Development of an Intelligent Road Safety Mobile Navigation Application Using Machine Learning for Real-Time Accident Hotspot and Severity Prediction: A Case Study in Sri Lanka

JMCT Jayamanna[1,] , Pradeep Kalansooriya[2]*

[1,2] Faculty of Computing, General Sir John Kotelawala Defence University, Sri Lanka
38-bcs-0012@kdu.ac.lk*

Abstract. In Sri Lanka as well as other countries, traffic accidents are a leading cause of fatalities. Many factors contribute to the high mortality rate in developing countries. One of the reasons that road accidents occur is that people are unaware of common accident locations. The government has already enforced other tactics, like traffic signals and fines, to reduce these incidents, but they have been ineffective. But if people have a way to change their driving patterns based on the common accident locations, so-called hotspots, we can decrease the road traffic accidents that happen on a daily basis. This study focuses on improving road safety by developing an intelligent road safety mobile navigation application that provides real-time alerts on accident hotspots and severity levels to drivers during their journey on the road. This system uses supervised machine learning for building a most suitable model for accident hotspots and severity prediction that is tailored for Sri Lankan accident data by conducting an in-depth review of existing machine learning techniques and analysis of historical accident data. According to the review, comparison, and evaluation, the final predictive machine learning model was implemented using an XGBoost regression model for the prediction of accident hotspots and a random forest classification model for accident severity prediction. By integrating the implemented predictive model, this study proposes a mechanism to gather real-time accident data and the development of the mobile application, which provides the foundation to revolutionize road safety and enables drivers to make safer decisions on the road.

***Keywords**: machine learning, supervised learning, regression algorithms, classification algorithms, predictive analytics, hotspots identification, severity analysis*

1. Introduction

In recent decades, the number of vehicles on roads has notably risen, which results in congested highways. This huge traffic load has pushed up the road accident graph, increasing injuries and deaths globally. Accidental deaths or impairments in children have a profound impact on families and society. According to the World Health Organization (WHO), approximately 1.19 million people lose their lives in road traffic accidents each year, and those accidents claim the lives of up to 3,700 people daily.[1] Between 20 and 50 million non-fatal injury cases occur each year, many of which lead to disability. The Ministry of Transport and Highways reports that 21,953 traffic incidents occurred in Sri Lanka in 2022.[2] And not only have accidents caused more deaths or injuries, but they have also increased the financial damage to public property.

One of the reasons that road accidents occur is that people are unaware of common accident locations. While many accident causes appear to be similar and may be managed together, the actual causes of traffic accidents vary from state to state based on their level of development. As a result, it is critical to reduce the number of traffic accidents that occur every day. The government has already enforced other tactics, like traffic signals and fines, to reduce these

https://doi.org/10.31705/ ICAHETS.2024.8

incidents, but they have been ineffective. But in order to decrease road traffic accidents that happen on a daily basis, people must change their driving patterns.

In both developed and developing nations, road accidents are acknowledged as a major contributor to human and economic losses. Every year, Sri Lanka has seen a sharp increase in these instances, which have led to substantial expenses in terms of both money and human lives. These occurrences pose an extensive danger to public welfare, thus necessitating efficacious measures for their prevention and alleviation. As stated before, since current tactics taken by the government have been ineffective, if people have a way to change their driving patterns based on the common accident locations, so-called hotspots with severity, we can decrease the road traffic accidents that happen on a daily basis. While looking for a solution to that problem, existing research studies around the world have proposed predictive machine learning models for accident hotspots and severity. But there were not any existing studies that proposed a solution that was suitable for Sri Lankan accident data. And also, the current approaches have been ineffective and lack real-time road accident analysis and geospatial visualization, hindering the ability to respond proactively to road safety issues, and there is no precise practical application to identify accident-prone areas, so-called 'Hot Spots' and severity in real time.

The aim of this research study is to develop a mobile navigation application by integrating machine learning models for prediction of accident hotspots and severity tailored for Sri Lankan accident data that enriches road safety by presenting up-to-date real-time alerts on accident hotspots and severity to drivers during their journey on the road. The objectives of this study for the proposal of an intelligent road safety mobile navigation application are given below:

- To analyze and finalize the features from the collected traffic accident data and identify the correct machine learning techniques for developing and training the data to predict target variables (accident hot spots, severity).
- To test the performance of trained data by using new data, assess accuracy, and create mechanisms to get new accident data from relevant authorities in real time to the system to make new predictions.
- To complete the development phase of the mobile application by finalizing the backend architecture, integrate the trained model into the application, and conduct testing to ensure functionality and reliability.

2. Literature Review

2.1 Existing studies on implementing predictive models for accident hotspots and severity

Existing studies on implementing predictive models for accident hotspots and severity that are specifically tailored for their own accident data are discussed below, demonstrating that machine learning techniques performed better as predictors than statistical techniques,

The goal of a study by Jayesh Patil et al. is to create a map-connected application that provides a user with a clear picture of places that are prone to accidents, issues a warning about the state of the roads, and offers safety tips. Using the unsupervised learning technique known as k-means clustering, which analyzes traffic accidents by considering various factors like potholes in the road, sharp turns, and weather conditions, and then provides appropriate and preventative measures to avoid accidents by representing them on a map and building an understandable model for everyone, the study aims to reduce the mortality rate by developing a prediction model. The expected model's accuracy was 81% [3].

The authors of a paper by Aklilu Elias Kurika et al. talked about how a machine learning system was used to forecast auto accidents. The model evaluation, performance measurement, and experimental results showed that the Random Forest tree did poorly, but the F-measure of the J48 and Rep tree classifiers was equal [4].

An accident prediction system can be used to estimate the risk of an event and evaluate potential safety concerns, per a study by Teres Augustine et al. Several machine learning algorithms were compared in order to ascertain which model could help with more precise accident prediction. The dataset used in this work was the government database of accidents that occurred in a particular Indian area. The machine learning models utilized in this work to predict accidents are Support Vector Machine, K-Nearest Neighbor, Random Forest, Decision Tree, XGBoost and Logistic Regression. The Random Forest method yielded the highest accuracy, 80.78%, when the accuracies of the machine learning models were compared [5].

Accident data features such as location, weather, road type, and collision type were taken into consideration for forecasting traffic accidents in the article by Mahendra G et al. Based on collision information from several Indian states, a machine learning model that includes a decision tree and random forest regressor is being developed and is presently being used to predict collisions. The findings of the relative analysis show that the Random Forest Regressor (RFR) model performs better than the Decision Tree Regressor (DTR) model [6].

