Predicting Injury Severity in Traffic Accidents using Statistical and Machine Learning Models

Dev Pratap Singh[1*], Anuj Sharma[2], Bharat Bharat[3], Aditya Singh[4] ,Nidhi Singh[5]
[1,2,3,4,5] Krishna Institute of Engineering & Technology(KIET),Ghaziabad, Delhi-NCR, Uttar Pradesh, India, [1*]dev.2226cseai1059@kiet.edu, [2]anuj.2226cseai1017@kiet.edu, [3]bharat.singh745484@gmail.com, [4]aditya.2226cseai1032@kiet.edu, [5]nidhi.singh@kiet.edu

Abstract. — Road traffic accidents cause hundreds of thousands of injuries, deaths, and economic costs annually, and hence the prediction of severity is significant for effective rescue resource allocation and policy intervention [1]. Machine learning (ML) has been shown to be a valuable tool in achieving the patterns in massive databases of road traffic accidents that are not visible to human observation. ML algorithms process different types of information such as weather, road conditions, traffic, time, and geographical position, which are all factors in the prediction of accident severity [7]. Ensemble learning algorithms like Random Forest, XGBoost, and LightGB are known to be particularly effective in leveraging the strengths of multiple decision trees to uncover complex relationships between variables [6]. The authors enhance these algorithms using stacking and automated parameter tuning. In addition to being accurate, it is essential for an algorithm to be interpretable. This is achieved by algorithms like SHAP (SHapley Additive exPlanations), which can reveal the variables that contribute most to the predictions [7]. Another issue is the problem of class imbalance because the more severe accidents are not as common, and this issue is handled by methods such as SMOTE (Synthetic Minority Oversampling Technique), which generates new samples of the minority but crucial class [11]. With the combination of ensemble learning, interpretability, and balancing, reliable and interpretable models for traffic crash severity prediction can be built, which can result in safer roads with reduced traffic injuries [12].

Keywords—Traffic Crash Severity Prediction, Ensemble Machine Learning, Stacking Classifier, SHAP, Feature Importance, SMOTE

1 INTRODUCTION

1.1 Background and Motivation

Road traffic accidents remain one of the most pressing issues in public health and safety. According to the World Health Organization (WHO), millions of people are injured or killed each year due to road traffic accidents, and traffic accidents are among the leading causes of premature deaths [1]. Apart from the loss of lives, accidents also have a substantial economic burden in terms of treatment costs, lost productivity, and property damage. It is crucial to not only understand where and when accidents occur but also how serious they are in order to maximize emergency response and transport safety strategies. Traditional statistical models are only able to point out a few variables and are not capable of detecting the non-linear relationships between many variables. With the availability of traffic data, machine learning has the capability of detecting hidden patterns and relationships between many risk factors, which can be utilized to make roads safer [7], [12].

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_89

1.2 Role of Machine Learning in Accident Severity Prediction

Machine learning is very effective in dealing with the complexities involved in the data of road accidents. Machine learning does not require any assumptions about the data distribution. Machine learning can learn from large datasets and find the underlying patterns and relationships between variables that are not easily visible to human analysts [6]. The severity of an accident depends on various variables that are interlinked, such as weather, road conditions, time of day, driving behavior, and traffic. Machine learning models process all these variables simultaneously to find the subtle differences between minor and fatal accidents. Ensemble models such as Random Forest, XGBoost, and LightGBM have been found to be significantly better than others by leveraging the strengths of multiple decision trees to minimize errors and maximize accuracy [6]. However, interpretability is also an important requirement. SHAP values assist researchers and policymakers in identifying the most influential variables that affect the severity of an accident [7].

