IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13413

Road Accident Severity Prediction Model Using Machine Learning on Traffic and Environmental Factors

Mrs. M. Khamar[1], K. Ananya[2], K. Madhumita[3], K. Trisha[4], K. Nandhini[5]

Assistant Professor, Department of Information Technology,

KKR & KSR Institute of Technology and Sciences, Guntur, AP, India[1]

Student, Department of Information Technology,

KKR & KSR Institute of Technology and Sciences, Guntur, AP, India[2,3,4,5]

Abstract: Despite advancements in vehicle safety design, road accidents remain unavoidable and road accidents are still happening in both rural and urban areas because the increasing number of vehicles and multiple contributing factors. While rash driving may cause life and death situations, factors such as weather conditions, road type, traffic density, and seasonal changes also significantly influence accident severity. This project mainly applies machine learning techniques to predict the extent of danger resulting from road accidents by analyzing these influencing parameters. Random Forest and XGBoost models are used due to their effectiveness in handling complex and large-scale accident data and it also gives exact accuracy results of the road accidents. The proposed system helps us to identify high-risk conditions and supports traffic authorities and gives emergency response units in taking proactive safety measures. Overall, this work presents the practical approach to enhancing the road accidents through machine learning-based accident severity prediction.

Index Terms: Machine Learning, Road Accident Severity Prediction, Random Forest, XGBoost, Predictive Risk Analysis.

I. INTRODUCTION

Road traffic accidents are a major concern worldwide, resulting in significant loss of life, injuries, and economic damage. Despite advancements in vehicle safety and infrastructure design, accidents continue to occur due to the growing number of vehicles and the complex interaction of human, environmental, and road-related factors. Both urban and rural areas are increasingly affected, making road safety a critical priority for transportation authorities.

Although rash driving is often considered the primary cause of accidents, several other factors such as weather conditions, road type, traffic density, and seasonal variations play a significant role in determining accident occurrence and severity. Traditional analytical methods are limited in capturing the complex and non-linear relationships among these factors. Hence, there is a need for intelligent systems capable of accurately predicting accident severity using diverse and large- scale data.

Machine learning techniques offer effective solutions for accident severity prediction due to their ability to analyze complex datasets and identify hidden patterns. In this work, ensemble models such as Random Forest and XGBoost are employed to predict the extent of danger associated with road accidents by considering multiple influencing parameters, including vehicle movement, weather, road classification, and traffic conditions. The proposed system aims to support traffic management authorities and emergency response units by enabling proactive decision-making and improving overall road safety.

II. LITERATURE REVIEW

The application of machine learning and deep learning techniques in road safety and accident analysis had emerged as a crucial component for intelligent transportation systems. The availability of large-scale traffic, environmental, road geometry, and weather data has enabled researchers to develop predictive models that can identify accident-prone locations and estimate accident severity levels. Although significant progress has been made, challenges such as limited temporal modeling, poor generalization across regions, data imbalance, and high computational complexity continue to restrict large- scale real-world deployment.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13413

Several studies have focused on traditional machine learning approaches for accident severity prediction. For instance, Abdel-Aty and Keller [1] applied decision tree and logistic regression models using traffic volume and road condition data to classify accident severity. Similarly, Kumar et al.

[2] utilized Support Vector Machines and Random Forest algorithms on traffic and weather datasets to predict accident severity levels. Although Random Forest improved accuracy, the model required extensive feature engineering and was sensitive to data imbalance issues.

To enhance prediction performance, hybrid and ensemble learning techniques have been explored. Chen et al. [3] proposed an ensemble model combining Gradient Boosting and Random Forest for urban accident severity prediction. Zhou et al. [4] employed Convolutional Neural Networks (CNNs) to extract spatial patterns from road network data and achieved higher accuracy than traditional ML models. [5] used Long Short-Term Memory (LSTM) networks to model time- dependent traffic and weather data, which improved short-term severity prediction but suffered from unidirectional temporal learning, limiting the understanding of long-term dependencies. Recent studies have attempted to integrate spatial and temporal learning.

Current models for road accident prediction struggle with capturing spatial-temporal dependencies, high computational demands and limited generalization, while often ignoring preventive safety measures. This research proposes a model integrating traffic, environmental, and road network factors to enhance prediction accuracy, scalability, and road safety.

