Differential Analysis of Discerned Motorcycle Risk Factors on Traffic Control and Law in Developing Country's Urban Context

Md. Gausul Azam Noman[a]
Department of Civil and Resource Engineering
Dalhousie University, Halifax, NS B3H 4R2, Canada
Email: nomanazambd@gmail.com

Md. Mushtaque Tahmid[a]
Department of Civil, Environmental and Construction Engineering
University of Central Florida, Orlando, FL 32816, United States
Email: sayeemss21@gmail.com

Md Asif Raihan*
Accident Research Institute (ARI)
Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh,
Email: raihan@ari.buet.ac.bd

Md. Shamsul Hoque
Department of Civil Engineering
Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
Email: shamhoque84@gmail.com

**Corresponding Author*
[a] *Equal Contributors*

Word Count: 6,201 words

ABSTRACT

As a developing country, Bangladesh is burdened by the rising number of motorcycles on the roads, which make up most vehicles and cause the bulk of traffic accidents. To address this critical issue, this paper aims to address the lack of comprehensive research on motorcycle accidents by exploring the impact of the perception of motorcycle crashes on traffic control and law enforcement. Prior studies have concentrated on real-time crash prediction and identifying significant contributors to crashes, neglecting simultaneous consideration of factors like driving environment, weather conditions, traffic control, driving behavior, and pedestrian-related characteristics. To bridge this gap, the study employs advanced machine learning algorithms, namely the Random Forest and CNN 1D algorithm, to analyze distinctive responses from various types of road users: motorcycle riders, motorcycle users (females and males) with different graphical illustrations. This research endeavors to establish a robust model of precursors by gathering ratings from road users, enabling effective design, management, planning, and implementation of policies to improve traffic control and law enforcement to reduce motorcycle accidents. ‘Traffic movement,’ ‘sign marking and lighting,’ and ‘driving environment’ related features were proved to be the most significant precursors. Furthermore, the study has deployed the machine learning model on a public server, providing policymakers and users with a user-friendly interface to predict the target variable ‘traffic control and law’ based on their input. Users can enter ratings on relevant risk factors using a 1 to 5 scale to generate predictions.

Keywords: Random Forest, Deep Learning, Policy Implications.

BACKGROUND

As a developing nation Technology-driven ride-sharing services did not exist in Bangladesh until 2015. Accessibility, affordability, enhanced short-distance mobility, use of ride-sharing applications (such as Uber, Pathao, Shohoj, Obhai etc.), increased imports, local manufacturing, and policies that are supportive of business are predecessors for instigating the motorcycle revolution in this South-Asian developing country. According to a recent study by Policy Research Institute, Bangladesh's ridesharing market is estimated to make up 23% of the country's transportation economy (*1*). Moreover, bikes' accessibility was improved by E-commerce companies during and after the COVID epidemic with price reduction, when there was also a serious unemployment problem.

Meanwhile, with economic growths, urban regions of Bangladesh are facing an influx of automobiles, greater traffic congestion and these are overwhelming the country's existing infrastructure and traffic management systems. Due to corruption, limited research, lack of political will, insufficient funding, poor risk perception and poor law enforcement, Bangladesh has the world's greatest death rate from motorcycle accidents (*2*). Fifty percent of bikers in Bangladesh do not have licenses (*3*). Majority of the deceased are from young generation. Not only because of riders' traffic offences and irresponsibility but also riding between lanes, right of way dispute, visibility problem in blind spots, poor road condition and drainage problems, poor intersection management, poorly managed roads, inadequate sign, marking, lighting, unsafe vehicles, lack of training and awareness, lack of comprehensive data collection and targeted interventions are responsible for increased safety threat from motorcycles. Therefore, proper policy making regarding traffic control and law enforcement can offset motorbike involved crashes properly.

