IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Accident-Prone Zone Identification For Enhancing Road Safety with Real-Time Alerts

Akash Tarlekar[1], Aditya Jagtap[1], Omkar Kale[1] and Nikita Khawase[1]
[1]Department of Artificial Intelligence and Data Science, ISBM College of Engineering,
Savitribai Phule Pune University,
Pune, India

Abstract : Road traffic accidents remain one of the major causes of fatalities, injuries, and economic losses worldwide. This project aims to reduce the increasing number of road accidents by developing an AI-powered road safety system that proactively identifies accident-prone zones and provides real-time alerts to drivers. The proposed system utilizes five years of historical accident data (2021–2025) collected across twelve accident-related parameters, including fatalities, grievous injuries, accident intensity, and yearly accident trends. The accident records are mapped to precise geographical coordinates across Pune, Maharashtra, enabling spatial analysis and hotspot identification.The system integrates machine learning and geospatial technologies to improve road safety. K-Means clustering is employed to identify and visualize accident hotspots, achieving an optimal Silhouette Score of 0.5925 with three distinct clusters. Furthermore, an XGBoost classification model is trained using historical accident data from 2021–2023 to predict future accident-prone zones for 2024–2025. Experimental evaluation demonstrates strong predictive performance, achieving an accuracy of 92.08%, precision of 91.84%, recall of 100%, F1-score of 95.74%, and ROC-AUC of 97.78%.To enhance user awareness and safety, the identified accident-prone zones are displayed on an interactive Leaflet.js map as color-coded hotspots. Drivers receive visual alerts and audio notifications through the Web Speech API whenever they approach within a predefined distance of a hazardous zone. The system is further supported by a FastAPI backend and GPS-based location tracking for real-time operation. By combining accident hotspot detection, predictive analytics, and proactive driver alerts, the proposed framework contributes toward safer transportation networks, reduced accident risks, and improved decision-making for both drivers and traffic management authorities.

Keywords - Accident-Prone Zone Identification, Road Safety, K-Means Clustering, XGBoost, Leaflet.js, Machine Learning, FastAPI, GPS Tracking, Real-Time Alerts.

1. INTRODUCTION

Road accidents are one of the leading causes of fatalities and injuries worldwide, creating significant challenges for transportation authorities and urban planners. Rapid urbanization, increasing vehicle density, and growing traffic congestion have contributed to a rise in accident occurrences, particularly in metropolitan cities such as Pune, Maharashtra. Identifying accident-prone zones and understanding spatial accident patterns are essential for improving road safety and reducing accident-related losses.

Traditional navigation systems primarily focus on route optimization and travel time while providing limited information regarding road safety risks. Consequently, drivers often remain unaware of hazardous road segments until accidents occur. To address this issue, there is a need for intelligent systems capable of analyzing historical accident data, identifying high-risk locations, and providing proactive safety alerts.

This work presents an AI-powered accident-prone zone identification and risk prediction system for Pune city. The proposed framework utilizes five years of accident data (2021–2025) covering multiple accident-related parameters, including accident frequency, fatalities, grievous injuries, minor injuries, accident intensity, and persons involved. The collected data is mapped to geographical coordinates and analyzed using machine learning techniques to identify accident hotspots and predict future accident-prone regions.

K-Means clustering is employed to identify spatial accident hotspots, while the XGBoost classification algorithm is used to predict future accident risk based on historical accident patterns. The identified risk zones are visualized on an interactive Leaflet.js map using color-coded markers. In addition, the system provides route-based risk analysis and real-time alerts through GPS tracking and audio notifications, enabling users to make safer navigation decisions.

The primary objective of this study is to develop an intelligent accident-prone zone identification system capable of detecting high-risk locations and predicting future accident occurrences using historical accident data. The proposed framework integrates K-Means clustering for hotspot detection, XGBoost classification for risk prediction, and an interactive Leaflet.js-based visualization platform with real-time alerts to improve road safety and support informed decision-making.

The proposed solution aims to assist drivers, traffic authorities, and urban planners by providing data-driven insights for accident prevention, safer route planning, and improved transportation management.

