Road Traffic Accident Severity Prediction

221FA04295-T.Prem kumar
Department of Computer Science and Engineering
Vignan's Foundation for Science, Technology and Research
Guntur, India
221fa04295@gmail.com

221FA04316-Syed Sohail
Department of Computer Science and Engineering
Vignan's Foundation for Science, Technology and Research
Guntur, India
221fa04316@gmail.com

221FA04378-L.Dolasai
Department of Computer Science and Engineering
Vignan's Foundation for Science, Technology and Research
Guntur, India
221fa04378@gmail.com

221FA04381-M.Sujith
Department of Computer Science and Engineering
Vignan's Foundation for Science, Technology and Research
Guntur, India
221fa04381@gmail.com

***Abstract*— Road accidents pose a significant challenge globally, contributing to numerous fatalities and injuries every year. Accurate prediction of accident severity is crucial for enhancing emergency response management and improving road safety. This paper presents a machine learning approach to predict accident severity, utilizing algorithms such as Random Forest, Logistic Regression, Support Vector Machines (SVM), and Decision Trees. The dataset includes accident-related features such as the number of vehicles involved, number of casualties, and road conditions. To address the issue of class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was applied. Among the models tested, Random Forest outperformed the others, achieving the highest accuracy in predicting accident severity. The findings of this study can assist in better decision-making processes for emergency response teams, thereby saving lives and optimizing resource allocation.**

Keywords: Accident Severity Prediction, Machine Learning, Random Forest, Logistic Regression, SVM, Decision Trees, SMOTE, Class Imbalance, Emergency Response Management, Road Safety, Vehicle Involvement, Casualties, Road Conditions, Data Preprocessing, Model Performance, Resource Allocation

I.INTRODUCTION

Road Traffic Accidents represent a global public health issue, resulting in over 1.3 million deaths annually, according to WHO reports. These accidents often vary in severity, and predicting the seriousness of an accident can significantly aid emergency response teams and policymakers in resource allocation and accident prevention. Machine learning provides an effective approach for predicting accident severity by analyzing historical accident data and identifying patterns in the contributing factors

