International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Explainable AI-Based Road Accident Severity Prediction on Indian Highways

Dr. Rashmi Malvankar[1], Kainaat Shaikh[2]

[1]Assistant Professor, Department of Information Technology
[2]Student, Department of Information Technology
Shah & Anchor Kutchhi College of Engineering, Mumbai, India

Abstract - Road accidents are a global issue however India has recorded most fatalities which accounts for roughly 11% of all road-related deaths worldwide. Over 53% of all accidents and 59% of all fatalities occurred on highways. This study proposes a method to classify a road accident as Fatal, Grievous, Minor or No-Injury. The dataset used for this study comprises of road accident data collected from four distinct highway segments under the jurisdiction of the National Highways Authority of India (NHAI). To handle class imbalance, SMOTE with Tomek links were used. Various machine learning models were used such as Random Forest, XGBoost, LightGBM. The models achieved accuracies of 86.2%, 84.7%, and 83.4% respectively. As most of the machine learning models exhibit a common limitation: "black box" nature, results were interpreted using SHapley Additive exPlanations (SHAP) method. SHAP analysis reveals that accident location represented by highway chainage and the primary vehicle type consistently emerged as the dominant predictors for both Grievous and Fatal accident classifications. Moreover, accident cause exhibited greater prominence in the Fatal-class explanation, whereas temporal characteristics occupied higher ranks in the Grievous-class explanation.

Key Words: Road Accident Severity Prediction, Machine Learning, XGBoost, LightGBM, Explainable AI, SHAP, Road Safety.

1.Introduction

Road traffic injuries remain a leading and preventable cause of death and disability worldwide, with an estimated 1.3–1.35 million deaths annually and a rising burden projected to become the seventh leading cause of death by 2030 [1]. However, statistics from the year 2022 shows that road accidents in India were over 450,000 recorded accidents that resulted in over 150,000 fatalities [2].

Highways represent a crucial component of the road transportation system. However, their extensive use and high traffic volumes also make them particularly vulnerable to road crashes. Statistics show that even though national and state highways account for only around 5% of India's overall road network, they are disproportionately involved in road accidents, accounting for approximately 52% of all reported incidents [2]. Moreover, crashes occurring on highways tend to be more severe because vehicles generally travel at higher speeds, resulting in a greater probability of fatalities and serious injuries. Consequently, highway safety remains a major concern.

In the broader literature on crash severity, ensemble tree-based methods tend to outperform comparable benchmarks, with Random Forest delivering the best results in 29% of all studies in the literature review, and in 70% of cases where it was tested. However, XGBoost, SVM, and other approaches also perform well in some settings [3]. This pattern is echoed in applied studies: Random Forest was the best classifier in New Zealand crash data, achieving 81.45% accuracy, while it also outperformed seven alternatives in Thailand, though fatal-class recall remained poor at 0.198 [4]. Conversely, the previous Indian highway Random Forest study showed a sharp drop from 67% training accuracy to 41.47% test accuracy, explicitly suggesting dataset bias or imbalance and possible underestimation of severity [5].

Khanum, Humera et al. developed predictive models for accident severity on indian highways using Gradient Boosting and Random Forest with Random Forest demonstrating superior results of 86.65%. However, despite the high overall predictive accuracy, the dataset exhibits a substantial class imbalance, with the Fatal accident class accounting for only approximately 9% of the observations. This concern is supported by the reported class-wise results, where the RF model demonstrated high precision but comparatively lower recall for the Fatal class, indicating that a considerable proportion of actual fatal accidents may remain undetected [6]. This highlights the need for class-imbalance-aware learning and evaluation strategies, including resampling techniques such as SMOTE–Tomek and greater emphasis on macro-precision, macro-recall, macro-F1, and class-specific performance, rather than accuracy alone.

To understand how different factors combine to cause road accidents on highways, this study uses machine learning models to explore the links between accidents and their key contributors. Building on existing epidemiological and intervention evidence, the study is designed in two stages. First, it carries out a structured comparison of multiple models for road accident severity. Second, it adds explainable AI techniques to look inside the models and show how combinations of factors leads to fatal, grievous and minor or no-injury outcomes, instead of only producing a simple global ranking of variables. These findings can be turned into concrete measures aimed at reducing risk and injury severity.

