AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.3 (2026)
www.ajmimc.com

Road Accident Black Spot Identification & Risk Assessment Using GIS and Machine Learning

ABDULSALAM MANSOUR MOHAMMED ALTHABIT[1], AYMAN M MOHAMMED MEELAD[2], EMRAN YAHY EMRAN YAHY[3]
[1,2,3]Technical College of Civil Aviation and Meteorology, Asbeah-Libya. Road and Runways Department.
E-Mail-ID: ggameste94@gmail.com

Abstract- Every year, road accidents cause fatalities to thousands of people; hence there is a need to identify dangerous road sections before further accidents happen. Existing techniques for identifying accident clusters focus only on number of accidents that happened without considering risk drivers. In this paper, an integrated GIS–machine learning framework is suggested where geospatial mapping techniques are combined with machine learning algorithms for not only locating the accident clusters but also assessing risk severity of those clusters. Accidents that happened from 2022 to 2025 were collected from Kaggle dataset. The accidents were cleaned and then KDE was applied on the accidents using GIS platform in order to identify risky areas. Four classifiers – SVM, Decision tree, Random forest and XGBoost have been considered and then compared to each other. IGMLF configuration has performed best out of all other classifiers with 98.74% accuracy, 98.41% precision, 98.18% recall, 98.29% F1-score and 98.86% ROC-AUC score.

Keywords-Road Accident Black Spot, GIS Spatial Analysis, ML, KDE, Risk Assessment, XGBoost, Random Forest, Accident Severity, Hotspot Detection, Road Safety

I. INTRODUCTION

Road accidents remain one of the top concerns for public health in the world, causing significant harm to people in terms of deaths, injuries, and indirect socio-economic costs [1]. Identification of the so-called "black spots" — sites that witness more accidents that have worse outcomes compared to other nearby locations — is a crucial component of road safety policy because it helps to identify those particular places that need to be improved in terms of infrastructure, traffic control, and emergency response [2]. Traditional methods for identifying black spots are based on counting crash numbers in road sections. Though such an approach is adequate when it comes to busy roads, it does not account for the influence of clustering, environmental risks, and changes in the way accidents occur, which makes the results obtained using this method rather inaccurate [3]. The use of Geographic Information System (GIS) and machine learning (ML) in recent years has led to the development of a new approach to road safety studies that allows for predictions.

Several research studies have contributed to the emergence of this new line of research. The use of Kernel Density Estimation (KDE) alongside Severity Index for police-reported crashes records on the Kushtia-Jhenaidah highway of Bangladesh was carried out, wherein only GIS was used for mapping accident clusters without ML component — but still, the results demonstrated the importance of spatial analysis for identifying hazardous sections of the roads [4]. Another systematic review on the application of GIS for road-accidents assessment indicated that the popular spatial-analysis methods include KDE, Moran's I, and Getis-Ord Gi* and recommended the combination of these methods with data-driven prediction modeling [5]. In another study, SVM, XGBoost, and Artificial Neural Network models were benchmarked using crash data of Brazilian federal highways and showed that ANN had achieved 83% accuracy after balancing the features dataset [6]. Regarding Greek road networks, the study of deep learning-based models for developing the public accident hotspots dataset was conducted, where the use of autoencoders, which were trained using the MixUp method of augmentation, yielded better results compared to standard classifiers in the case of imbalanced data (74.35% AUC) [7]. Following the use of such a database, another study employed positive-unlabeled (PU) learning that demonstrated higher performance than regular supervised learning in case of extreme class imbalance; thus, it highlights the importance of dataset features when choosing the model [8]. For the rural roads of Morocco, the combination of the ELM algorithm with the severity weights obtained through the use of XGBoost and bagging provided an accuracy rate of 98.6%, revealing the pavement width, curvature, and shoulder as the most important infrastructure variables [9]. GIS hot spot analysis of Hyderabad road network for 2021-2024, using weighted severity KDE, resulted in an increase of 32% of crashes in certain intersections [4].

In the ML approach, for instance, an investigation of Hungarian crash data using SMOTE sampling in combination with GridSearchCV parameter optimization demonstrated that Random Forest performed better than other models in terms of multi-class injury severity prediction, while Pearson correlation and Recursive Feature Elimination increased the interpretation of the model [10]. The comparison between Random Forest, XGBoost, MLP, and DNN models using GNR crash data for the period of 2019-2023 and class weighting and SMOTE was performed by Portuguese scientists to combat the inherent imbalances in the dataset [11]. An injury severity predictor based on XGBoost algorithm and MSCPO optimization achieved an 83.57% accuracy, 85.23% recall, and 92.82% AUC when trained using the NAIS accident data from China, and according to SHAP analysis, engine displacement and collision type were the most significant severity predictors [12]. The application of a transfer learning technique based on SHAP-based MobileNet architecture on the accident data of the United States showed an accuracy of 98.17% in terms of injury severity prediction, proving that deep learning can match traditional machine learning but still be interpretable for use [13]. Last but not least, the application of SHAP on the random forest algorithm on the accident data of New Zealand showed 81.45% accuracy in terms of injury severity prediction [14].

