Geographic Information System And Machine Learning-Based Expressway Accident Modelling

Dilshara M.K.H. [1*], Koswatte S.[2]

[1*] Department of Remote Sensing & GIS, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka, Sri Lanka

[2] Professor in GIS, Department of Remote Sensing and GIS, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka, Sri Lanka

* dilsharakumarage@gmail.com

Abstract: *Although expressways play a significant role in the country's economy, increasing usage often leads to a rise in accidents. Research on expressway accidents in Sri Lanka has been limited, highlighting the need to investigate their causes and identify accident-prone areas proactively. This study aimed to utilise Geographic Information Systems (GIS) and Machine Learning (ML) techniques to model expressway accidents, with a specific focus on the Southern Expressway in Sri Lanka. The dataset used was collected from the Southern Expressway Operation Maintenance and Management Division (EOMMD) and underwent preprocessing, including encoding, oversampling, and feature selection. Machine Learning algorithms—Random Forest (RF), Support Vector Classifier (SVC), and Decision Tree (DT)—were used to identify accident-prone locations and assess the severity of accidents. The performance of the three models was evaluated using metrics such as the Receiver Operating Characteristic (ROC) curve, Area under the ROC Curve (AUC), Mean Absolute Error (MAE), and Mean Square Error (MSE). RF demonstrated the highest accuracy with an accuracy score of 81.19%, followed closely by SVC with 79.8% and DT with 69.7%. RF also had the lowest MAE and MSE values, and an impressive AUC value of 0.86, indicating superior prediction accuracy and strong discrimination capabilities. Maps were generated to visualise the results, and an operational dashboard was developed to facilitate data analysis and improve safety management on expressways. This study provides valuable insights into modelling expressway accidents using GIS and machine learning techniques, which can be used to enhance safety management practices and prevent accidents.*

Keywords: Expressway Accidents, GIS, Machine Learning, Road Safety Management

Introduction

Road traffic accidents have a significant societal impact, resulting in 1.35 million annual fatalities worldwide. Developing countries experience higher fatality rates due to inadequate infrastructure, enforcement, and urbanization, while developed countries have lower accident trends (WHO).

In Sri Lanka, there were 2,414 fatal accidents, 8,070 minor accidents, 6,401 critical accidents, and 5,434 accidents causing damages only reported in 2021 (Ministry of

Transport and Highways). These statistics highlight the number of lives lost in traffic accidents in the country.

With the establishment of Sri Lanka's expressway network in 2011, people have increasingly relied on expressways for their long commutes due to time, traffic, and convenience. As a result, the number of vehicles on expressways has risen, leading to a significant increase in traffic accidents in recent years (Kushan & Chandrasekara, 2020).

To maximize the economic benefits of the expressway system, Sri Lanka requires policies and measures to control and reduce the occurrence and severity of expressway accidents.The Road Development Authority (RDA) of Sri Lanka has implemented safety measures for the Southern Expressway, including median fences, guard fences, speed limits, lighting, road studs and reflectors, road markings and signs, and traffic patrols (Chinthanie & Lanka, 2015).

Between 2011 and 2018, a total of 3,724 accidents were reported on the three major expressways in Sri Lanka, resulting in 40 fatalities. The number of accidents has been increasing in recent years (Kushan & Chandrasekara, 2020).

Previous studies have identified human, environmental, and physical factors as influencing expressway accidents; some of them are described below(Table 1)(Dharmasena and Suresh, 2018 ; Chinthanie and Lanka, 2015; Kushan and Chandrasekara, 2020).

Table 1: Factors affecting accidents on Southern Expressway.

Human Factors	**Environmental Factors**	**Physical Factors**
• Drunk driving • Exceeding the speed limits • Reckless driving • Driving without seat belts • Competition between drivers	• Animal Crossing • Weather conditions:-clear, misty cloudy, rainy	• Road Geometry and design:-bends, narrow sections, street lights conditions • Vehicle defects:-failure of breaks, poor quality steering and tires,

• Lack of attention • Negligence • Fatigue • Overtaking		insufficient headlights, technical issues • Road conditions:-dry, wet, flooded with water, slippery • Traffic

Source: Previous studies

To reduce traffic fatalities and serious injuries on Highways, it is essential to review the characteristics of traffic accidents and identify hidden patterns in collision data. Machine Learning and GIS are suitable platforms for this purpose. Limited research and experience have hindered policymakers in Sri Lanka from developing effective strategies to reduce expressway accidents (Kumara & Walgampaya, 2021). This knowledge gap emphasizes the need for research and accurate data collection to inform policymaking and develop strategies that can address road safety issues. Identifying the concentration of road accidents is essential for reducing their occurrence and determining the exact reasons behind accidents leads to effective policymaking(Dharmasena and Suresh, 2018).

Accident modeling plays a crucial role in understanding contributing factors, predicting road safety conditions, and informing policymaking. Models are categorized into severity, risk, and frequency (Farhangi, Sadeghi-Niaraki, Razavi-Termeh, et al., 2021).

Accurate knowledge of the status and causes of road accidents is essential for effective policymaking and improving road safety. This research aims to utilise machine learning algorithms to identify accident-prone locations, assess accident severity, and model and identify underlying causes of accidents on the Southern Expressway to enhance safety management.

Literature Review

Somasundaraswaran (2006) highlights the increasing number of road accidents in Sri Lanka, a country with a history of significant development from the British era. The rise is attributed to the escalating number of vehicles and inadequate road infrastructure, resulting in an annual death rate of 12.1 deaths per 100,000 people.