2. Area of Study

To guarantee operational and geographic variety, four separate highway portions falling under NHAI control were selected. Refer Fig. 1 for highway stretches considered for this study. The four stretches are as follows :-

a) Pune–Solapur (Maharashtra): 2804 records

b) Barwa-Adda–Panagarh (Jharkhand & West Bengal Region) : 3710 records

c) Chengapally–Walayar (Tamil Nadu Region) : 422 records

d) Nagpur Region (Maharashtra): 1180 records

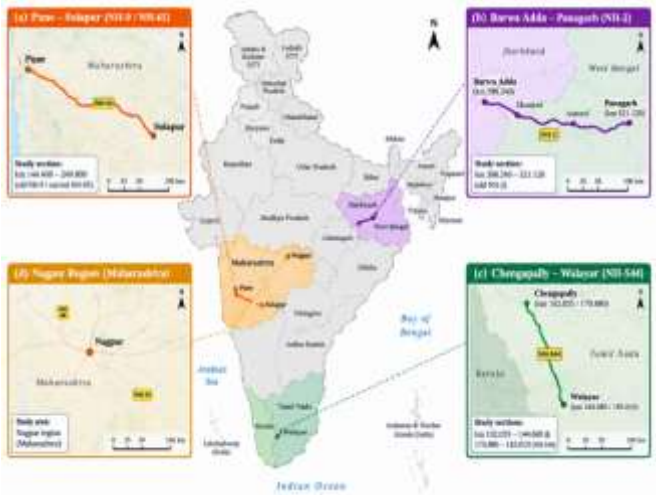


Fig -1: Study Area

3. Methodology

A. Overview of the Proposed Framework

Figure 2 shows the whole operational framework suggested in this study. To begin with, the source dataset undergoes preprocessing then the processed data are then divided using stratified sampling. During each fold of stratified cross-validation, the training portion of the data was isolated and SMOTE-Tomek was applied exclusively to the training fold. SMOTE-Tomek was selected because it combines minority-class oversampling with Tomek-link-based boundary cleaning, which is suitable for addressing both class imbalance and potential class overlap in multi-class accident severity data. For each hyperparameter combination, the model was trained using the balanced training fold and subsequently evaluated on the corresponding untouched validation fold. This process was repeated across all folds, and the validation performance was used to identify the optimal hyperparameter configuration. It is then followed by model training using Random Forest, XGBoost, and LightGBM. The models are evaluated on the test set. Finally, the best-performing model is selected and interpreted using SHAP analysis to identify the most important features and explain the model predictions.

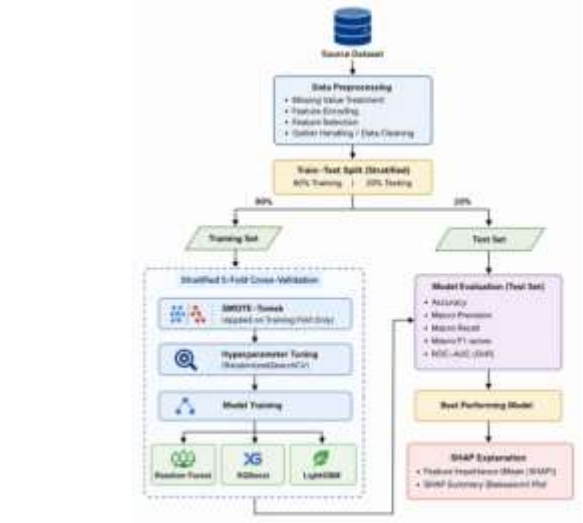


Fig -2: ML Pipeline

B. Dataset

The accident data used in this study is accessible via Zenodo. The dataset comprises of 8,116 accidents and the classes of accident severity are Fatal, Grievous, Minor and No-Injury. The percentage accident severity shown in Figure 3 demonstrates the class imbalance as Fatal class has minority samples.

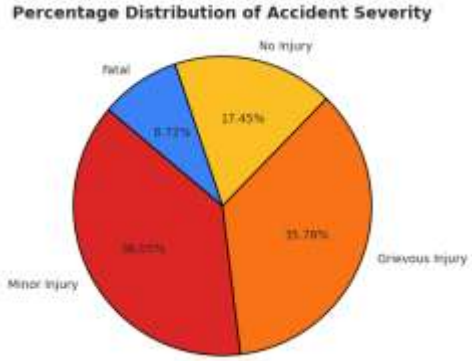


Fig -3: Class Distribution of Accident Severity

C. Data Preprocessing

First, the 'Date' attribute was converted into a standardized datetime format to ensure consistency. Two additional features, namely 'Accident_Month' and 'Accident_Year', were subsequently extracted from the standardized date attribute to capture potential seasonal and yearly variations in accident severity. Similarly, the 'Time_of_Accident' attribute was converted from its original string representation into a numerical feature, 'Accident_Hour', enabling the learning algorithms to capture time-dependent accident patterns. After extracting these derived features, the original 'Date' and 'Time_of_Accident' attributes were removed to eliminate redundancy and avoid potential processing errors. Furthermore, missing values in the 'Vehicle_Type_Involved_J_V2' attribute were replaced with zero, representing observations in which a second vehicle was not involved. These preprocessing steps produced a structured and machine-readable dataset suitable for subsequent model training and evaluation.