2. LITERATURE SURVEY

1. Smart Real-Time Detection of Risky Roads Using Vehicle Trajectories for Intelligent Transportation (2025)

Eldawy et al. proposed the RiskyMove framework, which uses vehicle trajectory data and the Minimum Adaptive Viterbi (MAV) algorithm to identify risky roads in real time. The study demonstrated effective hazardous road detection but highlighted challenges in handling abnormal trajectories and large-scale GPS datasets.

2. Revealing Crash Hotspots Concerning Google Traffic Maps Historical Data by Supervised and Ensemble Machine Learning Techniques (2025)

Mirzahossein et al. integrated Google Traffic Maps with machine learning models for crash hotspot prediction. Among the evaluated models, XGBoost achieved the highest prediction accuracy of approximately 88%, demonstrating the effectiveness of ensemble learning techniques.

3. Identifying Accident-Prone Areas and Factors Influencing the Severity of Crashes Using Machine Learning and Spatial Analyses (2024)

Khosravi et al. applied hierarchical clustering techniques along with KNN and Random Forest models to identify accident-prone areas. The study concluded that visibility, lighting conditions, and road infrastructure significantly influence accident severity.

4. Machine Learning for Predictions of Road Traffic Accidents and Spatial Network Analysis for Safe Routing on Accident and Congestion-Prone Road Networks (2024)

Berhanu et al. combined Random Forest algorithms with spatial network analysis to predict accident risks and recommend safer travel routes. The study successfully balanced route safety and travel efficiency.

5. Enhancing Road Safety: Predictive Modeling of Accident-Prone Zones with ADAS-Equipped Vehicle Fleet Data (2024)

Mishra et al. introduced a hotspot detection framework using ADAS-equipped vehicle data and Earth Mover's Distance (EMD) techniques. The proposed approach enabled high-resolution identification of accident-prone zones and improved large-scale road safety analysis.

6. Division and Analysis of Accident-Prone Areas Near Highway Ramps Based on Spatial Autocorrelation (2024)

Ye et al. utilized spatial autocorrelation methods to identify accident hotspots near highway ramps. Their findings highlighted the importance of localized spatial analysis for infrastructure planning and accident prevention.

7. Road Accident Hotspot Analysis and Prediction Using Machine Learning (2023)

Prasad et al. employed clustering and machine learning algorithms for accident hotspot analysis and prediction. The study emphasized the importance of data quality, scalability, and real-time information for improving prediction performance.

8. Overview of the Identification of Traffic Accident-Prone Locations Driven by Big Data (2023)

Dong et al. presented a comprehensive review of accident hotspot identification techniques, including statistical analysis, clustering algorithms, machine learning approaches, and big-data-driven transportation systems. The study highlighted the transition toward intelligent road safety management.

Research Gap

Although several studies have focused on accident hotspot detection and accident prediction, most existing approaches concentrate either on spatial analysis or predictive modeling independently. Very few systems integrate accident hotspot identification, future risk prediction, interactive visualization, and real-time driver alerts within a single framework. The proposed work addresses this gap by combining K-Means clustering, XGBoost classification, Leaflet.js-based visualization, and GPS-enabled real-time alerts to improve road safety and support proactive decision-making.

3. SYSTEM ARCHITECTURE

The proposed system is designed to identify accident-prone zones and provide real-time safety alerts using historical accident data and machine learning techniques. The architecture consists of multiple layers including data collection, preprocessing, risk analysis, machine learning, output generation, and visualization. K-Means clustering, DBSCAN clustering, and XGBoost classification are utilized to analyze accident patterns and identify high-risk locations. The generated results are displayed through an interactive Leaflet.js map and integrated with GPS-based voice alerts to enhance road safety and support informed navigation decisions.

