Intelligent Road Safety Navigation in Sri Lanka: A Review of Machine Learning Techniques and Proposal of a Model for Predicting Accident Hotspots and Severity

JMCT Jayamanna
Faculty Of Computing
General Sir John Kotelawala Defence University
Sri Lanka
38-bcs-0012@kdu.ac.lk

Pradeep Kalansooriya
Faculty Of Computing
General Sir John Kotelawala Defence University
Sri Lanka
pradeepkalansooriya@kdu.aclk

***Abstract* - Both public safety and the stability of the economy are seriously threatened by traffic accidents. Like many other regions, Sri Lanka is faced with the challenge of road accidents, which hinder the path to safer roads. One of the reasons that 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. In order to decrease road accidents, people must change their driving patterns. While looking for a solution to that problem, existing studies around the world have proposed predictive machine learning models for accident-prone locations known as hotspots and severity levels. But there were not any existing studies that proposed a solution that was suitable for Sri Lanka. This research seeks to address these challenges by conducting an in-depth review of existing machine learning techniques and proposing the most suitable model approaches for prediction of accident hotspots and severity in Sri Lanka based on the availability of the accident data. One of the main objectives is to identify the correct machine learning techniques. According to studies, 81% used the 'Random Forest' algorithm, which is a supervised machine learning algorithm for the prediction. And Random Forest performed better in approximately 69% of the studies. And this research is not just proposing suitable model approaches for predictions. It provides the foundation to revolutionize road safety through the development of an intelligent road safety mobile navigation application.**

Keywords - Predictive Models, Machine Learning (ML), Hotspots Identification, Severity Analysis

I. INTRODUCTION

In recent decades, the number of vehicles on the 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. Children who suffer injuries or pass away accidentally have a significant effect on society and families.

The World Health Organization (WHO) estimates that up to 3,700 people per day are killed in automobile accidents, accounting for around 1.19 million annual fatalities. [1] There are an additional 20 to 50 million non-fatal injury cases, many of which result in disability. According to statistics from the Ministry of Transport and Highways, in 2022, Sri Lanka recorded 21,953 road accidents.[2] 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 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 reducing the number of traffic accidents that occur every day. The government has already enforced other tactics, like traffic signals and fines, to reduce these incidents, but they have been ineffective. But in order to decrease accidents, people must change their driving patterns.

This research seeks to address these challenges by proposing an in-depth review of existing machine learning techniques for road accident hotspots and severity prediction in Sri Lanka. The primary objective is to identify the correct machine learning techniques for predicting accident hotspots and severity and provide the most suitable model approaches for the predictions.

Going beyond theoretical modeling, this study uses the insights from the evaluation of these predictive models as the foundation for an intelligent road safety mobile navigation application. This application is designed to enable immediate responses and interventions, enhancing road safety practices in Sri Lanka.

The rest of this paper is organized as follows: Section II introduces some previous work and existing machine learning techniques for road accident hot spot prediction. Section III describes the methodology. Section IV introduces the results of this study through discussion. Sections V and VI give the conclusion and future work.

II. LITERATURE REVIEW

A. *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,

A study by Jayesh Patil et al. aims to develop an application connected to maps that gives a user a clear idea of accident-prone areas, passes an alert about road conditions, and suggests certain ways to be safe. The goal of this study is to lower the mortality rate by developing a prediction model using the unsupervised learning technique known as k-means clustering, which analyzes traffic accidents by taking into account various factors such as potholes in the road, sharp turns, and weather conditions, and then offers appropriate and preventative measures to avoid accidents by representing them on a map and building an understandable model for everyone. The accuracy of the anticipated model was 81% [3] .

In a paper by Aklilu Elias Kurika et al., the authors discussed how predictions of vehicular accidents were determined using a machine learning system.The data for the projection was taken from Wolaita Zone's 12 districts and three municipal administrations. The experimental findings, model assessment, and performance measurement revealed that the J48 and Rep tree classifiers' F-measure was equivalent (decision tree classifiers (J48, Random Forest, and Rep Tree)), while the Random Forest tree performed poorly. Based on performance, the J48 tree was determined to be the best model [4].

