Enhancing Road Safety: Predicting Motorcycle Riders' Risk Avoidance Skills with Machine Learning

Ambareen Khan[1], Lilis Surienty[1*] and Chin Chin Tsun[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

Abstract: The training for motorcycle riders on road safety is done. The total number of participants is 400. In this training, the riders' risk avoidance skills are assessed by testing their plank, slalom, and emergency brake test results. First, the riders are given practice for the mentioned test, and then they are tested. These avoidance skills behaviour training results are used to train the random forest machine learning algorithm to predict the future riders' risk avoidance behaviour. The machine learning algorithm can predict this highly complex data with acceptable accuracy between 0.91 and 0.72.

1. Introduction:

In the WHO report 2023, it was found that around 1.19 million people died due to road traffic accidents. Road traffic injuries have become the leading cause of death among children and young adults (5-29 years). Over 50% of road traffic fatalities occur among vulnerable road users, such as pedestrians, cyclists, and motorcyclists. Road traffic accidents incur an economic cost equivalent to 3% of the gross domestic product in most nations (WHO, 2023). Approximately, 92% of global road fatalities occur in the low- and middle-income nations, although these countries account for only about 60% of the world's total automobiles. Malaysia is a middle-income country and motorcycle accidents are increasing. Malaysia ranks third in terms of road traffic mortality rate among countries in the WHO Western Pacific Region (WHO, 2021). Road traffic injuries are responsible for 14% of deaths among children aged 5-14 years in Malaysia, making it the primary cause of child mortality in the country (WHO, 2021). Hence, there is a urgent need to find the main cause of road crash in Malaysia.

Past studies have made different conclusions to causes of traffic fatalities following road crashes. According to the analysis done by Derma et al. (2017) of road traffic fatalities in Malaysia for the years 2000 to 2011, it was found that only 15.4% and 11.25% of deaths were due to road defects such as road shoulder edge drop-off and potholes. In comparison, 48.6% of deaths are due to a lack of street lighting provision. In addition, Manan et al. (2017) who did extensive analyses of road crashes involving motorcycle riders in Malaysia using police traffic crash reports for the years 2010 to 2012 found that most road traffic fatalities occurred between 7 and 12 evening. However, Hashim and Iqbal (2011) found that most of the injured group of motorcycle riders was primarily male which is commonly understood to display higher risk riding behaviour. Consistently, human error has been cited to be the major cause of traffic crashes (Bucsuházy et al., 2020; Dakota and Kim, 2024). There are multi causes of road crashes in Malaysia and the amount of crashes making it difficult to be objective in predicting the possible risks. Thus, the contribution of artificial intelligence can be effective in prioritising the main causes of road crashes.

The machine learning classifiers (like K-nearest neighbors, support vector machine, logistic regression, and random forest) are applied to the real-time Road traffic accidents dataset of Gauteng, South Africa, to do a comparative study by Bokaba et al. (2022). The empirical findings and analysis demonstrate that when combined with multiple imputations via chained equations, the classifier achieved the best performance compared to alternative combinations. Another study by Infante et al. (2022) compared statistical and machine learning models to study the severity of road traffic accidents in Portugal. The machine learning algorithm performed well with a high accident severity and large sample. A study by Santos et al. (2021) researched the prediction of factors influencing road traffic accidents. They applied a machine learning algorithm to predict the same. To classify road accidents in Punjab, Singh et al. (2019) used machine learning techniques such as a k-nearest neighbor, naïve Bayes, decision tree, and support vector machine to evaluate road accidents. They found that the decision tree performed the best with 86.25 % accuracy. Another study by Labib et al.(2019) applied the k-nearest neighbor, AdaBoost, naïve Bayes, and decision tree to know the intensity of road traffic accidents in Bangladesh. They classified the severity of accidents into Fatal, Grievous, Simple Injury, and Motor Collision. And found that AdaBoost gives the best performance. Random Forest performs best for predicting the severity of traffic accidents with 75.5% accuracy, as AIMamlook et al. (2019) studied.