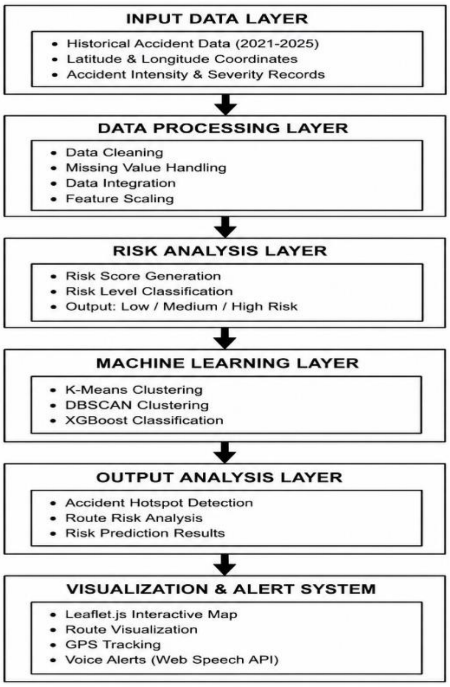

Figure 1: System Architecture of the Accident-Prone Zone Identification For Enhancing Road Safety with Real-Time Alerts

The system architecture consists of six major layers. The Input Data Layer collects historical accident records, geographical coordinates, and accident severity information. The Data Processing Layer performs data cleaning, integration, and feature scaling to prepare the dataset for analysis. The Risk Analysis Layer computes risk scores and categorizes accident zones into Low, Medium, and High risk levels. The Machine Learning Layer applies K-Means clustering, DBSCAN clustering, and XGBoost classification to identify accident hotspots and analyze risk patterns. The Output Analysis Layer generates hotspot detection results and route risk assessments. Finally, the Visualization and Alert System presents the identified risk zones on an interactive Leaflet.js map and provides GPS-based voice alerts to users approaching hazardous locations.

4. PROPOSED METHODOLOGY

The proposed system identifies accident-prone zones using historical accident records, spatial clustering techniques, and route-based risk analysis. The methodology consists of seven major stages.

4.1 Data Collection and Preprocessing

The dataset consists of 12 CSV files containing accident records from Pune district between 2021 and 2025. The collected parameters include Total Accidents, Persons Killed, Persons Involved, Grievous Injury Accidents, Minor Injury Accidents, No Injury Accidents, Accident Intensity, and yearly accident statistics.

The datasets were cleaned by removing unnecessary metadata columns, handling missing values, standardizing column names, and merging all records using Grid ID, Latitude, and Longitude coordinates. After preprocessing, a master dataset containing 501 geographical grids was created for further analysis.

Table 1: Accident Dataset Parameters

Category	Parameters
Accident Statistics	Total Accidents, Accident Intensity
Severity Indicators	Persons Killed, Grievous Injury Accidents
Casualty Information	Persons Involved, Minor Injury Accidents, No Injury Accidents
Historical Records	Accident Counts (2021–2025)
Spatial Features	Grid ID, Latitude, Longitude

4.2 Grid-Based Spatial Structuring

The Pune region was divided into geographical grid cells. Each grid was assigned a unique Grid ID along with its corresponding latitude and longitude coordinates.
For every grid:
Grid = {Grid ID, Latitude, Longitude, Accident Features}
This structure enables efficient spatial analysis and hotspot identification.

4.3 Risk Score Generation

A composite Risk Score was computed for each grid using accident frequency, severity, recent accident trends, and accident intensity.
Risk Score = (0.35 × Historical Frequency)

- (0.30 × Severity Impact)
- (0.25 × Recent Activity)
- (0.10 × Accident Intensity)

where,
Historical Frequency = Accidents(2021) + Accidents(2022) + Accidents(2023)
Severity Impact = Persons Killed + Grievous Injury Accidents + Minor Injury Accidents
Recent Activity = Accidents(2023) + Accidents(2024) + Accidents(2025)
The resulting score is normalized to a scale of 0–100.

Table 2 : Risk Score Components

Component	Weight
Historical Frequency	35%
Severity Impact	30%
Recent Activity	25%
Accident Intensity	10%

Risk categories were assigned using percentile-based thresholds.