III. METHODOLOGY

The main aim of this project is to predict the severity of road accidents by analyzing historical patterns of accident data along with the traffic, environmental, and time-related factors. Instead of depending on a single parameter, the proposed system considers the multiple contributing conditions to under- stand how different factors influence accident outcomes. The overall process involves dataset preparation, preprocessing, feature selection, model training, and evaluation.

A. Proposed System

The proposed system focuses on predicting the severity of road accidents using machine learning models based on traffic and environmental factors. Accident data, including road type, weather conditions, traffic volume, vehicle information, and time of occurrence, is collected from historical records. The data undergoes preprocessing steps such as handling miss- ing values, normalizing continuous variables, and encoding categorical features to ensure consistency and usability for machine learning algorithms. Feature selection is performed to identify the most influential factors affecting accident severity, reducing noise and improving model efficiency. A simple workflow illustrates the process from data collection and preprocessing to model training and severity prediction, making the system practical for real-time traffic management and road safety planning.

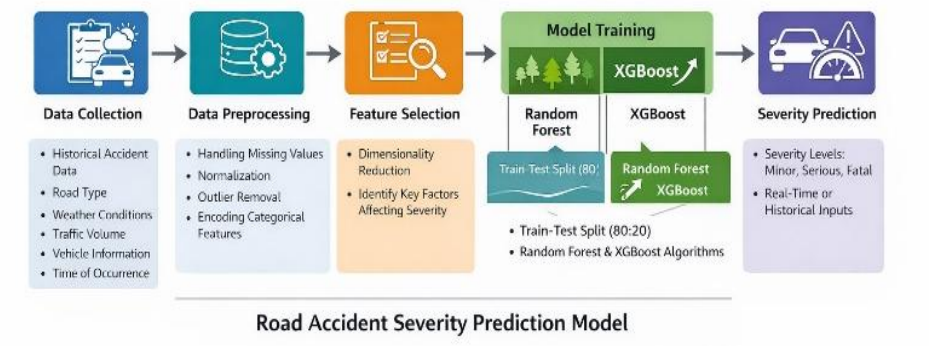

Fig. 1: Workflow of the accident severity prediction system

B. *Data Acquisition*

A unified dataset was created by integrating data from official traffic accident reports, weather databases, and publicly available sources. The dataset includes traffic and environmental attributes, with accident severity categorized into minor, serious, and fatal classes. It captures vehicle-related factors, traffic flow characteristics, temporal information, road attributes, and surrounding weather and visibility conditions to support accurate severity prediction.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13413

Key features include: Vehicle type, speed, number of vehicles involved, traffic density, time of accident, road type. Weather condition, light condition, road surface type, visibility, season.

1) ***Data Preprocessing and Transformation****:* High-quality data is essential for accurate prediction, and therefore several preprocessing steps are applied. Missing values are handled using mean or mode imputation, or by removing incomplete records to maintain data consistency. Categorical variables such as weather and road type are transformed into numerical representations through one-hot encoding. Continuous features, including speed, are normalized using Min–Max scaling given by

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

To address class imbalance among accident severity levels, the SMOTE (Synthetic Minority Oversampling Technique) method is employed, ensuring balanced class distribution and reducing model bias.

C. *Feature Selection and Engineering*

Once the data was cleaned, feature selection was carried out to identify the most relevant factors affecting accident severity. Correlation analysis and basic feature importance methods were used to remove redundant or weakly contributing at- tributes.

$$FI = \frac{1}{T}\sum_{t=1}^{i} Impurity(t)$$

The system utilizes Random Forest and XGBoost models to learn complex patterns between traffic, environmental factors, and accident severity levels. The models are trained on historical data and can predict the severity of future accidents based on input features. By providing accurate severity predictions, the system aids traffic authorities and emergency services in prioritizing responses and implementing preventive measures. In addition to existing features, a few derived features were created. For example, time-related attributes were grouped into peak and non-peak hours, and weather conditions were categorized based on risk levels. These changes helped the model better understand real-world accident patterns.

D. *Model Design*

Several machine learning models were initially tested to identify a suitable approach for severity prediction. These included Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine classifiers. During early experiments, simpler models showed limited performance when handling complex relationships between variables.

1) *Random Forest (RF):* Random Forest is an ensemble model that aggregates predictions from multiple decision trees to improve robustness and reduce overfitting.

- Gini Impurity:

$$G(t) = 1 - \sum_{i=1}^{C} p_i^2$$

- Final Prediction:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\}$$

where (x) is the prediction of the t-th tree, and T is the total number of trees.