1.3 Challenges in Accident Severity Prediction

However, there are still some challenges in predicting the severity of accidents despite the innovations in ML. The first and most important challenge is data imbalance, where data on less severe accidents is abundant, and severe accidents are rare [11]. This makes the model less accurate in predicting severe accidents. SMOTE overcomes this challenge by creating new instances of the minority class, allowing optimal learning [11]. The second challenge is interpretability. Ensemble models, which are highly accurate, are "black boxes." This makes policymakers and decision-makers less confident in using the predictions of these models. Emergency planners and transportation authorities need accurate predictions as well as explanations [12]. There may be missing values, noise, or geographical differences in accident data, which need preprocessing and feature engineering [12].

1.4 Research Objective and Contribution

The purpose of this research is to develop robust and interpretable models for traffic accident severity classification. The research combines accuracy and interpretability by using new ensemble methods together with interpretability and data balancing methods. Random Forest, XGBoost, and LightGBM are used and tuned to explore the complex relationships between the characteristics of accidents [6].SMOTE balances the data by providing sufficient data for dangerous and fatal accidents [11]. SHAP offers

explanations for feature contributions that can be understood by policymakers and researchers. The significance of this paper is that it shows how the integration of these techniques can create an interpretable and balanced accident severity prediction system with both theoretical and practical significance [1], [12].

2 RELATED WORK

2.1 Traditional Methods

Historical research on traffic accidents and their severity used statistical and rule-based methods. Logistic regression and decision trees were preferred due to their interpretability and simplicity [9], [10]. These methods are suitable when the relationships between the variables are simple and most of the variability can be explained by a few variables. Logistic regression models produce hard severity probabilities, and decision trees produce predictions in a step-by-step process. However, these models are not efficient in handling complex non-linear relationships in real-world traffic accident data. In addition, traffic accident data is imbalanced, as severe and fatal accidents are less common compared to minor ones. Traditional models are prone to overfitting the majority class, thus neglecting the rare but significant accidents [14]. Although these models have paved the way for traffic safety analytics, they are not efficient when handling large, diverse, and imbalanced data.

2.2 Machine Learning in Traffic Safety

With the availability of large-scale data on transportation, machine learning techniques have gained popularity in traffic safety studies. Random Forest, Gradient Boosting, and LightGBM are more effective in identifying non-linear and complex relationships between factors of accidents [7], [8]. These techniques are more superior to traditional techniques in that they take into account multiple variables simultaneously, handle large datasets, and provide accurate predictions. Ensemble techniques also take into account the marginal contribution of variables such as weather, traffic flow, and road geometry, which are not considered in traditional techniques. However, the major drawback of existing studies is that they give more importance to accuracy and less importance to interpretability [6], [16]. Black-box techniques provide highly accurate predictions but lack interpretability, making it difficult for policymakers to trust and act on their predictions.

2.3 Ensemble Learning Approaches

Trends include ensemble learning, which is the combination of the use of several models for better results. Unlike single models that are prone to ground bias or unrealistic inputs, ensemble learning utilizes the strengths of multiple algorithms. Stacked models enable the combination of the predictions of several base models, which in turn helps one algorithm to enhance the predictions of another algorithm for maximizing accuracy [5], [18]. LightGBM-TPE is an ensemble learning algorithm that combines gradient boosting and hyperparameter optimization for efficient training. SHAP values enhance

interpretability by providing important information on the ranking of features [6], [17]. Several comparisons have indicated that ensemble models like Random Forest and XGBoost are more accurate than baselines in accident severity prediction [7], [12].

2.4 Limitations in Existing Studies

Although there has been some progress in ML, there are some weaknesses in the literature. The first is the lack of interpretability of highly accurate models. Even highly accurate ensembles are "black boxes," and it is difficult for researchers to understand the contribution of features and use the results for decision-making purposes [6], [16]. The current visualizations are primarily explanatory and show correlations rather than actionable results. Another weakness is the inconsistent treatment of class imbalance. Although SMOTE has been available, not all studies use it consistently, resulting in suboptimal results for the detection of key severe cases [11], [15]. In addition, cross-validation on hold-out sets is not performed in most studies, and it is difficult to generalize over regions [7], [12]. Highly accurate causal models are still very limited, and most studies are association rather than causation studies [9], [14]. Models have to be accurate and reliable for practical traffic safety applications.