A research study by Teres Augustine et al. proposes an accident prediction system that can help analyze potential safety issues and predict whether an accident will occur or not. A comparative study of various machine learning algorithms was conducted to check which model could help predict accidents more accurately. The dataset used for this paper is the government record of accidents that occurred in a district in India. The machine learning models utilized in this study to predict accidents are Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbor, XGBoost, and Support Vector Machine. The Random Forest algorithm gave the highest accuracy of 80.78% when the accuracies of the machine learning models were compared [5].

Mahendra G et al.'s article forecasts traffic accidents by taking into account four factors: location, weather, road type, and collision type. A machine learning model that combines a decision tree and random forest regressor is being created and is currently in use to forecast collisions based on collision statistics from various Indian states. The Random Forest Regressor (RFR) model outperforms the Decision Tree Regressor (DTR) model, according to the results of the relative analysis.[6]

Daniel Santos et al.'s article seeks to create models that can identify a subset of important variables that can be utilized to categorize an accident's severity in order to enable a study of the accident data. Furthermore, based on historical data, this study suggests a predictive model for upcoming traffic incidents. Both supervised and unsupervised machine learning approaches such as hierarchical clustering and DBSCAN are employed, along with supervised machine learning strategies including decision trees, random forests, logistic regression, and naive Bayes. The predictive model's output indicates that the RF model might be a helpful tool for predicting accident hotspots.[7]

Study by Brunna de Sousa Pereira Amorim et al., Several machine learning methods were used in order to determine which feature combination would yield the best classifier for Brazilian federal road hotspots linked to either a severe or non-severe accident probability. SVM, random forest, and a multi-layer perceptron neural network were used for testing. A ten-year report on traffic accidents from the Brazilian Federal Highway Police is included in the dataset. The outcomes of the analysis are the following features set including as aspects of road types, the route, the direction, weather conditions, types of the accidents, spatial extent of the accident, day of the week, and the hour of the day when the accident occurred. Hence, the proposed neural network model gave the desired outcome with the accuracy: Accuracy: 83%, precision: 84%, recall: 83%& F1 Score: 82%. The results were encouraging. The results were encouraging. [8]

Study by R. Vanitha et al., They employed machine learning approaches, such as Decision Tree, Random Forest, and Logistic Regression, to build an accident prediction model. These classification schemes will be useful for both accident prediction and safety measure creation. 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 86.86% 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 weighted severity index (WSI), which in turn produces the severity scores to be allocated to specific road segments, is generated using the XGBoost model. After that, the latter are categorized using a four-class system (high, medium, low, and safe). The results of the simulation demonstrate that the suggested framework surpassed the reliable competing models, particularly in terms of accuracy, which was 98.6%, and correctly and efficiently recognized the black areas.[10]

Study by Rabia Emhamed et al., investigates models in order to identify a subset of significant variables and construct a model for categorizing the severity of injuries. These models are created using a variety of machine learning methods. Examples of the supervised machine learning algorithms like AdaBoost, Logistic Regression (LR), Naive Bayes (NB), and Random Forests (RF) constructed from the data of traffic accidents. When the data sets are unbalanced, a particular method known as the SMOTE method is followed. Based on the results of the study, this paper has revealed that RF model has the potential of being used to forecast on the severity of the accidents. As far as the accuracy is concerned, the RF is higher than the accuracy of LR (74. 5%), NB (73. 1%), AdaBoost (74. 5%). Accuracy of the RF algorithm is 75. 5%..[11]

Study by Jian Zhang et al., In this study, the predictive performance of various machine learning and statistical approaches with various modeling logics for crash severity analysis is compared in terms of variable importance estimation and prediction accuracy. Data on traffic flow, road shape, and crash severity were collected in Florida's highway diverge regions. The ordered probit (OP) model and the multinomial logit (MNL) model, two of the most widely used statistical models, were computed in addition to four widely used machine learning models: Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), and K-Nearest Neighbors (KNN). 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]

Study by Md. Kamrul Islam et al., In order to forecast the severity of road traffic crashes (R.T.C.s.), this study contrasts a logistic regression model with tree-based ensemble methods (random forest and gradient boosting). Random forest (R.F.) is used to identify significant traits that are significantly correlated with the severity of the R.T.C.s. The investigation's findings showed that the type of collision and its cause are the two main variables affecting how seriously people are injured in traffic accidents. The random forest (R.F.) strategy fared better than other models when using k-fold (k = 10) in terms of injury severity, individual class accuracies, and collective prediction accuracy, based on several performance metrics. [13]

