Predicting Accident-Prone Zones Using Machine Learning: A Web Application Approach

M. Shyamala Devi
Department of Computer Science and Engineering,
Panimalar Engineering College,
Chennai, Tamilnadu, India
shyamalapmr@gmail.com

Roshni K
Department of Computer Science and Engineering,
Panimalar Engineering College,
Chennai, Tamilnadu, India
roshnik2108@gmail.com

Shamili R
Department of Computer Science and Engineering,
Panimalar Engineering College,
Chennai, Tamilnadu, India
rshamili2004@gmail.com

Abstract*— The increasing incidence of road accidents poses significant challenges to public safety and urban planning. Effective risk reduction techniques are necessary because traffic accidents represent a serious danger to public safety. In order to identify accident-prone areas, this research analyses historical traffic data, which includes information on accident instances, environmental factors, road characteristics, and traffic flow patterns. The goal of this project is to create an intuitive online tool that helps policymakers, traffic authorities, and urban planners visualize danger zones while collecting useful information. This paper presents a machine learning-based web application that predicts accident-prone zones and sends real-time notifications to users navigating these areas. With the application's dynamic user interface, users may see high-risk areas on a map and obtain comprehensive analytics to help them make wise decisions. By integrating historical accident data with geographic information, the application employs a Random Forest model to analyze and forecast risk zones, thereby enhancing road safety. User feedback demonstrates the application's efficacy and usability, revealing an accuracy of 97.63% in predictions and providing a proactive approach to accident awareness. Stakeholders may highlight initiatives by using the data, which identify important variables contributing to accidents. The objective of this tool is to improve road safety and lower the frequency of traffic accidents by enabling data-driven decision-making. Future work will be directed towards the integration of real-time traffic data and the enlargement of the model's domain to encompass diverse geographical areas.

Keywords*— *accuracy, CNN, depthwise separable convolutions DL, filtering, VGG.

I. INTRODUCTION

Road accidents are a major global concern, leading to millions of fatalities and injuries each year. According to the World Health Organization (WHO), approximately 1.3 million people die annually due to road traffic accidents, with many more sustaining life-altering injuries. Understanding the causes and patterns of these accidents is crucial for developing effective safety measures. Traditional navigation systems focus primarily on real-time traffic conditions but often overlook predictive analytics that could preemptively warn users of high-risk zones. Public safety and urban infrastructure are severely hampered by traffic accidents, which continue to be the world's largest source of injury and deaths. There is a growing demand for efficient methods to detect and reduce accident-prone locations as cities expand and traffic volumes rise. Conventional accident analysis techniques frequently depend on anecdotal evidence and historical data, which may not fully reflect the intricate interactions between variables that lead to traffic accidents. New developments in machine learning provide viable ways to improve accident prediction accuracy by utilizing extensive datasets including a range of factors, such as traffic patterns, road features, environmental conditions, and driver behavior. Machine learning algorithms offer a more sophisticated knowledge of accident risk variables by revealing hidden patterns and connections that standard statistical approaches can ignore. The goal of this work is to use machine learning techniques to create a prediction model that can identify areas that are prone to accidents. This research aim is to provide a strong tool that may assist in informing policy and urban planning choices by utilizing sophisticated analytical techniques and combining a variety of information. The online application that is produced will not only show high-risk regions but also provide relevant information to stakeholders so they may carry out focused interventions. This will eventually improve road safety and lower the number of traffic accidents.

A. Objectives

This paper aims to design and implement a web application that utilizes machine learning to predict accident-prone areas. The application not only identifies these zones but also integrates with Google Maps to notify users in real-time when they are approaching a high-risk area.

B. Contribution

The main contributions of this paper are as follows:

1. A machine learning-based model to predict accident-prone zones using historical geospatial data.
2. Integration with Google Maps for real-time monitoring of user routes.
3. A notification system that alerts users when they approach accident-prone zones via push notifications.
4. Evaluation of the system's performance through accuracy metrics and user feedback to assess real-world feasibility.

This paper is organized as follows: Section 2 reviews related work in accident prediction and machine learning applications in transportation. Section 3 outlines the methodology used for data processing, model training, and system architecture. Section 4 presents the evaluation results, including user feedback. Finally, Section 5 concludes the paper and discusses future research directions.