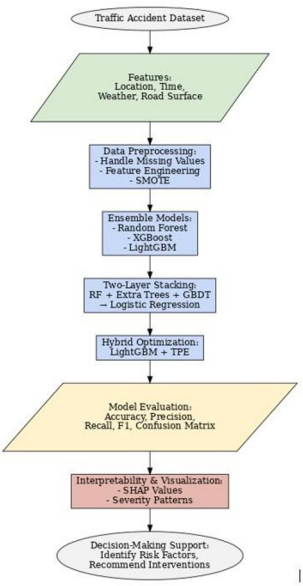
3 METHODOLOGY

3.1 Data Collection

The research uses a comprehensive dataset that includes complete information about traffic accidents. The most important features of the research include the geographical location of the accident in terms of its coordinates (latitude and longitude), the correct time of the accident, weather conditions, and the type of road surface. The severity of the accident, which is either minor, serious, or fatal, is the target variable for prediction [7], [8], [15].

3.2 Preprocessing

Data preprocessing helps the models learn from clean and meaningful data. Missing values are handled using the correct imputation methods, which include mode for categorical variables and median for numerical variables. Data with too much missing information is also discarded to ensure quality [11], [15]. Features that are highly correlated are also discarded to avoid redundancy. Time is further categorized into rush hour and off-peak hours. Coordinates are not altered.



The overall workflow of preprocessing, modeling, and evaluation is illustrated in Fig. 1.

FIGURE 1 : Workflow of preprocessing, modeling, and evaluation

3.3 Handling Class Imbalance with SMOTE

Severe accidents are less common but have the most effect, and they pose a challenge to predictive models learning from imbalanced data. SMOTE creates new instances of minority class events to handle the imbalance in the data [11]. This helps models to learn patterns associated with severe but less common accidents [15].

3.4 Machine Learning Models

Three robust ensemble models are developed and compared:

- Random Forest: Provides strength and robustness against overfitting, while also allowing interpretability through feature importance scores [7], [12].
- XGBoost: A boosting model that refines predictions incrementally, performing well in terms of accuracy and speed [6], [17].
- LightGBM: Designed for efficiency in big data, emphasizing high accuracy and fast training times [8], [18].

3.5 Stacking and Hybrid Frameworks

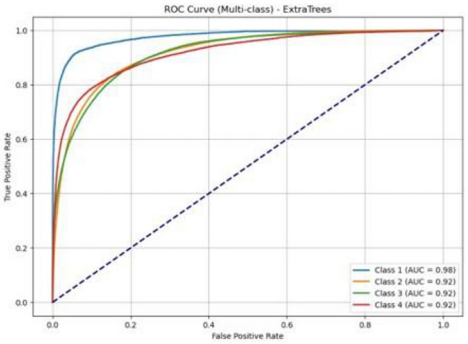
For further enhancement of the prediction accuracy, stacking methods and ensemble learning strategies are discussed. Stacking is performed using predictions from various base models, such as Random Forest, Extra Trees, and GBDT, with logistic regression as the meta-model [5], [18]. Furthermore, the LightGBM-TPE approach combines the boosting power of LightGBM with

automatic hyperparameter adjustment and explains the results using SHAP values [6], [17].

4 EXPERIMENTS AND EVALUATION

4.1 Evaluation Metrics

There are various evaluation metrics used to test the performance of the model. Accuracy is used to test the overall validity of predictions made by the model, but it may be deceptive in imbalanced datasets where small accidents are given priority [7], [12]. Precision, recall, and F1-score play a vital role in testing the efficiency of the model in classifying the minority class, such as fatal accidents, where accurate classification is of utmost importance [8], [15]. AUC-ROC and AUPRC help in understanding the capacity of the model to differentiate between severity levels at various thresholds [6], [16].