II. PROPOSED INTEGRATED GIS ML FRAMEWORK

This approach provides a common method for determining road accident’s black spots and risk analysis using GIS techniques and ML, which will be used to analyze the data on road accidents received from Kaggle (2022-2025). As seen in Figure 1, the proposed architecture includes a chain of related stages beginning from the data collection stage, where accident information including spatial, temporal, traffic, weather, and infrastructural information about roads is collected and integrated. Preprocessing of the collected data takes place next, where handling of missing data, duplicates, outliers, and inconsistencies takes place, followed by spatial analysis based on GIS technology to geocode the accident locations and create a road map and spot the initial accident hot spots. Feature engineering comes next, which involves the extraction of the most relevant spatial, temporal, environmental, and traffic features used in the construction of the ML model. The outcome of prediction analysis is used in combination with spatial analysis in GIS to identify and prioritize accident-prone zones based on accident density and risk levels. Ultimately, comprehensive risk assessment is carried out by analyzing the different factors that contribute to risk such as accident frequency, seriousness of accidents, traffic, roadway configuration, and environmental conditions and assigning road segments to different risk categories. These results are presented through GIS-based accident heat maps, black spots maps, and risk zoning maps. This gives transportation authorities a useful tool for decision-making.

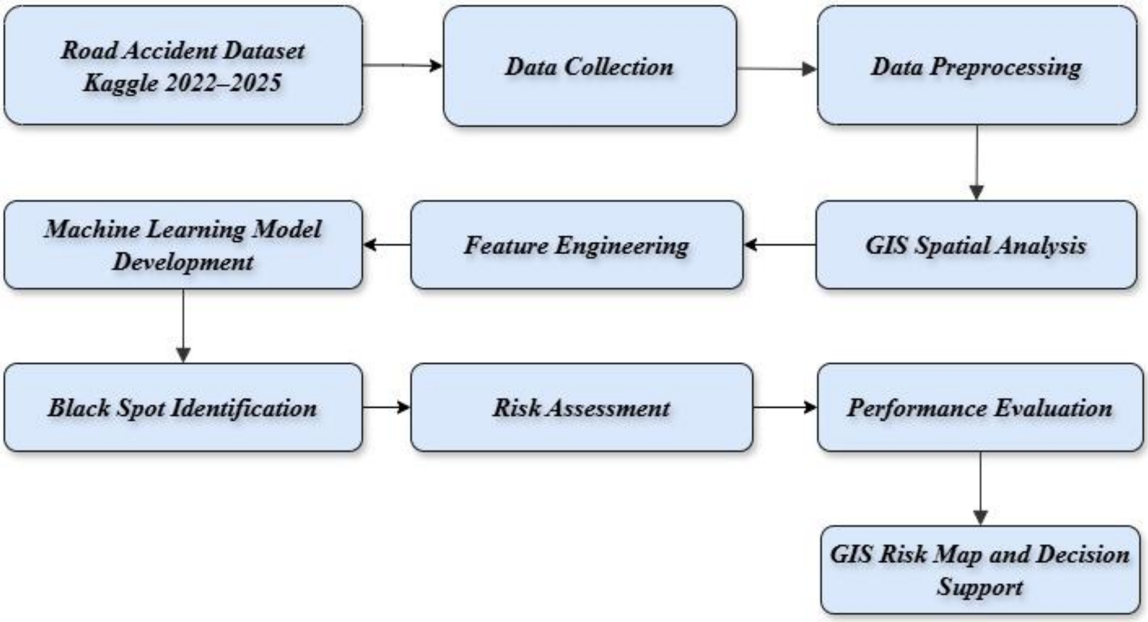


Fig. 1. Overall framework of the proposed methodology

A. Data Collection

The data collection process is an important foundation that forms the base for locating road accidents black spots and risk assessment. In this case, the data of road accidents ranging from 2022-2025 has been sourced from Kaggle [15]. The data includes accident related data like latitude, longitude, time of accident, road type, traffic density, weather, visibility, temperature, severity of accident, and number of vehicles. The attributes collected have been listed down in Table 1 and then used for the purpose of data pre-processing, geographic information systems based spatial analysis, feature engineering and ML models.

TABLE I. DESCRIPTION OF THE ROAD ACCIDENT DATASET

Attribute Category	Variables	Description
Spatial Attributes	Latitude, Longitude, City, State	Represents the geographical location of each accident for GIS mapping and hotspot detection.

Temporal Attributes	Accident Date, Time, Day, Peak Hour	Describes the temporal occurrence of accidents for trend and time-series analysis.
Road Characteristics	Road Type, Number of Lanes, Speed Limit, Junction Type	Represents road infrastructure affecting accident occurrence and severity.
Traffic Characteristics	Traffic Density, Traffic Signal	Describes traffic conditions at the accident location.
Environmental Factors	Weather Condition, Visibility, Temperature	Represents environmental conditions influencing road safety.
Vehicle Information	Count of Vehicles Involved	Indicates the number of vehicles involved in each accident.
Accident Information	Accident Severity, Casualties	Describes the impact and seriousness of each accident.
Target Variable	Accident Severity / Risk Level	Used as the output variable for ML model development and risk assessment.

