Machine Learning Models for Predicting Risk Avoidance in Motorcycle Riding: A Road Safety Approach

Ambareen Khan[1], Lilis Surienty[1*] and Kunitomo Asano[2]

[1]School of Management, Universiti Sains Malaysia, Gelugor, Pulau Pinang, Malaysia
[2] Boon Siew Honda Sdn Bhd, 721, Persiaran Cassia Selatan 1, Kawasan Perindustrian Batu Kawan, 14100 Simpang Ampat, Pulau Pinang. Malaysia

* Corresponding author

Abstract: In Malaysia, accidents are increasing daily (Azami et al., 2024). Motorcycle riders were given road safety training to reduce this increasing number of accidents. The overall number of participants is 400. During this training, the motorcyclists' risk-avoidance abilities are evaluated by testing their plank, slalom, and emergency braking tests. The riders are first given practice for the test, followed by the actual test. These avoidance skill training results are utilized for training the random forest machine learning algorithm, which predicts future riders' risk avoidance behaviour. The machine learning method can predict this highly complicated data with an acceptable accuracy of 0.91 to 0.72.

1. Introduction:

According to the WHO Report 2023, road traffic accidents killed approximately 1.19 million people. Road traffic injuries are now the top cause of death among children and young adults aged 5 to 29. Over half of all traffic fatalities include vulnerable road users such as pedestrians, cyclists, and motorcyclists. Road traffic accidents cost most countries 3% of their GDP (WHO, 2023). Low- and middle-income countries account for over 92% of worldwide road fatalities while having only roughly 60% of the world's total automobiles. Malaysia is a middle-income country, and motorbike accidents are increasing. Malaysia has the third highest road traffic death rate among countries in the WHO Western Pacific Region (WHO, 2021). Road traffic injuries account for 14% of deaths among Malaysian children aged 5 to 14, making them the country's leading cause of child mortality (WHO, 2021). As a result, there is an urgent need to identify the primary cause of road crashes in Malaysia.

Previous research has reached varied results about the causes of traffic fatalities after road crashes. Derma et al. (2017) revealed that only 15.4% and 11.25% of road traffic fatalities in Malaysia from 2000 to 2011 were caused by road defects such as road shoulder edge drop-offs and potholes, respectively. In comparison, 48.6% of deaths are caused by inadequate street lighting. Furthermore, Manan et al. (2017), who thoroughly analyzed motorbike crashes in Malaysia using police traffic crash information from 2010 to 2012, discovered that most road traffic fatalities happened between 7 and 12 p.m. However, Hashim and Iqbal (2011) found that most injured motorbike riders were male, which is generally associated with higher-risk riding conduct. Human mistakes have consistently been the leading cause of road crashes (Bucsuházy et al., 2020; Dakota and Kim, 2024). There are several reasons for road crashes in Malaysia, and the sheer number of crashes makes it difficult to anticipate potential dangers objectively. Thus, artificial intelligence can help to prioritize the critical causes of traffic accidents.

Bokaba et al. (2022) conducted a comparative analysis utilizing machine learning classifiers, including K-nearest neighbors, support vector machine, logistic regression, and random forest, on the real-time road traffic incidents dataset from Gauteng, South Africa. The empirical results and analysis indicate that the classifier attained optimal performance with multiple imputations through chained equations, surpassing alternative combinations. Infante et al. (2022) conducted a study comparing statistical and machine learning models to analyze the severity of road traffic accidents in Portugal. The machine learning system showed strong performance, elevated accident severity, and a substantial sample size. Santos et al. (2021) studied the predictive factors affecting road traffic accidents. They utilized a machine learning technique to forecast the same. Singh et al. (2019) employed machine learning approaches, including k-nearest neighbor, naïve Bayes, decision tree, and support vector machine, to identify road accidents in Punjab. The decision tree had the highest performance, achieving an accuracy of 86.25%. A further study by Labib et al. (2019) employed k-nearest neighbor, AdaBoost, naïve Bayes, and decision tree algorithms to assess the severity of road traffic incidents in Bangladesh. They categorized the severity of accidents as Fatal, Grievous, Minor Injury, and Motor Collision. AdaBoost demonstrates superior performance. According to the study by AIMamlook et al. (2019), Random Forest achieves optimal performance in forecasting the severity of traffic accidents, attaining an accuracy of 75.5%.