Safety riding training is conducted by Boon Siew Honda Sdn Bhd, a company that manufacture and sell Honda motorcycles in Malaysia in collaboration with Universiti Sains Malaysia to reduce road traffic fatalities among motorcycle riders in Batu Kawan Industrial Park (BKIP), Penang state in Malaysia, and to counter the above problem.

2. Methodology :

2.1 Random forest:

The random forest algorithm is applied to the training data collected from the trainings done at Boon Siew Honda in Batu Kawan, Penang, Malaysia. In this model, 70% of the data is designated for training, while the remaining 30% is reserved for testing. To ensure the reproducibility of the results, the random state for the train-test split is set to 42. Additionally, the maximum depth of the random forest is fixed at 100 to capture complex patterns in the data effectively. The procedural steps are highlighted as follows:

2.1.1 Data Collection: At the training, we collect data through in-depth surveys and research that observe people in action. The dataset is robust for training models, as it covers characteristics linked to risk factors and motorcycle riding behaviours.

2.1.2 Feature Selection: Features like speed and reaction time are selected that have a major influence on risk avoidance abilities. In order to train the model on the most significant and impactful factors, which will enhance its prediction abilities, this phase is crucial.

2.1.3 Model Tuning: To improve performance, the random forest model's hyperparameters are tuned. Fine-tuned parameters include the number of trees, the maximum number of characteristics considered for splitting, and the minimum number of samples needed for a node split.

2.1.4 Evaluation Metrics: Mean Absolute Percentage Error (MAPE) and R-squared (R^2) are used to assess the random forest model's accuracy and performance. R^2 shows how much of the dependent variable's variance can be explained by the independent variables, and MAPE shows how accurate the predictions are in terms of percentage error.

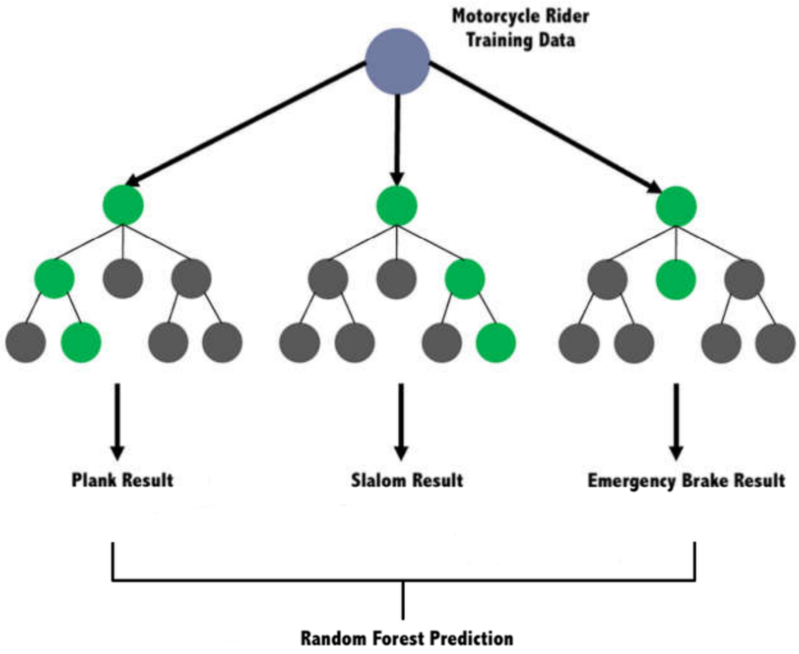

Figure 1. Random forest architecture.

Figure 1 shows the architecture of the random forest model for predicting motorcycle rider training results based on plank, slalom, and emergency brake tests. The process can be broken down into several key components:

2.2 Motorcycle Rider Training Data: The training data, which includes a variety of features pertaining to the behaviours and abilities of motorcyclists, is displayed at the top of the diagram. Training the random forest model requires these data.