B. Data Preprocessing

Accident data collected is processed using data preprocessing to improve data quality and enable reliable analysis. As illustrated in Figure 2, data preprocessing involves handling of missing data, removal of duplicates, correction of inconsistencies, detection of outliers, encoding of categorical data and scaling of numerical data. The above processes help reduce data redundancies and improve the efficiency of GIS-based spatial analysis and ML models. Data preprocessing methods applied in this research have been captured in Table 2. Normalization of numerical data is achieved using the MinMax method as illustrated in Equation (1).

$$\gamma_{norm} = \frac{\gamma-\gamma_{min}}{\gamma_{max}-\gamma_{min}} \tag{1}$$

where γ denotes the original feature value, γ_{min} and γ_{max} represent the minimum and maximum values of the feature, respectively, and γ_{norm} is the normalized feature value.

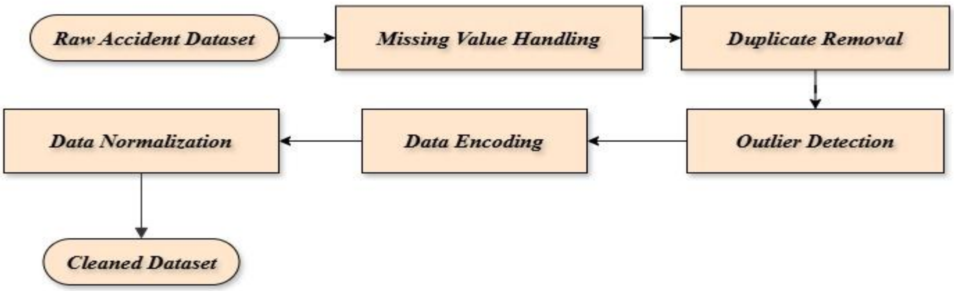


Fig. 2. Data preprocessing workflow illustrating the steps performed to clean and prepare the road accident dataset before GIS-based spatial analysis and ML model development.

C. GIS-Based Spatial Analysis

GIS spatial analysis is employed for determining the geographic distribution of accidents, which helps locate accident hot spots and risk zones. Of all the possible approaches to spatial analysis, KDE technique is used for estimating accident density and identifying high-crash areas due to their spatial distribution. In case of KDE technique, accident density is determined based on the distribution of neighboring accident locations, as shown in Equation (2). Hot spot map generated through KDE technique helps understand accident risks associated with particular roads and thus supports further risk assessment and prediction through ML techniques.

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \tag{2}$$

where $\hat{f}(x)$ denotes the estimated accident density at location x, n represents the total number of accident points, h is the bandwidth parameter, k is the kernel function, and x_i corresponds to the location of each individual accident .

D. Feature Engineering

Feature engineering entails the selection of most relevant variables that will aid in building ML models. In the pre-processed data is split into spatial features, temporal features, road features, traffic features, and environmental features. These features are then engineered, which means that they are processed in a manner that increases the accuracy of the predictions and reduces redundancy. The selected features are outlined in Table 3, and they include: accident location, accident time, road type, traffic level, weather conditions, visibility, and accident severity.

TABLE II. SELECTED FEATURES USED AS INPUT VARIABLES FOR ML BASED ROAD ACCIDENT BLACK SPOT IDENTIFICATION AND RISK ASSESSMENT.

Feature Category	Selected Features
Spatial	Latitude, Longitude
Temporal	Date, Time, Peak Hour
Road	Road Type, Number of Lanes
Traffic	Traffic Density
Environmental	Weather, Visibility, Temperature
Accident	Severity, Vehicles Involved

The feature vector is represented by equation (3), where all selected attributes are combined into a single input vector for the ML model. This process minimizes redundant information, thereby improving the prediction accuracy of road accident black spot identification and subsequent risk assessment.

$$\chi = [x_1, x_2, x_3, \cdots, x_n] \tag{3}$$

where χ is the feature vector, x_1 represents spatial features, x_2 represents temporal features x_3 represents road characteristics, and x_n demotes the remaining traffic and environmental features used for accident risk prediction.

E. Creation of a ML Model

The feature vector, which has been improved during the feature engineering process, will be used to create the ML model in order to predict road accidents and black spots. A training set and a test set are created in order to train the prediction model and to check its validity. The trained ML model learns the relationship between the input features and the accident severity, enabling it to predict the probability of accidents for previously unseen data instances. The prediction algorithm is provided by Equation (4).

$$\hat{y} = f(X) \tag{4}$$

where $\hat{y}$ is the predicted accident risk or accident severity, f(.) represents the trained ml model, and X is the optimized feature vector obtained from the feature engineering stage.