Table 3: Risk Categorization

Risk Level	Condition
Low Risk	Risk Score < 70th Percentile
Medium Risk	70th Percentile ≤ Risk Score < 95th Percentile
High Risk	Risk Score ≥ 95th Percentile

Table 4 : Risk Zone Distribution

Risk Level	Number of Zones
Low Risk	350
Medium Risk	125
High Risk	26

4.4 K-Means Clustering

K-Means clustering was applied to identify accident hotspot regions.
The clustering objective minimizes the within-cluster sum of squared distances:
$J = \Sigma ||x_i - \mu_j||^2$
where:
x_i = data point
μ_j = cluster centroid
Different values of K (3–8) were evaluated using the Silhouette Coefficient.
The highest score obtained was:
Silhouette Score = 0.5925 (K = 3)

Table 5 :K-Means Clustering Results

Cluster	Number of Zones
Cluster 0	138
Cluster 1	328
Cluster 2	35

4.5 DBSCAN Clustering

DBSCAN was used as a density-based clustering technique for identifying dense accident regions and eliminating noise.

Table 6: DBSCAN Parameters

Parameter	Description
eps	Search Radius
min_samples	Minimum Neighbor Points

DBSCAN groups nearby accident grids based on density connectivity and can identify outlier zones without requiring a predefined number of clusters.

4.6 XGBoost Classification

The accident prediction model was developed using XGBoost.

Training Features
• Total Accidents
• Accident Intensity
• Persons Killed
• Persons Involved
• Grievous Injury Accidents
• Minor Injury Accidents
• Accident Counts (2021–2023)
Target Variable:
Danger = 1
Safe = 0
Dataset Split:
Training Set = 80%
Testing Set = 20%

Table 7 : XGBoost Performance Metrics

Metric	Value
Accuracy	92.08%
Precision	91.84%
Recall	100%
F1 Score	95.74%
ROC-AUC	97.78%

4.7 Route Risk Analysis and Alert System

The user enters source and destination locations which are converted into geographical coordinates using the OpenStreetMap Nominatim API.
The OSRM Routing Engine generates the actual travel route. For every route coordinate, nearby accident grids within a specified radius are identified and analyzed.
The identified risk zones are displayed using Leaflet.js.

Table 8 : Risk Zone Visualization

Color	Risk Level
Red	High Risk
Amber	Medium Risk
Green	Low Risk

Real-time GPS monitoring is performed using the browser Geolocation API. Voice alerts are generated through the Web Speech API whenever the user approaches within 500 meters of a High or Medium risk zone.

5. RESULTS AND DISCUSSION

The proposed framework was evaluated using accident records collected from Pune district between 2021 and 2025. Experimental analysis was conducted to assess the effectiveness of risk score generation, accident hotspot detection, clustering performance, and accident risk prediction. The obtained results demonstrate the capability of the system to identify accident-prone zones and provide proactive safety insights.

5.1 Risk Score Analysis

The generated risk scores were normalized to a scale of 0–100 and categorized into Low, Medium, and High-risk zones using percentile-based thresholds. The classification process enabled prioritization of accident-prone locations based on historical accident frequency, severity indicators, recent accident trends, and accident intensity.

Table 9: Risk Zone Distribution

Risk Level	Number of Zones
Low Risk	350
Medium Risk	125
High Risk	26

5.2 K-Means Clustering Results

K-Means clustering was applied to identify spatial accident hotspots. Different cluster sizes ranging from K=3 to K=8 were evaluated using the Silhouette Coefficient. The highest clustering quality was achieved for K=3.