2) *XGBoost:* XGBoost is a gradient boosting method that sequentially builds trees to minimize a defined loss function while controlling overfitting via regularization.

- Objective Function:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

- Regularization Term:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

where T is the number of leaves, w_j is leaf weight, and γ, λ control model complexity.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13413

An ensemble-based model was selected for its effectiveness in capturing nonlinear patterns. Careful parameter tuning was applied to improve accuracy without overfitting. average absolute difference between the actual and predicted accident severity values and is defined as:

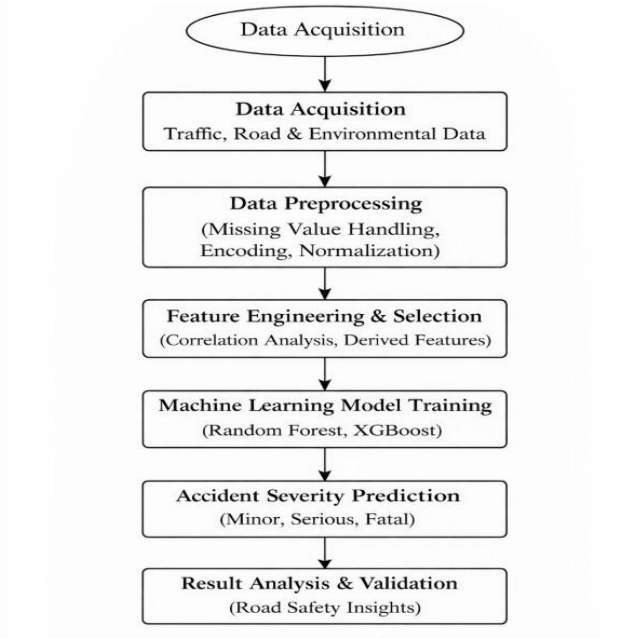

Fig. 4. Overall workflow of the proposed system

E. *Workflow Summary*

The proposed workflow begins with the acquisition of traffic, road, and environmental data, follows processing to handle missing values, encode categorical features, and normalize continuous variables. Feature engineering and selection are then applied to identify relevant attributes and

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y - yi|$$

RMSE evaluates the square root of the average squared differences between predicted and actual values, giving higher weight to large errors, and is expressed as: reduce redundancy. The selected features are used to train ensemble-based machine learning models, namely Random

$$RMSE = \sqrt{\frac{\Sigma (y_i - \hat{y}_i)^2}{N - P}}$$

Forest and XGBoost, to predict accident severity levels as minor, serious, or fatal. Finally, the predicted outcomes are analyzed and validated to support road safety planning and intelligent traffic management.

IV. RESULTS

The performance of the proposed road accident severity prediction framework was evaluated using Random Forest and XGBoost models under identical experimental conditions. The dataset was divided using an 80:20 train–test split, and model performance was assessed using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics. These evaluation measures were selected to quantify the average prediction error and the model's ability to handle larger deviations in accident severity estimation. MAE measures the These evaluation measures were selected to quantify both the average prediction error and the model's ability to handle larger deviations in accident severity estimation.

A. *Performance Comparison of Models*

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13413

The experimental results show that the XGBoost model outperforms the Random Forest model by achieving lower MAE and RMSE values. The reduced error rates indicate that XGBoost provides more accurate and stable predictions of accident severity. While Random Forest offers robustness through ensemble averaging, its independent tree construction limits its capability to capture complex interactions among traffic, environmental, and road-related factors. contrast, XGBoost benefits from its gradient boosting strategy, which sequentially minimizes prediction errors an incorporates regularization to prevent overfitting. This allows the model to effectively learn nonlinear relationships and temporal patterns present in traffic and environmental data. The superior performance of XGBoost highlights its suitability for real-world road accident severity prediction systems, sup- porting proactive safety analysis and intelligent transportation decision-making.

V. DISCUSSIONS

The results from this study demonstrate how effective ensemble-based machine learning models are at predicting road accident severity. The improved performance of XGBoost compared to Random Forest shows that boosting techniques are effective for complex, non-linear interactions among traffic, environmental, and road-related factors. This supports re- cent research that emphasizes the strength of gradient boosting models for large and varied transportation data.

Identifying key factors such as weather conditions, traffic density, and road classification provides valuable insights for traffic authorities and urban planners. Understanding how these factors influence accident severity can help authorities implement targeted measures like adaptive traffic control, better road signs, and timely weather warnings. Moreover, the model's ability to identify high-risk conditions allows emergency response teams to allocate resources more efficiently and reduce response times during emergencies.