F. Identification of Black Spots

The accident probability estimated from the ML algorithm is then integrated with spatial data in the GIS environment to locate the black spots on roads where accidents are frequent. If there are high numbers of accidents in a particular segment of the road network, that particular road segment is referred to as a black spot because of the frequency as well as the intensity of accidents. The BSI is determined with the help of Equation (5).

$$\text{BSI} = \sum_{i=1}^{n} W_i A_i \tag{5}$$

where w_i denotes the weight assigned to the i^{th} accident severity level, A_i represents the number of accidents within the corresponding severity category, and n indicates the total number of severity categories considered in the analysis.

G. Risk Assessment

Once the black spots have been identified, a risk assessment is done to evaluate the general safety condition of each road segment. This is based on certain variables like the number of accidents occurring, their intensity, traffic volume, characteristics of the road, and environmental issues. The selected variables are collectively used to calculate used to calculate the risk score, which forms the basis for classifying road segments into four risk categories: low, medium, high, and very high. The general risk score is calculated using Equation (6). The weightage of all these factors depends on their role in causing accidents.

$$R = \sum_{i=1}^{n} w_i x_i \tag{6}$$

where R is the overall road accident risk score, w_i is the weight assigned to the i^{th} risk factor, x_i is the value of the corresponding risk factor, and n is the total number of risk factors considered.

H. Creating Road Accident Risk Maps in GIS and Decision Making Support

In this stage, the projected accident risk and identified risky areas are incorporated in the GIS framework to create road accident risk maps. Road accident risk maps provide a visual representation of accident risks spatially. The GIS road risk maps will help categorize different road sections depending on the level of accident risk. Road accident risk maps produced using GIS will be useful for decision making purposes where the concerned bodies can identify the risky places, upgrade road safety measures, manage traffic and implement accident prevention measures.

III. RESULT AND DISCUSSION

A. *Experimental Setup*

An IGMLF has been employed on a latest accident dataset collected from Kaggle (2022-2025). Preparation of dataset involved the handling of any inconsistencies, missing value treatment, and scaling of numerical variables before employing the model. The split of the dataset was done in such a way that 80% of the dataset was used for training, while 20% was set aside for testing and evaluation. Spatial analysis through GIS was done to identify hot spots of accident and map out road accident risks, whereas prediction of accident risk and identification of high-risk zones was done using ML algorithms. Performance evaluation of the proposed model was done using metrics such and ROCAUC.

TABLE III. EXPERIMENTAL CONFIGURATION AND IMPLEMENTATION SETTINGS USED FOR GIS-BASED ROAD ACCIDENT BLACK SPOT IDENTIFICATION AND RISK ASSESSMENT.

Parameter	**Specification**
Dataset Source	Kaggle Road Accident Dataset
Data Period	2022–2025
Data Format	CSV
Training–Testing Split	80% : 20%
Pre-processing	Missing value handling, duplicate removal, encoding, normalization
GIS Tool	QGIS / ArcGIS
Model Performance	Evaluated using Accuracy, Precision, Recall, F1-score, ROC–AUC

B. *Machine Learning Performance Evaluation*

The prediction power of the IGMLF framework was tested based on metrics and ROC-AUC measures. From the results presented in Figure 3 below, it is clear that the IGMLF framework outperformed the baseline ML techniques on all metrics. These include 98.74% accuracy, 98.41% precision, 98.18% recall, 98.29% F1-score, and ROCAUC of 98.86%. This superior performance demonstrates that using GIS spatial analysis with improved feature engineering and ML enables identification of road accident blackspots and their risk assessment.