II.LITERATURE REVIEW

In the field of road safety management, building models to predict the severity of traffic accidents is crucial. This research examines the various factors that contribute to accident severity, including vehicle characteristics such as speed and size, as well as road features like design and traffic volume [1]**.**Road accidents have posed a significant threat to both developed and developing nations. Ensuring road safety has been a global concern for many years, with traffic congestion and reckless driving being prevalent worldwide [2]**.** In recent years, road accidents have emerged as a global issue and are now ranked as the ninth leading cause of death worldwide. Bangladesh, in particular, faces a significant challenge due to the high number of road accidents each year, making it a critical issue for the country[3]**.** Traffic accidents happen at an alarming rate across the globe. In India alone, over 151,000 lives were lost in road accidents in 2019. Identifying the key factors that lead to traffic collisions is essential to reducing both the human and financial toll [4]**.** Road accidents are on the rise globally, leading to millions of deaths each year and imposing substantial financial and economic burdens on society. While previous research has often treated road accident prediction as a classification task—focusing on whether an accident will occur—few studies have delved into the complex relationships between the factors that contribute to accidents. [5]. Engineers and researchers in the automotive industry have worked to design safer vehicles, but traffic accidents remain inevitable. However, by developing accurate prediction models that can automatically classify the severity of injuries in various traffic accidents, we can identify patterns in high-risk crashes[6]**.** In recent years, road traffic accidents (RTA) have emerged as a major global health concern, particularly affecting children and young people. RTA have become a leading cause of death in these groups [7]**.**In today's growing demand for data analytics in the drug discovery field, handling complex, high-dimensional drug datasets presents a significant research challenge. High dimensionality refers to datasets with more columns than rows, often containing irrelevant features that add little value [8]. Accidents are considered one of the most alarming issues in many countries today, leading to numerous fatalities worldwide, primarily due to road accidents during traffic congestion[9].We propose leveraging machine learning to explore the complexities of road accidents, with the goal of classifying them by severity and uncovering the key contributing factors [10]. Traffic accidents pose a significant risk and contribute to a high number of fatalities globally. Reducing these incidents is crucial for saving lives and fostering sustainable cities and communities. Machine learning and data analysis techniques can help identify the causes of car accidents and suggest ways to mitigate them [11]**.** Road accidents pose a significant threat in both developed and developing countries. Globally, road safety has been a major concern for years, with efforts ongoing to address this issue[12]**.** Road accidents are a pressing global issue, resulting in numerous fatalities and injuries each year. This project aims to predict the severity of road accidents using machine learning techniques to reduce their frequency and mitigate related risks[13]**.** The primary goal of this research was to develop a method for predicting the extent of damage from accidents using machine learning. The study created a prediction framework and utilized three different machine learning algorithms—Random Forest, Logistic Regression, and Decision Tree—to assess the potential impact of accidents[14]**.** Traffic accidents have increasingly become a critical global issue, impacting both lives and economic stability. The US Accidents dataset from 2016 to 2023 provides a comprehensive record of incidents across the United States, including detailed information about the accidents and their environmental contexts[15]. Traffic accidents have emerged as a significant global problem, leading to substantial human and financial losses. The World Health Organization estimates that road traffic accidents impact between 20 and 50 million people each year. [16]. Traffic accidents are a major global issue, leading to numerous casualties, injuries, fatalities, and significant economic losses every year. Understanding and predicting the factors that contribute to these accidents can help mitigate their impact and severity. [17]. Recently, road traffic accidents have emerged as a significant global issue, ranking as the ninth leading cause of death worldwide. This problem is particularly severe in Kerala, where the high number of road traffic accidents each year underscores a critical need for effective intervention. Addressing this issue is both urgent and essential.This research focuses on analyzing traffic accidents in Kerala using machine learning techniques to assess accident severity. The study aims to identify key factors that significantly influence road accidents and offer practical recommendations for improvement.We applied three supervised learning methods—Decision Tree, K-Nearest Neighbors (KNN), and Naive Bayes—to categorize the severity of accidents into three levels: fatal, grievous, and minor. The results indicate that Naive Bayes provided the best performance among the methods tested[18] . Traffic accidents continue to be a major cause of fatalities, injuries, and disruptions on roadways. Understanding the factors leading to these accidents is crucial for improving road safety. Recent research highlights the benefits of predictive modeling in identifying these factors. However, there has been limited focus on elucidating the inner workings of complex machine learning and deep learning models and how various features affect accident prediction. This lack of transparency risks making these models seem like "black boxes," potentially undermining stakeholder trust.This study aims to build predictive models using various transfer learning techniques and to clarify the impact of different features on these models through Shapley values. We utilize a range of models including Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Residual Networks (ResNet), EfficientNetB4, InceptionV3, Extreme Inception (Xception), and MobileNet to predict injury severity in accidents. Among these, MobileNet achieved the highest accuracy at 98.17%. By understanding how different features influence prediction models, researchers can better identify the factors contributing to accidents and develop more effective prevention strategies[19].Road accidents are a significant global issue, leading to numerous fatalities, injuries, and various economic losses. To address this problem, countries and international organizations have developed various technologies, systems, and policies. Leveraging big traffic data and artificial intelligence (AI) offers a promising approach to predicting and reducing the risk of road accidents. While many studies have focused on the effects of road geometry, environmental factors, and weather conditions on accidents, human factors such as alcohol consumption, drug use, age, and gender are often overlooked in assessments of accident severity.