Sri Lanka's expressway system has been a subject of significant interest in research, with studies focusing on accident patterns and injury severity. Since the first expressway was built in 2011, measures have been developed to reduce accidents and improve road safety, including stronger traffic enforcement, public awareness campaigns, and infrastructure improvements.

Senadeera (2016) and Farhangi et al. (2021) highlight the importance of accident modelling in understanding the nature and impact of contributing factors and forecasting road safety conditions. Three categories of models are severity modelling, risk modelling, and frequency modelling. Severity models illustrate the severity of incidents and identify contributing variables, while risk models measure elements that increase the likelihood of accidents. Frequency models estimate the frequency of accidents in a given area or section of road, identifying high-risk areas and implementing safety measures. Santos, Dias, and Amado (2022) highlight the severity index as a key factor in predicting accidents, assessing the level of seriousness of accidents and assigning appropriate weights to different levels of injuries and damages. This information helps researchers and policymakers develop targeted strategies to improve road safety and reduce the negative effects of accidents.

Machine learning (ML) algorithms can be categorized into supervised and unsupervised learning. Supervised learning includes classification and regression algorithms. Examples of classification algorithms used for accident analysis include decision trees (DT), random forests (RF), support vector machines (SVM), and artificial neural networks (ANN).

A literature review analyzed 56 studies on crash injury severity prediction and found that Random Forest performed best in 70% of applications, followed by Support Vector Machine (SVM) and Decision Tree. Bayesian networks and K-nearest neighbors also showed good performance in a small percentage of studies (Santos et al., 2022).

Geographic Information Systems (GIS) are widely used in accident analysis due to their ability to efficiently organize, analyze, and display spatial data. GIS allows for spatial analysis of accidents, identifying hotspots, providing visualization tools for understanding accident distribution, risk assessment, and supporting evidence-based decision-making in accident prevention and road safety planning. It also helps in identifying high-risk areas, prioritizing resources, and implementing targeted interventions to mitigate accidents. Overall, GIS plays a crucial role in accident prevention and road safety planning.

GIS technologies, such as Ordinary Kriging, Kernel Density Estimation, Inverse Distance Weighting, and Nearest Neighbor Interpolation, are widely used for hotspot analysis. However, non-spatial models struggle to visually identify hotspots and address spatial correlation issues, while spatial models can address these issues (Le, Liu, Lin, 2022).

QGIS, an open-source GIS system software, is available from reliable sources for free. It offers various methods for analyzing spatial patterns, such as hotspot analysis based on kernel density estimation (KDE). Researchers use the KDE heat map plugin in QGIS to identify hotspots, utilizing its sophisticated algorithmic nature (Nalaka et al., 2021).

Another study focused on the Colombo-Katunayake Expressway in Sri Lanka and used statistical tests to identify factors contributing to accidents. Machine learning models based on the Naive Bayes algorithm and Probabilistic Neural Network (PNN) achieved classification accuracies of 72.14% and 74.29%, respectively (Kushan & Chandrasekara, 2020).

A study in Qazvin province, Iran, aimed to model accidents caused by driver drowsiness using decision tree (DT), random forest (RF), and support vector regression (SVR) algorithms in a GIS environment. The RF model showed the best overall performance, followed by SVR and DT (Farhangi, Sadeghi-Niaraki, Nahvi, et al., 2021).

Decision trees are fast and easy to use for classification tasks. Random Forest is an ensemble algorithm that combines multiple decision trees to make predictions. Support Vector Machines find optimal hyperplanes to separate different classes of data.Data preprocessing steps include cleaning, handling missing values, and treating outliers. Techniques like random under-sampling, oversampling, or hybrid sampling can address class imbalance. Feature selection helps choose informative features, and hyperparameter tuning optimizes model performance.

Farhangi et al. (2021) evaluated tree-based machine learning algorithms, including Bagged Decision Trees (BDTs), Extra Trees (ETs), and Random Forest (RF), for accident risk mapping caused by driver lack of alertness. The study used 159,735 road traffic accidents from the Iran Road Maintenance and Transportation Organization (IRMTO). Results showed that BDT and RF models had slightly higher accuracy, with an AUC of 0.846 compared to the ET model. The study emphasizes the importance of traffic volume data in improving accident-risk mapping models.

The study by Al-Radaideh and Daoud (2018) explores the use of data mining techniques to predict traffic accident severity. The study used three classification techniques: Decision trees, Artificial Neural Networks (ANN), and Support Vector Machines (SVM). The Random Forest technique achieved the highest accuracy rate of 80.6%, followed by ANN at 61.4% and SVM at 54.8%. This research highlights the effectiveness of data mining methods in predicting accident severity. The generated model was used to develop a decision system, assisting decision-makers in predicting accident severity and improving decision-making processes. The study provides valuable insights into traffic accident analysis and prevention(Al-Radaideh and Daoud, 2018).

The study (Zhang, Waller and Jiang, 2020) presents a framework for modeling traffic crash frequency using ensemble machine learning (EML) techniques. The framework includes Random Forest (RF) and Extra-random Trees (ERTs) averaging methods, as well as AdaBoost (AdaB) and Gradient Boosting (GTB) boosting methods. The models are optimized through k-fold cross-validation and sensitivity tests, and the Gini diversity index is used to assess the importance of features in crash frequency analysis.