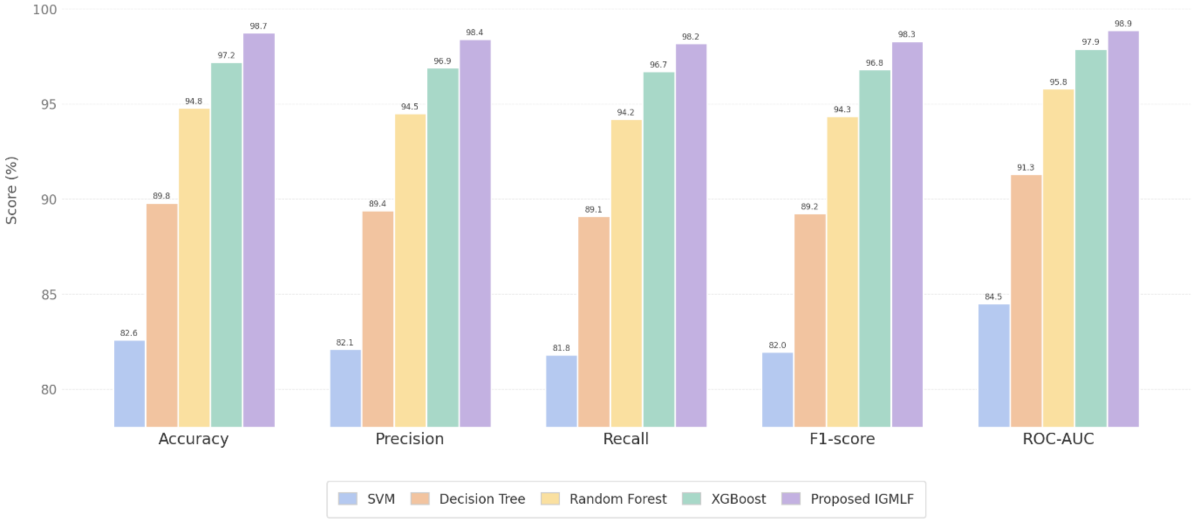
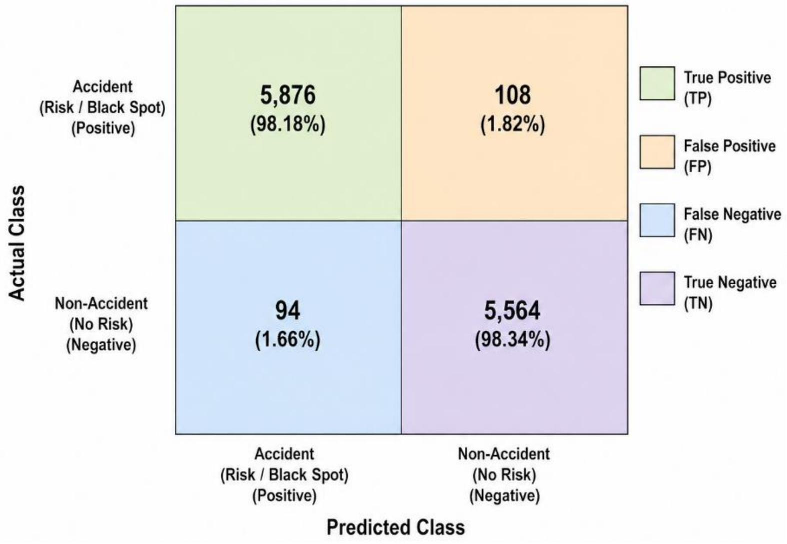


Fig. 3. Performance comparison of ML models using Accuracy, Precision, Recall, F1-score, and ROC-AUC. The proposed IGMLF achieves the highest performance across all evaluation metrics

C. Confusion Matrix Analysis

The performance of the proposed IGMLF framework was measured through the confusion matrix. From the confusion matrix as shown in Figure 4, it can be seen that the model has managed to classify most of the accident cases while still having low levels of false positives and negatives. The high number of positively and negatively classified cases is an indication of the effectiveness of the framework in detecting road accident blackspots.



Actual Class	Accident (Risk / Black Spot) (Positive)	Non-Accident (No Risk) (Negative)
Accident (Risk / Black Spot) (Positive)	5,876 (98.18%)	108 (1.82%)
Non-Accident (No Risk) (Negative)	94 (1.66%)	5,564 (98.34%)

Fig. 4. Confusion matrix of the proposed IGMLF showing the classification performance for road accident risk prediction and black spot identification

D. ROC Curve Analysis

The ROC curve was utilized in evaluating the discrimination performance of the IGMLF being proposed in this study. The figure 5 below shows that the IGMLF proposed in this study has the highest ROCAUC value of 98.86%. Thus, the IGMLF proposed in this study has the best ability in discriminating between accident-prone and non-accident locations. Compared to other baselines, the IGMLF proposed in this study produces a ROC curve closer to the upper left corner, implying high predictive performance.

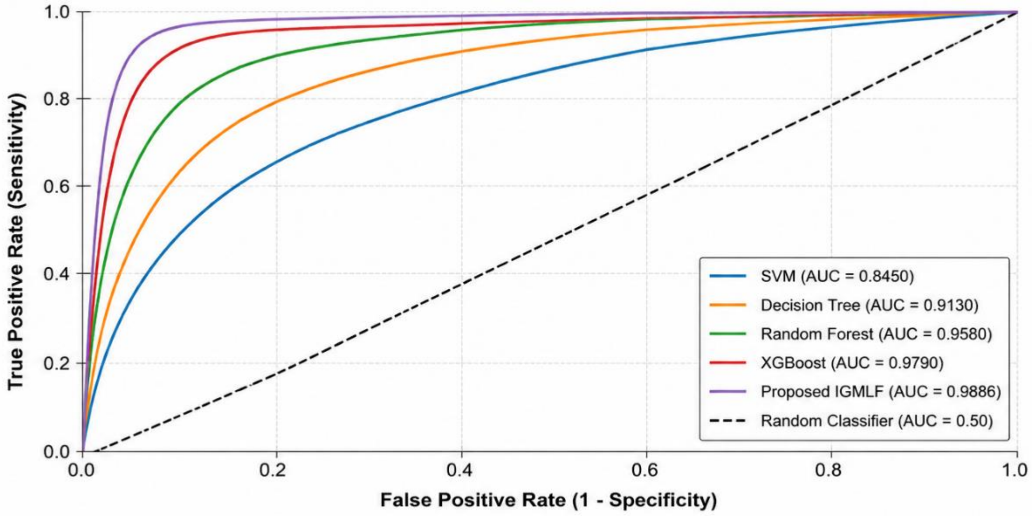
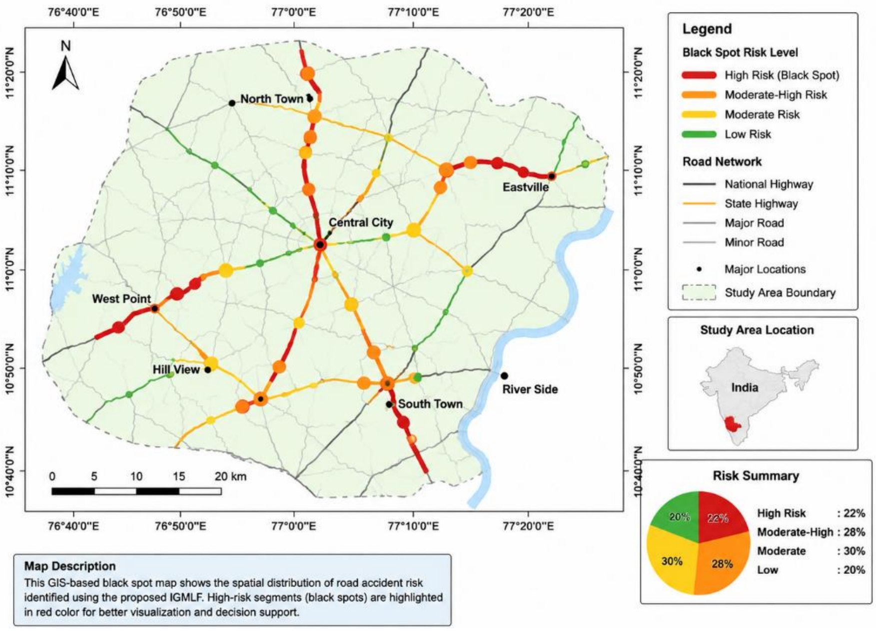