2.3 Decision Trees in the Random Forest: The riders' performance is depicted in the diagram in several decision trees. Attention is drawn to the three trees shown here by:

2.3.1 Plank Result: The ability of the riders to retain their balance while standing on a plank is evaluated by this tree, which provides insight into their ability to maintain control and balance.

2.3.2 Slalom Result: The ability of the riders to navigate through a slalom course is evaluated by this tree, which reflects the riders' agility and their ability to manoeuvre obstacles.

2.3.3 Emergency Brake Result: The riders' level of ability in applying an emergency brake is evaluated by this tree, which also serves to demonstrate their response time and braking abilities.

2.4 Random Forest Prediction: The random forest model's output is shown in the bottom portion of the diagram. Each decision tree adds to the overall prediction by assessing the riders' abilities. The combined results from all of the trees are used to produce a final forecast about motorbike riders' risk avoidance ability.

The model can capture complicated patterns and relationships in the data by utilising a random forest composed of many decision trees. This strategy improves the model's predicted accuracy and resilience, ensuring a thorough evaluation of the riders' risk-avoidance abilities.

Using separate trees for different performance indicators allows the random forest to take into account diverse aspects of riding ability, resulting in a more complete prediction. This technology is especially useful in cases where several factors influence the result since it uses the benefits of ensemble learning to improve prediction accuracy.

3. Results and Discussion:

As the fatality rate is high in the Penang state of Malaysia, motorcycle riders were given hands-on practical training on road safety. The motorcycle riders were given this training in collaboration with Boon Siew Honda Safety Riding Centre. During the training, 400 riders were given this training in the course of time. There are two levels of training. In level 1, the riders have explained the importance of road safety for themselves and others. They explained the riders' common mistakes on the road and measured them to prevent such errors. Then, the team collected the hazard-prone area on the way to work. Ultimately, their driving skills were assisted by a virtual reality test of riding.

In level 2, the riders were given on-road practical training. During this training, they were given training on how to avoid the plank effectively, how to slalom on the road to prevent accidents, and how to apply emergency brakes effectively within 10 m.

3.1 Plank Avoidance Training:

In this training, the participants are given 1 hour 45 minutes of practice to avoid plank. After the practice test, the team conducts two tests. In both tests, plank avoidance skills are tested. Plank avoidance should be completed within 15 seconds. If the time to avoid plank is less than 15 seconds, the rider is considered incompetent in balancing their motorcycles which is critical during slow speed circumstance.

3.2 Slalom Training:

During this training, the practice runs for 1 hour. In order to become competent, the riders are trained to avoid obstacles by not using any brakes. Instead, they are trained to control their speed by releasing the throttle and this consequently act as a brake. And they have to complete the slalom distance in less than 31 seconds to be considered as safe competent rider.

3.3 Emergency brake Training:

Emergency brake training is given to the participants to train them to apply it within 6 meter at a speed of 40 km/hour. The main objective behind this training is to be able to able to do emergency stop in the shortest distance safely. They are taught to apply all three types of brakes to achieve this competency.

With the application of machine learning, the effectiveness of this training is determined. The random forest applied to the training data from 400 participants is used to train the model. The model can predict the training data with acceptable accuracy.