The study aimed to develop an improved prediction model for avoiding road collisions in Bangladesh using machine-learning algorithms. Among the examined algorithms, Decision Tree achieved the best results, achieving an impressive accuracy of 99.77% for accident prediction and 99.80% for severity prediction, with F1 scores of 98.68% and 99.80%, respectively(Paul et al.,2020) .

Geyik and Kara (2020) used data mining classification algorithms to predict traffic accident injuries severity levels. Data were extracted from UK traffic accident data from 2010 to

2012. The study used algorithms like Multi-layer Perceptron (MLP), Decision Tree, Random Forest, and Naive Bayes. The results showed that the decision tree algorithm achieved the highest accuracy at 86.67%. The study used Python programming and the Spyder IDE. The severity attribute was categorized into fatal, serious, and slight levels .

The study developed an integrated approach that uses the kernel density estimation (KDE) algorithm and spatial autocorrelation analysis to identify and evaluate traffic accident hotspots. The study analyzed data from Hanoi, Vietnam, from 2015-2017, aiming to improve accident patterns and optimize traffic authorities' resource allocation. The GIS-based KDE algorithm successfully identified hotspots, and Moran's I statistic indices assessed their significance. The approach also proposed a validation process to prioritize hotspots based on severity. The findings provide a robust methodology for identifying statistically relevant traffic accident hotspots, enabling proactive management and resource allocation(Le, Liu and Lin, 2022).

The study evaluated various GIS interpolation techniques to identify accident hotspots on the Southern Expressway in Sri Lanka. Data from 966 accident records from 2015 to 2017 was used. The interpolation methods included Ordinary Kriging, Kernel Density Estimation (KDE), Inverse Distance Weighting (IDW), and Nearest Neighbour Interpolation. Multiple hotspots were detected in the right lane, but the PAI values were slightly higher in the left lane. The IDW and KDE algorithms were found to be the best-fit models for demarcing hotspot locations. These findings can help improve road safety measures and decision-making processes for mitigating accidents on the Southern Expressway(Nalaka et al., 2021).

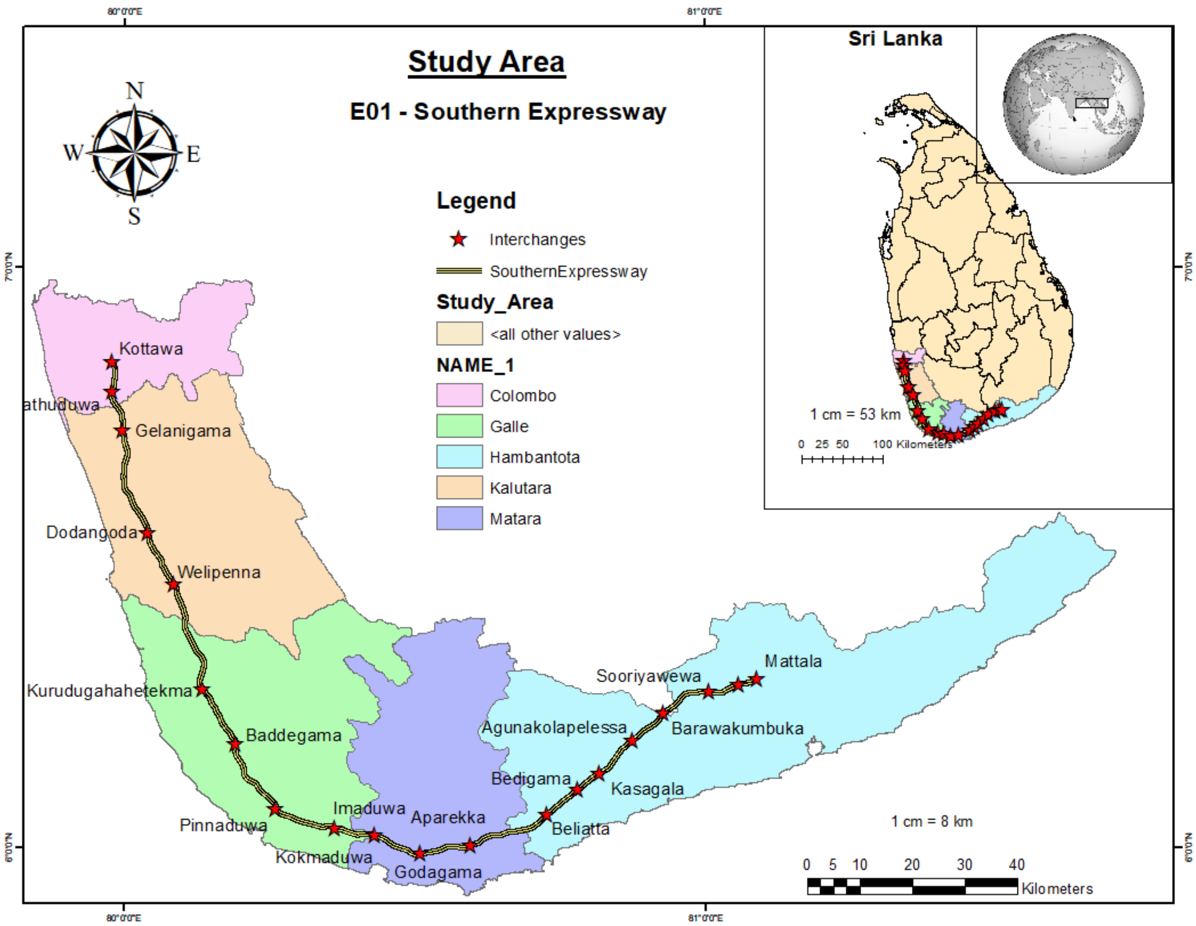
Droj et al (2022) conducted a GIS-based survey to explore public transport strategies for sustainable urban traffic planning, using ArcGIS Online for data collection, including real-time traffic information from the Living Atlas .