In this study, we explored the impact of these human factors on predicting accident severity. We evaluated a range of machine learning (ML) methods, both single and ensemble models, to determine their effectiveness in predicting accident severity. The evaluation was based on several metrics, including prediction accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUROC).We approached the problem as a classification task, dividing it into binary classification (e.g., grievous vs. non-grievous) and multiclass classification (e.g., fatal, serious, minor, and non-injury). Our results indicate that the Random Forest (RF) algorithm outperformed other methods, including logistic regression (LR), K-nearest neighbors (KNN), naive Bayes (NB), extreme gradient boosting (XGBoost), and adaptive boosting (AdaBoost), achieving accuracy rates of 86.64% for binary classification and 67.67% for multiclass classification.Single-mode ML methods such as LR, KNN, and NB performed similarly in both binary and multiclass classifications. However, ensemble methods, particularly RF, XGBoost, and AdaBoost, demonstrated superior accuracy in predicting accident severity compared to single-mode methods. The insights gained from this study can help identify critical factors contributing to road accidents and improve the accuracy of severity predictions, ultimately aiding in the development of more effective safety measures [20].

III. METHODOLOGY

Input Dataset and Features

The dataset used in this project contains detailed records of road accidents, providing a comprehensive set of features to analyze and predict accident severity. Key attributes of the dataset include:

Number of Vehicles Involved: This feature records the total number of vehicles involved in each accident.

Number of Casualties: This includes the total number of injuries or fatalities in the accident.

Weather Conditions: Weather at the time of the accident (e.g., clear, rainy, foggy).

Road Alignment: The geometry of the road at the accident location, such as curves, straight paths, or inclines.

Accident Severity: The target variable, categorized into three classes—minor, serious, and fatal—indicating the severity of each accident.

The dataset was carefully selected to provide a wide range of attributes that can affect the severity of an accident, allowing for more nuanced predictions.

Data Preprocessing

Before training the machine learning models, several preprocessing steps were taken to clean and prepare the dataset for analysis:

Handling Missing Values: Missing data in key features such as "Driving Experience" and "Type of Vehicle" were addressed by filling the missing entries with the mode (most frequent value) of the respective feature. This ensures that no information is lost due to incomplete data.

Class Imbalance Handling: The target variable (Accident Severity) was imbalanced, with a majority of accidents classified as minor and only a few classified as serious or fatal. To address this, the **Synthetic Minority Over-sampling Technique (SMOTE)** was applied to the training set. SMOTE generates synthetic samples for the minority class (serious and fatal accidents) by creating new instances that are similar to existing ones. This step was crucial to ensure that the model was not biased towards predicting the more common minor accidents.

Feature Extraction

To identify the most important features that influence accident severity, various exploratory data analysis (EDA) techniques were applied:

Correlation Matrices: Correlation analysis was performed to understand the relationships between different variables. Features with high correlation to the target variable (Accident Severity) were identified as key predictors.

Visualizations: Scatterplots, boxplots, and heatmaps were generated to visually analyze relationships between features, such as the number of vehicles involved versus the number of casualties. Heatmaps helped highlight the strength of the relationships between variables, providing insights into how certain conditions (e.g., weather or road alignment) affect accident severity.

Outlier Detection: Boxplots were used to detect outliers in features like "Number of Vehicles Involved" and "Number of Casualties," which could skew the model's predictions. Identifying these outliers helped improve the model's performance.

Model Building

Several machine learning algorithms were trained to predict accident severity. Each model has unique strengths, allowing for a comparison of their performance:

Logistic Regression: This linear model was used as a baseline for classifying accident severity. It is favored for its simplicity and interpretability, making it easy to understand how changes in input features affect the output prediction.

***Decision Tree**: A non-linear model that splits the data into decision nodes based on feature values. Decision trees are useful for capturing complex interactions between features and can easily model non-linear relationships.

Support Vector Machine (SVM): SVM was chosen for its robustness in handling high-dimensional data. It works well in cases where the number of features is large relative to the number of observations. The SVM model uses hyperplanes to classify data points and can capture intricate boundaries between different classes.