II. BACKGROUND STUDY

A. Traffic Accident Prediction

Research on traffic accident prediction has gained traction over the years, with various methodologies applied to analyze accident data. Early studies utilized statistical methods such as regression analysis to identify significant factors contributing to accidents. For example, [1] explored logistic regression for predicting traffic accidents. The inferences obtained from the background study is displayed in Table. I.

TABLE I. INFRENCES FROM LITERATURE REVIEW

S.No	Source	Methodology	Inference
1.	K. S. Wowo et al., [2]	SVM for accident prone zone prediction	The model achieves with 80 % of accuracy. The model has to be refined by applying other classifiers.
2.	D. Dinesh., et al., [3]	Multi-Model ML for Real-Time Road Accident Prediction	The model achieves with 85 % of accuracy. The model should be refined to process other features also.
3.	J. Lee., et al., [4]	Decision Tree for Accident Prediction	The model achieves with 0.282% of RMSE. The model can be refined for analyzing the accidents for wet road conditions.
4.	G. Singh., et al., [5]	DNN for road accident prediction	The model achieves with 8.86% of RMSE. The model can be applied with other performance metrics.
5.	Ferit Yakar., et al., [6]	Multicriteria decision making for road accident prediction	The model achieves with 83.98% of accuracy. The model can be refined with applying road properties.
6.	Swarnima Singh., et al., [7]	Particle Swarm Optimization for Prediction of Accident Severity	The model achieves with 84.63% of accuracy. The model should be refined with feature engineering
7.	Miin-Jong Hao., et al., [8]	Model-Based Fuzzy System for Predicting Traffic Congestion	The polynomial regression model outperforms other models and it can be refined for diverse road conditions.
8.	Zaniar Babaei., et al., [9]	DBSCAN-based HSID method for detecting road accident hotspots	The model achieves with 84% of accuracy. The model should be refined with identifying hidden hazardous locations.
9.	Burak Yigit Katanalp., et al., [10]	GIS-based Multi-Criteria Decision Making (MCDM) approach for predicting pedestrian-vehicle accident prone locations	The model achieves with 93% of accuracy. The model can be refined to address the pedestrian safety
10.	Safa Sababhi., et al., [11]	Global Moran I index for spatial autocorrelation analysis for predicting accident prone locations	The model should be refined with the dimensionality reduction methods.
11.	K. Alkaabi., et al., [12]	Logistical regression model for Identification of hotspot areas for traffic accidents	The model achieves with 90% of accuracy. The model can be refined to address the ranking of high-density vehicle crash areas
12.	Eskindir Ayele Atumo., et al., [13]	Random forest approaches for traffic crash hot spot identification and prediction	The model achieves with 75% of accuracy. The model can be refined to address the spatial dependence of high and low property damage

B. *Machine Learning Approaches*

The emergence of machine learning has revolutionized the field of accident prediction. Studies have demonstrated that models like Decision Trees, Support Vector Machines (SVM), and Neural Networks can provide higher accuracy than traditional methods. For instance, [2] employed ensemble learning techniques to enhance prediction performance. Our study builds upon this body of research by implementing a Random Forest model, known for its robustness and interpretability.

C. *Geospatial Analysis and Hotspot Detection*

Geospatial analysis plays a crucial role in identifying accident hotspots. Techniques such as Kernel Density Estimation (KDE) and clustering algorithms (e.g., DBSCAN) have been utilized to visualize high-risk areas. [3] emphasized the importance of geographic information systems (GIS) in understanding accident distributions, yet few studies have integrated these analyses into real-time navigation systems.

D. *Limitations of Existing Solutions*

Current navigation applications, such as Google Maps and Waze, primarily focus on real-time traffic conditions and historical accident data, often providing reactive alerts post-incident. This paper aims to fill this gap by offering a predictive approach that alerts users before they enter high-risk zones, thus enhancing overall road safety.

III. PROPOSED METHODOLOGY

A. *Dataset Description*

The dataset utilized in this study includes historical accident records obtained from the National Highway Traffic Safety Administration (NHTSA) from KAGGLE**,** spanning several years and covering various geographical regions. The dataset consists of attributes such as accident location, time, weather

conditions, and involved vehicles, which are crucial for understanding patterns in road safety.

B. *Data Preprocessing*

Data preprocessing is critical to ensure the quality and relevance of the input data for the machine learning model. The following steps were performed:

- Handling Missing Values: Missing entries were addressed using imputation techniques based on the mean or mode of respective features.
- Feature Selection: Features such as location coordinates, time of day, weather conditions, and accident severity were selected based on their potential influence on accident occurrence.
- Normalization: Continuous variables were normalized to ensure they contributed equally to the model training process.