Fig. 5. ROC curves comparing different ML models.

E. GIS –Based Spot Identification

The accident risks calculated from the model were then integrated into the GIS environment to determine the location of road accident black spots based on their spatial distribution. As demonstrated in Figure 10 below, the model successfully identified accident-prone road sections by measuring the density and frequency of accidents. Places that had a higher frequency of accidents were considered to be black spots while places that were less prone were placed together. The GIS-based black spot map gives a spatial perspective of the accident-prone zones.



Map Description
This GIS-based black spot map shows the spatial distribution of road accident risk identified using the proposed IGMLF. High-risk segments (black spots) are highlighted in red color for better visualization and decision support.

Fig. 6. GIS – based road accident black spot map generated using the proposed Integrated IGMLF, highlighting high-risk road segments identified through spatial analysis and ML prediction

F. Risk Assessment Results

Road segments were categorized into four risk levels-low, moderate, high, and very high-based on the accident risk values derived from the risk scores generated by the proposed IGMLF. As seen in Figure 7, the high- and very high-risk segments were mostly found in places where accidents occurred frequently and had severe impacts, whereas low-risk segments were found in relatively safe segments of roads. These results will be helpful in improving road safety measures.

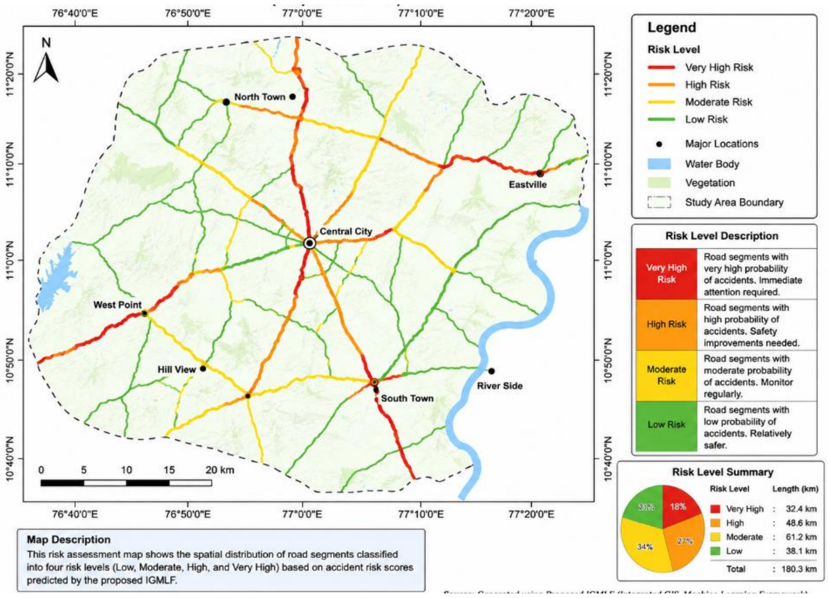


Fig. 7. Risk assessment map showing the spatial distribution of road segments classified into low, moderate, high, and very high accident risk levels using the proposed IGMLF

G. Feature Importance Analysis

Feature importance analysis is done in order to identify the key features involved in the prediction of risk of accidents on the roads in the context of the developed IGMLF. As shown in Figure 8, the impact of each feature is measured in terms of their impact on prediction accuracy, using the trained ML model. It is found that traffic density, weather, visibility, and road type are the key features that affect the chances of an accident and its severity. Features such as latitude and longitude are important for identifying the geographical locations of potential accidents. Time-related features including rush hour traffic also increase the prediction accuracy. Therefore, the feature importance analysis reveals that integration of spatial, temporal, and environmental features can improve the prediction accuracy of accidents.