Safety riding training is organized by a company in Batu Kawan Industrial Park(BKIP) in Penang, in partnership with Universiti Sains Malaysia, aimed at diminishing road traffic fatalities among motorcycle riders in Batu Kawan Industrial Park (BKIP), Penang, Malaysia, and addressing this issue.

2. Methodology :

2.1 Random forest:

The random forest method is utilized to analyze the training data obtained from the sessions conducted at Boon Siew Honda in Batu Kawan, Penang, Malaysia. In this model, 70% of the data is allocated for training, while the remaining 30% is allocated for testing. The random state for the train-test split is established at 42 to guarantee result repeatability. The maximum depth of the random forest is set at 100 to capture intricate patterns in the data efficiently. The process steps are delineated as follows:

2.1.1 Data Acquisition: We gather data via comprehensive questionnaires and studies that monitor individuals' activities during the course. The dataset is comprehensive for model training, encompassing attributes associated with risk factors and motorcycle riding behaviors.

2.1.2 Selection of Features: Attributes such as speed and response time are chosen for their significant impact on risk mitigation capabilities. This phase is essential for training the model on the most critical and impactful factors to improve its predictive capabilities.

2.1.3 Model Optimisation: The hyperparameters of the random forest model are optimized to enhance performance. The optimized parameters encompass the number of trees, the maximum number of features evaluated for partitioning, and the lowest sample size required for a node division.

2.1.4 Assessment Criteria: The Mean Absolute Percentage Error (MAPE) and R-squared (R^2) are utilized to evaluate the accuracy and performance of the random forest model. R^2 indicates the proportion of variance in the dependent variable that the independent variables can elucidate, while MAPE quantifies the accuracy of predictions regarding percentage error.

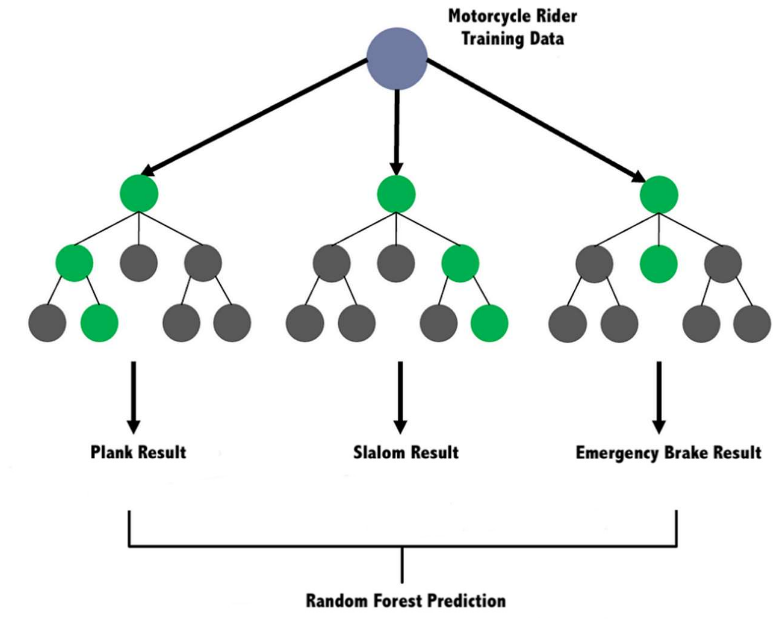

Figure 1. Random forest architecture.

Figure 1 illustrates the architecture of the random forest model utilized for forecasting motorcycle rider training outcomes based on plank, slalom, and emergency braking assessments. The procedure can be delineated into numerous essential components.:

2.2 Collection of Motorcycle Rider Training Data: At the top of the diagram is the training data, which includes much information about how riders behave and their skills. These data are needed to train the random forest model.

Figure 2. Motorbike riders are given instructions before the start of every training

Figure 3. Motorbike riders are given plank avoidance training

In Figure 3, the motorbike riders are given plank avoidance training. During this, the time is recorded for the riders to complete this training. The riders should be able to complete this training within at least 15 seconds.

Figure 4. Riders are given slalom training.

Figure 4 shows the slalom training. During this training, the riders learn how to avoid accidents within 35 seconds.

Figure 5. Riders practicing emergency braking during the training session

Figure 5 shows the emergency braking training. During this training, the riders learn to apply the emergency brake within 10 minutes while riding at 40 m/s.