Research review by Kenny Santos et al., provides a current assessment of the research on the techniques used in injury severity modeling from traffic crashes. 56 research that span the years 2001 to 2021 and take into account more than 20 distinct statistical or machine learning methods are included. Findings: Of all the algorithms tested, Random Forest produced the best outcomes in 29% of the studies and 70% of the applications. In 53% and 31% of cases, as well as in 16% and 14% of all trials, Support Vector Machine and Decision Tree had the best performance, respectively. Although they were employed 67% and 40% of the time, respectively, Bayesian Networks and K-Nearest Neighbors performed best in only 4% and 7% of the studies that were examined. [14]

Study by Buket Geyik et al., uses prior knowledge as training data to classify data objects into groups, which is useful for us to work with. The models that we used are Multilayer Perceptron (MLP), decision tree classifier, random forest classifier, Naïve Bayes classifier. Three categories of accident severity were applied to the dataset: These include minor infection, serious infection and death. The goal of the study is to establish models to predict the accident severity levels of traffic accident injury records for potential accidents by using some data mining classification methods. The findings of the evaluated classification methods show that the MLP model has an accuracy of 86.67%, the decision tree algorithm has an accuracy of 80.74%, the random forest classifier has an accuracy of 85.19%, and the Naive Bayes strategy has an accuracy of 83.40%.[15]

Study by Imad EL MALLAHI et al., concentrate on predicting the severity of traffic accidents, as this is a critical step in managing accidents on the road. In order to categorize and forecast the severity of traffic accidents, the Random Forest, Support Vector Machine, and Artificial Neural Network methods are used in this suggested study and compared. They divide the severity prediction of traffic accidents into three classes using accident data: pedestrian, vehicle or pillion passenger, and driver or rider. At the precision recall level, they have 93.82% compared to 82.22% for SVM and 87.88% for ANN, and they have 93% accuracy for random forest. [16]

Study by Tariq Al-Moqri et al., Their main goal is to evaluate the role of the main contributors to the severity of injuries. Traditional statistical methods were contrasted with the chosen machine learning algorithms. Three types of injury severity were identified in the accident data: Severe, Serious, and Minor. A synthetic minority oversampling approach (SMOTE) was used to achieve balance. The five machine learning classifiers, Naïve Bayes (NB), J48 Decision Tree, Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP), were compared with the Multinomial Logit Model (MNL).With an accuracy rate of 94.84%, RF is the best classifier out of five.[17]

B. *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 Performin g ML Technique**	**Predicti on 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

III. METHODOLOGY

This section describes the research design, study population, data collection methods, model development.

Quantitative and qualitative approaches will be used for this research, and the Mount Lavinia region of the Colombo district of the western province of Sri Lanka will be the study population.

A. *Data collection*

The data collection methods will be past police accident reports and interviews. The past police accident reports will be used to collect road accident data. The interviews will be used to collect ideas and thoughts from the officers of the police stations and relevant authorities.

Based on the analysis of the data that were collected from the existing studies, the accident data is categorized as given below in table 2 and will be hoped to collect in partnership with the mount Lavinia police station.

Table 2 Accident data

Road accident locations coordinates (latitude and longitude)
Date (Day of the week)
Time (Hour of the day)
Accident location severity levels (fatal, critical, minor, damages only)
Weather condition
Vehicle types involved
Number of people injured
Age of the Driver
Gender of the Driver
Reason for the accident
The total number of reported road traffic accident cases, deaths (fatal), critical, minor, and damages only, for the years 2020, 2021, 2022, and 2023

B. *ML Model Development*

:

1) *Proposed model solution for prediction of accident hotspots*

According to the collected data and the critical review of the existing research papers, the following proposed model will be use for the prediction of accident hotspots.

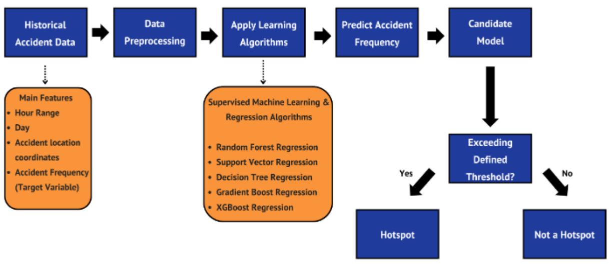

Fig. 1 Proposed model for prediction for hotspots.