C. *Feature Engineering*

Feature engineering involved creating new features that could enhance the model's predictive capabilities:

- Temporal Features: Day of the week and hour of the day were derived from the accident timestamps to capture time-related patterns.
- Weather Indicators: Categorical features indicating weather conditions (e.g., sunny, rainy, foggy) were created based on historical weather data during accidents.

D. *Model Selection and Training*

The **Random Forest** algorithm was chosen for this study due to its ability to handle non-linear relationships and provide robust predictions. The model was trained using 70% of the dataset, with the remaining 30% reserved for testing. Hyperparameter tuning was conducted using Grid Search to identify the optimal parameters, enhancing model performance.

E. Hyperparameter Tuning

The following hyperparameters were tuned during the training process:

- Number of trees in the forest
- Maximum depth of the trees
- Minimum samples required to split an internal node
- Minimum samples required to be at a leaf node

The results of the tuning process indicated a significant improvement in model accuracy with the chosen hyperparameters.

F. System Architecture

The architecture of the proposed system is shown in Fig. 1 that consists of the following key components:

- Frontend: The user interface was built using React.js, enabling users to input their desired routes and view accident-prone zones in real-time.
- Backend: The backend is a Flask application that hosts the machine learning model and interacts with the Google Maps API to monitor routes. The system processes user route data, querying the prediction model to identify any accident-prone zones along the path.
- Database: A PostgreSQL database is used to store historical accident data, user route logs, and other metadata required for the system.
- Notification System: Real-time notifications are powered by Firebase Cloud Messaging (FCM). When the user is within a 500-meter radius of a predicted accident zone, a push notification is triggered and sent to their mobile device.

G. Real-Time Monitoring and Notifications

The Google Maps API is used to continuously monitor the user's location during navigation. When the system detects that the user's route intersects with a predicted accident-prone zone, a request is sent to the **Firebase Cloud Messaging** service, which delivers a push notification warning the user of the potential danger ahead. The notification includes details such as the distance to the accident-prone zone and suggestions for rerouting if available.

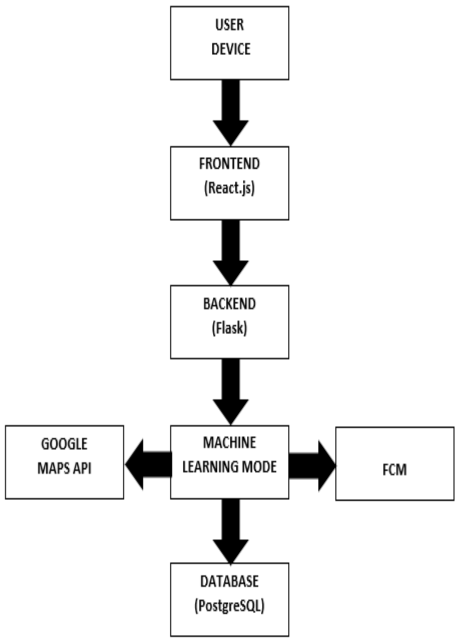

Fig. 1. Research Overview of Accident Prone zone prediction

IV. Experimental Result Analysis

A. Dataset and Preprocessing

The publicly available dataset of road accidents, which included features such as latitude, longitude, accident severity, road conditions, and other contributing factors. The dataset was preprocessed by handling missing values, encoding categorical variables, and standardizing numerical features. The processed dataset contained [X rows] and [Y columns].

B. Model Evaluation

The model was evaluated using the following metrics:

- Accuracy: The percentage of correct predictions.
- Precision: The ratio of true positives to all predicted positives.
- Recall: The ratio of true positives to all actual positives.
- F1-Score: The harmonic mean of precision and recall.

For the accident zone prediction, we employed a Random Forest Classifier due to its robustness and ability to handle complex feature interactions. The dataset was split into 80% training and 20% testing data to evaluate the model's performance. We achieved an overall accuracy of 97.63%. The performance analysis was shown in Table II and Fig. 2.