2.3 Decision Trees in the Random Forest: The diagram shows the riders' achievement in the form of several decision trees. The three trees shown here are called out by:

2.3.1 Plank Result: The riders' ability to keep their balance while standing on a log is tested at this tree. That shows how well they can keep their balance and control.

2.3.2 Slalom Score: This tree measures how well the riders can move through a slalom run by showing how agile they are and how well they can handle obstacles.

2.3.3 Emergency Brake Outcome: This tree assesses the motorcyclists' proficiency in executing an emergency brake while illustrating their response time and braking capabilities.

2.4 Random Forest Prediction: The output of the random forest model is displayed in the lower section of the diagram. Each decision tree contributes to the overall prediction by evaluating the riders' competencies. The aggregated outcomes from all trees are utilized to generate a conclusive prediction regarding motorbike riders' risk aversion capability.

The program can identify intricate patterns and relationships in the data by employing a random forest of several decision trees. This technique enhances the model's predictive accuracy and robustness, guaranteeing a comprehensive assessment of the riders' risk-avoidance capabilities.

Employing distinct trees for various performance indicators enables the random forest to consider multiple facets of riding proficiency, culminating in a more comprehensive prediction. This approach is particularly advantageous in scenarios where various factors affect the outcome, as it leverages the advantages of ensemble learning to enhance predictive accuracy.

3. Results and Discussion:

Due to the elevated death rate in Penang, Malaysia, motorbike riders received practical training in road safety. The motorbike riders received this instruction in partnership with Boon Siew Honda Safety Riding Centre. Throughout the training period, 400 riders received this instruction. Training consists of two tiers. In level 1, the riders have articulated the significance of road safety for themselves and others. They elucidated the prevalent blunders made by riders on the road and assessed them to avert such mishaps. The team subsequently gathered data on the hazard-prone area en route to work. Ultimately, their driving proficiency was enhanced via a virtual reality riding assessment.

During level 2, the participants received practical training on the road. During this training, participants were instructed to skilfully avoid obstacles, execute slalom maneuvers to avert accidents and apply emergency brakes within a distance of 10 meters.

3.1 Plank Avoidance Training:

During this training, learners are allotted 1 hour and 45 minutes of practice to circumvent the plank exercise. After the practice test, the team administers two assessments. Both assessments evaluate plank avoidance abilities. The plank avoidance must be executed within 15 seconds. If the duration to avoid a plank is under 15 seconds, the rider is deemed inept in motorcycle balance, essential in low-speed situations.

3.2 Slalom Training:

The practice sessions last for one hour during this training. To achieve proficiency, the motorcyclists are instructed to navigate obstacles without employing brakes. Instead, they are educated to regulate their speed by disengaging the throttle, which functions as a brake. They must complete the slalom course in under 31 seconds to be deemed a proficient and safe rider.

3.3 Emergency brake Training:

Participants receive emergency brake training to execute it within 6 meters at 40 km/h. The primary purpose of this training is to safely execute an emergency stop within the smallest distance. They are instructed to utilize all three braking kinds to get this proficiency.

The application of machine learning assesses the efficacy of this training. The random forest is employed to train the model using the training data from 400 participants. The model can forecast the training data with satisfactory precision.

