ISSN 2277-2685
IJESR/April-June. 2026/ Vol-16/Issue-2/1142-1146
Mohammed Salman Ateef *et. al.*, / International Journal of Engineering & Science Research
Full Length Article

Road Accident Analysis and Accident Severity Prediction

Mohammed Salman Ateef[1], Mohd Abdul Danish[2], Mohd Abdul Jameel[3], Dr. K. Nagi Reddy[4]
[1,2,3]B.E. Students, Department of IT, Lords Institute of Engineering and Technology, Hyderabad, India.
[4]Professor and Head, Department of IT, Lords Institute of Engineering and Technology, Hyderabad, India.
Mail Id; k.nagireddy@lords.ac.in[4]
Accepted 28-04-2026

Abstract
Due to the exponentially increasing number of vehicles on the road, the number of accidents occurring on a daily basis is also increasing at an alarming rate. With the high number of traffic incidents and deaths these days, the ability to forecast the number of traffic accidents over a given time is important for the transportation department to make scientific decisions. In this scenario, it will be good to analyze the occurrence of accidents so that this can be further used to help us in coming up with techniques to reduce them. Even though uncertainty is a characteristic trait of majority of the accidents, over a period of time, there is a level of regularity that is perceived on observing the accidents occurring in a particular area. This regularity can be made use of in making well informed predictions on accident occurrences in an area and developing accident prediction models. In this Work, we have studied the inter relationships between road accidents, condition of a road and the role of environmental factors in the occurrence of an accident. We have made use of data mining techniques in developing an accident prediction model using Apriori algorithm and Support Vector Machines. Bangalore road accident datasets for the years 2014 to 2017 available in the internet have been made use for this study. The results from this study can be advantageously used by several stakeholders including and not limited to the government public work departments, contractors and other automobile industries in better designing roads and vehicles based on the estimates obtained.

Keywords: *Road Accident Prediction, Data MIning, Apriori Algorithm, Support Vector Machines, Bangalore Accident Dataset, Traffic forecasting.*

INTRODUCTION

The alarming rate of increase of accidents in India is now a cause for serious concern. According to some recent statistics [1], India accounts for roughly six percent of global road accidents while owning only one percent of the global vehicle population. There are a lot of accident cases reported due to the negligence of two-wheelers, whereas over-speeding is also another contributing factor. Accidents caused while under the influence of alcohol or during general traffic violations are also common. In spite of having set regulations and the highway codes, the negligence of people towards the speed of the vehicle, the vehicle condition and their own negligence of not wearing helmets has caused a lot of accidents. While the major cause of road accidents is attributed to the increasing number of vehicles, the role played by the condition of the roads and other environmental factors cannot be overlooked. The number of deaths due to road accidents in India is indeed a cause for worry. The scenario is very dismal with more than 137,000 people succumbing to injuries from road accidents. This figure is more than four times the annual death toll from terrorism.

Accidents involving heavy goods vehicles like trucks and even those involving commercial vehicles used for public transportation like buses are some of the most fatal kind of accidents that occur, claiming the lives of innocent people. Weather conditions like rain, fog, etc., also play a role in catalysing the risk of accidents. Thus, having a proper estimation of accidents and knowledge of accident hotspots and causing factors will help in taking steps to reduce them. This requires a keen study on accidents and development of accident prediction models. To implement a well-designed road framework management system for looking into road security aspects, it is often desired to have an optimized accident prediction model which can analyze potential issues arising due to infrastructure fallbacks and to estimate the effect of existing models in reducing the occurrence of accidents.

The main challenges involved in the creation of such a model include the evaluation of the weight that can be attributed to the impact of each variable in contributing to the accident and assessing how the model can be best designed to incorporate the effects of all such variables. we have investigated the inter-relationship between the occurrences of

road accidents and the roles played by the underlying road conditions and environmental factors in contributing to the same. Since such a study requires us to cover several aspects affecting accidents, we can make use of data mining techniques to analyze this data to extract relevant details from them, as these huge volumes of data would otherwise be meaningless without the right interpretation applied to them.