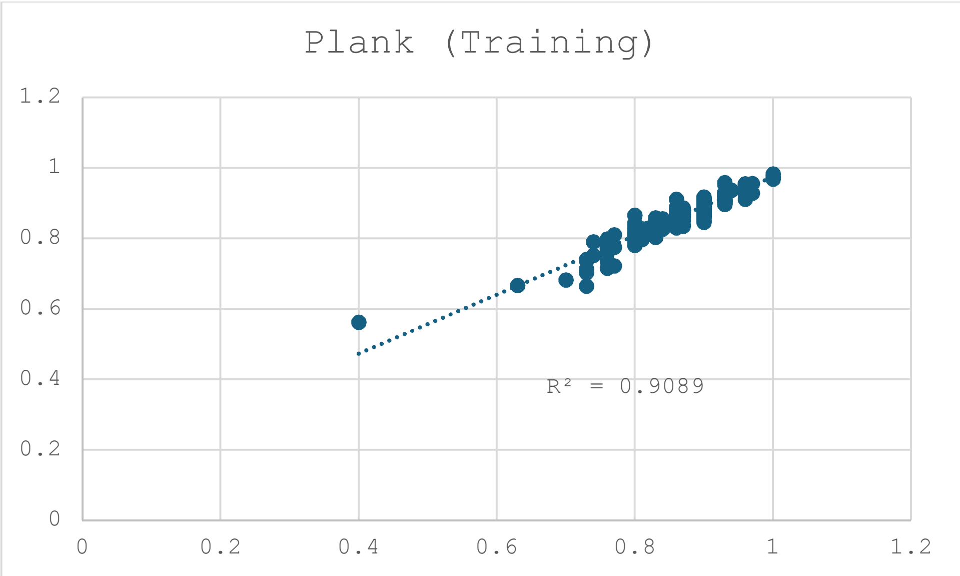

Figure 2. Plank Training Result

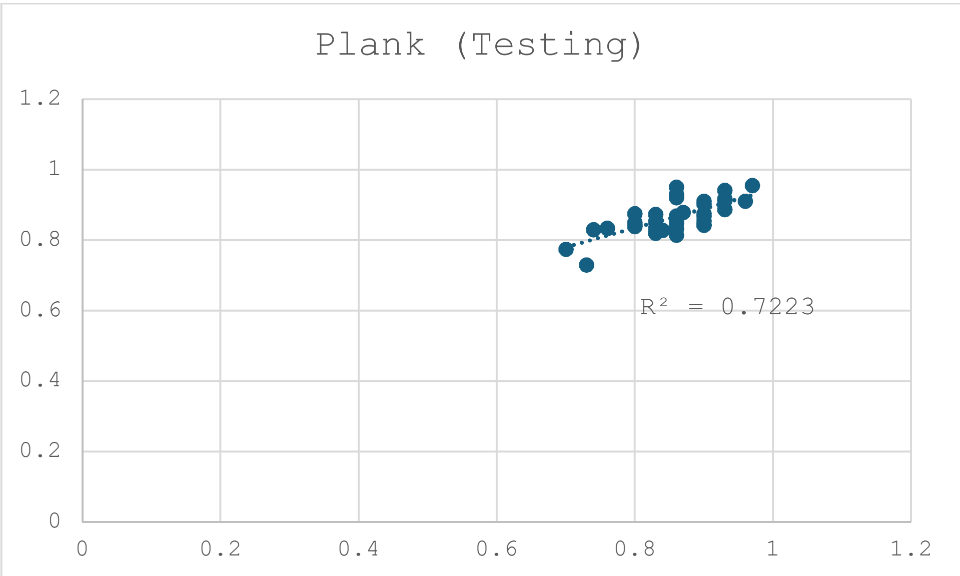

Figure 3. Plank Testing Result

During the plank training, the R^2 for the training is 0.90; for the testing, it is 0.72, as shown in Figures 2 and 3. The mean absolute percentage error was 0.016 when the random forest algorithm was taught. Meanwhile, for testing, the mean absolute percentage error is 0.039.