The goal of accident frequency prediction 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. By comparing the accident frequency with the defined threshold value, accident hotspots can be predicted.

As shown in Fig. 1, 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 that is shown in Tables 6 and 7, hour-range, day, coordinates of the accident location, and accident count were the features that were used for this hotspot prediction.

Right after the selection of the features for the prediction, the data preprocessing will be conducted. Cleaning and preparing data for a machine learning model requires data preprocessing, which also increases the model's efficacy and accuracy.

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 shown in Table 1, 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.

2) *Proposed model solution for prediction of accident severity*

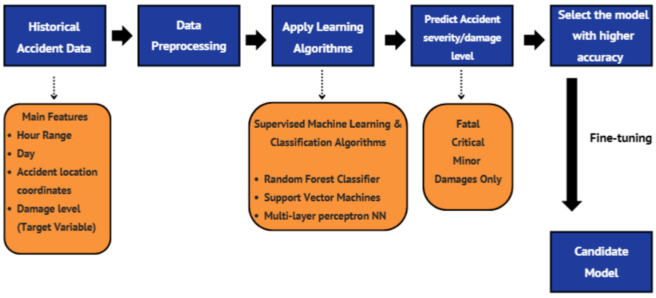

Fig. 2 Proposed model for prediction of severity.

As shown in Fig. 2, 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 that is shown in Table 3, hour-range, day, coordinates of the accident location, and damage level were the features that were used for this severity prediction. This model also conducted the data preprocessing right after the selection of the features for the prediction.

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 shown in Tables 6 and 7, classification models such as random forest classifiers, XGBoost, multi-layer perceptron NN, and support vector machines will be used to predict the damage level and 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.

IV. Results & Discussion

A. *Data collection and analysis*

In this study, a comprehensive evaluation of the data collected from the Mount Lavinia region in Colombo, Sri Lanka was conducted.

The past police accident reports are used to collect road accident data, and accident data collected from the Mount Lavinia police station is given below in table 3.

Table 3 Collected Accident Data

Road accident locations coordinates (latitude and longitude)
Date (Day of the week)
Time (Hour of the day)
Accident location 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

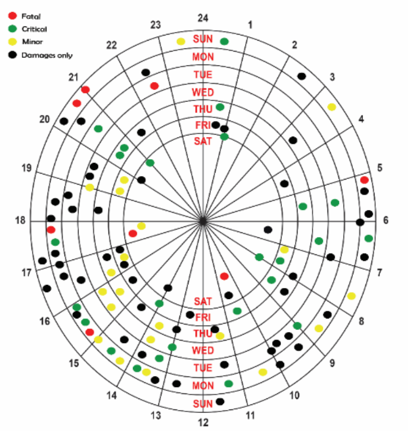

Fig. 3 Time Clock with damage severity.

Fig. 3 shows the days and times when traffic accidents occurred in the Mount Lavinia region in 2023 (with damage severity).

Fig. 4 Road accident hot spots.

Fig. 4 is a map showing traffic accident locations, also called "hot spots," in the Mount Lavinia region in 2023. These accident hotspots are divided into four levels of damage severity (fatal, critical, minor and damage). only). According to Figure 3, most traffic accidents occurred near intersections.

Table 4

Number of Road Accidents in Mount Lavinia Police Area				
Damage Level	2020	2021	2022	2023
Fatal	02	10	13	09
Critical	38	36	36	31
Minor	40	30	26	22
Damages only	64	42	48	51
Total	144	118	123	113
Deaths	02	10	13	09

According to the collected data from the police station, Table 4 shows the number of road accidents and deaths in the Mount Lavinia region based on the severity of the damage.