4. Results & Discussion

The performance of the classification models is influenced by the selection of key hyperparameters, which were carefully configured to balance accuracy and generalization. The heatmap shown in Figure 4 compares the performance of the Random Forest, XGBoost, and LightGBM models using Macro Precision, Macro Recall, and Macro F1-score. Among the evaluated models, Random Forest achieved the best overall performance, with a Macro Precision of 0.858, Macro Recall of 0.856, and Macro F1-score of 0.856. The optimized hyperparameters for Random Forest model are 'classifier__n_estimators': 200, 'classifier__min_samples_split':2,'classifier__max_features': log2, 'classifier__max_depth': 30. XGBoost showed comparable performance, whereas LightGBM obtained relatively lower scores. Overall, the results indicate that the Random Forest model provided the most balanced classification performance across all the accident severity classes.

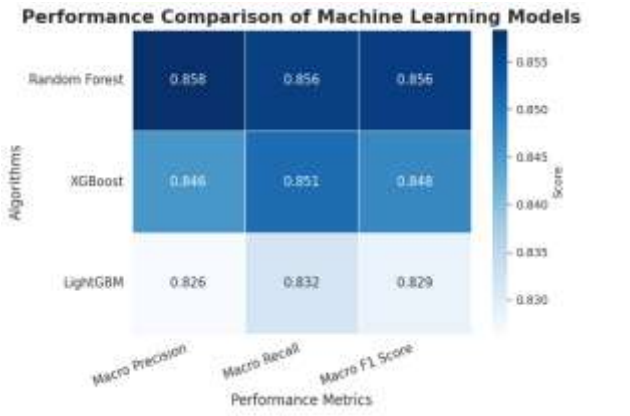


Fig -4: Performance Comparison

The multi-class ROC curves demonstrate the discriminative ability of the classification model using a One-vs-Rest (OvR) approach. The curves for Fatal, Grievous Injury, Minor Injury, and No Injury classes are positioned close to the upper-left corner, indicating strong classification capability. The macro-average ROC curve for Fatal class achieved an AUC of 0.97 as class imbalance was handled using Smote-Tomek. These results indicate that the model has a high ability to distinguish between the different accident severity categories.

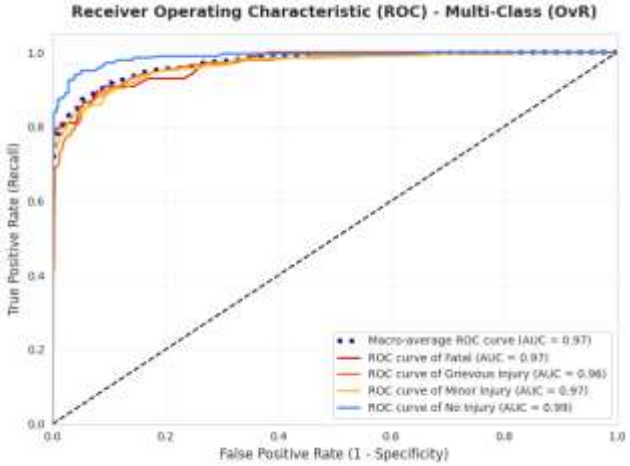


Fig -5: ROC Curve

A. *SHAP Analysis*

The above findings suggest that Random Forest classifier to be the winner model. Hence, we explain the factor contributions of predictions using SHAP. Figure 6 represents a SHAP summary (beeswarm) plot for Fatal class, which illustrates how each feature contributes to Fatal class predictions across all data points. The x-axis represents the SHAP values, indicating the impact of a feature on the prediction; values toward the right increase the likelihood of higher accident severity, while values toward the left reduce it. The y-axis lists the features in descending order of importance. Each point corresponds to an individual observation, and the color gradient represents the feature value, where red denotes higher values and blue indicates lower values.