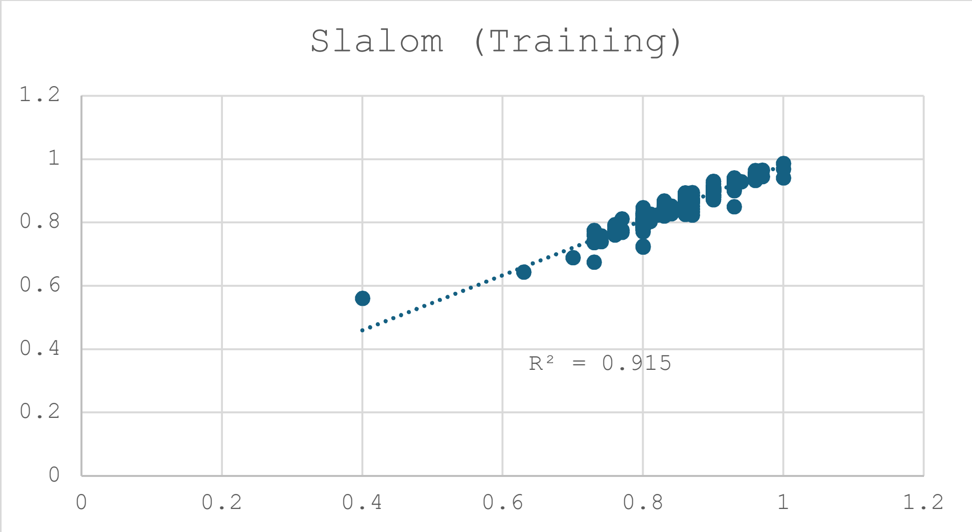

Figure 4. Slalom Training Result

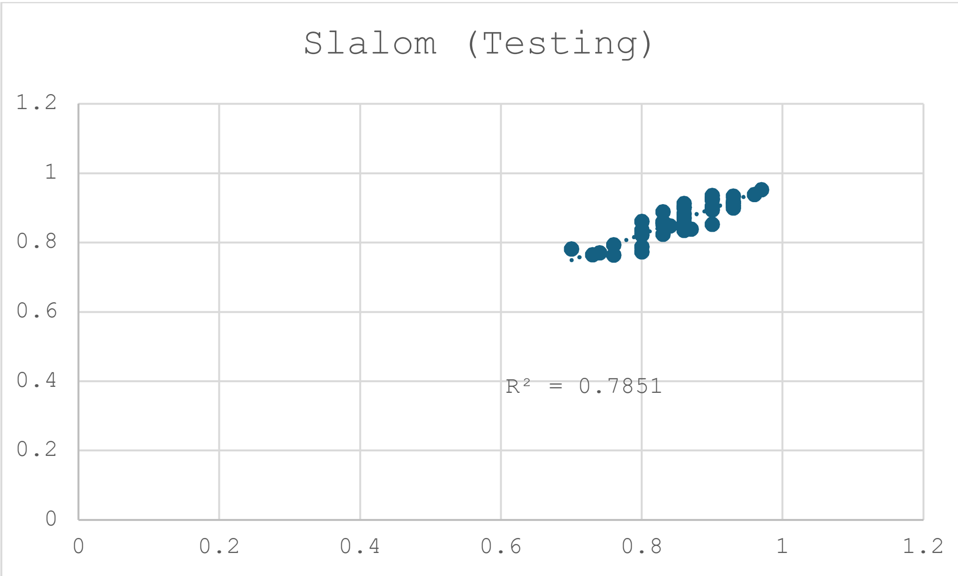

Figure 5. Slalom Testing Result

The R^2 value for slalom during the random forest model training is 0.91, whereas, during testing, it is 0.78 (shown in Figures 4 and 5). The mean absolute percentage error is 0.014 during training of the model and 0.032 during testing.