Numerous studies have investigated how various factors influence motorcycle safety. Ozkan et al. (2012) attempted to determine the structure of the Motorcycle Rider Behavior Questionnaire (MRBQ) among Turkish riders and to examine the relationships between various categories of rider behavior, active and passive crashes, and motorcyclist offenses. The results indicated that MRBQ consisted of five factors: traffic errors, control errors, speed violations, unusual performance, and use of safety equipment (*4*). But this study doesn't include the road condition, weather condition etc. Dapilah et al. (2017) administered questionnaires to riders and conducted interviews with traffic and safety institution officials. This study examined the relationship between motorcyclist characteristics, how they behave in traffic, and traffic accidents. However, the traffic management system and other potential accident-causing factors are omitted from this investigation also (*5*). In another study nearly 30,000 current motorcyclists were asked to fill out a survey that looked at the relationship between accident risk and things like yearly mileage, age, experience, trip type, training, personal characteristics of the riders, and the riders' self-reported behaviors and attitudes (*6*). But this study did not consider the impact of external factors such as road design, traffic conditions, and traffic control system etc on motorcycle accident risk. At multiple road intersections in Sargodha, Pakistan, another questionnaire survey was conducted. The data were subsequently analyzed using multiple regression analysis, and three risk compensation hypothesis-based models were developed. Abid et al. (2002) evaluated the physical and behavioral effects of helmets on cyclists and riders. Other factors that can affect accidents, such as the condition of the roads, the weather, and visibility, are not considered in this study (*7*).

Moreover, various studies have been conducted on traffic safety. A Bayesian logistic regression model to figure out crashes (*8*), deep learning techniques to find likelihood of crashes (*9*), real-time crash risk prediction in arterials (*10*), variations in drivers' road safety perceptions based on sociodemographic traits and driving experience (*11*), driver behavior chain of violating traffic regulations, effect of traffic flow, traffic operation on violation behavior (*12*), Vietnam pedestrian safety attitudes, risk perceptions, and pedestrian behavior connections (*13*), adverse effects of climate change and acute weather on road traffic safety, leading to severe accidents (*14*), impact of extreme temperatures on medical attendances due to motorcycle crashes in Iran (*15*), impact of pavement condition, road geometry, and roadside conditions on accidents in Jordan (*16*), effects of cell phone usage , audio devices, chatting with passengers, smoking, eating, and drinking (*17*) were conducted.

However, most of the studies identified the primary contributors or characteristics in real-time crash prediction, examined the relationship between various real-time factors and crash incidents, and compared the efficacy of various models for classification. Meanwhile, only a small number of studies were conducted to determine the perceived risk scale (*18*, *19*, *20*, *21*). In prior studies, driving environment, weather, traffic control, driving behavior, and pedestrian-related characteristics have received less attention simultaneously. Incorrect recording in developing countries is another serious problem. Importing pre-accident factors from road users can have a significant impact on reducing the accident rate. By gathering the ratings of precursors from road users, a reliable model of precursors can be created that can be implemented in the design, management, planning, and policy for traffic control and law to offset motorbike crashes. This research bridges the research gap by exploring how different motorcycle risk factors can impact traffic control and law enforcement, how differently bike users, drivers, female, male perceive the causes of motorbike accidents on traffic control and law utilizing random forest algorithm and CNN 1D algorithm. Machine learning and deep learning algorithm were utilized for capturing complex relationships between data and better predictions. This study also developed deployment of machine learning model in public server so that policy makers and users can use the interface to predict 'traffic control and law' after input of rating on other features. Moreover, policy implications have been developed considering graphical interpretations and relationship between different attributes. Eventually, ameliorating such precursors may significantly offset motorcycle risk, prevent accidents, facilitate rapid results and sustainable transportation in line with SDG objectives.