In this Work, we are discussing the effects of such an accident prediction model in identifying the risks involved in road accident scenarios. The next section discusses the prior works done with respect to analyzing the different accidents that have taken place over the years. This is followed by a summarized description of the methodology used in this work. Further, the different components of implementation including the system architecture, software and languages used, simulation, user interface and screenshots of the developed application are discussed. Finally, the discussion and conclusions derived from the present study and the future scopes are outlined in the last two sections. The results from this study have been used to propose a model that can be used as a tool to estimate the possibility of road accidents in a particular area chosen by the user.

Related Work

The work titled "AI-based prediction of traffic crash severity for improving road safety and transportation efficiency" presents an advanced framework that utilizes artificial intelligence techniques to model complex relationships among traffic, environmental, and behavioral factors. The study emphasizes the capability of AI models to outperform traditional statistical approaches in predicting crash severity and highlights their usefulness in improving real-time traffic management and safety interventions. [1]. Another work titled "A methodological framework for road accident severity prediction for Indian highways using machine learning models" proposes a structured approach tailored to Indian traffic conditions. The study incorporates data pre-processing, feature selection, and classification techniques to address heterogeneous traffic patterns and infrastructure variability, demonstrating improved prediction performance and relevance for localized safety planning. [2]. Another research work titled "Road Crash Injury Severity Prediction Using Ensemble Learning" focuses on the application of ensemble techniques such as bagging and boosting to enhance prediction accuracy. The study shows that combining multiple classifiers improves robustness and effectively handles imbalanced datasets, which is a common challenge in accident severity prediction. [3]. The study titled "Meta-feature-based traffic accident risk prediction: A novel approach to forecasting severity and incidence" introduces a feature engineering approach where higher-level meta-features are derived from raw data. This enhances the learning capability of predictive models and improves generalization, making it effective in forecasting both accident occurrence and severity levels. [4]. Another significant work titled "Real-time crash risk prediction on freeways using deep learning" applies deep neural networks to capture spatiotemporal traffic patterns. The study demonstrates that deep learning models can effectively predict crash risk. [5]. The work titled "Severity prediction of road accidents using machine learning techniques" evaluates various classification algorithms and compares their effectiveness in predicting accident severity. The findings indicate that machine learning approaches provide higher accuracy and better performance than traditional statistical models. [6]. Another study titled "A review on road accident detection using data mining techniques" provides a comprehensive overview of different data mining methods applied to accident analysis. It discusses classification, clustering, and association techniques while highlighting key challenges such as data imbalance and feature selection. [7]. The research work titled "Analysis of road accidents in India using data mining classification algorithms" applies multiple classification techniques to identify patterns in accident data. The study demonstrates how data mining approaches can uncover hidden relationships and support decision-making in road safety management. [8]. Another work titled "Use of accident prediction models in road safety management – an international inquiry" reviews various predictive models used globally for accident analysis. It highlights their role in policy-making and emphasizes the importance of standardized modelling approaches for improving road safety outcomes. [9]. The study titled "Predictive model for incident occurrences in steel plant in India" focuses on identifying risk factors and predicting incident likelihood using data-driven techniques. Although applied in an industrial context, the methodology is relevant and adaptable to road accident severity prediction. [10]. Another research work titled "Text mining analysis of railroad accident investigation reports" utilizes natural language processing techniques to extract insights from unstructured accident reports. The study shows how textual data can be used to identify contributing factors and enhance accident analysis models. [11]. Finally, the work titled "Text mining to decipher free-response consumer complaints: Insights from the NHTSA vehicle owner's complaint database" explores the use of text mining to analyze unstructured complaint data. It highlights how integrating textual information with structured datasets can improve the

understanding and prediction of accident-related issues. [12]

Research Methedology

The proposed system for road accident severity prediction follows a structured approach that includes data collection, preprocessing, model training, and evaluation. The methodology used in this Work is described as follows.

1. Dataset Identification

The Road Traffic Accident dataset from Kaggle was used in this study. The dataset was prepared from manual records of road traffic accidents from the years 2017 to 2020. Sensitive information was excluded during data encoding. The dataset consists of 32 features and 12316 instances related to accident severity.