Feizizadeh et al. (2022) conducted a GIS-based spatiotemporal modelling of urban traffic accidents in Tabriz City during the COVID-19 pandemic. The study found that integrating GIS analysis, Comap, and Sever Index efficiently analyzed hotspot patterns of road accidents. The Global Moran's index and Kernel Density Estimation were used to develop the spatiotemporal pattern of accidents. The study highlights the effectiveness of GIS-based techniques in analyzing traffic accident patterns during the pandemic.

Considering the aforementioned research findings, it is evident that developing countries are prone to a higher incidence of accidents, and Sri Lanka, being a developing nation, is no exception. Consequently, the necessity for a robust safety management system becomes imperative. Despite possessing abundant expertise and resources, the dearth of studies about road safety management poses a significant challenge. Therefore, harnessing the potential of AI technology for conducting such analyses emerges as a valuable solution. Through this study, aimed to contribute towards bridging this existing gap and making a meaningful impact.

Methodology

In this research methodology for GIS and machine learning-based expressway accident modelling, several key steps were followed. Firstly, an extensive literature review was conducted to gather comprehensive knowledge about road traffic accidents and machine learning algorithms. This review helped identify the factors that significantly impact accident occurrences. The research gap and scope were determined by the findings of previous studies. In the second step, a spatial database was created from accident points and effective criteria in the study area. In the third step, the data preprocessing method was used to prepare the data. In the fourth step, accident severity was performed using DT, RF, and SVR data mining algorithms. In the fifth step, the models were evaluated using the ROC curve and AUC, MAE, and MSE metrics. By using prediction results, maps were created for each model. ArcGIS Online (AGOL) operational Dash Board were created relevant to chainage on southern expressway.



Source: by Author

Figure 1: Study Area.

Southern Expressway is the first link of Sri Lanka's expressway network, travelling from Kottawa to Mattala (200.451 km). It has twenty interchanges and a 4-lane capacity, with a maximum operating speed of 100kmph (Figure 1).

Data were taken from the Southern Expressway Operation Maintenance and Management Division (EOMMD) for this study, which covers 2018 January through 2022 August. 5070 accidents were reported during this period, and after preprocessing, 4435 observations were used in the analysis. The Southern Expressway road layer was digitized through Google Earth Pro with 10meter accuracy.

Table 2: Effective Criteria.

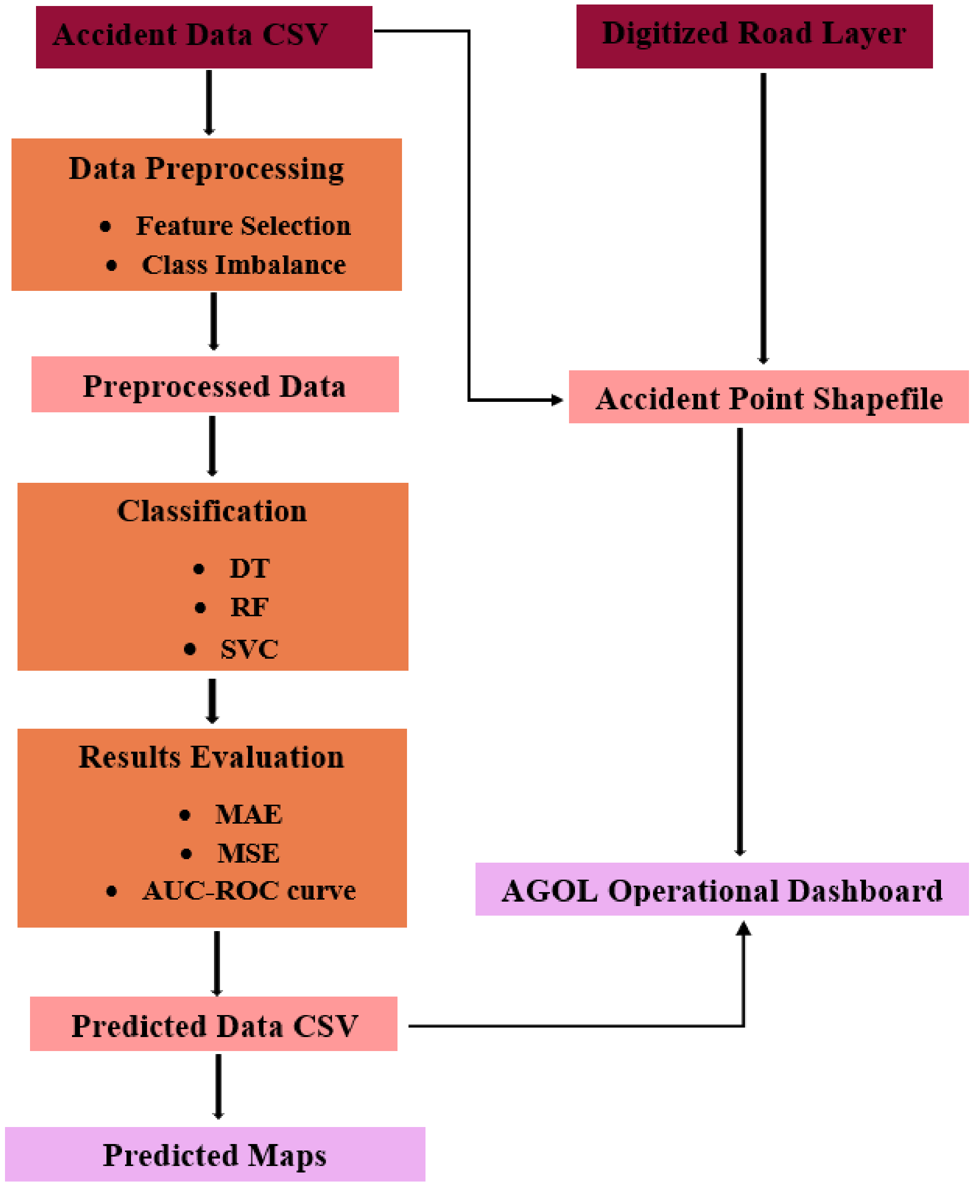
Variable	Values	Description
Month	January -December	Months of the year
Day	Monday -Sunday	Days of the week within the weekend
Time Period	(0-6) h and (20-24)h (6-9)h (9-16)h (16-20)h	Night Morning Peak Day Evening peak
Location km post	0-125 km	Round off chainage values
Direction	LHS RHS	Mattala to Kottawa Kottawa to Maththala
Damage Level	F NG PDO	Fatal Non-Grievous Property damage only
Reason	Human Factor NEG FAT OVT SS Physical Factor MF OFV OOS	 Negligence Fatigue Overtaking Sudden Stop Mechanical Fault Obstacles from front vehicles Obstacles from outside