In order to analyze the accident data, Daniel Santos et al.'s article developed models that can identify a subset of significant characteristics that may be used to categorize an accident's severity. Additionally, this study proposes a forecast model for future traffic events based on historical data. In addition to supervised machine learning techniques like decision trees, random forests, logistic regression, and naive Bayes, unsupervised machine learning techniques like DBSCAN and hierarchical clustering are also used. The prediction model's findings imply that the RF model might be a practical tool for locating accident hotspots [7].

Brunna de Sousa Pereira Amorim and colleagues' study focuses on to find the best classifier for Brazilian federal road hotspots associated with either a severe or non-severe accident probability, a variety of machine learning techniques were applied. Testing was done using a multi-layer perceptron neural network, random forest, and SVM. The dataset includes a 10-year report on road accidents from the Federal Highway Police of Brazil. Weather conditions, accident type, road type day of the were features used for analysis. As a result, the suggested neural network model accurately produced the intended result: 83% accuracy, 84% precision, 83% recall, and 82% F1 score [8].

Research by R. Vanitha and associates, they employed machine learning methods such as random forest, decision tree, and logistic regression to develop an accident prediction model. Both accident prediction and the development of safety measures will benefit from these classification schemes. Road surface conditions, weather, vehicle condition, and lighting conditions are some of the factors that are used to forecast traffic accidents. Additionally, the RF model offered 86.66% greater accuracy [9].

Study by Abdelilah Mbarek et al., created a model with the purpose of identifying, categorizing, and analyzing black patches on Moroccan highways. Ordinal regression is used to examine the infrastructure factors once the extreme learning machine (ELM) technique has been used to identify these areas. The XGBoost model is used to construct the weighted severity index (WSI), which in turn yields the severity scores to be assigned to certain road segments. After that, the latter are categorized using a four-class system (high, medium, low, and safe). The simulation results show that the proposed framework outperformed the trustworthy competing models, especially in terms of accuracy (98.6%), and effectively and accurately identified the black areas [10].

Rabia Emhamed et al.'s work looks into models to find a selection of important characteristics and build a model for classifying injury severity. These models are built using traffic accident data using supervised machine learning methods including AdaBoost, Random Forests (RF), Logistic Regression (LR), and Naive Bayes (NB). A specific technique called the SMOTE approach is used when the data sets are out of balance. In light of the study's findings, this work has demonstrated that the RF model may be utilized to predict accident severity with an accuracy of over 74.5% [11].

Jian Zhang and colleagues' study, this study compared the predictive performance of different statistical and machine learning techniques with different modeling logics for crash severity analysis in terms of prediction accuracy and variable importance estimate. In Florida's highway diverging regions, information on crash severity, road shape, and traffic flow was gathered. Four popular machine learning models were calculated, including Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), and K-Nearest Neighbors (KNN). Two of the most popular statistical models, the ordered probit (OP) model and the multinomial logit (MNL) model, were also computed. The findings demonstrated that machine learning techniques performed better as predictors than statistical techniques. For both overall and severe crashes, the RF technique produced the best forecast, while the OP method produced the lowest estimate [12].

In order to forecast the severity of traffic accidents, Md. Kamrul Islam and colleagues' research compares a logistic regression model with tree-based ensemble techniques (random forest and gradient boosting). Random forest is used to identify significant traits that are significantly correlated with the severity of traffic accidents. The analysis's findings showed that the type of collision and its cause are the two main factors affecting the degree of injuries received in traffic accidents. When using k-fold (k = 10), the random forest strategy outperformed other models in terms of injury severity, individual class accuracies, and collective prediction accuracy, according to a variety of performance parameters [13].

Kenny Santos et al.'s research review offers a current evaluation of the methods employed in traffic crash injury severity modeling. Included are 56 studies that cover the years 2001–2021, accounting for over 20 different statistical or machine learning techniques. According to the results, Random Forest yielded the best results in 70% of the applications and 29% of the studies. Support Vector Machine and Decision Tree performed best in 53% and 31% of cases, respectively, and in 16% and 14% of total trials. K-Nearest Neighbors and Bayesian Networks fared best in only 4% and 7% of the trials that were reviewed, although being used 67% and 40% of the time, respectively [14].

The study by Buket Geyik et al. uses prior knowledge as training data to classify data objects into groups, which is useful for our work. Among our models are the Multilayer Perceptron (MLP), Naïve Bayes, decision tree, and random forest classifiers. Three accident severity categories such as mild infection, serious infection, and death were applied to the dataset. The goal of the study is to develop models that predict the accident severity levels of traffic accident injury records for upcoming incidents using data mining classification approaches. 83.40% precision. The accuracy of the MLP model was 86.67%, the decision tree algorithm was 80.74%, the random forest classifier was 85.19%, and the Naive Bayes strategy was 83.40%, based on the findings of the classification techniques that were evaluated [15].

Predicting the severity of traffic accidents is a crucial step in managing accidents on the road, according to a study by Imad EL MALLAHI et al. This proposed study compares the Random Forest, Support Vector Machine, and Artificial Neural Network approaches for classifying and predicting the severity of traffic incidents. They use accident data to categorize the severity prediction of traffic accidents into three classes: driver or rider, vehicle or pillion passenger, and pedestrian. They have 93% accuracy for random forest and 93.82% precision recall at the precision recall level, compared to 82.22% for SVM and 87.88% for ANN [16].

Research by Tariq Al-Moqri and colleagues, Evaluating the major factors that contribute to the severity of injuries is their primary objective. The selected machine learning algorithms were compared to conventional statistical techniques. The accident data revealed three categories of injury severity: minor, serious, and severe. To attain balance, a synthetic minority oversampling technique (SMOTE) was applied. The Multinomial Logit Model (MNL) was used to compare the five machine learning classifiers: Naïve Bayes (NB), J48 Decision Tree, Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP).Among the five classifiers, RF has the highest accuracy rate (94.84%) [17].

2.2 Best performing machine learning technique for accident hotpots and severity prediction analysis

Table 1. Existing machine learning techniques used for the predictions.