DATA AND RESEARCH METHODOLOGY

Survey Administration & Descriptive Statistics

Online and offline based questionnaire survey were done for data collection. 'Google forms' was used for the preparation of the questionnaire online. It was supplied through online media: WhatsApp, QR code, Mail, Microsoft Teams, Outlook, Facebook, Messenger, and other social media platforms. Additionally, it gave respondents the freedom to complete the form online whenever it was convenient for them. Although there are certain issues with this technique, including difficulties with comprehension and interpretation, these issues have been attempted to be resolved by illustrative figures, drawings, formatting, styles, and translations into the native language. Additionally, a survey form for offline data collection through face-to-face surveys was developed.

The attributes were chosen from discussions with road users, press articles, academic studies, Road Safety Audit Guideline, interviews with motorcyclists, faculties in the Civil Engineering Department of Bangladesh University of Engineering and Technology (BUET), and policymakers. There were questions regarding 38 potential precursors associated with motorcycle accident risks. Translations were also provided for overcoming language problems for the nontechnical respondents. Figures were also provided for better understanding and interpretation. In the last section, there was space for writing personal opinions. After filtering, the remaining sample size was 1559.

The percentage of male and female were 82.10 % and 17.90 % respectively. Here in a developing country's situation female participation is lesser because of lesser road usage, reluctance to the motorbike related survey from uninterest. Moreover, ground reality also represents the unequal exposure and mobility due to conservativeness and local peculiarities (*22*). On the other hand, intentionally majority of the data was also received from 18-25 and 26-35 age ranged people. Because those are more concerned about risk and better data structure can be gotten from the young generation as they can access online platforms well. Responses were from the different parts of Bangladesh. So, the dataset can be representative of the real scenario. Among the total sample there were 809 users but not rider, 549 were nonusers and 201 numbers were riders of motorbike. There were 1199 samples from Dhaka/Mymensingh, 143 from Rajshahi/Rangpur, 103 from Chittagong, 92 from Barishal/Khulna, 22 from Sylhet regions of Bangladesh.

Random Forest

The Random Forest (RF) model is a widely used ensemble method employed for both classification and regression tasks (*23*, *24*, *25*). The choice rendered by a random forest model is contingent upon the collective decisions made by multiple decision trees within the ensemble. The RF algorithm is widely regarded as a suggested approach for achieving a steady selection of important components (*26*). In general, the process of constructing a decision tree can be broken down into three main steps:

1. Identification of the dependent variable and independent variables. The entirety of the data is centralized within the root node before to undergoing any division or separation.
2. The dataset is partitioned into child nodes by the utilization of a splitting method, such as the ID3 algorithm, which operates on attribute variables.
3. Each child node will be regarded as a parent node for subsequent splitting.

The splitting method aims to maximize homogeneity within the resulting child nodes. In this study, the ID3 algorithm was simulated using the entropy criterion in the Random Forest implementation to guide the splitting process. ID3 relies on entropy and information gain to select the best attribute for splitting and determine whether further splits are necessary [(22)]. Entropy quantifies uncertainty, where higher entropy implies greater uncertainty [(23)]. Assuming the presence of a variable S with c unique values, the entropy E(S) of S is computed as **Equation 1**:

$$E(S) = \sum_{i=1}^{c} -p_i log_2 p_i \tag{1}$$

Where p_i represents the probability associated with a specific value. The variable "i" represents the index number of the available alternatives.

If variable S is partitioned into subsets based on a certain attribute, the expected entropy EH quantifies the anticipated level of uncertainty of the resulting c outcomes of variable S after the partitioning process. This expected entropy is computed as **Equation 2**:

:

$$EH = \sum_{i=1}^{c} \frac{a_i}{a} \times (-p_i log_2 p_i) \tag{2}$$

Where "a_i" represents the quantity of observations within each subset, and "a" represents the total quantity of observations within the parent node "S".

The Information Gain (IG) from splitting based on attribute A is the reduction in entropy, as shown in **Equation 3**:

$$IG(S, A) = E(S) - EH \tag{3}$$

There is no need for additional splitting at a node with zero information gain, which is regarded as a terminal node. The dataset contained within the terminal node exhibits the highest degree of homogeneity. The decision tree correctly and consistently classifies the training data, achieving 100% accuracy.