Random Forest: An ensemble learning method that combines multiple decision trees to improve accuracy and prevent overfitting. By averaging the predictions from multiple trees, Random Forest increases the stability and accuracy of the model. It is especially effective when dealing with complex, high-dimensional datasets.

Each model was trained on the training set after applying SMOTE to balance the data. Hyperparameter tuning was performed to optimize each model's performance.

Model Evaluation

After training the models, several evaluation metrics were used to assess their performance:

Accuracy: The overall percentage of correct predictions made by the model across all classes (minor, serious, fatal).

Precision: Precision was calculated for each class to measure the proportion of true positives out of all predicted positives. For example, precision for the "fatal" class represents the accuracy of predictions that classify an accident as fatal.

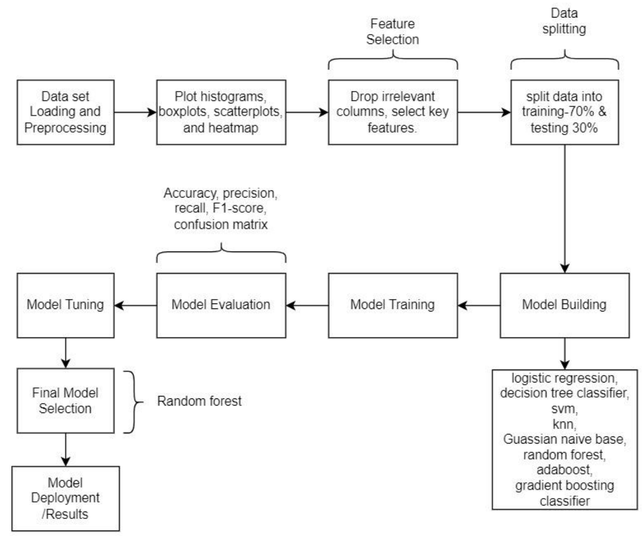
Recall: Recall measures the proportion of actual positives that were correctly predicted by the model. High recall for the "serious" and "fatal" classes was particularly important, as false negatives (severe accidents predicted as minor) can have significant real-world consequences.

F1-Score: This is the harmonic mean of precision and recall, providing a single metric to assess the model's balance between these two metrics.

Confusion Matrix: A confusion matrix was generated for each model to visualize the number of true positives, false positives, true negatives, and false negatives. This provided a more detailed understanding of where each model was making errors, particularly in distinguishing between minor and severe accidents.

Cross-Validation: To ensure model robustness, **k-fold cross-validation** was applied. This technique divides the data into k subsets, trains the model on k-1 subsets, and tests it on the remaining subset. The process is repeated k times, and the average performance across all folds is calculated.

Finally, the results from each model were compared, with Random Forest typically achieving the highest accuracy and robustness in handling class imbalance. Hyperparameter tuning, using techniques such as GridSearchCV, was applied to the Random Forest model to further improve its performance.



IV. IMPLEMENTATION

1.Environment Setup : The project was implemented in **Python** using a variety of libraries designed for data handling, machine learning, and visualization. The key libraries include:

Pandas for data manipulation, allowing easy reading, filtering, and transformation of the accident dataset.

NumPy for numerical operations, especially for array manipulation and handling mathematical computations.

Scikit-learn for implementing machine learning algorithms, data splitting, preprocessing, and model evaluation.

Seaborn and **Matplotlib** for creating informative visualizations, such as scatterplots, heatmaps, and histograms, which help in understanding the dataset's distribution and relationships between features.

imbalanced-learn for applying **SMOTE** (Synthetic Minority Over-sampling Technique) to balance the dataset by generating synthetic samples for the minority class.

Jupyter Notebooks was the development environment used for running code, visualizing outputs inline, and documenting the process.

2. Code for Preprocessing, SMOTE, and Model TrainingData Preprocessing:The dataset was first examined for missing values. Critical features such as **"Driving Experience"** and **"Type of Vehicle"** contained missing data, which were filled with the most frequent value (mode) in each column. This ensured that the models could utilize as much data as possible without dropping rows, thereby maintaining the dataset's integrity.