Author	Machine Learning Techniques Used	Best Performing ML Technique	Prediction Focus
Jayesh Patil et al. [3]	K-means Clustering	K-means Clustering	Accident Hotspot
Daniel Santos et al. [7]	Hierarchical clustering, DBSCAN, Decision trees, Random Forest, Logistic regression, Naive Bayes	Random Forest	Accident Hotspot
Aklilu Elias Kurika et al.[4]	J48, Rep Tree classifiers	J48	Accident Hotspot
Mahendra G et al.[6]	Decision Tree, Random Forest Regressor	Random Forest Regressor	Accident Hotspot
R.Vanitha et al. [9]	Decision Tree, Random Forest, Logistic Regression	Random Forest	Accident Hotspot

Abdelilah Mbarek et al. [10]	Extreme Learning Machine (ELM), XGBoost	XGBoost	Accident Hotspot
Jian Zhang et al. [12]	KNearest Neighbor, Decision Tree, Random Forest, Support Vector Machine	Random Forest	Accident Hotspot
Tao Lu et al. [18]	Logistic regression, Random Forest	Random Forest	Accident Hotspot
Kenny Santos et al. [14]	Random Forest, Support Vector Machine, Decision Tree, Bayesian Networks, K-Nearest Neighbors	Random Forest	Accident Hotspot
Rabia Emhamed et al. [11]	AdaBoost, Logistic Regression, Naive Bayes, Random Forests	Random Forest	Accident Severity
Brunna de Sousa Pereira Amorim et al. [8]	SVM, Random Forest, Multi-layer perceptron neural network	Multi-layer perceptron NN	Accident Severity
Teres Augustine et al. [13]	Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbor, XGBoost, Support Vector Machine	Random Forest	Accident Severity
Md. Kamrul Islam et al. [10]	Logistic Regression, Random Forest, Gradient Boosting	Random Forest	Accident Severity
Buket Geyik et al. [15]	Multilayer Perceptron (MLP), Decision Tree, Random Forest, Naive Bayes	MLP	Accident Severity
Imad EL MALLAHI et al. [16]	Random Forest, Support Vector Machine, Artificial Neural Network	Random Forest	Accident Severity
Tariq Al-Moqri et al. [17]	Naïve Bayes, J48 Decision Tree, Random Forest, Support Vector Machine, Multilayer Perceptron	Random Forest	Accident Severity

2.3 Selection of machine learning technique for the proposed models

Table 2. Usage of machine learning techniques in existing studies

Machine learning techniques	Usage Percentage (%)
Random Forest	81.25%
Decision Tree	50%
Support Vector Machine	37.5%
Logistic Regression	37.5%
Naïve Bayes	25%
K-Nearest Neighbor	18.75%
Multi-Layer perceptron neural network (MLP)	18.75%
XGBoost	12.5%
J48	12.5%
K-means Clustering	6.25%
Hierarchical Clustering	6.25%
DBSCAN	6.25%
Extreme Learning Machine (ELM)	6.25%
Bayesian Networks	6.25%
AdaBoost	6.25%
Artificial Neural Network	6.25%

Table 3. Usage of best-performing machine learning techniques in existing studies

Machine learning techniques	Performance Percentage (%)
Random Forest	68.25%
Multi-Layer perceptron neural network (MLP)	12.5%
XGBoost	6.25%
J48	6.25%
K-means Clustering	6.25%

3. Methodology

3.1 Overview of the project

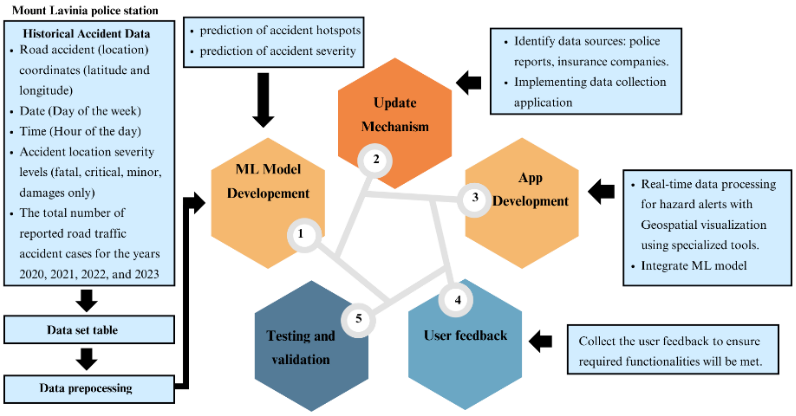


Fig. 1. Process of proposed methodology.

https://doi.org/10.31705/ ICAHETS.2024.8

As shown in Fig. 1, to reach the milestones, this proposed methodology approach primarily consists of five stages.

Stage 1: Data collection

This study's proposed methodology is a thorough strategy that combines quantitative and qualitative research methods. This means that the research will involve collecting and analyzing both qualitative and quantitative data, such as accident data reports and interviews. The study population will focus on the Mount Lavinia region of the Colombo district of the western province of Sri Lanka. The following data will be hoped to be collected in partnership with the Mount Lavinia police station.

- Road accident hot spot (location) coordinates (latitude and longitude)
- Days and times of accidents
- Accident hotspots (locations) severity levels (fatal, critical, minor, damages only)
- The total number of reported road traffic accident cases, deaths (fatal), critical, minor, and damages only, for the years 2020, 2021, 2022, and 2023
- Reasons for road accidents

When the data collection phase is over, the feature analysis will be done to provide the most suitable solutions for implementing the machine learning models. During the feature analysis, in order to train and test the machine learning models to make predictions, the gathered historical accident data should be converted into a tabulated data table, and the features and target variables should be preprocessed before starting the machine learning model development part of stage 1.

Stage 1 part 2: ML model development

The analysis of finding the correct machine learning techniques based on the features and implementing, training, and testing the machine learning model for both accident hotspot and severity prediction that is specifically tailored for Sri Lankan accident data is the main purpose of the stage.

Stage 2: Update mechanism

The implementation of the machine learning model will be done based on the historical accident data, and dealing with limited historical accident data and converting that raw data into tabulated data is time-consuming. In order to keep the mobile navigation application long-lasting and to provide the predictions with more accuracy, the models need to be retrained with new accident data. The implementation of a responsive data collection application to keep the model updated in a timely manner with new accident data is the purpose of this stage.

Stage 3: App development

Concurrently, this stage focuses on developing the final practical implementation of the study, the intelligent road safety mobile navigation application integration with machine learning model, and the required functionalities.

Stage 4: User Feedback Mechanism

Once the implementation is done as discussed from the previous stages, in order to do the future enhancement, this stage is focused on generating feedback on the application to make the application user-friendly and provide accurate results in the future.

Stage 5: Testing and validation

Along with the analysis and development, testing the machine learning models and validating the application with the integration of ML models will be done in this stage simultaneously.