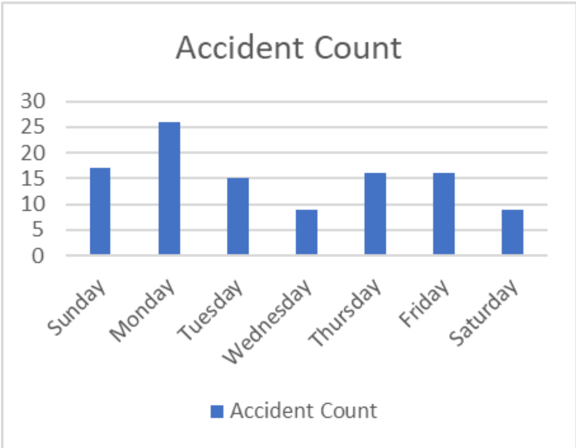

Fig. 5

According to Fig. 5, most traffic accidents occurred on Monday.

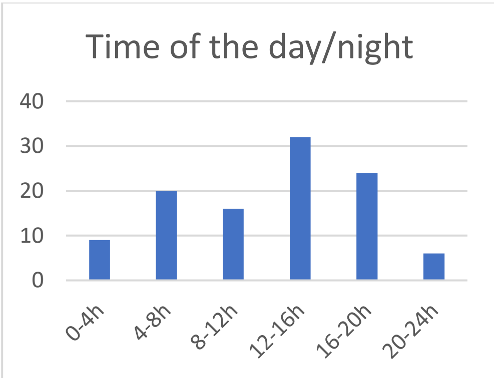

Fig. 6

According to Fig. 6, most traffic accidents occurred in between 12-to-16-hour range.

B. Input Features for the proposed models in Figure 1,2

Table 5 Selected features

Prediction Type	Features	Target Variable
Accident Hotspot	Hour Range	Accident Count
	Day	
	Latitude	
	Longitude	
Accident Severity	Hour Range	Damage Level
	Day	
	Latitude	
	Longitude	

C. Selection of machine learning technique for the proposed models

Table 6 Usage of machine learning techniques

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 7 Usage of Best performing machine learning techniques

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%

Random Forest is a supervised machine learning technique that was employed for prediction in 81.25% of the examined research publications, according to the analysis of tables 6 and 7. Labeled data is used to train supervised machine learning algorithms, which then learn to produce output according to the input features.

And among the supervised machine learning techniques, Random Forest performed better in terms of accuracy in approximately 69% of the studies.

Even though the Random Forest model did provide better accuracy for both accident hotspots and severity than the other techniques, since this study focused on implementing a predictive machine learning model that was specifically tailored for Sri Lankan accident data and there were no existing studies designed for Sri Lankan accident data, the Random Forest model will be compared with other machine learning techniques that were shown in existing studies in Tables 6 and 7. And the model that provides better accuracy will be chosen as the final model.

V. CONCLUSION

In conclusion, this research journey has shed light on the critical issue of road safety in Sri Lanka and similar regions, emphasizing the need for innovative approaches to accident prediction. The substantial human and economic losses resulting from road accidents necessitate immediate and data-driven solutions.

By methodically exploring and reviewing existing machine learning techniques for accident prediction and severity analysis, we have uncovered valuable insights into the potential for enhancing road safety practices. Our focus on the Mount Lavinia region of Colombo, Sri Lanka, and the comprehensive data collection methods, including accident hotspot coordinates, severity levels, day and time, accident reasons, and accident cases for multiple years, as well as the proposed machine learning models for the predictions, have provided a foundation for this research.

Most notably, a review of numerous predictive models confirms that the random forest model, a supervised machine learning technique, consistently outperforms other methods in terms of accuracy for predicting accident hotspots and severity.

And this proposed machine learning models are aimed at enhancing road safety through an "Intelligent Road Safety Mobile Navigation Application" that is going to provide alerts on accident hotspots and severity to drivers in real time. In Sri Lanka, no such kind of application has been developed yet. And it helps drivers adjust their driving patterns, which encourages the use of technology for greater excellence, will have a sizeable effect on avenue safety practices, and will doubtless save lives. This study isn't pretty much instructional overall performance; it's a sensible technique for an actual global problem that impacts countless human beings each day.

ACKNOWLEDGEMENT

The necessary and valuable instructions and guidance for the paper and the study model was given by Dr. Pradeep Kalansooriya.