	OOR PC SH TP FIRE Environmental Factor AC RW	Physical condition Stone Hitting Tire Punch Fire Animal Crossing Rainy Weather
Vehicle	Bowser Bus Cab Car Jeep Lorry Van Unknown	Vehicle type
Time of Day Description (Light Condition)	Dawn Dusk NGSL NNSL DL	Dawn Dusk Night with Good Street Light Night without street Light Daylight

Source: Southern Expressway Operation Maintenance and Management Division (EOMMD)

Table 2 shows effective criteria relevant to the study.

Several software packages were used in the data analysis and visualization process. The process involved the utilization of Google Earth Pro, QGIS Desktop 3.22.3 Version, ArcGIS Online, and ArcGIS. For executing code and facilitating data analysis and visualization, the Jupyter Notebook IDE was employed under the Anaconda Distribution platform. The data analysis tasks made use of the SciKit-Learn library. The computational resources consisted of machines equipped with an Intel i7 8th generation processor, 8 GB of RAM, and a 64-bit Windows 10 operating system.



Source: by Author

Figure 2: sketch of methodology.

Figure 2 shows the sketch of the methodology.

Data pre-processing is crucial in machine learning modelling, transforming raw data into an efficient format for accurate models. The study extracted data from Excel format, but found it disorganized and unstructured, reducing efficiency. To improve data quality, a comprehensive approach was adopted, including data cleaning, handling missing values, treating outliers, and encoding. These techniques ensure data purity and reliability before utilizing it in model building. The selected data was saved in CSV format after pre-processing.

The data was imported into QGIS using CSV format and a separate vector file containing the road layer. Two QGIS plugins, LRS (Linear Referencing System) and Q Chainage, were used to integrate the accident points onto the road layer.

Firstly, imported essential libraries for data manipulation, encoding categorical variables, hyperparameter tuning, data splitting, and addressing class imbalance.These libraries, including pandas, category_encoders, LabelEncoder, RandomizedSearchCV, train_test_split, and SMOTE .

The features and the target variable were separated from a data frame.The features were assigned to the variable 'X', while the target variable was assigned to the variable 'y'.All categorical values were converted to numerical values by using the LabelEncoder class.Class imbalance in the data set was reduced by applying SMOTE algorithm.20% of the data would be allocated for testing, while the remaining 80% would be used for training.

The Decision Tree Classifier class was imported from sklearn.tree module and hyperparameters were defined for the random search method. The best parameters were identified and the decision tree classifier was trained with them. The model was predicted using the 5-fold cross-validation method and a heat map was created with predicted values and locations.

After importing the Random Forest Classifier, the above steps were followed. Additionally, the Support Vector Classifier also underwent the same steps.

sklearn.metrics,sklearn.model_selection,matplotlib.pyplot, numpy, and scipy.interpolate were imported for evaluating models, analysing data, and generating visualisations. To define parameters, the parameter grid was used for each model.

The process of creating an operational dashboard involved generating chainage coordinate values through the field geometry calculator in QGIS and importing them into the ArcGIS Online platform as a feature layer. This enabled the development of an operational dashboard, which was a powerful tool for visualizing severity levels in conjunction with their respective locations.

Results and Discussion

Data manipulations revealed that there were numerous errors, particularly related to missing values. Three different techniques were employed to tackle the class imbalance in the dataset, but only the oversampling technique yielded improved results.

The dataset consists of categorical values with unknown relationships, so one hot encoding was used for the categorical features and label encoding for the target variable, increasing dimension.

The results of each model are described below.

The Parameter Grid used for the Decision Tree (DT) model included parameters such as max_depth, min_samples_split, min_samples_leaf, and max_features. These parameters were systematically varied to find the optimal configuration for the DT model. The best parameters were determined to be max_depth = None, min_samples_split = 10, min_samples_leaf = 1, and max_features = sqrt.

According to the evaluation of the importance of effective criteria in the DT model Vehicle type was the most important criterion, followed by reason (negligence, fatigue, and animal crossing) and light condition (night without street lights). Both directions have some probability of accidents.