Label Encoding: Categorical variables, which represent data in text form (such as vehicle type, road alignment, accident severity), were converted into numeric format using **LabelEncoder** from Scikit-learn. This step is crucial as machine learning models require numeric inputs to perform calculations.

Class Imbalance Handling Using SMOTE:One of the significant challenges with accident severity prediction is the imbalance in the target variable, where severe accidents are underrepresented compared to minor ones. **SMOTE** was applied to balance the classes by generating synthetic data for the underrepresented (serious and fatal) categories. This step ensures that the machine learning models are not biased towards predicting only minor accidents and can accurately classify all severity levels.

ModelTraining: After preprocessing, several machine learning models were built to classify the accident severity. These models were implemented using Scikit-learn and trained on the preprocessed data:

Logistic Regression: A linear model that served as the baseline. It is simple yet effective for binary or multi-class classification problems.

$$\text{Logistic Regression Hypothesis: } h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

Decision Trees: This non-linear model splits the data based on feature values, helping to capture complex relationships between the input features and accident severity.

Support Vector Machines (SVM): Known for its ability to handle high-dimensional data, SVM was used to separate the classes by maximizing the margin between different severity levels.

$$\min_{w,b} \frac{1}{2}||w||^2 \text{ subject to } y_i(w \cdot x_i + b) \geq 1$$

Random Forest: An ensemble learning method that builds multiple decision trees and aggregates their predictions to improve model accuracy and reduce overfitting. Random Forest is particularly effective with complex datasets containing both categorical and numerical data.

The dataset was split into training (70%) and testing (30%) sets using **train_test_split()** from Scikit-learn. After applying SMOTE to the training set, the models were trained using the training data.

3. Model Selection and Tuning

Random Forest Hyperparameter Tuning: **GridSearchCV** was used to find the optimal hyperparameters for the Random Forest model. The hyperparameters tuned included:

n_estimators: This parameter controls the number of trees in the forest. Testing different values helps find the balance between model performance and computation time.

criterion: The criterion defines the function used to measure the quality of splits (either **gini** impurity or **entropy**).

max_depth: This parameter restricts how deep the tree can grow, helping to prevent overfitting by limiting the model's complexity.

SVM Hyperparameter Tuning: Similarly, **GridSearchCV** was applied to optimize the SVM model. The key hyperparameters tuned included:

kernel: Various kernels (e.g., linear, radial basis function (RBF)) were tested to find the best fit for the data.

C: The regularization parameter that controls the trade-off between achieving a low error on the training data and maintaining generalization to unseen data.

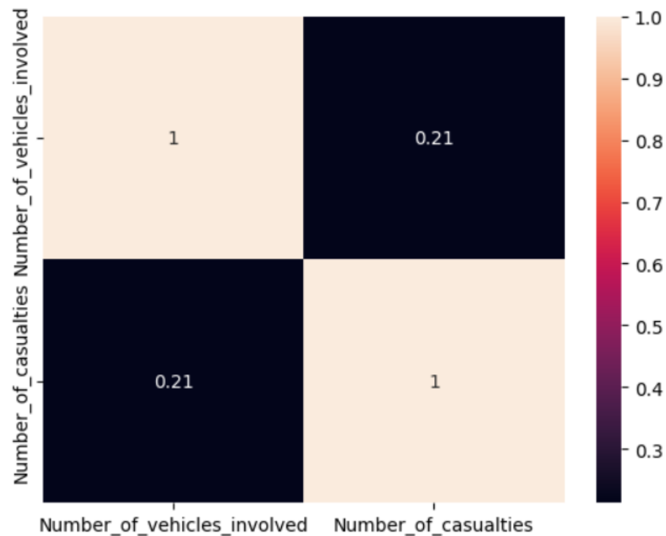
By testing a range of values for each hyperparameter, the models were optimized to improve their performance in predicting accident severity

V. EXPERIMENTATION AND RESULT ANALYSIS

Model Comparison:

The performance of each machine learning model was evaluated using the test set after training the models on the training set (which was balanced using SMOTE). The models compared included **Logistic Regression**, **Decision Tree**, **Support Vector Machine (SVM)**, and **Random Forest**. Each model's performance was measured using several key metrics, including **accuracy**, **precision**, **recall**, and **F1-score**.