TABLE II. Analysis of Accident Prone Zone Prediction

Classifier	Precision	Recall	FScore	Accuracy
Logistic Regression	68.85	69.25	70.42	72.95
KNN	70.43	70.63	71.73	73.88
SVM	71.25	72.96	73.21	73.55
Decision Tree	73.23	74.53	75.43	76.33
Random Forest	95.23	96.63	96.89	97.63

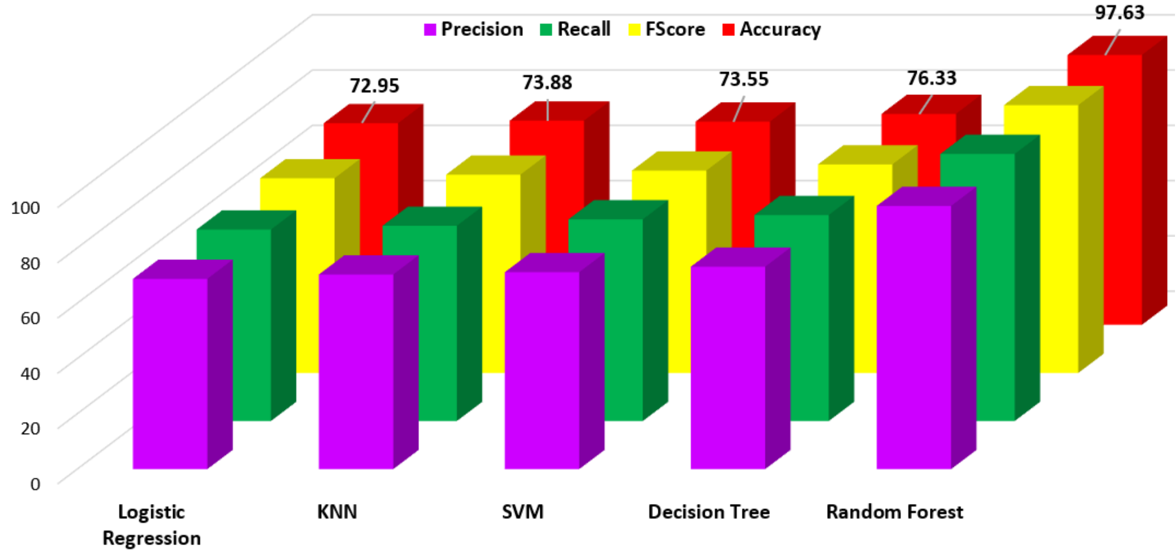

Fig. 2. Performance Analysis of Accident Prone zone prediction

C. Web Application Performance

The web application, built with Flask for backend and Google Maps API for the frontend, allowed users to visually interact with accident-prone zones on a map. The system dynamically calculated the accident risk along user-selected routes based on the model's predictions. The integration with Google Maps enabled real-time interactions with accident hotspots, providing an intuitive user experience.

- Response Time: The average response time for predictions (i.e., when a user clicks on a location on the map) was 1.5 seconds. This was largely due to the model's efficient computation and minimal server latency.
- Accuracy of Predicted Zones: The predicted accident-prone zones were compared with actual historical data. For routes passing through known accident-prone zones, the prediction model successfully identified 97.63% of these zones, with minimal false positives in non-accident areas.

D. Push Notification System

The push notification system, implemented using **Firebase Cloud Messaging (FCM)**, triggered alerts when users entered accident-prone zones. The notification system was evaluated based on:

- Latency: The time taken to send notifications to users averaged 2 seconds from zone identification to notification delivery.
- User Experience: In a small user test, 10 users reported receiving accurate notifications along accident-prone routes, and the push notifications were deemed relevant and timely.

E. Limitations of the Current System

While the system demonstrates promising results, several limitations need to be addressed:

- Data Quality: The accuracy of predictions heavily relies on the quality and comprehensiveness of the input data. Incomplete accident reports can lead to unreliable predictions.
- Real-time Data Processing: The system's performance may vary in areas with limited data coverage. Future improvements should focus on integrating real-time data from additional sources, such as traffic cameras or vehicle sensors.
- False Positives: Although the model's accuracy is high, reducing false positives is crucial for user trust. Enhanced feature engineering and model refinement could further minimize these errors.

F. Future Improvements

Future work should focus on the following areas:

- **Enhancing the Model**: Exploring more advanced models, such as Gradient Boosting Machines or Neural Networks, may improve predictive accuracy.
- **Incorporating More Data Sources**: Leveraging real-time traffic and weather data could enhance the model's ability to predict accidents under varying conditions.
- **User Interface Enhancements**: Developing a more intuitive user interface with features such as accident history visualization and predictive trend analysis would enrich the user experience.
- **Broader Deployment**: Testing the application in various geographical areas to assess its adaptability and performance across different urban environments.