2. Data Preprocessing and Exploratory Data Analysis

Exploratory Data Analysis was performed to clean the dataset and identify patterns and relationships between various parameters. The following steps were performed:

Removal of duplicate values
Filling missing values
Detection of outliers
Transformation of categorical features using label encoding
Conversion of numerical values into nominal values
Dropping unnecessary attributes
Balancing the dataset using SMOTE

3. Model Development

The preprocessed dataset was split into training and testing datasets. Feature selection was performed using the Chi-Squared method. The dataset was then fed into multiple machine learning classification algorithms such as:

- Random Forest
- Decision Trees
- Support Vector Machine
- Logistic Regression
- K-Nearest Neighbour
- Gaussian Naïve Bayes
- AdaBoost
- Gradient Boosting

The performance of each algorithm was evaluated and compared.

4. Model Training and Evaluation

After comparing different algorithms, the best performing model was selected. Hyperparameter tuning was performed using GridSearchCV to improve the performance of the selected model. The final classifier was created using the Random Forest algorithm.

The performance of the models was evaluated using accuracy, classification reports, and confusion matrices. The final model was deployed using Flask for prediction and analysis.

Result

In this Work, multiple machine learning algorithms were applied to the preprocessed Road Traffic Accident dataset in order to predict accident severity. The dataset was first preprocessed by detecting outliers, imputing missing values, encoding categorical data, splitting the data into training and testing sets, and performing feature selection using the Chi-Squared method. The processed dataset was then fed i–nto various machine learning algorithms.

The performance of different algorithms was evaluated using accuracy. The tuned Random Forest classifier was selected as the final model, which achieved an accuracy of 91.68%, outperforming other models in predicting accident severity. The confusion matrices and classification reports were used to evaluate the performance of each model. The results indicate that the proposed system can effectively predict the severity of road accidents with high accuracy .

Table 1: Key Findings

Model	Accuracy %
Decision Trees	69.19%
Random Foreset	84.54%
Support Vector Machine	84.38%
Logistic Regression	84.37%
KNN	82.63%
Gradient Boosting Classifier	84.9%
AdaBoost	84.25%

Fig.1. Accuracy of Model

Discussion

An accident can change the lives of many people. It is up to each of us to bring down this increasing number. This can be made possible by adopting safe driving measures to an extent. Since all instances of accidents cannot be attributed to the same cause, proper precautionary measures will also need to be exercised by the road development authorities in designing the structure of roads as well as by the automobile industries in creating better fatality reducing vehicle models. One thing within our capability is to predict the possibility of an accident based on previous data and observations that can aid such authorities and industries.

This Work was successful in creating such an application that can help in efficient prediction of road accidents based on factors such as types of vehicles, age of the driver, age of the vehicle, weather condition and road structure, so on. This model was implemented by making use of several data mining and machine learning algorithms applied over a dataset for Bangalore and has been successfully used to predict the risk probability of accidents over different areas with high accuracy.

The model can be further optimized in future to include several constraints that have been left out in the current study. These optimized models can be efficiently utilized by the government to reduce road accidents and to implement policies for road safety. Another scope of this work would be to develop a mobile app that will help the drivers in choosing a route for a ride. A call out to the driver through the maps service can also be implemented that would also announce the risk probability in a chosen route along with the directions. This can then be implemented by service provider companies such as Uber, Ola and so on in future. This will also be useful in having a better surveillance of accident prone areas and providing emergency services in the event of an accident.

