. Recent Advances in Traffic Accident Analysis and Prediction: A Comprehensive Review of Machine Learning Techniques\* arXiv:2406.13968v1(2024).
- [2] Koteswararao Kodepogu1\*,Vijaya Bharathi Manjeti2, Atchutha Bhavani Siriki3. Machine Learning for Road Accident Severity Prediction[https://doi.org/10.56578/mits020403\(2023\)](https://doi.org/10.56578/mits020403(2023)).
- [3] Izuchukwu Chukwuma Obasi \*,Chizubem Benson. Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents. Heliyon 9 (2023) e18812.
- [4] Jiming Tang 1, Yao Huang 1,2,\* , Dingli Liu 1,\* , Liuyuan Xiong 1 and Rongwei Bu 1. Research on Traffic Accident Severity Level Prediction

Model Based on Improved Machine Learning.<https://doi.org/10.3390/systems13010031>(2025).

- [5] Mohammad Yousuf Sulaiman AlHashmi."Using Machine Learning for Road Accident Severity Prediction and Optimal Rescue Pathways"(2024).
- [6] 1Seela Vidya Kumari, 2A. Gautami Latha."ROAD ACCIDENT SEVERITY PREDICTION USING MACHINE LEARNING". ISSN 2457-0362(2024).
- [7] Liridon Sejdiu 1, Tomaz Tollazzi 2, Ferat Shala 1, Halil Demolli 3\*\*"Analysis of Traffic Safety Factors and Their Impact Using Machine Learning Algorithms"(E-ISSN: 2476-3055; ISSN: 2676-6957)(2024).
- [8] 1\*Tushar Goud, 2Pallavi Kumari, 3Vedant Patil, 4Rishabh Tiwari"ROAD ACCIDENT SEVERITY PREDICTION USING MACHINE LEARNING"2024 JETIR April 2024, Volume 11, Issue 4.
- [9] Artur BUDZYŃSKI, Mikołaj FEDEROWICZ, Adam JABŁOŃSKI,Wesam HASAN,Jan GORSZANÓW, Taras NIKITISHYNA machine learning approach for predicting road accidents10(2) (2024).
- [10] Md. Asif Khan Rifata, Armana Huqb "An Explainable Machine Learning Approach to Traffic, Accident Fatality Prediction"246 (2024).
- [11] Shah Alam1, Faiaz Rahat Chowdhury2,Md Sakibul Hasan3, Saddam Hossain4,Tanaya Jakir,Arat Hossain6,Arifur Rahman7, Mohammad Nazmul Hossain8,Syed Nurul Islam9." Intelligent Streetlight Control System Using Machine Learning Algorithmsfor Enhanced Energy Optimization in Smart Cities" Volume: 4, No: 4, pp. 543 – 564(2025).