Decision trees have a tendency to exhibit over-fitting on the training data, leading to mediocre results when applied to the entire dataset (*27*). One notable benefit of the Random Forest (RF) algorithm is its ability to effectively address the issue of overfitting while maintaining a high level of prediction accuracy (*28*).

Four steps can typically be seen in the formation of an RF:

1. The bootstrapping method is employed to do random resampling of a sample, ensuring that the resampled dataset is of equal size to either the entire dataset or the training dataset. In this study, the Random Forest Classifier in scikit-learn implicitly handles bootstrapping for resampling data to build each tree.
2. The random subspace method is employed to choose K attributes from a pool of M attributes, where K << M (usually K is selected to represent the square root of M). The max_features parameter has been used here to control this.
3. The construction of a decision tree involves the utilization of a bootstrapping sample and the selection of attributes identified in steps 1 and 2. The ID3 algorithm is utilized in the construction of each tree. In this study, the criterion='entropy' parameter has been added to the

Random Forest Classifier to ensure that the entropy measure is used for splitting nodes, like the ID3 algorithm.

4. Repeating steps 1 through 3 to construct a large number of trees until the required RF is achieved. By specifying n_estimators=100, the code ensures that 100 trees are constructed. The Random Forest Classifier automatically handles the construction and aggregation of these trees.

The model was trained to predict the target variable traffic control and law enforcement, which was constructed from a set of four survey-derived attributes, each rated on a scale of 1 to 5. The Random Forest model classified the responses into five ordinal categories (1–5) representing increasing levels of perceived concern.

The dataset was split into training and testing sets (80% training, 20% testing). Model performance was evaluated using classification accuracy, confusion matrix, and a detailed classification report including precision, recall, and F1-score. Feature importance scores were extracted to identify the most influential predictors, such as traffic movement, biker driving behavior, sign marking and lighting, and others.

CNN 1D

To analyze the output, another model was used in this current study which is Convolutional Neural Network (CNN). In this study, the model has 5 Dense layers,1 MaxPooling1D layer and 2 output layers. The rectified linear unit (ReLU) function (*29*) was employed to activate the outputs of the convolutional neural network (CNN) and dense layers, facilitating a non-linear modification of the extracted features (**Equation 4**):

$$f(z) = \begin{cases} z, z > 0 \\ 0, z \leq 0 \end{cases} \tag{4}$$

Where, z is a feature computed by a CNN or dense layer.

The SoftMax function (*30*) was employed in the output layer to calculate the probability of the input belonging to one of the predetermined classes in this model. To maximize the weights of the model, the feedforward and backpropagation approach (*31*) was employed during the training process. In the feedforward process, a collection of samples is inputted into the model, which subsequently generates an output. This output is then evaluated using a loss function to determine the extent of error in the prediction. The employed model utilized the categorical cross-entropy loss function (**Equation 5**):

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N\_classes}\{j = Y^{(j)}\}\log(y_j^{(i)}) + \frac{\lambda}{2n}\sum_{k=1}^{m} w_k^2 \tag{5}$$

Where N is the number of samples, Y is the actual output (ground truth or measured output) and y is the model prediction. The second component in the loss function incorporates L2 regularization, a technique employed to mitigate overfitting in the model. The variable w represents the weight of the model, while λ is a tunable parameter. This parameter determines the extent to which the regularization term contributes to the categorical cross-entropy. The optimization technique known as Adaptive Moment Estimation (Adam), proposed by Kingma and Ba (2014), was employed to minimize the categorical cross-entropy loss function throughout the backpropagation process. It is important to acknowledge that the adjustable parameters in this study include the quantity of hidden layers, the number of feature maps within each layer, the sizes of the convolutional neural network (CNN) kernels, the pooling and striding sizes, and the regularization parameters. These parameters were optimized by empirical methods to get the most optimal values. The experimental technique employed in this study involved tweaking one parameter while keeping the remaining parameters fixed.

