GJR Publication
OPEN ACCESS

Global Journal of Research in Engineering & Computer Sciences
ISSN: 2583-2727 (Online)
Volume 06 | Issue 04 | July-August | 2026
Journal homepage: https://gjrpublication.com/gjrecs/

Research Article

Predicting Road Traffic Accident Severity Using Machine Learning Algorithms in Kano State, Nigeria

*[1]Bello Abubakar Imam, [2]Kabir Dalha Kabir, [3]Shehu Hassan Ayagi and [4]Sadiq Bello Abubakar

[1,2,3,4] Department of Computer Science Kano State Polytechnic, Kano State of Nigeria.

DOI: 10.5281/zenodo.21283245 **Submission Date: 25 May 2026 | Published Date: 09 July 2026**

***Corresponding author: Bello Abubakar Imam**
Department of Computer Science Kano State Polytechnic, Kano State of Nigeria.

Abstract

Road traffic accidents (RTAs) are a major global public health and safety crisis, leading to millions of deaths and injuries annually. This study focuses on predicting the severity of RTAs (fatal, serious, or minor) in Kano State, Nigeria, by addressing a key gap in prior research: the limited consideration of all three primary contributing factors (human, environmental, and vehicle) simultaneously. This research utilizes a dataset of 1,368 RTA instances with 19 attributes from the Federal Road Safety Corps (FRSC), Kano State command, covering the years 2019 to 2021. After pre-processing, including handling missing values and applying the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance, we developed and evaluated five machine learning classifiers: Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Neural Network (NN), and Random Forest (RF). The models were implemented using Python and Scikit-learn. Experimental results demonstrate that the Random Forest algorithm achieved the highest prediction accuracy of 97%, significantly outperforming other methods (KNN: 91%, DT: 91%, NN: 70%, NB: 58%). This high-accuracy model provides a robust framework for road safety authorities to predict accident severity based on causative factors, enabling more targeted and proactive interventions to reduce fatalities.

Keywords: *Road Traffic Accident, Accident Severity Prediction, Machine Learning, Random Forest, SMOTE, Kano State.*

1. INTRODUCTION

Road traffic accidents remain one of the most pressing challenges of modern transportation. According to the World Health Organization (WHO), approximately 1.35 million people die each year on the world's roads, with 20 to 50 million suffering non-fatal injuries often result in long-term disability [1]. In Nigeria, the situation is particularly dire, with a fatality rate 47 times higher than in Great Britain [2]. Kano State, one of Nigeria's most populous states, experiences a high daily rate of traffic accidents, with a single accident often resulting in multiple fatalities and injuries [3].

Traffic accidents are rarely caused by a single factor. The primary contributors are generally categorized into three groups: human factors (e.g., speeding, loss of control, driving under the influence), environmental factors (e.g., poor road conditions, bad weather), and vehicle factors (e.g., brake failure, tire burst) [4]. While numerous studies have attempted to predict accident severity, many have focused on only one or two of these factor categories, leading to suboptimal predictive accuracy. Furthermore, research specifically applying a comprehensive set of machine learning algorithms to Kano State's RTA data is scarce.

This study aims to fill these gaps by:

1. Constructing a comprehensive dataset that integrates all three causative factors (human, environmental, vehicle).
2. Developing and evaluating five distinct machine learning models (KNN, DT, NB, NN, RF) to classify accident severity as fatal, serious, or minor.
3. Identifying the most accurate model to serve as a predictive tool for road safety management in Kano State.

2. RELATED WORK

The application of machine learning for RTA severity prediction has gained significant traction. Various studies have employed different algorithms with mixed results.

Bahiru et al. [5] used data mining classifiers (J48, ID9, CART, and Naïve Bayes) and found J48 to have the highest accuracy for injury crash prediction. Jadhav et al. [6] explored Decision Tree, KNN, Naïve Bayes, and AdaBoost, reporting that AdaBoost performed best with 80% accuracy, but their study used a limited set of factors. Kumeda et al. [7] compared Fuzzy-FARCHD, Random Forest, and Naïve Bayes, achieving an 85.4% accuracy with Fuzzy-FARCHD, but noted that only a few key factors (e.g., lighting, road class) were considered.