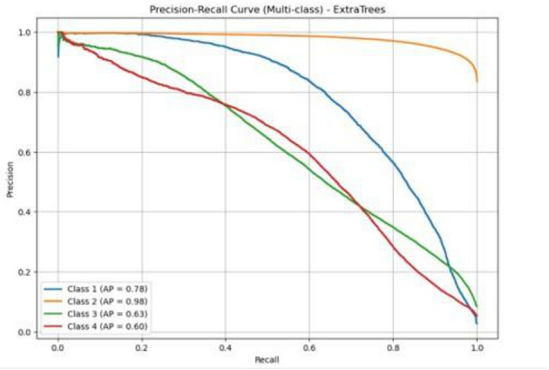


The ROC curves comparing the performance of different models are shown in Fig. 2.

FIGURE 2 : ROC curves for model comparison

FIGURE 3 : Precision-Recall curves

The precision-recall curves highlighting model performance on imbalanced data are presented in Fig. 3.



4.2 Training and Validation

The data is then split into training and testing sets, with the standard split being 70-80% for training and the remaining for testing purposes to avoid any bias in the testing process [8], [15]. To ensure that the results obtained are as accurate as possible, k-fold cross-validation with stratified sampling is used, which preserves the original class distribution in each fold. This ensures that the output variability is reduced and that the models are able to generalize well, particularly in the case of rare severe accidents.

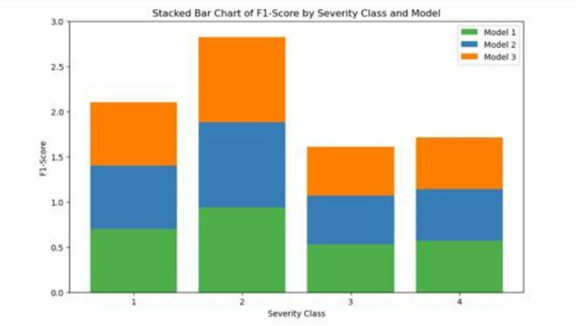
4.3 Implementation Details

Each experiment relies on widely known open-source Python libraries. Data processing and handling are done using pandas and numpy. Handling imbalanced classes is done using the SMOTE algorithm from imblearn [11]. Training of models is done using the scikit-learn, XGBoost, and LightGBM libraries. SHAP is used for explaining predictions of models and obtaining insights into feature importance. Plotting of evaluation curves is done using matplotlib and seaborn libraries [6], [18].

5 RESULTS AND DISCUSSION

5.1 Model Performance

The stacking ensemble model had the best overall performance with a macro F1-score of 0.69, proving its strength in dealing with different classes of severity [6], [7]. The LightGBM-TPE approach had the best recall performance, especially in the minority class of fatal, proving its strength in detecting rare but serious severe accidents.



A comparison of F1-scores across severity classes is depicted in Fig. 4.

FIGURE 4 : F1-score comparison across severity classes

TABLE 1
SUMMARY OF MODEL PERFORMANCE WITH KEY FEATURES AND NOTES

B. Confusion Matrix Analysis

Recall values for different severity classes are illustrated in Fig. 5.

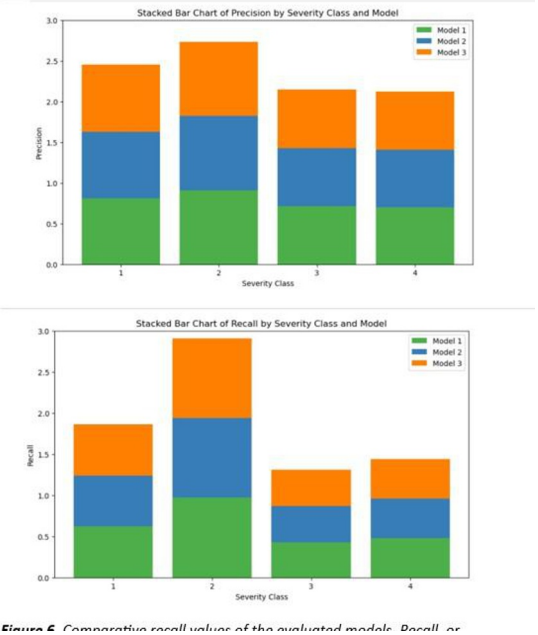


FIGURE 5 : Recall values by severity classes

The relatively low recall value for fatal accidents again emphasizes the problem of class imbalance, which has been a concern for some time and requires the continued use of techniques such as SMOTE and tailored modeling [11], [15].