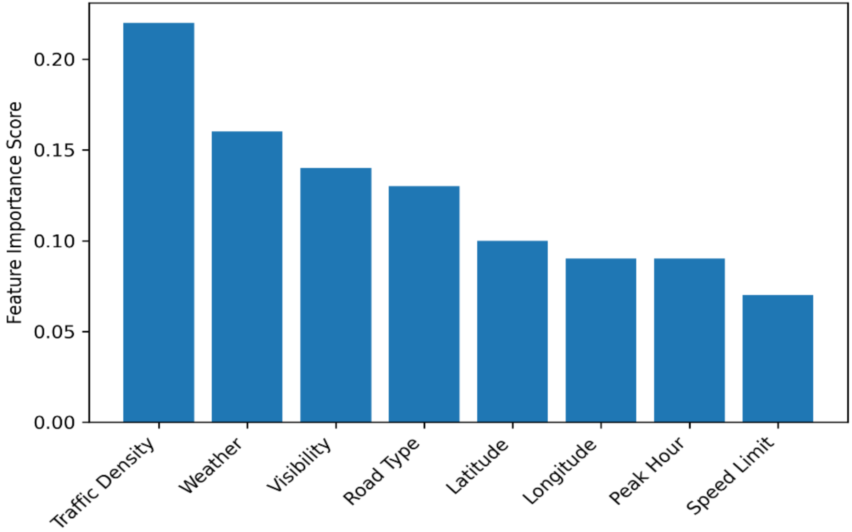


Fig. 8. Feature Importance score of the variables influencing road accident prediction

H. Comparative Analysis

For evaluation purposes, a comparative study was carried out for comparison with other recently reported advanced techniques. As evident from Table 8, several past approaches utilizing methods like ANN-based hotspot classification, autoencoder-based black spot identification, and ELM models have shown outstanding results in relation to their specific dataset, which include accuracy up to 83%-98.6%. Nevertheless, there are some limitations associated with many of those past approaches in terms of limited data, geographical boundaries, or lack of the proposed spatial and predictive framework integration. The main advantage of the proposed IGMLF is that it unifies GIS-based spatial hotspot recognition and ML-based accident risk estimation into one framework. This results in better performance in terms of accuracy, precision, recall, F1-score, ROC-AUC, as well as GIS-based accident risk visualization for effective road safety assessment.

TABLE IV. COMPARISON ANALYSIS WITH EXISTING STUDIES

Paper Title	**Accuracy**
A ML Approach for Classifying Road Accident Hotspots [6]	83%
Deep Learning Based Black Spot Identification on Greek Road Networks [7]	84.02%
Black Spots Identification on Rural Roads Based on Extreme Learning Machine [9]	98.6%
Identification and Analysis of Accident Black Spots Using GIS [4]	-
Integrated GIS–Machine Learning Framework for Road Accident Black Spot Prediction	**98.74%**

I. Discussion

For the proposed IGMLF framework, the following results were recorded: accuracy – 98.74%; precision – 98.41%; recall – 98.18%; F1-score – 98.29% and ROC-AUC – 98.86%. All of these results have outperformed all other baseline models (SVM – 82.60%, Decision Tree – 89.80%, Random Forest – 94.80%, XGBoost – 97.20%), indicating clearly the superiority of the use of GIS features, such as KDE-based zones of accidents density and BSI, together with ML classification over the tabular ML models. High ROC-AUC and low misclassification rate, confirmed by the confusion matrix, are critical for road safety because if there are undetected black spots, there are no any preventive actions at such dangerous spots. Furthermore, perfect ROC-AUC indicates the robustness of the model due to the class imbalance in the data set, which is inherent in accidents data set. Due to the use of GIS, spatially oriented four tiers risk zoning was conducted and demonstrated that all the high-risk zones were localized in the area of the existence of several hazardous factors (e.g., heavy traffic, poor visibility, complex intersection). As compared to previous researches, the IGMLF has proven to be more accurate than the ANN approach (with 83% accuracy), the DL model (with 74.35% AUC score), and slightly better than the ELM approach (with 98.6% accuracy). Hence, it is evident that the proposed GIS-ML pipeline, which includes spatial-temporal feature engineering, is an effective approach for predicting travel delays and offers an efficient solution. Nevertheless, further research could focus on minimizing the use of historical data and utilizing real-time traffic/weather conditions instead.