- **Random Forest** achieved the highest accuracy of **88%**, significantly outperforming the other models. This result can be attributed to the ensemble nature of Random Forest, which combines the predictions of multiple decision trees to achieve higher accuracy and reduce overfitting. The model's ability to capture non-linear relationships and handle a mix of categorical and numerical data made it highly effective for the given dataset.

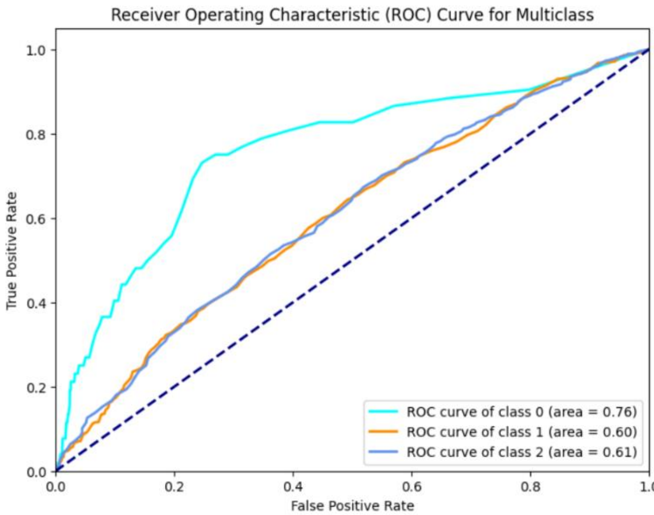


- **Logistic Regression** performed adequately but was limited by its assumption of linear relationships between input features and the target variable. While it served as a useful baseline, its accuracy was lower than that of the more complex models like Random Forest and SVM.
- **Support Vector Machine (SVM)** also demonstrated strong performance, especially in handling high-dimensional data, but it could not outperform Random Forest in this case. This was likely due to the complexity of the relationships in the data, which were better captured by the ensemble approach of Random Forest.
- **Decision Tree** offered reasonable accuracy but was prone to overfitting, particularly before hyperparameter tuning. It worked well for smaller datasets but lacked the robustness seen in Random Forest, which combines multiple trees for improved generalization.

The results clearly indicated that **Random Forest** was the best model for predicting accident severity based on the available features.

Impact of SMOTE on Class Balancing:

Class imbalance is a common issue in accident severity prediction, where the majority of accidents are categorized as "minor," while the more critical classes, "serious" and "fatal," are underrepresented. Without addressing this imbalance, machine learning models tend to favor the majority class, resulting in poor performance when predicting severe accidents.



To tackle this issue, the **SMOTE (Synthetic Minority Over-sampling Technique)** algorithm was applied to the training data. SMOTE works by creating synthetic instances of the minority classes (serious and fatal accidents) to balance the class distribution. This led to a significant improvement in model performance, especially in the prediction of severe accidents.

- **Before applying SMOTE**, models like Logistic Regression and Decision Trees struggled with low recall for the minority classes, often misclassifying severe accidents as minor. This led to high false negative rates, which can be dangerous in real-world scenarios where underestimating accident severity may lead to inadequate emergency response.
- **After applying SMOTE**, the models, especially Random Forest, showed marked improvements in **recall** for the severe classes. This means the models became much better at correctly identifying severe accidents. In particular, SMOTE helped to reduce false negatives, ensuring that more serious accidents were correctly classified.