The performance comparison of all models is summarized in Table 1.

Model	Accuracy	Macro F1	Features Used	Notes
Random Forest	0.85	0.66	Location,time,weather,road surface	Robust baseline;interpretable through feature importance
XGBoost	0.87	0.67	Same as RF	Strong accuracy with iterative error correction
LightGBM	0.88	0.68	Extended weather, lighting	Fast training; ideal for large datasets
Stacking	0.89	0.69	Combined features from base models	Best balanced accuracy and F1 score
LightGBM-TPE	0.88	0.70	LightGBM features with hyerparameter tuning	Highest recall for fatal minority class

5.2 Impact of Preprocessing and Class Balancing

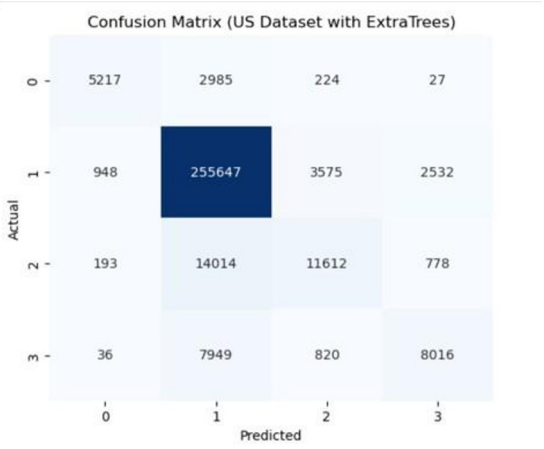
SMOTE and feature engineering helped improve the detection rate of the minority class and the stability of the model [8], [11], [15]. The preprocessing helped the model's ability to learn the rare but severe accidents by reducing the imbalance problem that normally exists in traffic severity prediction.
A smaller value of D indicates a better match between the generated images and the ground truth, proving the effectiveness of the frequency-based evaluation.

5.3 Relation to Existing Literature

Our results support previous research that recommended ensemble learning, balanced sampling techniques, and interpretability methods for creating interpretable and accurate models of accident severity predictions [5], [7], [12]. They support that a good model, interpretability, and data balance are well combined together.

6 VISUALIZATION AND INTERPRETABILITY

Visualization can also be used to interpret complex traffic crash data. Heat maps can be used to identify areas that are prone to traffic crashes or are dangerous, and safety researchers can make informed decisions about where to act. Visualization methods over time, such as time charts or box plots, can be used to identify trends such as the number of accidents being higher at night or during inclement weather conditions, making it easier to interpret the findings [5].
Beyond the simple visualizations, the application of interpretability methods such as SHAP can be used to provide a clear understanding of the predictions made by machine learning models. SHAP summary plots can be used to determine the factors, whether location, weather, or time, that the predictions of accident severity are most sensitive to [6].



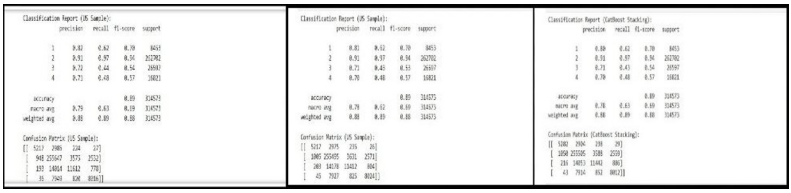
The confusion matrix illustrating prediction errors is shown in Fig. 7.