Feature Engineering & Model Deployment

This research applied feature engineering by aggregating 38 individual precursors into 10 major categories. The aggregation was performed by averaging the ratings of relevant variables to create new, representative features. These ratings were collected using a Likert scale, where participants evaluated different risk factors on a scale from 1 (Very Low/Strongly Disagree) to 5 (Very High/Strongly Agree). The numerical responses were then used as inputs in the machine learning mode. Feature engineering was applied targeting more accuracy of the model. Detailed feature engineering is portrayed in Figure 1. Moreover, Figure 2 reflects the overall workflow of the study. Deployment of machine learning model in a public server was accomplished for user interface allowing road users and planners' access to generate predictions for new data points. In Figure 3, user interface of the deployment is represented.

Online and Offline Questionnaire Survey: https://drive.google.com/drive/u/0/folders/1OvJzr-6V4_9MVPobaTCC_NOf2dHoiflp

Sample distribution and descriptive Statistics:
https://drive.google.com/drive/u/0/folders/1YMdswRArDxeWSyFSfhFZTnWB0lXhxm9B

The Deployment Server Link is https://noman74.pythonanywhere.com/

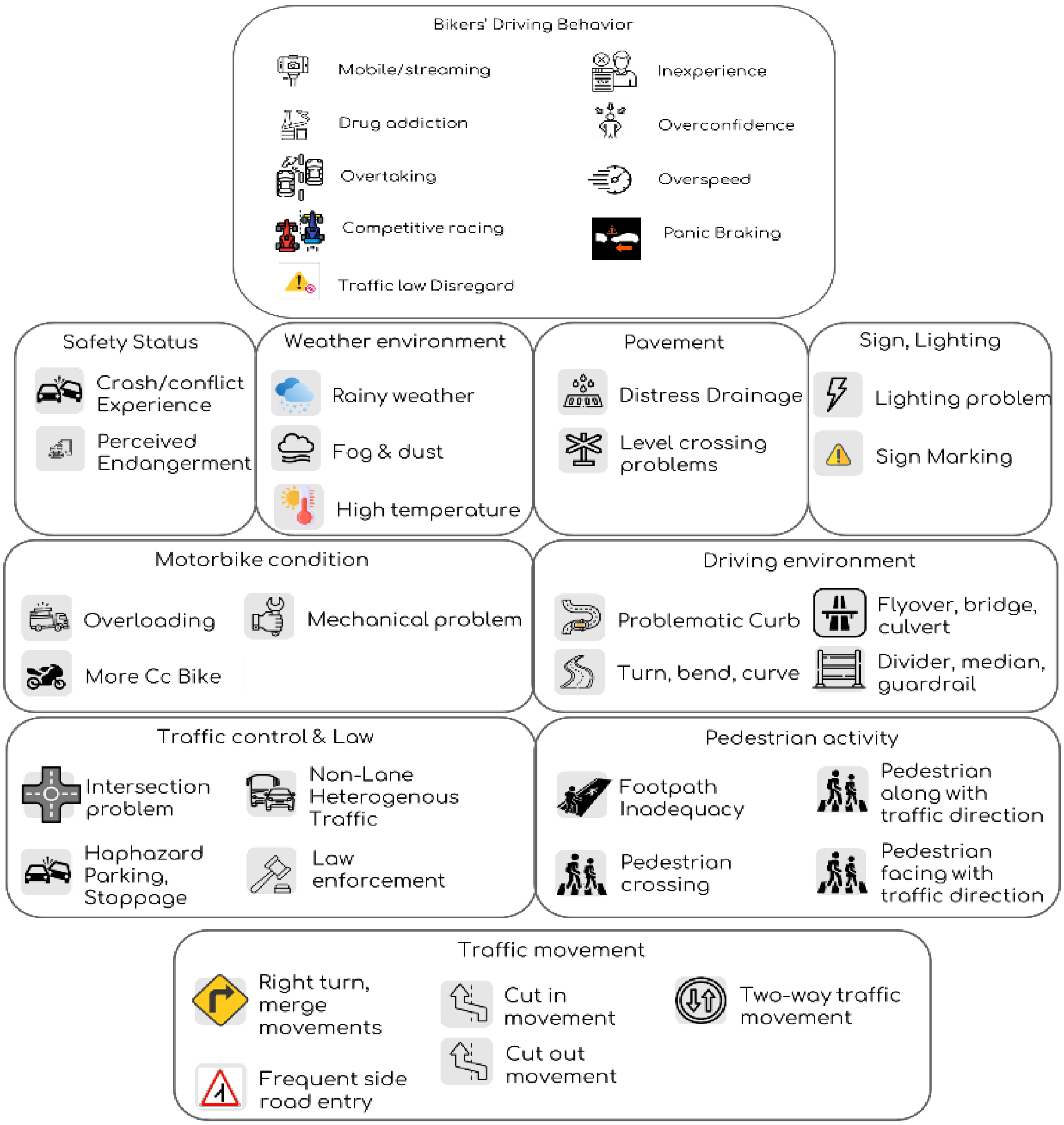


Figure 1 Feature Engineering

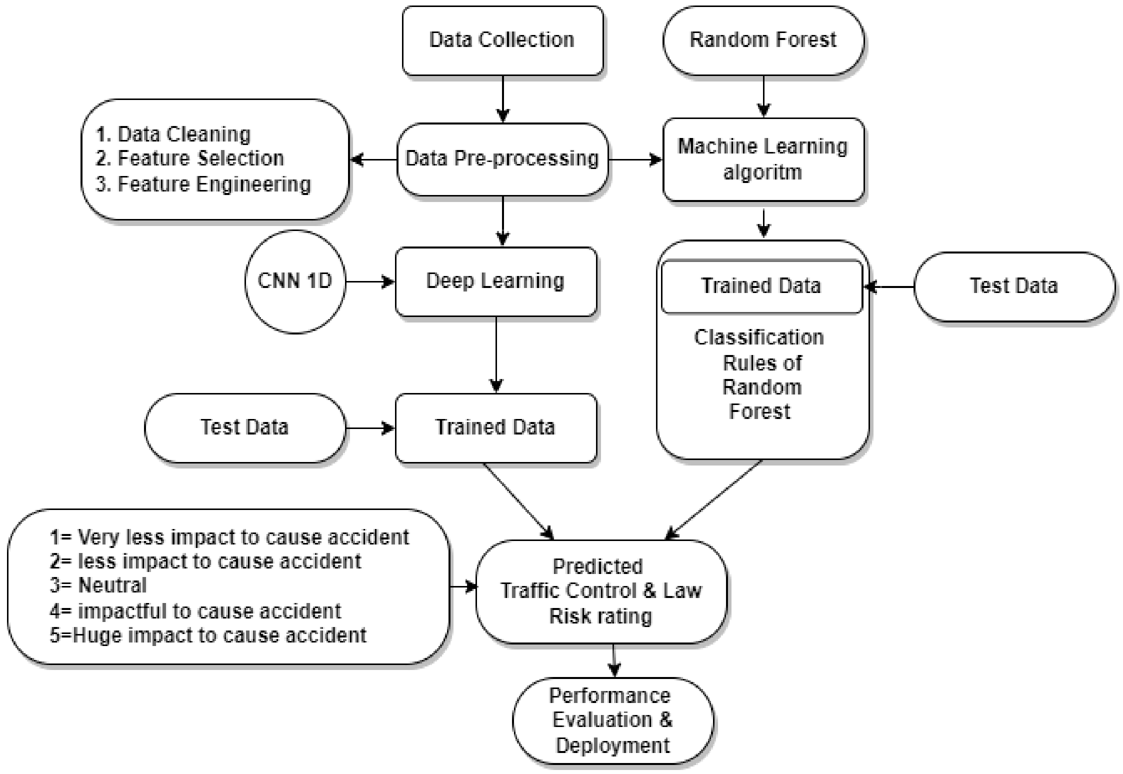


Figure 2 Methodology

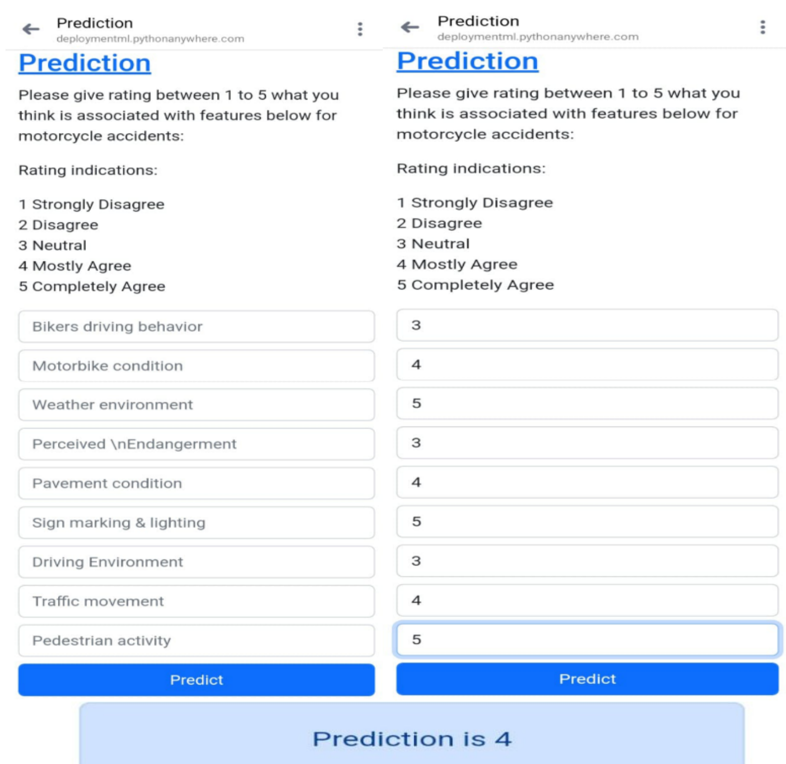
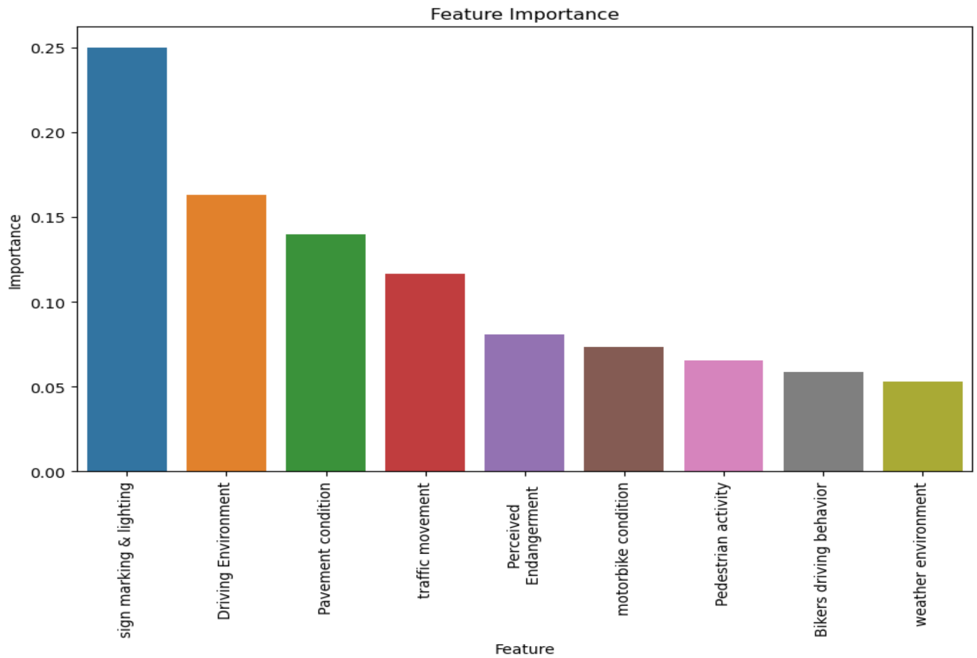