3.2 Functional Requirements

- Machine Learning Model Development: Implemented machine learning models should predict and provide accident hotspots with severity.
- Machine Learning Model Integration: The predicted accident hotspots and severity from the machine learning model should be integrated with the application and displayed on the hotspots map.
- Real-Time Alerts: The application should provide alerts on accident hotspots to the users along the route in real time.
- View the hotspot map: The user should be able to view the accident hotspot map through the application.
- User Navigation: The user should be able to enter the start and destination locations to start the navigation through the application.
- Accident Data Insertion: The users should be able to enter the accident details through the data collection app, and the data should be transmitted to the database.

3.3 Non-Functional Requirements

- Scalability: Application integration with ML models needs to be developed to handle new accident data that comes from the data collection application.
- Performance: The application needs to be optimized to provide user alerts in real-time without any delay.
- Reliability: Accurate hotspots with severity need to be provided for the user from the application.
- Data Quality: Frequent analysis and preprocessing should be done with the new accident data.

3.4 Technology

- Machine Learning (ML) with Python: Jupyter Notebook.
- Flutter Framework for Mobile App Development.
- Database Management: Firebase Firestore.
- Geospatial Visualization Tools: MapBox API
- Machine Learning Model Integration: Flask API

4. Proposed Solution

4.1 Machine learning model Development

Prediction of accident hotspots

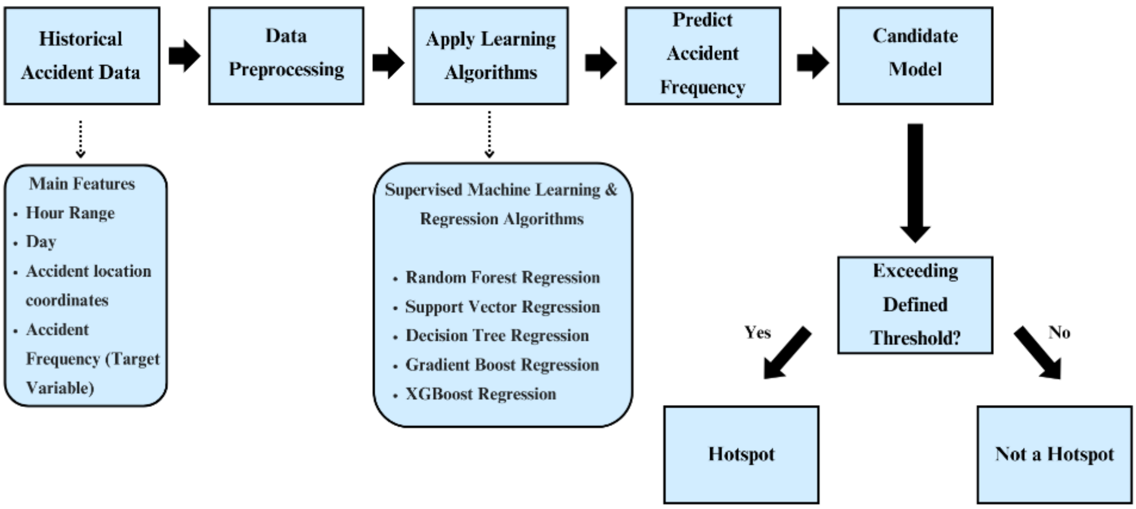


Fig. 2. Accident hotspots prediction solution design

The goal of accident frequency prediction, as seen in Fig. 2, is to project the probability that accidents will occur in a given area in the future. An accident hotspot is a location, whether it be a connection or a node, that has unusual crash frequencies or rates.

For the model selection, supervised machine learning and regression algorithms will be used to train the models since the collected data consist of a labeled data set and has a continuous target variable called accident frequency (accident count). Based on the collected historical accident data stated in the methodology, hour-range, day, coordinates of the accident location, and accident count were the features that were used for this hotspot prediction. Right after finalizing the features used for the prediction, the data processing should be conducted in order to maintain the data quality and to get an understanding of the data set.

After completion of data preprocessing, it will come to the next stage that focuses on applying machine learning algorithms. Based on the usage of machine learning techniques that were analyzed in Tables 2 and 3, regression models such as random forest regression, gradient boost regression, support vector regression, decision tree regression, and XGBoost regression will be used to predict the accident frequency. And the model that provides better accuracy will be chosen for further fine-tuning to increase accuracy. Once the fine tuning is done with the chosen model, it'll be considered as the candidate model.

After successfully predicting the accident frequency, the threshold value for accident frequency will be obtained by calculating the mean of the accident count. And the coordinates (locations) exceeding this threshold value are considered hotspots, which will be the final prediction of the candidate model.

Prediction of accident severity

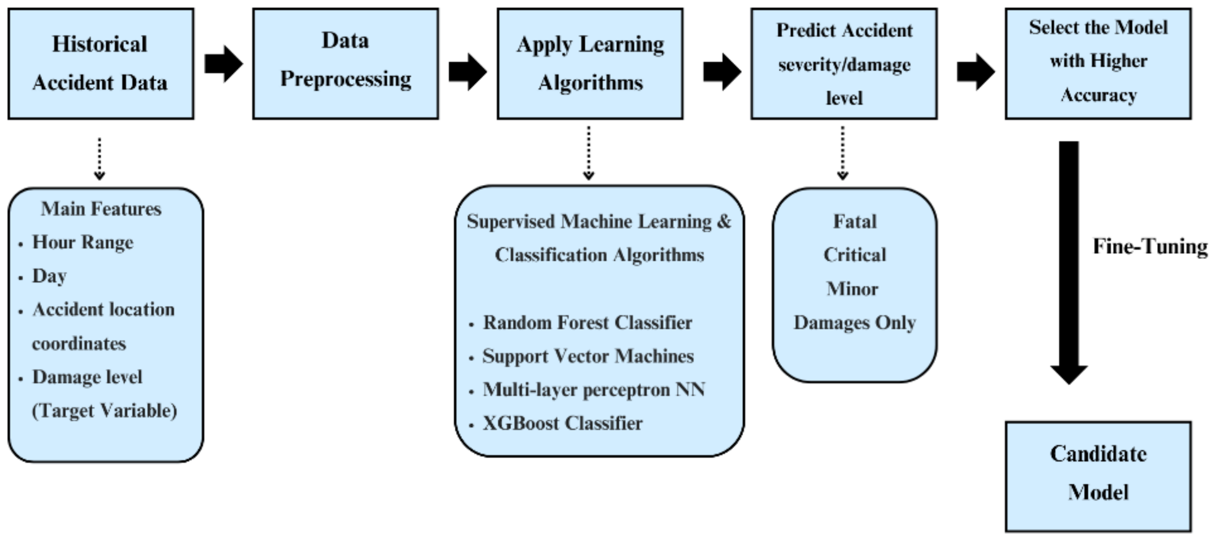


Fig. 3. Accident severity prediction solution design.