Table 10: K-Means Clustering Evaluation

K Value	Silhouette Score
3	0.5925
4	0.5897
5	0.5725
6	0.5602
7	0.5606
8	0.5643

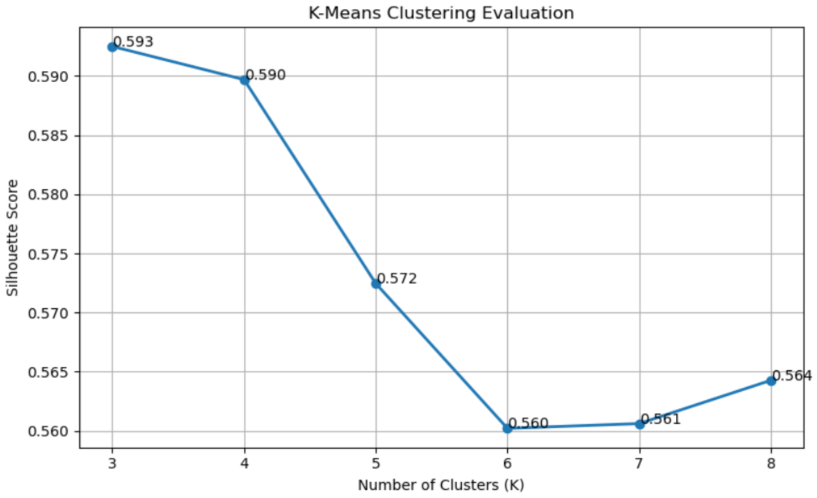

Figure 2: K-Means Evaluation Graph

The performance of K-Means clustering was evaluated using the Silhouette Coefficient for different values of K ranging from 3 to 8. The highest Silhouette Score of 0.5925 was obtained when K=3, indicating the best cluster separation and compactness. Therefore, three clusters were selected for accident hotspot identification and further spatial analysis.

Table 11: K-Means Cluster Distribution

Cluster	Number of Zones
Cluster 0	138
Cluster 1	328
Cluster 2	35

5.3 DBSCAN Clustering Results

DBSCAN clustering was evaluated as a density-based hotspot detection technique. Multiple parameter combinations were tested using different search radii. However, the algorithm was unable to identify significant density-based clusters within the available accident dataset.

5.4 XGBoost Classification Results

The XGBoost classifier was trained using historical accident features and evaluated using an 80:20 train-test split. Performance was measured using Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

Table 12 : XGBoost Performance Metrics

Metric	Value
Accuracy	92.08%
Precision	91.84%
Recall	100.00%
F1-Score	95.74%
ROC-AUC	97.78%

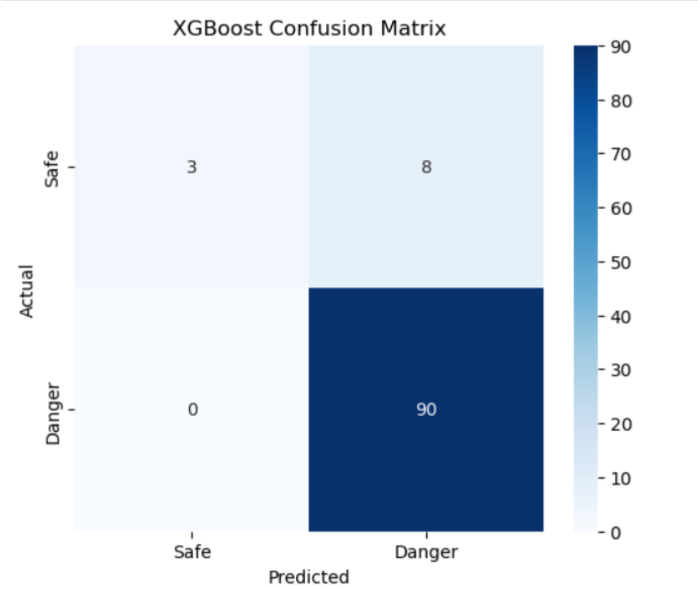

Figure 3: Confusion Matrix

This figure presents the confusion matrix of the XGBoost classification model. The model correctly classified 90 accident-prone (Danger) zones and 3 safe zones, while only a small number of samples were misclassified. The results demonstrate the strong predictive capability of the proposed model for identifying accident-prone locations.