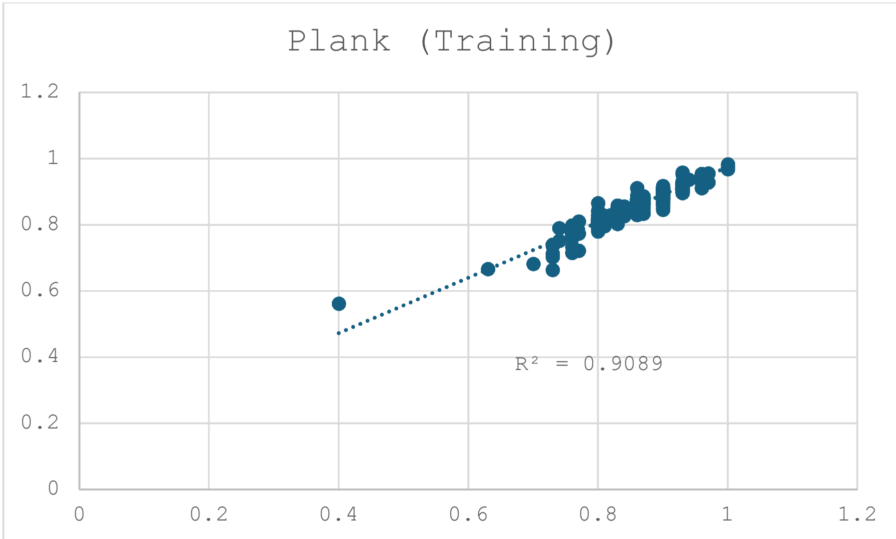

Figure 2. Plank Training Result

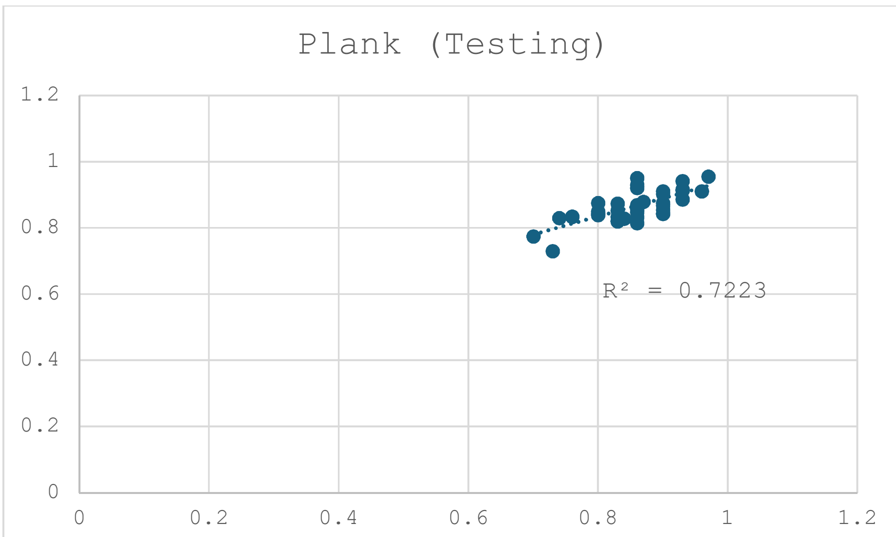

Figure 3. Plank Testing Result

In the plank training, the R2 value for the training set is 0.90, whereas for the testing set, it is 0.72, as illustrated in Figures 2 and 3. The mean absolute percentage error was 0.016 when the random forest method was trained. During testing, the mean absolute percentage error is 0.039.