In summary, SMOTE enhanced the models' ability to accurately predict severe accidents, thus making the predictions more balanced and reliable.

Performance Metrics:

Each model's performance was evaluated using a range of metrics, including **precision**, **recall**, **F1-score**, and **confusion matrix**. These metrics provided a more comprehensive evaluation beyond just accuracy, ensuring that the models were effective at predicting all classes, especially the minority ones.

- **Random Forest** yielded the highest **F1-score** of **0.85**, indicating a good balance between precision and recall. The **F1-score** is particularly useful in cases of class imbalance, as it accounts for both false positives and false negatives, ensuring that the model performs well across all accident severity levels.
- **Precision** and **Recall**: Random Forest demonstrated high precision, particularly for the "fatal" class, meaning that when the model predicted a fatal accident, it was correct most of the time. Additionally, the recall for severe accident cases was significantly improved after applying SMOTE, ensuring that the model was able to detect a greater proportion of actual severe accidents.

```
0.7713125845737483
              precision    recall  f1-score   support

           0       0.07      0.04      0.05        52
           1       0.24      0.19      0.21       552
           2       0.85      0.89      0.87      3091

    accuracy                           0.77      3695
   macro avg       0.39      0.37      0.38      3695
weighted avg       0.75      0.77      0.76      3695
```

- **Confusion Matrix**: The confusion matrix for Random Forest revealed a more balanced classification across all three severity classes (minor, serious, fatal) after applying SMOTE. The number of correct predictions for serious and fatal accidents increased, while misclassifications (false positives and false negatives) were reduced. The confusion matrix illustrated the model's ability to distinguish between the different accident severity levels more effectively after the dataset was balanced.

The combination of these performance metrics indicated that Random Forest, when combined with SMOTE for class balancing, was the most robust and reliable model for predicting accident severity.

VI. CONCLUSION

KeyFindings

Random Forest emerged as the most accurate model for accident severity prediction. The use of SMOTE was essential in addressing data imbalance and improving the prediction of severe accidents. Hyperparameter tuning further enhanced model performance.

FutureEnhancements

Future work could explore the use of more advanced models like XGBoost and deep learning methods such as LSTM for time-series data. Additionally, incorporating real-time data such as traffic conditions and weather reports could improve the accuracy of accident severity predictions.

VII. REFERENCES

[1] K. Kodepogu, V. B. Manjeti, and A. B. Siriki, “Machine Learning for Road Accident Severity Prediction,” *Mechatronics Intell. Transp. Syst.*, vol. 2, no. 4, pp. 211–226, 2023, doi: 10.56578/mits020403.

[2] J. Patil, M. Prabhu, D. Walavalkar, and V. B. Lobo, “Road Accident Analysis using Machine Learning,” *2020 IEEE Pune Sect. Int. Conf. PuneCon 2020*, no. 03, pp. 108–112, 2020, doi: 10.1109/PuneCon50868.2020.9362403.

[3] M. F. Labib, A. S. Rifat, M. M. Hossain, A. K. Das, and F. Nawrine, “Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh,” *2019 7th Int. Conf. Smart Comput. Commun.*

ICSCC 2019, no. September, pp. 1–5, 2019, doi: 10.1109/ICSCC.2019.8843640.

[4] A. Jadhav and R. Pawar, "Road accident analysis and prediction," *AIP Conf. Proc.*, vol. 2853, no. 1, pp. 740–747, 2024, doi: 10.1063/5.0197414.

[5] S. Ahmed, M. A. Hossain, S. K. Ray, M. M. I. Bhuiyan, and S. R. Sabuj, "A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance," *Transp. Res. Interdiscip. Perspect.*, vol. 19, no. April, p. 100814, 2023, doi: 10.1016/j.trip.2023.100814.

[6] M. Chong, A. Abraham, and M. Paprzycki, "Traffic accident analysis using machine learning paradigms," *Inform.*, vol. 29, no. 1, pp. 89–98, 2005.