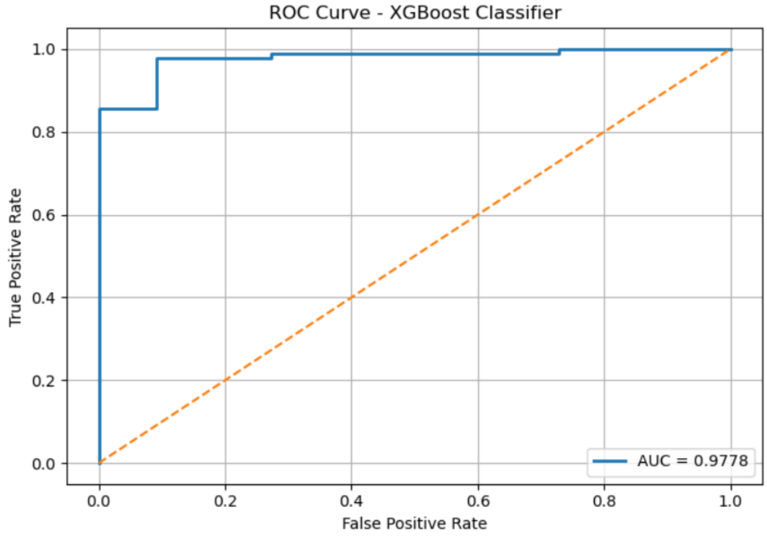

Figure 4: ROC Curve

This figure shows the ROC curve of the XGBoost classifier. The model achieved an AUC score of **97.78%**, indicating excellent classification performance and strong capability in distinguishing accident-prone zones from safe zones.

5.5 Route Risk Analysis and Visualization

The route analysis module was evaluated using multiple source-destination combinations across Pune city. The system successfully identified accident-prone zones located near the travel route and provided risk-aware navigation assistance.

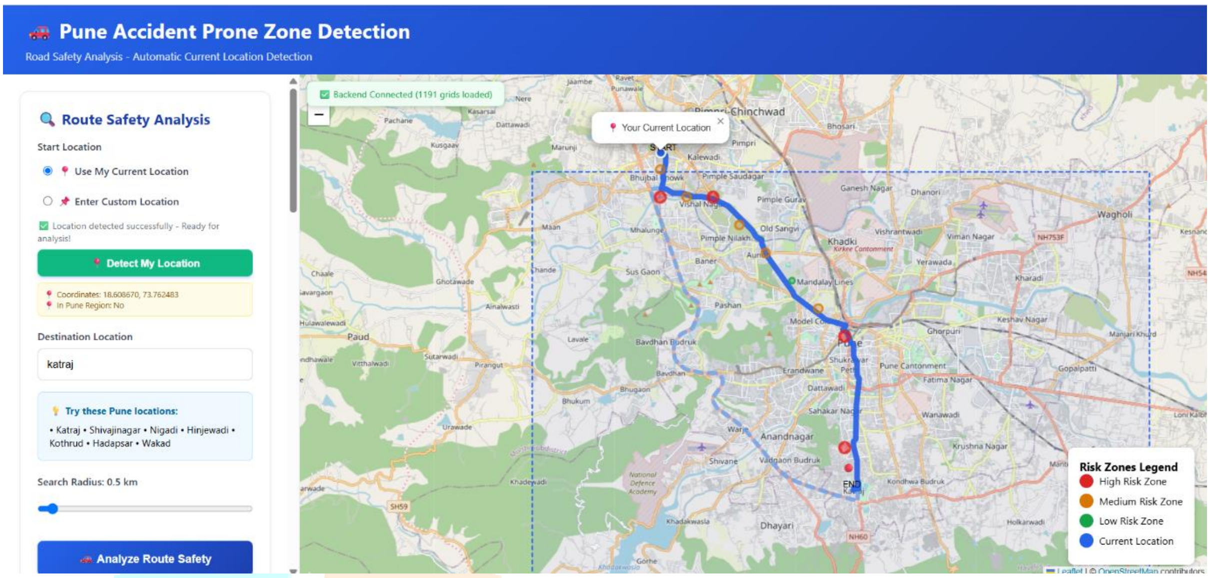

Figure 5: Route Risk Analysis Interface

The route risk analysis interface allows users to specify source and destination locations and evaluate road safety along the selected route. The system identifies accident-prone zones located near the route and visualizes them on the interactive map.