The parameter grid used for tuning the Random Forest (RF) model included values for n_estimators ranging from 100 to 300, max_depth options of None, 3, 5, 7, and 10, min_samples_split and min_samples_leaf values of 2, 3, 5, 7, and 10, max_features options of 'auto','sqrt', and 'log2', bootstrap choices of True and False, and criterion options of 'gini' and 'entropy'. The optimal values were determined to be n_estimators=300, min_samples_split=10,min_samples_leaf=2,max_features=sqrt, max_depth=None, criterion=entropy, and bootstrap=True. The Random Forest model has found that vehicle type, reason for accident (negligence,animal crossing) ,lighting condition(Night time without street light) and time of day(morning and evening peak time) were important factors in accident prediction.

The SVC model used a parameter grid of C, kernel, gamma, degree and C. The best parameters were found to be kernel: rbf, gamma: 1, degree: 4, and C: 100. By selecting the rbf kernel, setting the degree to 4, and choosing C as 100, the model achieved improved results in classification tasks.

Evaluations of Results are described below.

Table 3: Accuracy of the model.

Model	**MAE**	**MSE**	**Accuracy(%.)**
DT	0.2604	0.2789	69.7
RF	0.1521	0.1589	81.19
SVC	0.1797	0.1924	79.8

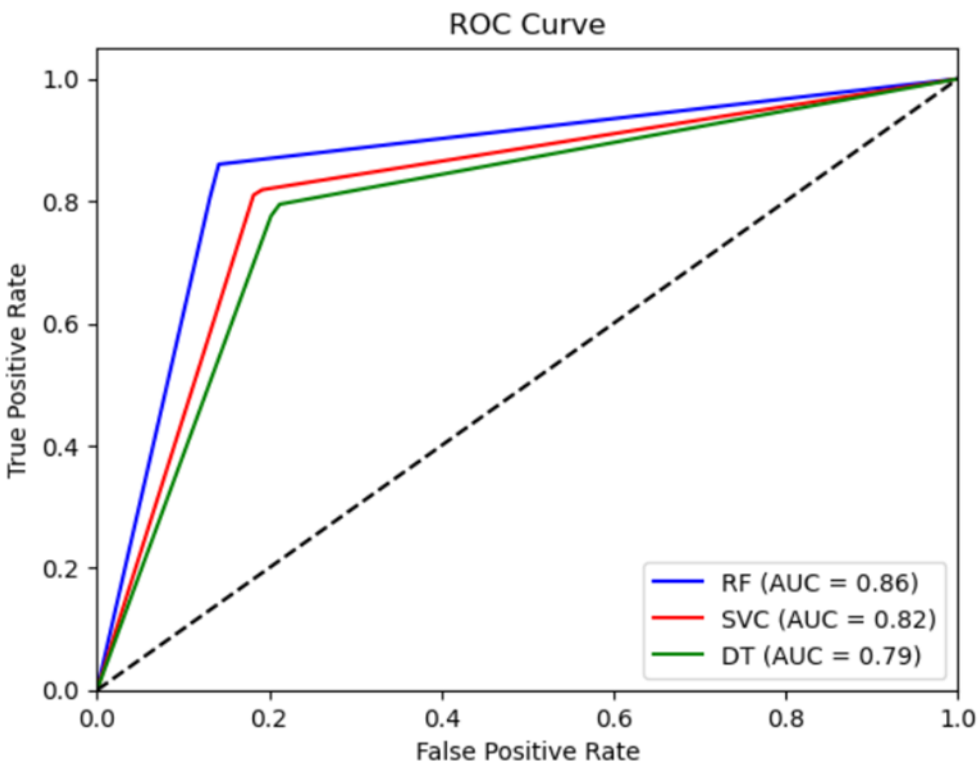
Source: by Author

When comparing the three models (Table 3), it was observed that RF achieved the highest accuracy score of 81.19%, followed by SVC with 79.8% and DT with 69.7%. Additionally, the mean absolute error (MAE) and mean squared error (MSE) were considered as evaluation metrics. The models were assessed using a 5-fold cross-validation method(Table 4).

Table 4: Cross Validation results.

Fold	DT	RF	SVC
1	0.7148	0.7435	0.7813
2	0.7054	0.7693	0.7967
3	0.6710	0.6459	0.7746
4	0.6631	0.7709	0.7968
5	0.7205	0.8411	0.8202

Source: by Author



Source: by Author

Figure 3: ROC Curve Results.

The discrimination capabilities of the models were evaluated using the area under the receiver operating characteristic curve (AUC)(Figure 3). Random Forest (RF) demonstrated the highest performance with an AUC of 0.86, followed by Support Vector Classifier (SVC) and Decision Tree (DT). RF had the highest accuracy and lowest MAE and MSE values, making it the best performance when taking accuracy, MAE, and AUC values into account.