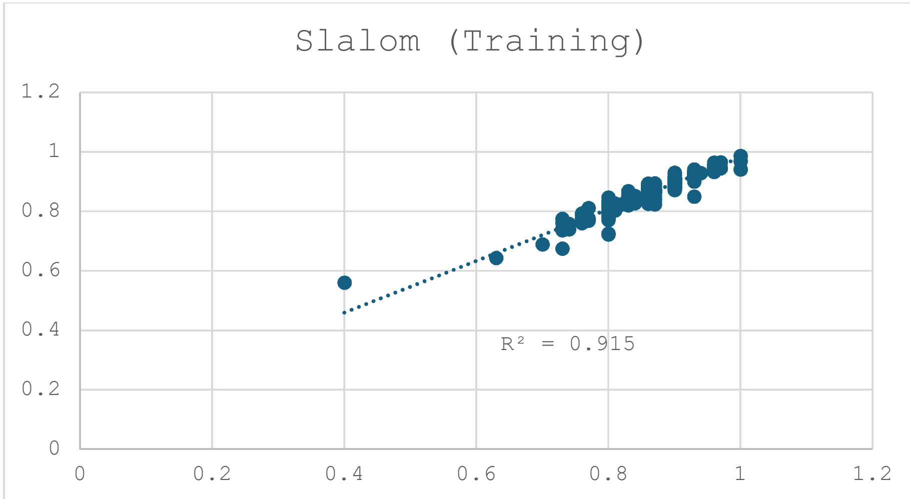

Figure 4. Slalom Training Result

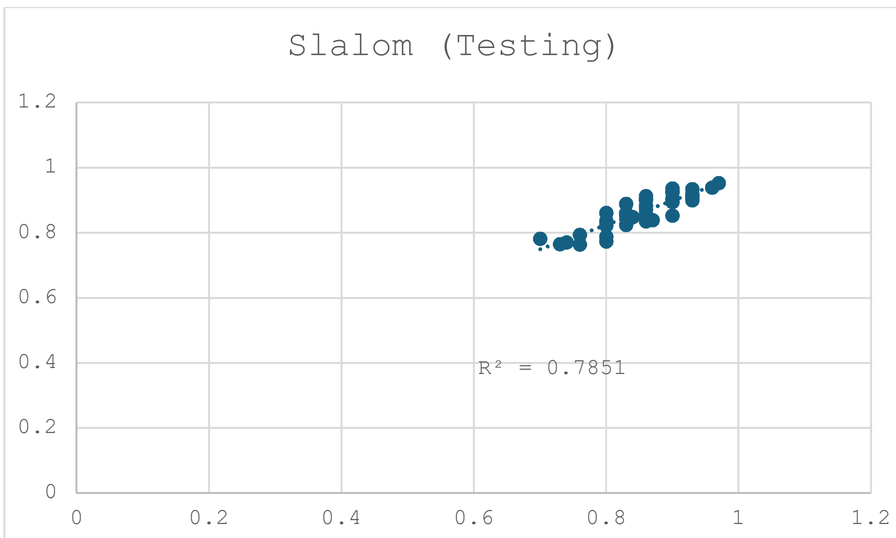

Figure 5. Slalom Testing Result

The R2 value for slalom in the random forest model training is 0.91, whereas in testing, it is 0.78 (shown in Figures 4 and 5). The mean absolute percentage error is 0.014 during model training and 0.032 during testing.

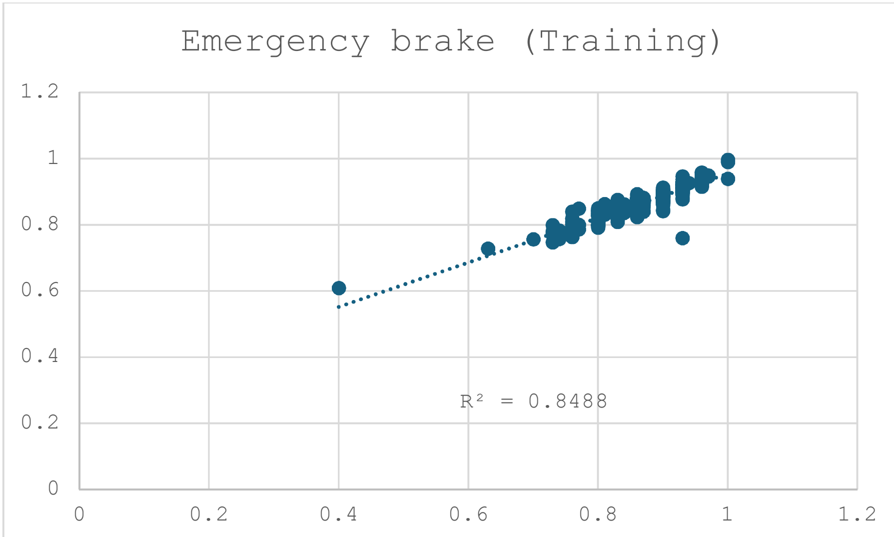

Figure 6. Emergency Brake Training Result

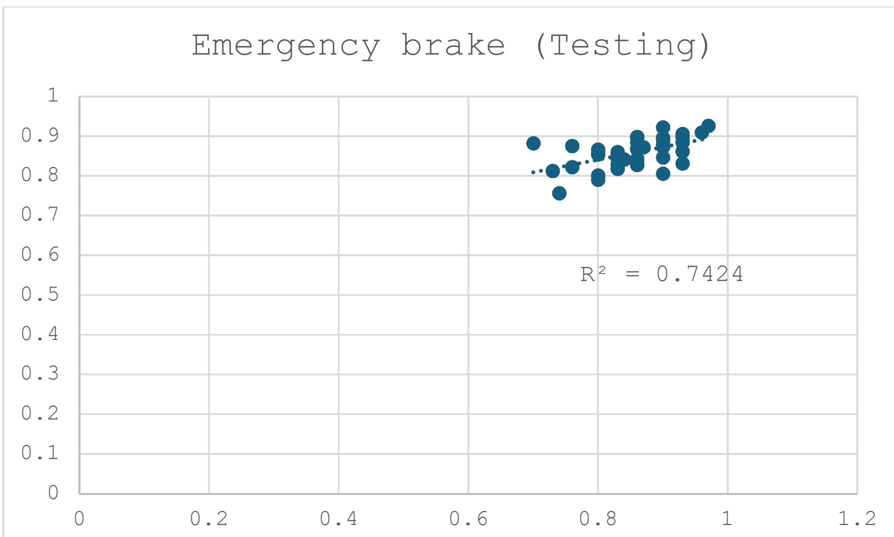

Figure 7. Emergency Brake Testing Result

The R2 value during emergency brake practice is 0.84 when training the random forest model and 0.74 when testing the same model (Figures 6 and 7). The average absolute percentage error during training is 0.021, while during testing, it is 0.045.