[7] Z. Farhart*, A. Karouni, B. Daya, P. Chauvet, and N. Hmadeh, "Traffic Accidents Severity Prediction using Support Vector Machine Models," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 7, pp. 1345–1350, 2020, doi: 10.35940/ijitee.f4393.059720.

[8] A. Dadhwal and M. Gupta, "Analysis and Prediction of Drugs using Machine Learning Techniques," *Proc. - 2021 3rd Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2021*, vol. 9, no. 8, pp. 21–27, 2021, doi: 10.1109/ICAC3N53548.2021.9725675.

[9] S. K. Reddy Koduru, "Prediction of Severity of an Accident Based on the Extent of Injury using Machine Learning," *Int. J. Comput. Trends Technol.*, vol. 70, no. 9, pp. 43–49, 2022, doi: 10.14445/22312803/ijctt-v70i9p106.

[10] A. Pradesh, "ROAD ACCIDENT ANALYSIS AND CLASSIFICATION BY USING MACHINE LEARNING," vol. XIII, no. Iv, pp. 1770–1778.

[11] S. Pourroostaei Ardakani *et al.*, "Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis," *Sustain.*, vol. 15, no. 7, 2023, doi: 10.3390/su15075939.

[12] J. Patil, M. Prabhu, D. Walavalkar, and V. B. Lobo, "Road Accident Analysis using Machine Learning," *2020 IEEE Pune Sect. Int. Conf. PuneCon 2020*, vol. 8, no. I, pp. 108–112, 2020, doi: 10.1109/PuneCon50868.2020.9362403.

[13] R. Arunachalam, S. Peararulselvi, M. Saraswathi, and M. Saraswathi, "Road Accident Severity Prediction using Machine Learning," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 631–636, 2023, doi: 10.48175/ijarsct-9629.

[14] D. V. Shah, "Prediction of Accident Severity Using Machine Learning Algorithms," 2023.

[15] W. Zhou, "Traffic Accident Severity Prediction Based on Data Cleaning and Machine Learning (Random Forest / Xgboost)," *Highlights Sci. Eng. Technol.*, vol. 85, pp. 376–388, 2024, doi: 10.54097/h8cq6864.

[16] T. T. Adeliyi, D. Oluwadele, K. Igwe, and O. J. Aroba, "Analysis of Road Traffic Accidents Severity Using a Pruned Tree-Based Model," *Int. J. Transp. Dev. Integr.*, vol. 7, no. 2, pp. 131–138, 2023, doi: 10.18280/ijtdi.070208.

[17] D. Santos, J. Saias, P. Quaresma, and V. B. Nogueira, "Machine learning approaches to traffic accident analysis and hotspot prediction," *Computers*, vol. 10, no. 12, 2021, doi: 10.3390/computers10120157.

[18] T. Varghese, A. R. Varghese, and S. M. Abraham, "Road Accident Analysis And Severity Prediction in Kerala using Machine Learning Techniques," vol. 8, no. 1, pp. 225–229, 2021.

[19] O. I. Aboulola, "Improving traffic accident severity prediction using MobileNet transfer learning model and SHAP XAI technique," *PLoS One*, vol. 19, no. 4 April, pp. 1–18, 2024, doi: 10.1371/journal.pone.0300640.

[20] S. Ahmed, M. A. Hossain, M. M. I. Bhuiyan, and S. K. Ray, "A Comparative Study of Machine Learning Algorithms to Predict Road Accident Severity," *Proc. - 2021 20th Int. Conf. Ubiquitous Comput. Commun. 2021 20th Int. Conf. Comput. Inf. Technol. 2021 4th Int. Conf. Data Sci. Comput. Intell. 2021 11th Int. Conf. Smart Comput. Networking, Serv. IUCC/CIT/DSCI/SmartCNS 2021*, pp. 390–397, 2021, doi: 10.1109/IUCC-CIT-DSCI-SmartCNS55181.2021.00069.