As shown in Fig. 3, the proposed solution for accident severity prediction is: Since the severity level of the accident hotspot is classified into four levels, such as "fatal," "critical," "minor," and "damages only," which was the target variable for the model selection, supervised machine learning and classification algorithms will be used rather than regression algorithms. Based on the collected historical accident data stated in the methodology, hour-range, day, coordinates of the accident location, and damage level were the features that were used for this severity prediction. Right after finalizing the features used for the prediction, the data processing should be conducted in order to maintain the data quality and to get an understanding of the data set.

After completion of data preprocessing, it will come to the next stage, which focuses on applying machine learning algorithms. Based on the usage of machine learning techniques that were analyzed in Tables 2 and 3, classification models such as random forest classifier, XGBoost, multi-layer perceptron NN, and support vector machines will be used to predict the damage level/severity. And the model that provides better accuracy will be chosen for further fine-tuning to increase accuracy. Once the fine tuning is done with the chosen model, it'll be considered as the candidate model.

4.2 Update mechanism

Since dealing with limited historical accident data and converting that raw data into tabulated data is time-consuming, while developing machine learning models, A separate mobile application will be developed for the accident data collection from the relevel authorities so that

https://doi.org/10.31705/ ICAHETS.2024.8

the models can be trained from time to time with the new accident data and transformed into the main app that provides alerts on accident hotspots and severity. Regularly evaluating and potentially retraining your model with new data can improve its effectiveness over time. This process will provide the solution for the update mechanism, and it'll save time as well. The Flutter framework with Firebase and Flutter Map will be used for the development of the data collection app.

4.3 App Development

As shown in Fig. 4, the main app called the intelligent road safety mobile navigation app consists of five main components that need to be focused on while selecting the feasible technologies to successfully accomplish the project.

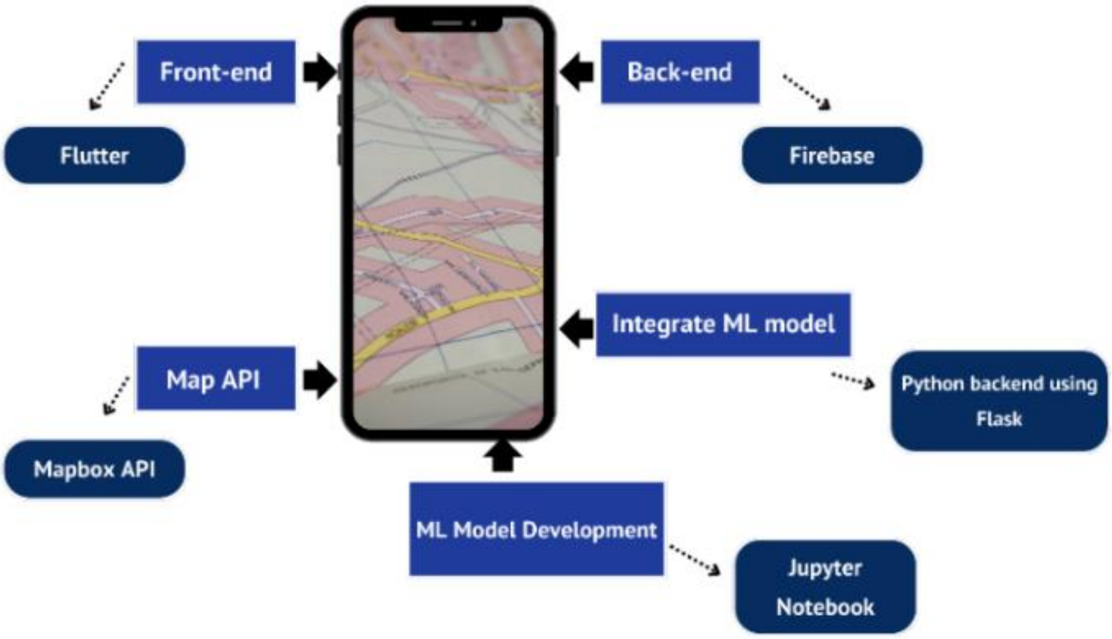


Fig. 4. Main Components of the Application

Front-end & back-end development

Choosing the Flutter framework as the front end and Firebase as the back end is very feasible for accomplishing the main navigation app, especially when working with machine learning and with the features described in the methodology. Cross-platform development, native performance, rich UI components, hot reload, and integration with Firebase provided with Flutter and the Firebase

Integrating maps APIs

This proposed application requires a map API that allows for custom overlays, real-time data updates, and custom marker placement for navigation software that includes features like real-time notifications for accident hotspots and severity levels. Mapbox API will be used to accomplish this requirement.

ML model development and integrating ML models

Jupyter Notebook is one of the major IDEs known as web-based computing platforms that is specifically used for machine learning model development. Right after developing the ML model for prediction of accident hotspot and severity, the integration of those ML models into the application can be achieved using the Flask API.

5. Results and Discussion

5.1 Data analysis

- Converted the historical accident data into tabulated form suits for the predictions.
- Selected four features/variables
  - "hour_range": 1-2 AM, 2-3 AM, etc.
  - "day": day of the week.
  - "damage_level": Details about the extent of damage caused by accidents, which could be used as an indicator of severity (fatal, critical, minor, or damage-only).
  - "coordinates": the latitude and longitude coordinates.

5.2 Accident hotspot prediction

- Finalized the data set table structure to train and test the models using regression algorithms that were stated in the proposed solution.

	hour_range_numerical	day_numerical	latitude	longitude	accident_count
0	1.5	4	6.809750	79.882724	2
0	1.5	4	6.829248	79.884956	2
1	10.5	4	6.830007	79.867472	1
2	10.5	0	6.809814	79.883027	2
2	10.5	0	6.840984	79.884568	2
3	10.5	5	6.834691	79.867351	2
3	10.5	5	6.808229	79.883881	2
4	10.5	1	6.836735	79.867436	1
5	11.5	4	6.834176	79.866375	1

Fig. 5. Final Data set table structure

- After finalizing the data set table, data preprocessing was conducted through exploratory data analysis (EDA).