The random forest examines respondents' risk avoidance skills, distinguishing training and testing results. That is based on predicting the behavior of motorbike riders. However, predicting human behavior is challenging. Given the intricacy of the data, the random forest model can produce satisfactory results.

Conclusion:

The random forest machine learning method can predict the data with an accuracy of 0.91 to 0.72. The data used in the prediction is based on motorbike riders' risk avoidance skills. Despite the complexity of human behavior, the random forest model can accurately anticipate the data.

Acknowledment:

Acknowledgement to “Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme with Project Code: FRGS/1/2018/SS03/USM/02/8”and CAMS Project PRO-S-1020-I145”.

References:

1. Azami, M. F. A. M., Misro, M. Y., & Hamidun, R. "Road crash dynamics in Malaysia: Analysis of trends and patterns". Heliyon (2024), 10(18).
2. World Health Organization. Global Status Report on Road Traffic Injuries 2023. World Health Organization; 2023. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
3. The World Health Organization Launches the Global Plan for the Decade of Action for Road Safety 2021 – 2030, World Health Organization;2021. https://www.who.int/malaysia/news/detail/28-10-2021-the-world-health-organization-launches-the-global-plan-for-the-decade-of-action-for-road-safety-2021-2030
4. Darma, Yusria, Mohamed Rehan Karim, and Sulaiman Abdullah. "An analysis of Malaysia road traffic death distribution by road environment." Sādhanā 42 (2017): 1605-1615.
5. Manan, Muhammad Marizwan Abdul, András Várhelyi, Ali Kemal Çelik, and Hizal Hanis Hashim. "Road characteristics and environment factors associated with motorcycle fatal crashes in Malaysia." IATSS research 42, no. 4 (2018): 207-220.
6. Hashim, Hasnah, and Syed Iqbal. "Motorcycle accident is the main cause of maxillofacial injuries in the Penang Mainland, Malaysia." Dental traumatology 27, no. 1 (2011): 19-22.
7. Bucsuházy, Kateřina, Eva Matuchová, Robert Zůvala, Pavlína Moravcová, Martina Kostíková, and Roman Mikulec. "Human factors contributing to the road traffic accident occurrence." Transportation research procedia 45 (2020): 555-56 McCarty,
8. Dakota, and Hyun Woo Kim. "Risky behaviors and road safety: An exploration of age and gender influences on road accident rates." PLoS one 19, no. 1 (2024): e0296663.
9. Bokaba, Tebogo, Wesley Doorsamy, and Babu Sena Paul. "Comparative study of machine learning classifiers for modelling road traffic accidents." Applied Sciences 12, no. 2 (2022): 828.
10. Infante, Paulo, Gonçalo Jacinto, Anabela Afonso, Leonor Rego, Vitor Nogueira, Paulo Quaresma, José Saias et al. "Comparison of statistical and machine-learning models on road traffic accident severity classification." Computers 11, no. 5 (2022): 80.
11. Santos, Daniel, José Saias, Paulo Quaresma, and Vítor Beires Nogueira. "Machine learning approaches to traffic accident analysis and hotspot prediction." Computers 10, no. 12 (2021): 157.
12. Singh, Jaspreet, Gurvinder Singh, Prithvipal Singh, and Mandeep Kaur. "Evaluation and classification of road accidents using machine learning techniques." In Emerging Research in Computing, Information, Communication and Applications: ERCICA 2018, Volume 1, pp. 193-204. Springer Singapore, 2019.
13. Labib, Md Farhan, Ahmed Sady Rifat, Md Mosabbir Hossain, Amit Kumar Das, and Faria Nawrine. "Road accident analysis and prediction of accident severity by using machine learning in Bangladesh." In 2019 7th international conference on smart computing & communications (ICSCC), pp. 1-5. IEEE, 2019.

14. AlMamlook, Rabia Emhamed, Keneth Morgan Kwayu, Maha Reda Alkasisbeh, and Abdulbaset Ali 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), pp. 272-276. IEEE, 2019.