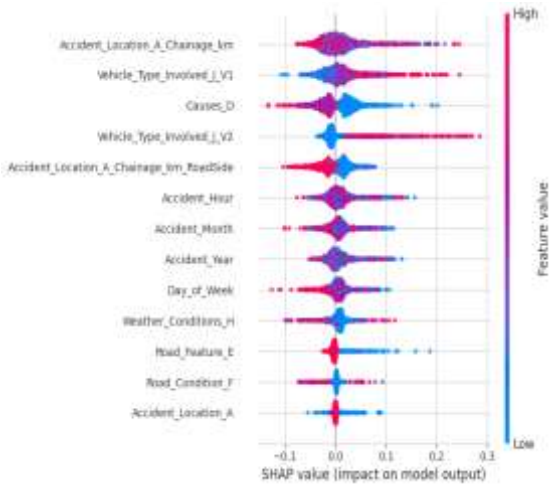


Fig -6: SHAP Summary plot of Fatal Class

The SHAP summary plot indicates that Accident_Location_A_Chainage_km is the most influential predictor of fatal accident classification, exhibiting the widest distribution of SHAP values and substantial positive contributions for several observations. This suggests that fatal accident patterns vary considerably across highway locations. Vehicle-related characteristics, particularly Vehicle_Type_Involved_J_V1 and Vehicle_Type_Involved_J_V2, also demonstrated substantial predictive influence, indicating that the characteristics and configuration of vehicles involved in an accident contribute meaningfully to the model's severity classification. Causes_D showed a heterogeneous SHAP distribution, with different cause categories producing both positive and negative contributions, highlighting variation in their association with fatal accident predictions. Temporal variables, including Accident_Hour, Accident_Month, Accident_Year, and Day_of_Week, were also among the influential predictors.

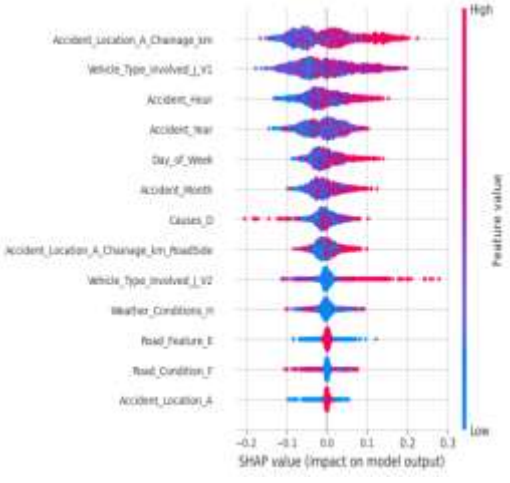


Fig -7: SHAP Summary plot of Grievous Class

Comparison between Fatal and Grievous class explanations reveals that Accident location represented by highway chainage and the primary vehicle type consistently emerged as the dominant predictors for both Grievous and Fatal accident classifications. However, the relative importance of other factors varied between severity classes. Accident cause exhibited greater prominence in the Fatal-class explanation, whereas temporal characteristics, occupied higher ranks in the Grievous-class explanation.

5. Conclusion & Future Work

This study presented a hybrid explainable machine learning framework for predicting road accident severity by utilizing the strengths of ensemble and boosting models with interpretability techniques. The results demonstrated that Random Forest model is highly effective in capturing complex patterns within accident data, while explainable AI approaches provide valuable insights into how different factors influence severity outcomes. By moving beyond simple prediction and offering interpretable results, the study contributes towards more transparent and reliable decision-support systems for road safety.

Looking ahead, several avenues can further enhance this work. Future research can focus on incorporating real-time data streams instead of relying solely on static or historical datasets, enabling dynamic and more responsive prediction systems. In addition, counterfactual explainable AI methods, such as those implemented through tools like DiCE, can be explored to provide actionable "what-if" scenarios, helping stakeholders understand how changes in specific factors could alter accident outcomes.

REFERENCES

[1] H. Rosen, I. Bari, and N. Paichadze, "Global road safety 2010–2018: An analysis of Global Status Reports," *Injury*, vol. 56, 2022.

[2] "ROAD ACCIDENTS IN INDIA 2022.", Highways Accidents Report, Ministry of Road Transport and Highways, 2022. www.morth.nic.in.

[3] Santos, K., et al. "A literature review of machine learning algorithms for crash injury severity prediction.." *Journal of safety research*, vol. 80, 2021, pp. 254-269, https://doi.org/10.1016/j.jsr.2021.12.007.

[4] Ahmed, Shakil, et al. "A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance." *Transportation Research Interdisciplinary Perspectives*, 2023, https://doi.org/10.1016/j.trip.2023.100814.

[5] Khanum, Humera, et al. "Accident severity prediction modeling for road safety using random forest algorithm: an analysis of Indian highways." *F1000Research*, vol. 12, 2023, https://doi.org/10.12688/f1000research.133594.2.

[6] Khanum, Humera et al. "A methodological framework for road accident severity prediction for indian highways using machine learning models." *MethodsX*, vol. 15, pp. 103728, 19 Nov. 2025, doi:10.1016/j.mex.2025.103728.