- Prepared the data as X (features) and Y (target variable) and split the data set into training and testing sets.
- Five selected regression machine learning models named random forest regres-sion, gradient boost regression, support vector regression, decision tree regres-sion, and XGBoost regression were trained, tested, and evaluated with mean squared error (MSE) and mean absolute error (MAE) to select the candidate model for further fine-tuning.

Table 4. Results of Machine Learning Models for The Prediction of Accident Hotspots

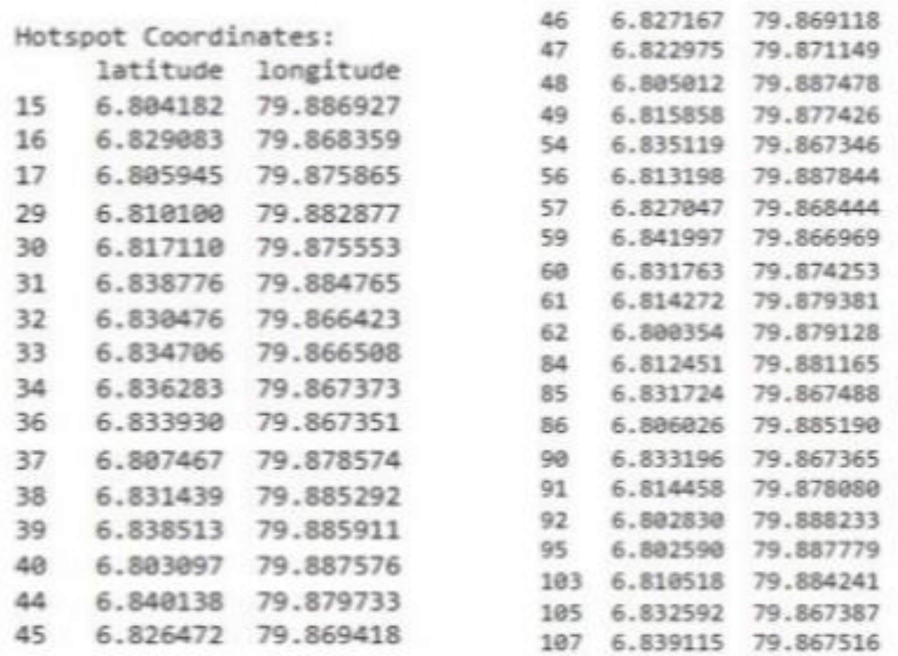
Model	Mean Squared Error	Mean Absolute Error
Random Forest Regression	0.54	0.58
Support Vector Regression (SVR)	1.20	0.70
Gradient Boosting Regression	0.53	0.58
XGBoost Regression	0.44	0.53
Decision Tree Regression	0.91	0.73

- Fine-tuned XGBoost regression model used for the application

Table 5. Results of Fine-Tuned XGBoost Regression Model

Model Name	Mean Squared Error	Mean Absolute Error
XGBoost Regression	0.40	0.50

As shown in Table 5, The fine-tuned XGBoost model was provided with a mean square error of 0.40 and a mean absolute error of 0.50, which indicates that the regression model achieved 91.7 percent accuracy within the target range of 6. Based on that conclusion, this fine-tuned model is selected as the final model for the prediction of accident hotspots.


Hotspot Coordinates:

	latitude	longitude
15	6.804182	79.886927
16	6.829083	79.868359
17	6.805945	79.875865
29	6.810100	79.882877
30	6.817110	79.875553
31	6.838776	79.884765
32	6.830476	79.866423
33	6.834706	79.866508
34	6.836283	79.867373
36	6.833930	79.867351
37	6.807467	79.878574
38	6.831439	79.885292
39	6.838513	79.885911
40	6.803097	79.887576
44	6.840138	79.879733
45	6.826472	79.869418
46	6.827167	79.869118
47	6.822975	79.871149
48	6.805012	79.887478
49	6.815858	79.877426
54	6.835119	79.867346
56	6.813198	79.887844
57	6.827047	79.868444
59	6.841997	79.866969
60	6.831763	79.874253
61	6.814272	79.879381
62	6.800354	79.879128
84	6.812451	79.881165
85	6.831724	79.867488
86	6.806026	79.885190
90	6.833196	79.867365
91	6.814458	79.878080
92	6.802830	79.888233
95	6.802590	79.887779
103	6.810518	79.884241
105	6.832592	79.867387
107	6.839115	79.867516

Fig. 6. Predicted Hotspot coordinates from the model

- A threshold (mean of accident count) was set for hotspot identification and predicted accident hotspots based on the predicted count. The predicted accident hotspot coordinates (latitude, longitude) are shown below Fig. 6.

5.3 Accident severity prediction

- Finalizing the data set table structure to train and test the models using classification algorithms that were stated in the proposed solution. Damage level classified and encoded as damages only (0), minor (1), critical (2), and fatal (3).

	data_point	hour_range_numerical	day_numerical	latitude	longitude	damage_level
0	0	1.5	4	6.809750	79.882724	2
1	0	1.5	4	6.829248	79.884956	0
2	1	10.5	4	6.830007	79.867472	0
3	2	10.5	0	6.809814	79.883027	0
4	2	10.5	0	6.840984	79.884568	1
5	3	10.5	5	6.834691	79.867351	0
6	3	10.5	5	6.808229	79.883881	3
7	4	10.5	1	6.836735	79.867436	0
8	5	11.5	4	6.834176	79.866375	0
9	6	11.5	0	6.821862	79.877699	2

Fig. 7. Final Data set table structure

- After finalizing the data set table, data preprocessing was conducted through exploratory data analysis (EDA).
- Prepared the data as X (features) and Y (target variable) and split the data set into training and testing sets.
- Three selected classification machine learning models named random forest classification, XGBoost classification, and multi-layer perceptron classification were trained and tested and evaluated with accuracy, precision, recall, and F1-score to select the candidate model for further fine-tuning.

Table 6. Results of Machine Learning Models for The Prediction of Accident Severity

Model	Average Accuracy Score	Average Precision Score	Average Recall Score	Average F1-Score
Random Forest Classification	36%	29%	36%	32%
MLP Classification	36%	27%	36%	31%
XGBoost Classification	31%	38%	31%	33%

- Fine-tuned random forest classification model used for the application

Table 7. Results of Fine-Tuned Random Forest Classification Model

Model	Average Accuracy Score	Average Precision Score	Average Recall Score	Average F1-Score
Random Forest Classification	68 %	67%	68%	67%

As shown in Table 7, the limited number of features available for accident severity prediction is one of the limitations to providing more accurate results. However, it still provided a good overall accuracy of 68.75 percent after fine-tuning the model.