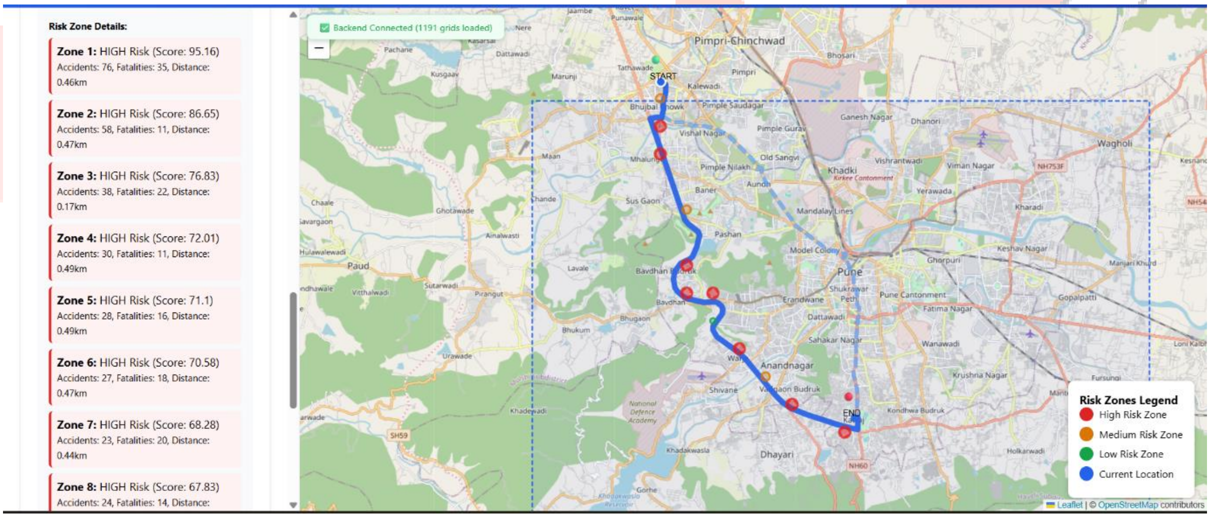

Figure 6: Accident Hotspot Visualization on Leaflet Map

Accident-prone zones are displayed using color-coded markers, where red, orange, and green indicate High, Medium, and Low-risk zones respectively. The visualization enables users to quickly identify hazardous road segments and accident hotspots.