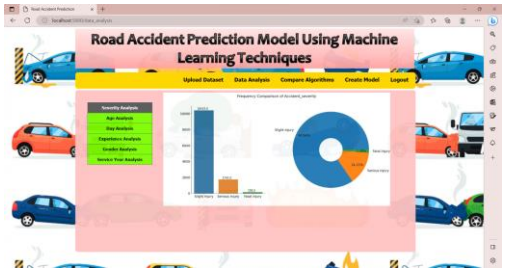

Fig 2: Severity Analysis

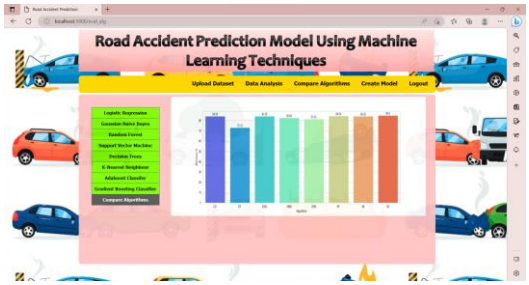

Fig 3: Compare Algorithm Summary

Conclusion

An accident can change the lives of many people. It is up to each of us to bring down this increasing number. This can be made possible by adopting safe driving measures to an extent. Since all instances of accidents cannot be attributed to the same cause, proper precautionary measures will also need to be exercised by the road development authorities in designing the structure of roads as well as by the automobile industries in creating better fatality reducing vehicle models. One thing within our capability is to predict the possibility of an accident based on previous data and observations that can aid such authorities and industries. This Work was successful in creating such an application that can help in efficient prediction of road accidents based on factors such as types of vehicles, age of the driver, age of the vehicle, weather condition and road structure, so on. This model was implemented by making use of several data mining and machine learning algorithms applied over a dataset for Bangalore and has been successfully used to predict the risk probability of accidents over different areas with high accuracy.

The model can be further optimized in future to include several constraints that have been left out in the current study. These optimized models can be efficiently utilized by the government to reduce road accidents and to implement policies for road safety. Another scope of this work would be to develop a mobile app that will help the drivers in choosing a route for a ride. A call out to the driver through the maps service can also be implemented that would also announce the risk probability in a chosen route along with the directions. This can then be implemented by service provider companies such as Uber, Ola and so on in future. This will also be useful in having a better surveillance of accident prone areas and providing emergency services in the event of an accident.

References

[1] A. Mohamed Mostafa, B. Aldughayfiq, M. Tarek, A. S. Alaerjan, H. Allahem, M. K. Elbashir, M. Ezz, and E. Hamouda, “AI-based prediction of traffic crash severity for improving road safety and transportation efficiency,” Scientific Reports, vol. 15, pp. 27468, 2025.

[2] H. Khanum, A. Garg, M. I. Faheem, and R. Kulkarni, “A methodological framework for road accident severity prediction for Indian highways using machine learning models,” MethodsX, vol. 15, pp. 103728, 2025.

[3] S. Krishnaveni, T. Vaishnavi, P. R. M, R. M, and V. R. D, “Road Crash Injury Severity Prediction Using Ensemble Learning,” in Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE), 2024.

[4] W. Sun, L. N. Abdullah, P. S. Sulaiman, and F. Khalid, “Meta-feature-based traffic accident risk prediction: A novel approach to forecasting severity and incidence,” Vehicles, vol. 6, no. 2, pp. 728–746, 2024.

[5] Y. Li, M. Abdel-Aty, Y. Yuan, and J. Cheng, “Real-time crash risk prediction on freeways using deep learning,” Accident Analysis & Prevention, vol. 135, pp. 105371, 2020.

[6] A. P. Pradhan, S. Bagchi, and A. Mukherjee, "Severity prediction of road accidents using machine learning techniques," in International Conference on Intelligent Transportation Systems (ITSC), IEEE, pp. 1–6, 2019.
[7] N. ArunPrasath and M. Punithavalli, "A review on road accident detection using data mining techniques," International Journal of Advanced Research in Computer Science, vol. 9, pp. 881–885, 2018.
[8] E. Suganya and S. Vijayarani, "Analysis of road accidents in India using data mining classification algorithms," in International Conference on Inventive Computing and Informatics (ICICI), pp. 1122–1126, 2017.
[9] G. Yannis, A. Dragomanovits, A. Laiou, T. Richter, S. Ruhl, F. La Torre, L. Domenichini, D. Graham, N. Karathodorou, and H. Li, "Use of accident prediction models in road safety management – an international inquiry," Transportation Research Procedia, vol. 14, pp. 4257–4266, 2016.
[10] S. Sarkar, V. Pateshwari, and J. Maiti, "Predictive model for incident occurrences in steel plant in India," in International Conference on Communication, Computing and Networking Technologies (ICCCNT), IEEE, pp. 1–5, 2015.
[11] T. Williams, J. Betak, and B. Findley, "Text mining analysis of railroad accident investigation reports," in Joint Rail Conference, American Society of Mechanical Engineers, 2015.
[12] M. Ghazizadeh, A. D. McDonald, and J. D. Lee, "Text mining to decipher free-response consumer complaints: Insights from the NHTSA vehicle owner's complaint database," Human Factors, vol. 56, no. 6, pp. 1189–1203, 2014.