More recently, Gan et al. [8] compared Deep Forest, RFs, KNN, and XGboost, reporting that Deep Forest achieved 93% accuracy. Pourroostaei Ardakani et al. [9] used Decision Tree, Random Forest, and Multinomial Logistic Regression, all achieving approximately 85.5% accuracy. A common weakness across many studies, particularly concerning data from developing nations, is the reliance on limited datasets that do not holistically integrate human, vehicle, and environmental variables. This research directly addresses this limitation by using a dataset that combines all three factor categories to improve predictive performance for Kano State.

3. METHODOLOGY

3.1. Dataset Description and Pre-processing

The dataset for this study was obtained from the Federal Road Safety Corps (FRSC), Kano State Sector Command. It comprises 1,368 instances of road traffic crashes recorded between 2019 and 2021. Nineteen attributes were used as predictors, covering all three causative categories (see Table 1). The target variable is accident severity, classified into three classes: Fatal, Serious, and Minor.

Table 1: Key Road Crash Features

Factor Category	Attributes (Examples)
Human	Speed violation, Loss of control, Dangerous driving, Sleeping on steering, Driving under alcohol, Use of phone.
Environmental	Route violation, Bad Road, Road obstruction violation, Poor weather.
Vehicle	Tire burst, Brake failure, Mechanically deficient vehicle, Overloading.

Pre-processing steps included:

- **Handling Missing Data:** Rows and columns with null or irrelevant values were removed.
- **Label Encoding:** Categorical text data was converted into numerical format using Scikit-learn's Label Encoder.
- **Handling Class Imbalance:** The initial dataset had a skewed distribution of severity classes. To prevent the "accuracy trap," the Synthetic Minority Oversampling Technique (SMOTE) was applied, generating synthetic samples to balance the classes, resulting in the final dataset of 1,368 instances.

3.2. Data Splitting and Models

The pre-processed dataset was split into 70% for training and 30% for testing. Five machine learning classification algorithms were implemented using Python (Jupyter Notebook) and the Scikit-learn library:

1. **K-Nearest Neighbor (KNN)**
2. **Decision Tree (DT)**
3. **Naïve Bayes (NB)**
4. **Neural Network (NN) - using MLPClassifier.**
5. **Random Forest (RF)**

3.3. Evaluation Metrics

The performance of each model was evaluated using a standard classification report containing:

- **Accuracy:** The overall percentage of correct predictions.
- **Precision:** The proportion of positive identifications that were actually correct.
- **Recall:** The proportion of actual positives that were identified correctly.
- **F1-Score:** The harmonic mean of precision and recall.
- **Confusion Matrix:** To visualize correct and incorrect predictions per class.

4. RESULTS AND DISCUSSION

The models were trained and tested under identical conditions. Table 2 summarizes the performance of the five algorithms.

Table 2: Performance Comparison of ML Algorithms

Algorithm	Class	Precision	Recall	F1-Score	**Overall Accuracy**
KNN	Fatal	0.93	0.86	0.90	**91%**
	Minor	0.83	0.86	0.84	
	Serious	0.89	0.97	0.93	
Decision Tree	Fatal	0.96	0.88	0.89	**91%**
	Minor	0.82	0.98	0.89	
	Serious	0.88	0.97	0.93	
Naïve Bayes	Fatal	0.97	0.18	0.31	**58%**
	Minor	0.58	0.74	0.65	
	Serious	0.53	0.94	0.68	
Neural Network	Fatal	0.86	0.47	0.61	**70%**
	Minor	0.10	0.54	0.70	
	Serious	0.61	0.97	0.75	
Random Forest	Fatal	0.98	0.95	0.96	**97%**
	Minor	0.92	0.98	0.95	
	Serious	0.97	0.98	0.98	

The results clearly show that the **Random Forest (RF) algorithm achieved the highest overall accuracy of 97%**, significantly outperforming the other models. The Naïve Bayes classifier performed poorly, with an accuracy of only 58%, largely due to its inherent assumption of feature independence, which is unlikely to hold true for complex, interrelated RTA data.