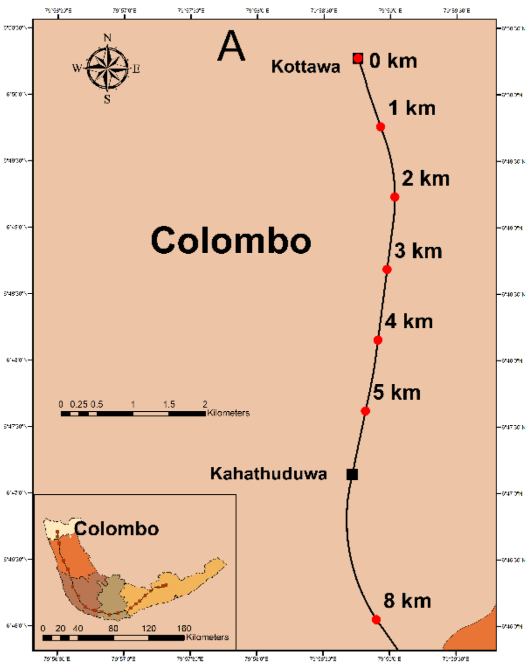
By analyzing actual database results and prediction results relevant to each model, accident-prone locations, their range, prominent reasons, and severity were identified (Table 5 , Table 6 ,Figure 4 , Figure 5 and Figure 6).

Table 5: Accidents Prone Locations with Reasons and Severity Level.

Location km post	Count	Reasons percentage	Severity Level Percentage
44-45	84	NEG: 56%	PDO: 94.05%
		AC: 13 %	NG: 3.57%
		SS: 11%	F: 2.38%
		OVT: 6%	
		OTHERS 14%	
4 - 5	75	NEG: 31%	PDO: 90.67%
		FAT: 16%	NG: 8.00%
		AC: 15%	F: 1.33%
		OOS: 13%	
		OVT: 11%	
		OTHERS 14%	
89 -90	72	AC: 28%	PDO: 97.22%
		FAT: 19%	NG: 2.78%
		RW: 13%	
		HR: 10%	

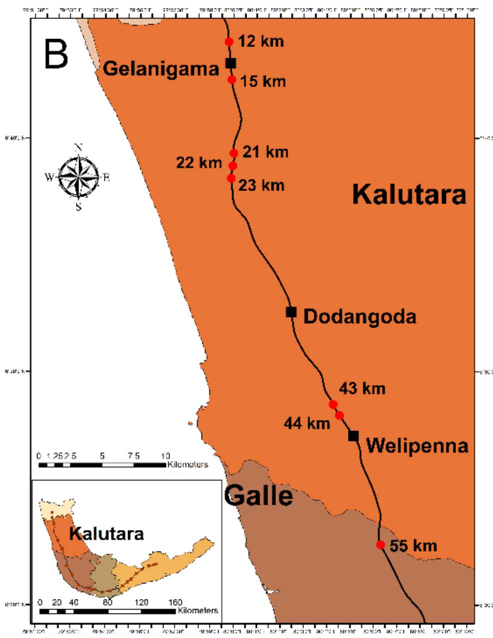
		OTHERS 30%	
1 – 2	68	NEG: 34%	PDO: 85.29%
		OVT: 15%	NG: 14.71%
		OOS: 12%	
		FAT: 10%	
		TP: 9%	
		OTHERS 20%	
2 - 3	64	NEG: 32%	PDO: 89.06%
		OVT: 16%	NG: 7.81%
		FAT: 13%	F: 3.12%
		OOS: 11%	
		AC: 10%	
		OTHERS 18%	

Source: by Author



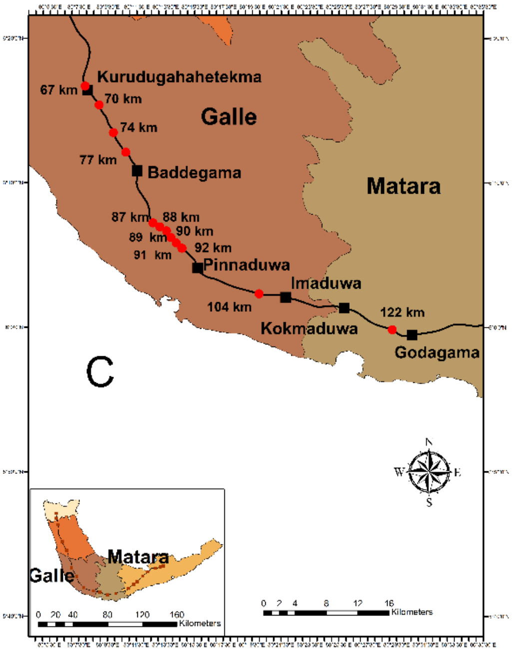
Source: by Author

Figure 4: . Accidents Prone Locations



Source: by Author

Figure 5: . Accidents Prone Locations.



Source: by Author

Figure 6: . Accidents Prone Locations.

Table 6: Structural and Spatial Characteristics in Identified Locations.

Location km	**Structural and Spatial Characteristics in Identified Locations of Southern Expressway**
0,1,2,3,5	Road section curve , Green surroundings open space ,Low variation of solid-voids ,Lack of variations of form ,Significant elevation
8,12,15	Narrow road section with curves, Green surroundings open space,Low variation of solid-voids ,Lack of variations of form, mixed with enclosed and open spaces.
21,22,23	Narrow road section with, Green surroundings mixed solid rock surfaces, lack of variations of forms, mostly includes lower proportion sections with open spaces.
43,44	Narrow road section curve, Green surroundings mixed rock surfaces, lack of variations of forms, composition of enclosed and open areas
55	Road sections with more curves, Green surroundings, lack of variations of forms, includes narrow road sections, lower proportion and no variation of solid- voids towards the Matara lane. Enclosed spaces towards to Kottawa lane.