Figure 7. Normalized confusion matrix showing prediction errors

7 CONCLUSION

This paper verifies that the integration of machine learning algorithms, class balancing methods like SMOTE, and interpretability methods like SHAP provides reliable and interpretable outcomes for the prediction of the severity of traffic accidents [6], [7]. The ensemble learning method recognizes complex patterns between traffic accident data, SMOTE handles the significant issue of class imbalance, and SHAP bridges technology and human knowledge.

Enhanced severity prediction helps inform policy decisions, which focus on high-risk situations and regions. The emergency department maximizes response efforts, and road safety programs can be developed to counteract risk factors using interpretable models [3], [4]. These enhancements help in minimizing road injuries and deaths.



D. Impact of Preprocessing and Class Balancing

8 FUTURE WORK

However, the relevance of this study is limited by the use of data from one region, which may influence the degree of generalization of the study to other regions. In addition, the relevance of this study is anchored on the relationship between the variables and the degree of severity of the results, as opposed to the cause-and-effect relationship between the variables [2]. Future studies should aim at the use of data from other regions and other sources of information, including real-time data from IoT sensors, to enhance the relevance and responsiveness of the model. Causal modeling techniques would also provide more information on the dynamics of accidents. Real-time analytics platforms would also provide more opportunities for the actionability of predictive knowledge in traffic safety control [7].

REFERENCES

[1] World Health Organization. Global status report on road safety 2019.

[2] K. Bhalla et al., "Building national estimates of the burden of road traffic injuries in developing countries," Int. J. Inj. Contr. Saf. Promot., 2020.

[3] S. M. Lee, A. Al-Mansour, "Traffic accidents trends and safety improvements in Saudi Arabia," International Journal of Injury Control and Safety Promotion, 2020.
[4] A. Al-Tit et al., "Factors affecting road accident severity in the Gulf Cooperation Council (GCC) countries," Int. J. Ind. Syst. Eng., 2020.
[5] J. Tang et al., "Crash injury severity prediction using two-layer ensemble machine learning model," J. Adv. Transp., 2019. [6] Z. Li, C. Xu, Y. Liu, "A LightGBM-TPE model for traffic accident severity prediction," Accid. Anal. Prev., 2022.
[7] A. Jamal et al., "Injury severity prediction of traffic crashes with ensemble machine learning techniques," Int. J. Inj. Contr. Saf. Promot., 2021.
[8] K. Santos et al., "A comparative analysis of machine learning algorithms for accident severity prediction," J. Saf. Res., 2022.
[9] N. Ullah et al., "Statistical and machine learning crash severity modeling for road safety," 2021.
[10] M. Zahid, C. Chen, A. Jamal, M. Al-Ahmadi, "Crash injury severity analysis using advanced machine learning and statistical methods," 2020.
[11] N.V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," J. Artif. Intell. Res., 2002.
[12] O.E. Obasi, et al., "Evaluating Machine Learning algorithms for Injury Severity Prediction of Road Crashes," Journal of Safety Science and Resilience, 2023.
[13] M.E. Mostafa et al., "Predicting crash severity on roadways using machine learning models," *Scientific Reports*, 2024.
[14] T. Çelik et al., "A comparison of Logistic Regression and XGBoost for traffic accident severity prediction," *Turkish Journal of Computing*, 2024.
[15] N. Sakib et al., "Using ensemble methods to identify key factors in fatal traffic accidents with imbalanced data," *Elsevier Journal of Transport Safety*, 2024.
[16] W.K. Sum et al., "Analyzing motorcycle crash severity in urban settings with Random Forest and SHAP," *Transportation Research*, 2024.
[17] Y. Chen et al., "A hybrid XGBoost model for improved traffic accident severity prediction," *Accident Analysis and Prevention*, 2024.
[18] K. Deepak et al., "Enhancing accident severity prediction through a stacking ensemble approach with SMOTE," *Journal of Safety Research*, 2024.
[19] N. Singh and M. Kumar, "A framework for identifying and classifying accident hotspots using gradient boosting and spatial clustering," *Earth Science Informatics*, vol. 18, no. 1, p. 168, 2024.