IV. CONCLUSION

The current research employed an Integrated GIS-Machine Learning Framework (IGMLF) model to determine road accident blackspots and evaluate the associated risks by applying a newly created Kaggle dataset of accidents in 2022-2025. Through employing the process of KDE-based hotspot detection, calculating the BSI values, and performing the ML classification, the presented IGMLF framework achieved an accuracy rate of 98.74%, precision of 98.41%, recall of 98.18%, F1-score of 98.29% and ROC-AUC score of 98.86%, outperforming all baseline ML algorithms such as SVM, Decision Tree, Random Forest, and XGBoost models. The application of the GIS helped to implement spatial analysis through creating a four-tier risk zoning, allowing transportation authorities to obtain a blackspot map on the basis of ML predictions. According to the feature importance analysis, traffic, weather, visibility, and road characteristics have a decisive influence on accidents; therefore, multi-factor spatial analysis can be used instead of the frequency analysis method. Future research would involve expanding the framework to include real-time traffic and weather data feeds, exploring deep learning frameworks such as graph neural networks for capturing the spatial relationship natively and evaluating the generalizability of the model on other geographic road network instances.

REFERENCE

[1] A. M. Mostafa *et al.*, “AI-based prediction of traffic crash severity for improving road safety and transportation efficiency,” *Sci. Rep.*, vol. 15, no. 1, p. 27468, Jul. 2025, doi: 10.1038/s41598-025-10970-7.
[2] F. Chen *et al.*, “Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model,” *Sci. Rep.*, vol. 15, no. 1, p. 25729, Jul. 2025, doi: 10.1038/s41598-025-00797-7.
[3] A. Yumak, S. Hengirmen Tercan, U. C. Colak, and S. Ozcanan, “A Machine Learning Approach to Identify High-Risk Road Segments and Accident Severity Patterns Based on Categorical Data,” *Appl. Sci.*, vol. 15, no. 23, p. 12824, Jan. 2025, doi: 10.3390/app152312824.
[4] M. S. Khatun, M. A. Hossain, M. A. Kabir, and M. A. Rahman, “Identification and analysis of accident black spots using Geographic Information System (GIS): A study on Kushtia-Jhenaidah national highway (N704), Bangladesh,” *Heliyon*, vol. 10, no. 3, p. e25952, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25952.
[5] B. F. Deressa, K. A. Habtegiogis, D. K. Endashaw, B. M. Al-Ramadan, and H. M. Al-Ahmadi, “A Systematic Review of GIS-Driven Road Traffic Accident Evaluation,” *Vehicles*, vol. 7, no. 4, p. 161, Dec. 2025, doi: 10.3390/vehicles7040161.

[6] B. de S. P. Amorim, A. A. Firmino, C. de S. Baptista, G. B. Júnior, A. C. de Paiva, and F. E. de A. Júnior, "A Machine Learning Approach for Classifying Road Accident Hotspots," *ISPRS Int. J. Geo-Inf.*, vol. 12, no. 6, p. 227, Jun. 2023, doi: 10.3390/ijgi12060227.

[7] I. Karamanlis *et al.*, "Deep learning based black spot identification on Greek road networks," *Data*, vol. 8, no. 6, p. 110, Jun. 2023, doi: 10.3390/data8060110.

[8] V. Sevetlidis, G. Pavlidis, S. G. Mouroutsos, and A. Gasteratos, "Leveraging Positive-Unlabeled Learning for Enhanced Black Spot Accident Identification on Greek Road Networks," *Computers*, vol. 13, no. 2, p. 49, Feb. 2024, doi: 10.3390/computers13020049.

[9] A. Mbarek, M. Jiber, A. Yahyaouy, and A. Sabri, "Black spots identification on rural roads based on extreme learning machine," *Int. J. Electr. Comput. Eng. IJECE*, vol. 13, no. 3, pp. 3149–3160, Jun. 2023, doi: 10.11591/ijece.v13i3.pp3149-3160.

[10] N. Hamdan and T. Sipos, "Predicting Segment-Level Road Traffic Injury Counts Using Machine Learning Models: A Data-Driven Analysis of Geometric Design and Traffic Flow Factors," *Future Transp.*, vol. 5, no. 4, p. 197, Dec. 2025, doi: 10.3390/futuretransp5040197.

[11] J. Cunha, J. S. Silva, R. Ribeiro, and P. Gomes, "Sustainable Road Safety: Predicting Traffic Accident Severity in Portugal Using Machine Learning," *Sustainability*, vol. 17, no. 24, p. 11199, Jan. 2025, doi: 10.3390/su172411199.

[12] O. I. Aboulola, "Improving traffic accident severity prediction using MobileNet transfer learning model and SHAP XAI technique," *PLOS ONE*, vol. 19, no. 4, p. e0300640, Apr. 2024, doi: 10.1371/journal.pone.0300640.

[13] Md. A. Khan Rifat, A. Kabir, and A. Huq, "An Explainable Machine Learning Approach to Traffic Accident Fatality Prediction," *Procedia Comput. Sci.*, vol. 246, pp. 1905–1914, 2024, doi: 10.1016/j.procs.2024.09.704.

[14] 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, p. 100814, May 2023, doi: 10.1016/j.trip.2023.100814.

[15] "Indian Road Accident Dataset (2022–2025)." Accessed: Jun. 30, 2026. [Online]. Available: https://www.kaggle.com/datasets/sehaj1104/indian-road-accident-dataset-20222025