67,70	Narrow road sections with curve, Green surroundings ,Hard solid rock surfaces towards Kottawa lane, composition of enclosed and open spaces.
87,88,89,90,91,92	Narrow road sections with curve , Green surroundings, lack of variations of forms, mostly includes lower proportion sections, More Open spaces towards the Matara lane, Hard solid rock surfaces towards Kottawa lane, Random void spaces towards the Matara lane(Animal crossing)
104	Road sections with curve, Green surroundings
122	Road sections with curve, Green surroundings

Source: by Author

Note: The structural and spatial characteristics of the identified locations were considered within a radius of 1 km.

The study revealed that the left lane and weekends had higher accident severity, with increased severity during January, March, and April. Cars were the most involved vehicles, and accidents were more frequent during peak time periods. Specific sections, such as Galanigama-Dodangoda, Welipenna-Kurudugahahetekma, and Baddegama-Pinnaduwa, experienced a higher frequency of accidents. Heat maps effectively visualized the accident distribution, and the Random Forest (RF) model outperformed the Decision Tree (DT) and Support Vector Classification (SVC) models. Major factors contributing to accidents were identified as human negligence, fatigue, and animal crossings, emphasizing the need to address these issues for road safety planning.

The concentration of accidents was observed in road sections with bends and narrow widths, particularly between 54 and 55 kilometers. The combination of narrow shoulders and bends increased the risk of crashes, especially at high speeds. Certain sections were designated as critical areas, marked with warning signs, to mitigate the effects of bends and narrow road sections during nighttime or rainy conditions. However, the limited coverage of anti-reflective boards and the absence of a continuous street light system on the Southern Expressway posed additional risks, particularly at night and during rainy weather. The buffer line on the expressway was not wide enough to provide sufficient safety margins in emergencies or breakdown situations.

The driving behavior of inexperienced drivers, particularly those recently obtaining their licenses, and certain driving behaviors of female drivers, such as signaling, turning, and sudden stopping, were observed as contributing factors to accidents. The presence of drunk drivers also posed a significant risk. Excessive vehicle load, unfamiliarity with exits and entrances, and fatigue-related driving without taking breaks were common causes of expressway accidents. Addressing these factors and implementing appropriate measures are crucial to improving expressway safety.

Conclusion and Recommendation

This comprehensive study focused on analyzing and modeling accident risk on Sri Lanka's Southern Expressway using advanced machine learning techniques. Decision Tree (DT), Support Vector Classifier (SVC), and Random Forest (RF) algorithms were utilized, with RF demonstrating superior accuracy and reliability in predicting accident risk compared to DT and SVC. The study also identified key factors contributing to accidents, such as driver-related factors, adverse weather conditions, negligence, fatigue, animal crossings, overtaking, and tire punctures. Practical solutions and policy implications were derived from these findings to enhance expressway safety and mitigate the severity and occurrence of accidents. The developed risk models, insights into accident causes, and visualization tools contribute to the effective management and identification of accident-prone areas on the Southern Expressway and similar road networks.

Recommendations:

1. Implement policies targeting negligence, fatigue/drowsiness, drunk driving, and overtaking to reduce expressway accidents.

2. Install speed tracking systems and enforce limits using mobile police units to promote safer driving speeds.

3. Improve road characteristics by addressing bends, narrow sections, anti-reflective boards, and street lighting to prevent accidents and minimize their impact.

4. Utilize the Survey 123 application for data collection and modeling to identify accident causes and factors. Apply findings to enhance safety management on other expressways.

5. Enhance accident modeling and safety management by utilizing the Survey 123 application, real-time operational dashboards, and considering external factors.

References

Chinthanie, R. P. D., & Lanka, S. (2015). Accident Analysis of Southern Expressway. September.

Dharmasena, S. R., & Suresh, E. A. T. (2018). A methodology to analyze road landscape in accident Black-Spots: The case of Southern Expressway, Sri Lanka. Archnet-IJAR, 12(2), 347–357. https://doi.org/10.26687/archnet-ijar.v12i2.1547

Farhangi, F., Sadeghi-Niaraki, A., Nahvi, A., & Razavi-Termeh, S. V. (2021). Spatial modelling of accidents risk caused by driver drowsiness with data mining algorithms. Geocarto International, 0(0), 000. https://doi.org/10.1080/10106049.2020.1831626

Farhangi, F., Sadeghi-Niaraki, A., Razavi-Termeh, S. V., & Choi, S. M. (2021). Evaluation of tree-based machine learning algorithms for accident risk mapping caused by driver lack of alertness at a national scale. Sustainability (Switzerland), 13(18). https://doi.org/10.3390/su131810239

Kumara, S. D. D. R., & Walgampaya, C. K. (2021). Identification of Severity Factors and Risk Areas of Southern Expressway Accidents. Engineer: Journal of the Institution of Engineers, Sri Lanka, 54(3), 61. https://doi.org/10.4038/engineer.v54i3.7460

Kushan, M. A. K., & Chandrasekara, N. V. (2020). Identification of factors and classifying the accident severity in Colombo - Katunayake expressway, Sri Lanka. Proceedings - International Research Conference on Smart Computing and Systems Engineering, SCSE 2020, September, 166–173. https://doi.org/10.1109/SCSE49731.2020.9313036

Santos, K., Dias, J. P., & Amado, C. (2022). A literature review of machine learning algorithms for crash injury severity prediction. Journal of Safety Research, 80, 254–269. https://doi.org/10.1016/j.jsr.2021.12.007

Somasundaraswaran, A. (2006). ACCIDENT STATISTICS IN SRI LANKA. IATSS Research, 30(1), 115–117. https://doi.org/10.1016/s0386-1112(14)60162-x