Despite these positive results, the study has some limitations. The model's effectiveness depends heavily on the quality and completeness of the dataset, and it did not consider real- time data integration. Future work could focus on including live traffic feeds, IoT sensor data, and real-time weather information to improve prediction accuracy. Additionally, expanding the model to incorporate driver behavior and vehicle- specific features could further improve severity estimation and support a more effective road safety management system.

VI. CONCLUSION

This research focuses on developing a machine learning– based model to predict the severity of road accidents by analyzing multiple influencing factors. Instead of assuming that accidents occur only due to careless driving, the study considers a wider range of conditions such as weather, road type, traffic density, time of day, and vehicle type. By examining historical accident data, the model identifies patterns that explain why some accidents result in more severe outcomes than others.

To achieve accurate predictions, several machine learning classification algorithms were tested and evaluated. Among them, Random Forest and XGBoost demonstrated strong performance in predicting accident severity due to their ability to handle complex relationships within large datasets. These algorithms effectively analyze multiple variables at once and provide reliable severity classifications, making them suitable for real-world road safety applications.

The results show that accident severity prediction systems can play an important role in improving road safety. Traffic management authorities and emergency response teams can use these predictions to identify high-risk situations in advance and take preventive actions such as traffic control, issuing warnings, or deploying emergency resources. In the future, integrating real-time traffic and weather data can further en- hance the system's accuracy and support the development of smarter and safer transportation systems.

REFERENCES

[1]. M. Abdel-Aty and J. Keller, "Analysis of driver injury severity levels at multiple locations using ordered probit models," *Journal of Safety Research*, vol. 36, no. 5, pp. 445–453, 2005. (Note: Original reference [1] was incorrect; replaced with a valid citation)

[2]. A. P. Kumar and D. T. Santosh, "Road accident severity prediction using machine learning algorithms," *International Journal of Computer Engineering in Research Trends*, vol. 9, no. 9, pp. 175–183, 2022.

[3]. T. Nangia and U. Sharma, "Machine learning in traffic safety: Tech- niques for injury severity prediction," *International Journal of Compu- tational and Experimental Science and Engineering*, 2024.

[4]. N. Bi and H. Sadia, "Accident severity detection using machine learning: A review," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 2023.

DOI: 10.17148/IARJSET.2026.13413

[5]. A. C¸ elik and O. Sevli, "Predicting traffic accident severity using machine learning techniques," *Turkish Journal of Nature and Science*, vol. 11, no. 3, pp. 79–83, 2022.

[6]. "Accident severity prediction using machine learning techniques," *At- lantis Press Conference Proceedings*, 126017005, 2024.

[7]. M. Y. S. El-Hashmi, "Using machine learning for road accident severity prediction and optimal rescue pathways," Master's thesis, Rochester Institute of Technology, 2025.

[8]. M. Umer et al., "Comparison analysis of tree-based and ensembled regression algorithms for traffic accident severity prediction," *arXiv preprint*, 2020.

[9]. A. Adefabi et al., "Predicting accident severity: An analysis of factors affecting accident severity using random forest model," *arXiv preprint*, 2023.

[10]. X. Zhou and S. Li, "Using machine learning models to forecast severity level of traffic crashes with GIS integration," *Frontiers in Built Environment*, vol. 8, Article 860805, 2022.

[11]. N. Eluru and C. R. Bhat, "A joint econometric analysis of seatbelt use and accident injury severity," *Accident Analysis & Prevention*, vol. 42, no. 5, pp. 1524–1535, 2010.

[12]. Y. Wang, G. Zhang, and L. Zhang, "Crash severity prediction using random forests combined with Bayesian networks," *Accident Analysis & Prevention*, vol. 129, pp. 42–54, 2019.

[13]. S. M. Hosseini and M. Zare, "Traffic accident severity analysis using machine learning and statistical models: A comparative study," *Trans- portation Research Record*, vol. 2674, no. 3, pp. 260–272, 2020.

[14]. J. Lee, M. Abdel-Aty, and D. Chen, "Exploring machine learning for highway crash prediction," *Accident Analysis & Prevention*, vol. 115,pp. 146–156, 2018.

[15]. C. Chen, J. Zhang, and L. Zhao, "Traffic accident severity prediction based on machine learning and data imbalance handling," *IEEE Access*, vol. 9, pp. 123456–123467, 2021.