5.4 Data collection app (Update Mechanism)

Integrated with Flutter Map to allow the user to mark accident locations through the map.

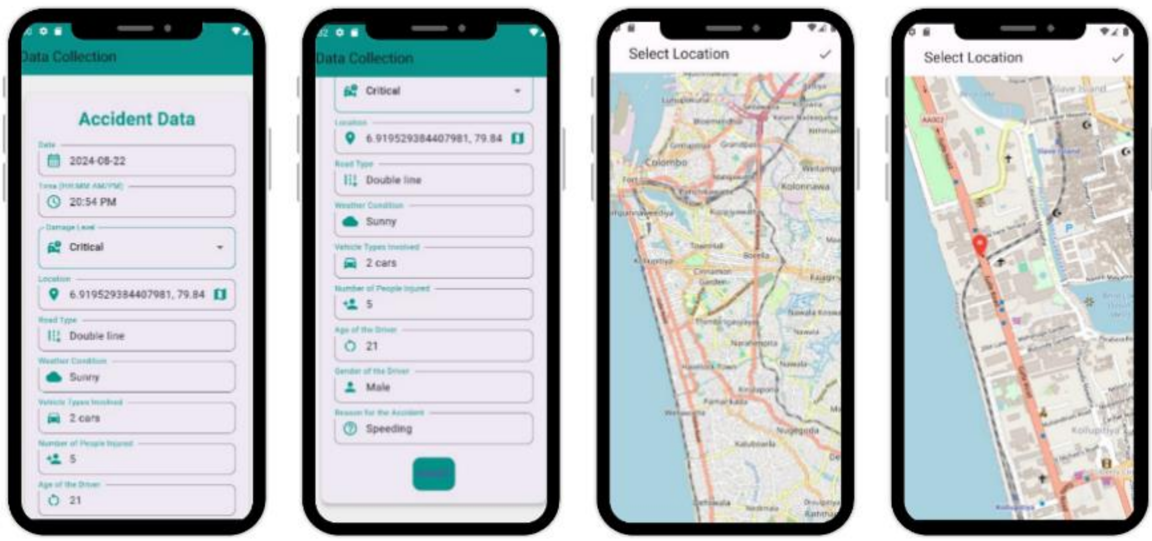


Fig. 8. Interfaces of the accident data collection application

Accident data that was identified in existing studies was implanted through application interfaces shown in Fig. 8 and provided with the option to enter the accident location by selecting from the map. The coordinates of the particular selected accident location will be stored in the location field as latitude and longitude.

1. **Proposed Intelligent Road Safety Mobile Navigation Application**

As shown in Fig. 9, the two machine learning models for the prediction of accident hotspots and severity are combined to implement the final model. The user can either view the hotspot map with color-coded severity levels or start their journey. And when the user starts a journey by entering the start location and destination to the system, the user will get sound alerts on accident hotspots with color-coded severity along the route in real time with the help of the machine learning model integrated with the application. And the data collection app allows to retrain the machine learning model with new accident data.

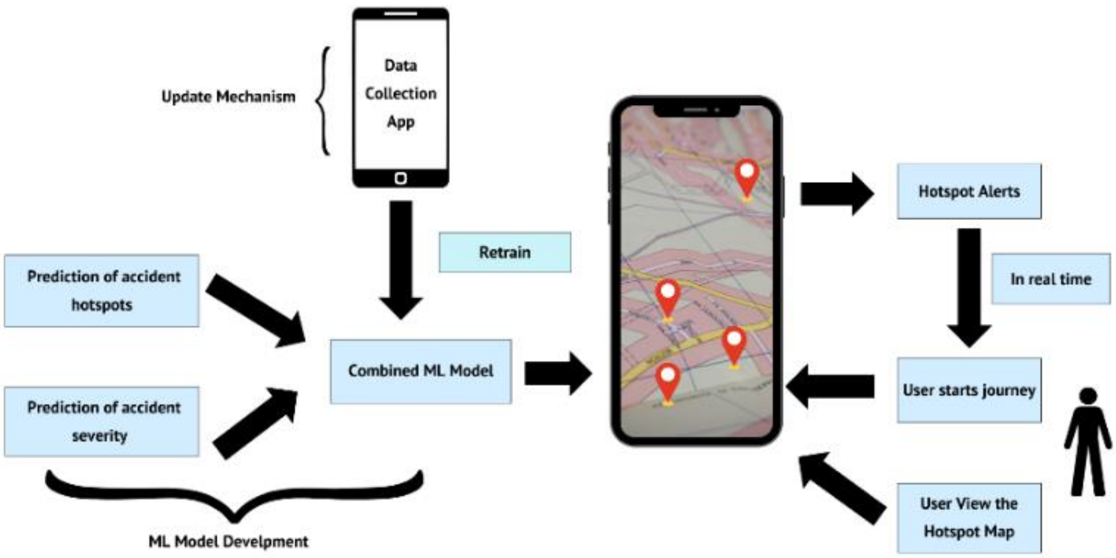


Fig. 9. Flow of the overall system

And the fully developed application shown in Fig. 10. consists of three main screens, such as the home screen where the user can view the hotspot map or start the navigation, the hotspot map screen where the user can view the hotspots with a color-coded severity map, and the navigation screen where the user can enter the current location and destination and start navigation. The simulate route button was implemented for testing purposes only.