REFERENCES

[1] "Road traffic injuries." Accessed: Apr. 18, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[2] "National Council for Road Safety." Accessed: Jul. 06, 2024. [Online]. Available: https://www.transport.gov.lk/web/index.php?option=com_content&view=article&id=29&Itemid=149&lang=en#road-accidents

[3] J. Patil, V. Patil, D. Walavalkar, and V. B. Lobo, "Road Accident Analysis and Hotspot Prediction using Clustering," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, Jul. 2021, pp. 763–768. doi: 10.1109/ICCES51350.2021.9489074.

[4] I. Ganie, "Predicting Factors of Vehicular Accidents using Machine Learning Algorithm," vol. 9, Oct. 2020, doi: 10.30534/ijeter/2020/46892020.

[5] T. Augustine and S. Shukla, "Road Accident Prediction using Machine Learning Approaches," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Apr. 2022, pp. 808–811. doi: 10.1109/ICACITE53722.2022.9823499.

[6] M. G and R. H. R, "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)*, Feb. 2023, pp. 1–6. doi: 10.1109/ICICACS57338.2023.10099519.

[7] D. Santos, J. Saias, P. Quaresma, and V. B. Nogueira, "Machine Learning Approaches to Traffic Accident Analysis and Hotspot Prediction," *Computers*, vol. 10, no. 12, Art. no. 12, Dec. 2021, doi: 10.3390/computers10120157.

[8] 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, Art. no. 6, Jun. 2023, doi: 10.3390/ijgi12060227.

[9] R. Vanitha and S. M, "Prediction of Road Accidents using Machine Learning Algorithms," May 18, 2023, *Rochester, NY*: 4460883. Accessed: Sep. 30, 2023. [Online]. Available: https://papers.ssrn.com/abstract=4460883

[10] 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, Art. no. 3, Jun. 2023, doi: 10.11591/ijece.v13i3.pp3149-3160.

[11] R. E. AlMamlook, K. M. Kwayu, M. R. Alkasisbeh, and A. A. Frefer, "Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, Apr. 2019, pp. 272–276. doi: 10.1109/JEEIT.2019.8717393.

[12] J. Zhang, Z. Li, Z. Pu, and C. Xu, "Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods," *IEEE Access*, vol. PP, pp. 1–1, Oct. 2018, doi: 10.1109/ACCESS.2018.2874979.

[13] M. K. Islam, I. Reza, U. Gazder, R. Akter, M. Arifuzzaman, and M. M. Rahman, "Predicting Road Crash Severity Using Classifier Models and Crash Hotspots," *Appl. Sci.*, vol. 12, no. 22, Art. no. 22, Jan. 2022, doi: 10.3390/app122211354.

[14] K. Santos, J. P. Dias, and C. Amado, "A literature review of machine learning algorithms for crash injury severity prediction," *J. Safety Res.*, vol. 80, pp. 254–269, Feb. 2022, doi: 10.1016/j.jsr.2021.12.007.

[15] B. Geyik and M. Kara, "Severity Prediction with Machine Learning Methods," in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Jun. 2020, pp. 1–7. doi: 10.1109/HORA49412.2020.9152601.

[16] I. E. Mallahi, A. Dlia, J. Riffi, M. A. Mahraz, and H. Tairi, "Prediction of Traffic Accidents using Random Forest Model," in *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, May 2022, pp. 1–7. doi: 10.1109/ISCV54655.2022.9806099.

[17] T. Q. S. Abdullah, H. Xiao, J. P. Namahoro, E. Alfalahi, and I. Al-Wesabi, "Exploiting Machine Learning Algorithms for Predicting Crash Injury Severity in Yemen: Hospital Case Study," *Appl. Comput. Math.*, vol. 9, pp. 155–164, Sep. 2020, doi: 10.11648/j.acm.20200905.12.

[18] T. Lu, Z. Dunyao, Y. Lixin, and Z. Pan, "The traffic accident hotspot prediction: Based on the logistic regression method," in *2015 International Conference on Transportation Information and Safety (ICTIS)*, Jun. 2015, pp. 107–110. doi: 10.1109/ICTIS.2015.7232194.

AUTHOR BIOGRAPHY

JMCT Jayamanna is a final year undergraduate at General Sir John Kotelawala Defence University. Reading the BSc. (Hons) Computer Science.

Dr. Pradeep Kalansooriya is Dean of the Faculty of Computing and senior lecturer within the Department of Computer Science, Faculty of Computing, General Sir John Kotelawala Defence University.