The strong performance of Random Forest can be attributed to its ensemble nature, which combines multiple decision trees to reduce overfitting and variance. Furthermore, RF is robust to noisy data and provides high accuracy when dealing with both categorical and numerical features, as present in our dataset. The 97% accuracy represents a notable improvement over many previous studies, where accuracies typically ranged from 70% to 93%. This improvement is likely a direct result of this study's comprehensive approach of integrating a wide array of features from all three causative factor categories.

The confusion matrix for the Random Forest model confirmed its robustness, correctly predicting 152 fatal accidents and 45 minor accidents out of the test set, with very few false predictions.

5. CONCLUSION AND FUTURE WORK

This study successfully developed and evaluated machine learning models for predicting road traffic accident severity in Kano State, Nigeria. By integrating human, environmental, and vehicle-related factors into a single predictive framework and employing SMOTE to handle class imbalance, we achieved a high-accuracy model. The Random Forest algorithm outperformed KNN, Decision Tree, Neural Network, and Naïve Bayes classifiers, attaining a 97% accuracy in classifying accident severity as fatal, serious, or minor.

This model offers a valuable tool for the FRSC and other road safety authorities. By inputting contributing factors, the model can predict the likely severity of an accident, enabling proactive resource allocation, targeted safety campaigns (e.g., focusing on specific human or vehicle factors in high-risk areas), and more informed policy decisions to reduce fatalities.

Future work will focus on:

1. **Expanding the Dataset:** Incorporating data from multiple states in Northern Nigeria to improve generalizability.
2. **Exploring More Algorithms:** Testing additional advanced algorithms like XGboost, LightGBM, and Deep Forest for potential accuracy gains.
3. **Geospatial Analysis:** Integrating GPS coordinates to develop location-specific severity prediction maps, allowing for targeted interventions on high-risk roads.
4. **Real-time Prediction:** Developing a framework for real-time severity prediction using data from IoT sensors or driver-monitoring systems.

REFERENCES

1. World Health Organization. (2018). *Global status report on road safety 2018*.
2. A. O. Atubi and P. C. Onokala, "Contemporary Analysis of Variability in Road Traffic Accidents in Lagos State, Nigeria," *African Geographical Review*, vol. 28, no. 1, pp. 11–41, 2009.
3. I. Khalil Umar, S. Bashir, and uf M. Gora, "Road Traffic Accident in Kano Nigeria: A Case Study of Kano Metropolitan," *Scholars Journal of Engineering and Technology*, vol. 5, no. 12, 2017.
4. A. Evwiekpaefe and U. Umar, "A Review of Factors Contributing to Road Traffic Accidents in Nigeria," *Journal of Transportation Technologies*, 2022.
5. T. K. Bahiru, D. K. Singh, and E. A. Tessfaw, "Comparative Study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity," in *Proc. ICICCT*, 2018, pp. 1655–1660.
6. A. Jadhav, S. Jadhav, A. Jalke, and K. Suryavanshi, "Road accident analysis and prediction of accident severity using machine learning," *Int. J. of Scientific & Technology Research*, vol. 7, no. 12, 2020.
7. B. Kumeda et al., "Classification of Road Traffic Accident Data Using Machine Learning Algorithms," in *Proc. ICCSN*, 2019, pp. 682–687.
8. J. Gan, L. Li, D. Zhang, Z. Yi, and Q. Xiang, "An Alternative Method for Traffic Accident Severity Prediction: Using Deep Forests Algorithm," *Journal of Advanced Transportation*, vol. 2020.
9. S. Pourroostaei Ardakani et al., "Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis," *Sustainability*, vol. 15, no. 7, 2023.

CITATION

Imam, B. A., Kabir, K. D., Ayagi, S. H.& Abubakar, S. B. (2026). Predicting Road Traffic Accident Severity Using Machine Learning Algorithms in Kano State, Nigeria. In Global Journal of Research in Engineering & Computer Sciences (Vol. 6, Issue 4, pp. 26–29). https://doi.org/10.5281/zenodo.21283245