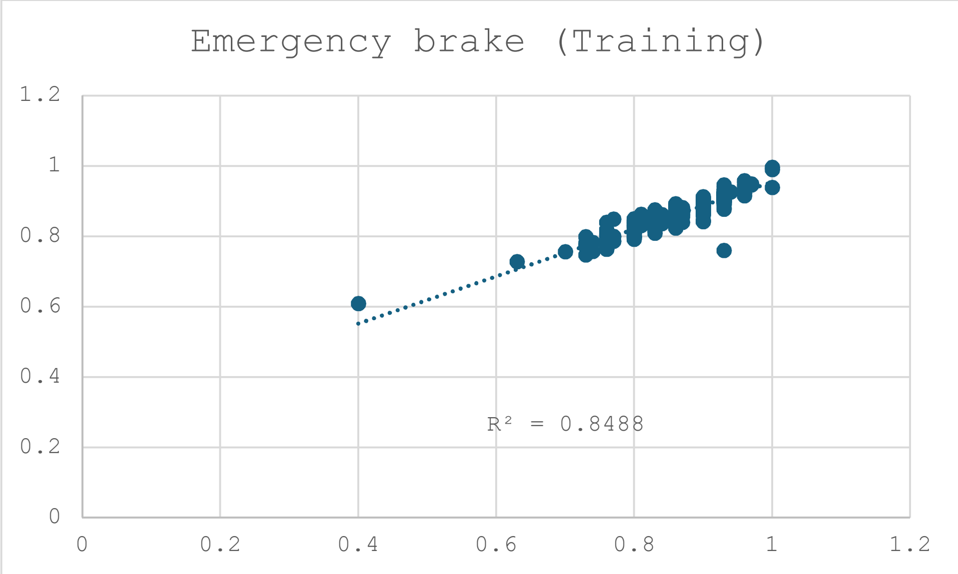

Figure 6. Emergency Brake Training Result

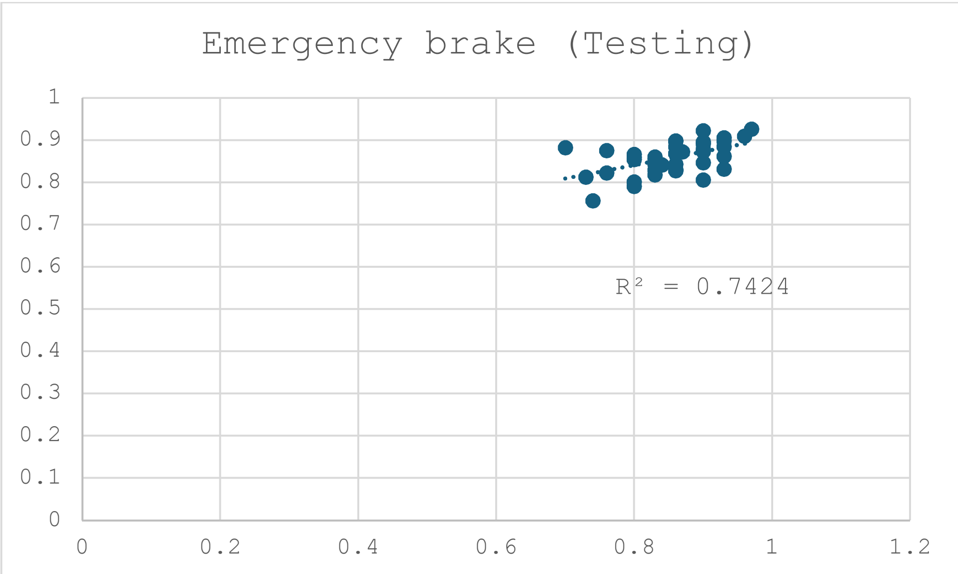

Figure 7. Emergency Brake Testing Result

The R^2 value during the emergency brake practice is 0.84 during training of the random forest model and 0.74 during testing of the same model (shown in Figures 6 and 7). The mean absolute percentage error during training is 0.021, and during testing is 0.045.

The difference between training and testing results is that the random forest is trying to assess the risk avoidance skills of the respondents. This is based on the prediction of the behavior of the motorcycle riders. The behavior of humans is, however, difficult to predict. However, seeing the complexity of the data, the random forest model can give satisfactory results.

Conclusion:

Applying the random forest machine learning algorithm can predict the data with an accuracy between 0.91 and 0.72. The data used for the prediction is based on the risk avoidance skills of the motorcycle riders. As human behavior is very complex, the random forest model can still predict the data with acceptable accuracy.

References:

1. 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
2. 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
3. Darma, Yusria, Mohamed Rehan Karim, and Sulaiman Abdullah. "An analysis of Malaysia road traffic death distribution by road environment." Sādhanā 42 (2017): 1605-1615.
4. 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.
5. 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.
6. 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,
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.