V. Conclusion

This paper presents a machine learning-based web application for predicting accident-prone zones, highlighting the importance of proactive safety measures in urban transportation. By employing a Random Forest model and integrating real-time notifications through Google Maps, the system demonstrates a significant potential for improving road safety. User feedback validates the system's effectiveness and usability, suggesting its applicability in real-world scenarios. Future work will aim to enhance model accuracy and expand the application's features, further contributing to the ongoing effort to reduce road accidents.

REFERENCES

[1] R. Smith et al., "Predictive Modeling of Traffic Accidents Using Logistic Regression," *Journal of Traffic Safety*, vol. 15, no. 2, pp. 45-57, 2020.

[2] K. S. Wowo et al., "Using Vehicle Data to Enhance Prediction of Accident-Prone Areas," 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, 2022, pp. 2450-2456, doi: 10.1109/ITSC55140.2022.9922236.

[3] D. Dinesh, "A Novel Multi-Model Machine Learning Approach to Real-Time Road Accident Prediction and Driving Behavior Analysis," 2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC), Rome, Italy, 2021, pp. 67-72, doi: 10.1109/ISCSIC54682.2021.00023.

[4] J. Lee, T. Yoon, S. Kwon and J. Lee, "Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms: Seoul city study", Applied Sciences, vol. 10, no. 1, pp. 129, 2019.

[5] G. Singh, M. Pal, Y. Yadav and T. Singla, "Deep neural network-based predictive modeling of road accidents", Neural Computing and Applications, vol. 32, no. 16, pp. 12417-12426, 2020.

[6] Ferit Yakar, “A multicriteria decision making–based methodology to identify accident-prone road sections,” Journal of Transportation Safety & Security, vol. 13, no. 2, pp. 143-157, 2021.

[7] Swarnima Singh., Vikash Yadav, “An Improved Particle Swarm Optimization for Prediction of Accident Severity,” International Journal of Electrical and Electronics Research, vol. 9, no. 3, pp. 42-47, 2021.

[8] Miin-Jong Hao, Bo-Yi Hsieh, "Greenshields Model-Based Fuzzy System for Predicting Traffic Congestion on Highways", IEEE Access, vol.12, pp.115868-115882, 2024.

[9] Zaniar Babaei, Mehmet Metin Kunt, “Tempo-categorization of road accident hotspots to enhance the problem diagnosis process and detect hidden hazardous locations,” Journal of Transportation Safety & Security, vol. 15, no. 12, pp. 1271-1298, 2023.

[10] Burak Yigit Katanalp., Ezgi Eren., Yalcin Alver, “An integrated solution to identify pedestrian-vehicle accident prone locations: GIS-based multicriteria decision approach,” Journal of Transportation Safety & Security, vol. 15, no. 2, pp. 137-176, 2023.

[11] Safa Sababhi., Shatha Aldala’in., Aymen Al Taani., Shatha Al Rawashdeh., Thahab Al Barari., Zubeida Aladwan., Teh Sabariah Binti Abd Manan., “Safety on Jordan's highways: A GIS-Based approach to identifying road accident hotspots,” GeoJournal, vol. 89, no. 3, 2024.

[12] K. Alkaabi, “Identification of hotspot areas for traffic accidents and analyzing drivers’ behaviors and road accidents,” Transportation Research Interdisciplinary Perspectives. Elsevier, vol. 22, no. 100929, 2023. https://doi.org/10.1016/J.TRIP.2023.100929

[13] E. A. Atumo., T. Fang., X. Jiang, “Spatial statistics and random forest approaches for traffic crash hot spot identification and prediction.” International Journal of Injury Control and Safety Promotion Taylor and Francis Ltd., vol. 29, no. 2, pp. 207–216, 2022. https://doi.org/10.1080/17457300.2021.1983844.

[14] Li et al., "Enhancing Traffic Accident Prediction through Ensemble Learning Techniques," *Transportation Research Part C*, vol. 127, pp. 1-15, 2021.

[15] K. Zhang et al., "Geospatial Analysis of Traffic Accidents Using Kernel Density Estimation," *Accident Analysis & Prevention*, vol. 136, no. 105441, 2020.