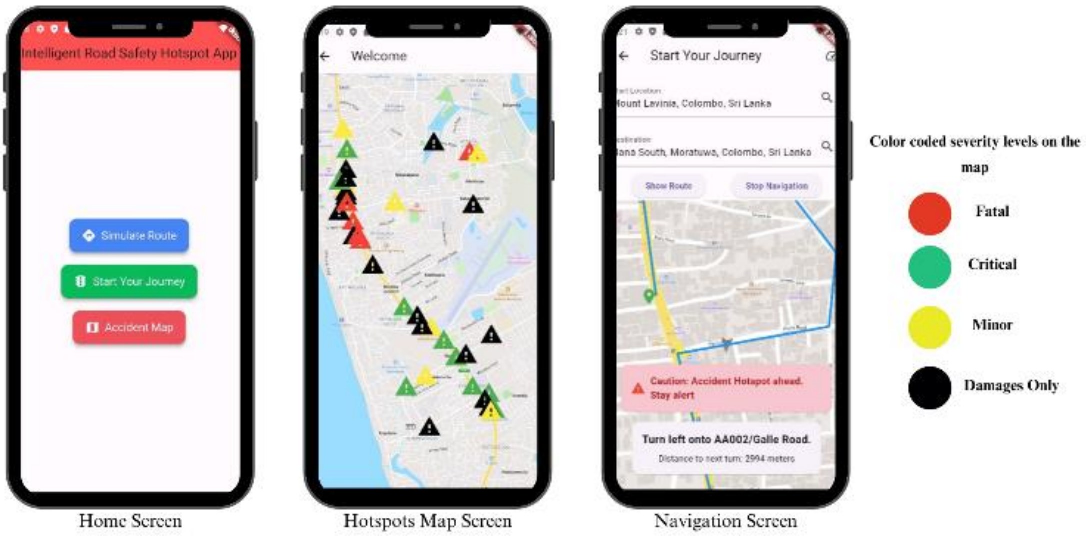


Fig. 10. Developed application interfaces

2. Conclusion & Future Work

"The Intelligent Road Safety Mobile Navigation Application" is an application bearing significant benefits in the enhancement of the existing road safety measures. A novel machine learning approach was presented with the application, which is considered to be the first machine learning model tailored for Sri Lankan accident data that combines hotspots and severity prediction into one integrated solution. And also, instead of traditionally recorded accident data on paper, this study presented an update mechanism through the development of an accident data collection application that allows to keep the model updated in a timely manner. Users are able to make decisions by being aware of accident-prone areas along with their intensity with the help of this navigation application that specifically provides hotspot alerts with text to speech and color-coded severity levels on the map. The natural ease of use and proactive accident prevention techniques of the program are capable of drastically reducing the extent of road traffic accidents, which is instrumental in protecting lives and enhancing social order. This application incriminates and encourages the relative use of technology in efforts to mitigate how these problems affect society in the making or implementation of this creative smartphone application. By adopting this progressive method, we move one step closer towards making the world's roads better and safer for everyone. With certain future improvements to the suggested system, the system's defined scope can be extended even further. Some potential future improvements, such as optimization of the user interface for better experience, connecting the application with the accident data collection application to keep the model updated, and expanding the application for whole Sri Lanka, can be proposed to the system.

References

1. Road traffic injuries, https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries, last accessed 2024/04/18.
2. National Council for Road Safety, https://www.transport.gov.lk/web/index.php?option=com_content&view=article&id=29&Itemid=149&lang=en#road-accidents, last accessed 2024/07/06.
3. Patil, J., Patil, V., Walavalkar, D., Lobo, V.B.: Road Accident Analysis and Hotspot Prediction using Clustering. In: 2021 6th International Conference on Communication and Electronics Systems (ICCES). pp. 763–768 (2021). https://doi.org/10.1109/ICCES51350.2021.9489074.
4. Ganie, I.: Predicting Factors of Vehicular Accidents using Machine Learning Algorithm. 9, (2020). https://doi.org/10.30534/ijeter/2020/46892020.
5. Augustine, T., Shukla, S.: Road Accident Prediction using Machine Learning Approaches. In: 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). pp. 808–811 (2022). https://doi.org/10.1109/ICACITE53722.2022.9823499.
6. G, M., R, R.H.: Prediction of Road Accidents in the Different States of India using Machine Learning Algorithms. In: 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS). pp. 1–6 (2023). https://doi.org/10.1109/ICICACS57338.2023.10099519.
7. Santos, D., Saias, J., Quaresma, P., Nogueira, V.B.: Machine Learning Approaches to Traffic Accident Analysis and Hotspot Prediction. Computers. 10, 157 (2021). https://doi.org/10.3390/computers10120157.
8. Amorim, B. de S.P., Firmino, A.A., Baptista, C. de S., Júnior, G.B., Paiva, A.C. de, Júnior, F.E. de A.: A Machine Learning Approach for Classifying Road Accident Hotspots. ISPRS

International Journal of Geo-Information. 12, 227 (2023). https://doi.org/10.3390/ijgi12060227.
9. Vanitha, R., M, S.: Prediction of Road Accidents using Machine Learning Algorithms, https://papers.ssrn.com/abstract=4460883, (2023).
10. Mbarek, A., Jiber, M., Yahyaouy, A., Sabri, A.: Black spots identification on rural roads based on extreme learning machine. International Journal of Electrical and Computer Engineering (IJECE). 13, 3149–3160 (2023). https://doi.org/10.11591/ijece.v13i3.pp3149-3160.
11. AlMamlook, R.E., Kwayu, K.M., Alkasisbeh, M.R., Frefer, A.A.: Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity. In: 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). pp. 272–276 (2019). https://doi.org/10.1109/JEEIT.2019.8717393.
12. Zhang, J., Li, Z., Pu, Z., Xu, C.: Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods. IEEE Access. PP, 1–1 (2018). https://doi.org/10.1109/ACCESS.2018.2874979.
13. Islam, M.K., Reza, I., Gazder, U., Akter, R., Arifuzzaman, M., Rahman, M.M.: Predicting Road Crash Severity Using Classifier Models and Crash Hotspots. Applied Sciences. 12, 11354 (2022). https://doi.org/10.3390/app122211354.
14. Santos, K., Dias, J.P., Amado, C.: A literature review of machine learning algorithms for crash injury severity prediction. Journal of Safety Research. 80, 254–269 (2022). https://doi.org/10.1016/j.jsr.2021.12.007.
15. Geyik, B., Kara, M.: Severity Prediction with Machine Learning Methods. In: 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). pp. 1–7 (2020). https://doi.org/10.1109/HORA49412.2020.9152601.
16. Mallahi, I.E., Dlia, A., Riffi, J., Mahraz, M.A., Tairi, H.: Prediction of Traffic Accidents using Random Forest Model. In: 2022 International Conference on Intelligent Systems and Computer Vision (ISCV). pp. 1–7 (2022). https://doi.org/10.1109/ISCV54655.2022.9806099.
17. Q. S. Abdullah, T., Xiao, H., Namahoro, J.P., Alfalahi, E., Al-Wesabi, I.: Exploiting Machine Learning Algorithms for Predicting Crash Injury Severity in Yemen: Hospital Case Study. Applied and Computational Mathematics. 9, 155–164 (2020). https://doi.org/10.11648/j.acm.20200905.12.
18. Lu, T., Dunyao, Z., Lixin, Y., Pan, Z.: The traffic accident hotspot prediction: Based on the logistic regression method. In: 2015 International Conference on Transportation Information and Safety (ICTIS). pp. 107–110 (2015). https://doi.org/10.1109/ICTIS.2015.7232194.