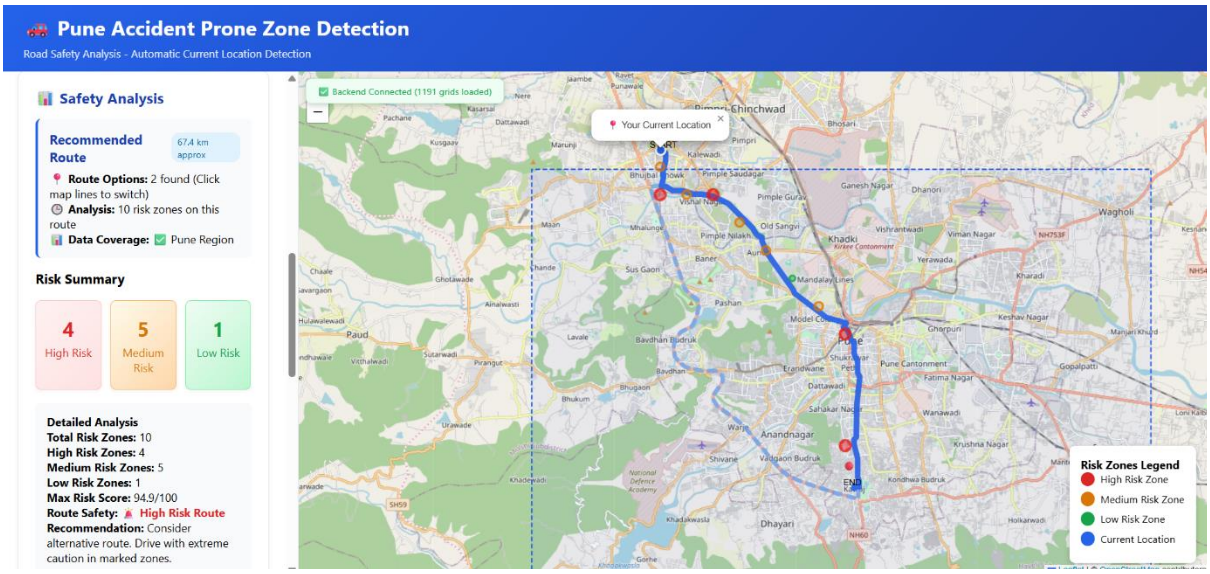

Figure 7: Alternative Route Comparison

The dashboard provides a summarized view of route safety by displaying the number of identified risk zones, maximum risk score, route safety status, and navigation recommendations. This assists users in making informed travel decisions.

5.6 Discussion

The experimental results demonstrate that the proposed framework effectively integrates accident hotspot detection, machine learning-based risk prediction, route analysis, and real-time safety alerts within a unified system. K-Means clustering successfully identified accident hotspot regions with a Silhouette Score of 0.5925, while XGBoost achieved 92.08% prediction accuracy and a ROC-AUC score of 97.78%. These results confirm the effectiveness of the proposed approach for supporting road safety management and accident prevention initiatives.

6. CONCLUSION

This paper presented an intelligent accident-prone zone identification system for enhancing road safety using historical accident records, machine learning techniques, and geospatial analysis. The proposed framework integrates risk score generation, K-Means clustering, XGBoost classification, route risk analysis, and real-time alert mechanisms to identify and visualize hazardous road segments across Pune city.

Experimental results demonstrated the effectiveness of the proposed approach, with K-Means successfully identifying accident hotspot regions and XGBoost achieving an accuracy of 92.08% and a ROC-AUC score of 97.78%. The integration of Leaflet.js visualization, GPS-based navigation, and voice alerts enables users to receive proactive warnings when approaching accident-prone zones. The developed system can support drivers, traffic authorities, and urban planners in improving road safety and reducing accident risks through data-driven decision-making.

7. FUTURE SCOPE

Several enhancements can be incorporated into the proposed system in future work. Real-time traffic data, weather information, and road condition monitoring can be integrated to improve prediction accuracy. Advanced deep learning models and spatio-temporal analysis techniques may be employed to enhance accident forecasting capabilities. The system can also be extended to support multiple cities and larger geographical regions. Furthermore, integration with mobile applications, emergency response systems, and smart city infrastructure can provide more comprehensive road safety solutions and real-time assistance to commuters.

REFERENCES

[1] Eldawy, E. O., Abdalla, M., et al. (2025). Smart real-time detection of risky roads using vehicle trajectories for intelligent transportation. Journal Name.

[2] Mirzahossein, H., Nobakht, P., Waller, T., and Lin, D. Y. (2025). Revealing crash hotspots concerning Google Traffic Maps historical data by supervised and ensemble machine learning techniques. Journal Name.

[3] Khosravi, Y., Hosseinali, F., et al. (2024). Identifying accident-prone areas and factors influencing the severity of crashes using machine learning and spatial analyses. Journal Name.

[4] Berhanu, Y., Schröder, D., Wodajo, B. T., and Alemayehu, E. (2024). Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing on accident and congestion-prone road networks. Journal Name.

[5] Mishra, R. S., Thakur, D. S., Subramanian, A., et al. (2024). Enhancing road safety: Predictive modeling of accident-prone zones with ADAS-equipped vehicle fleet data. Journal Name.

[6] Ye, Q., et al. (2024). Division and analysis of accident-prone areas near highway ramps based on spatial autocorrelation. Journal Name.

[7] Prasad, S., Kumar, A., et al. (2023). Road accident hotspot analysis and prediction using machine learning. Journal Name.

[8] Dong, C., Chang, N., et al. (2023). Overview of the identification of traffic accident-prone locations driven by big data. Journal Name.