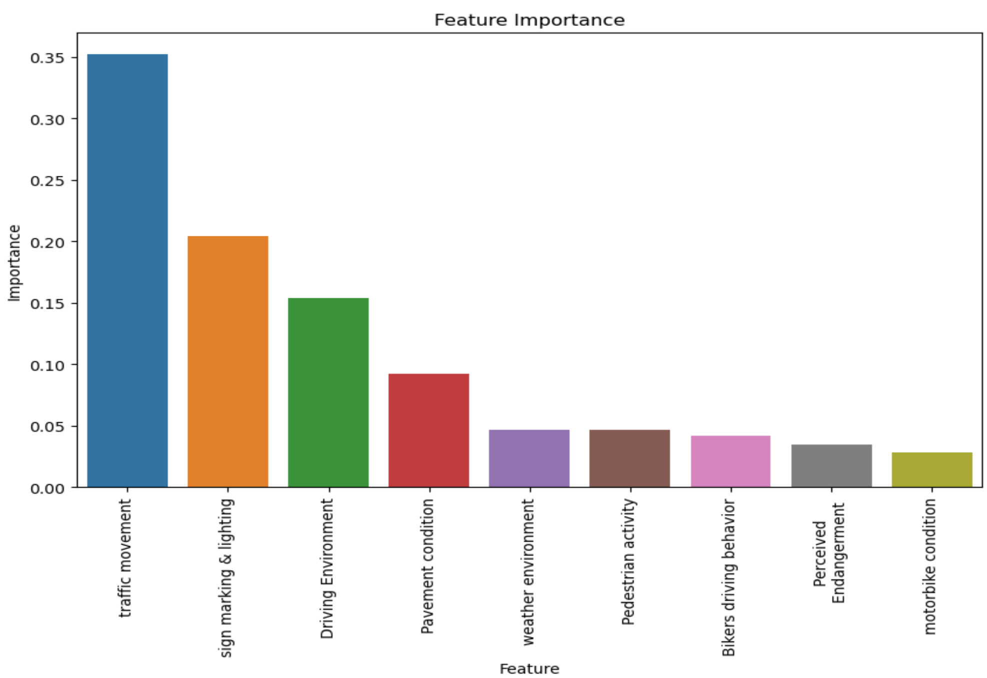
Figure 3 Server interface of Model Deployment

ANALYSIS, RESULT INTERPRETATION AND DISCUSSION

This research incorporated performance of Random Forest Classification and CNN 1D for target variable "Traffic control and law". Random Forest algorithm's accuracy was 74.36% after selecting 20% of the sample data as test data randomly and deep learning's accuracy was 69% after taking 30% of the sample data as test data. Maximum accuracy was confirmed after different trials with different percentage of test set.



(a)



(b)

Figure 4 Feature Importance with respect to (a) Riders (b) Users but not Riders

The relative contribution of features towards the ‘Traffic control & law’ can be explained by feature importance. Figures 4(a) and 4(b) show the relative importance of each parameter in descending order from motorcycle drivers’ and motorcycle users’ responses respectively.

Feature importance with respect to motorcycle riders (drivers) reveals that they consider ‘Sign marking & lighting’ as the most important feature in predicting the target variable ‘Traffic control & law’. This indicates that riders strongly associate clear road signs and proper lighting with effective traffic regulation. For them, visible infrastructure signals a managed and enforced environment, while its absence implies weak enforcement. The second most important feature is ‘Driving Environment’, reflecting concern for road design elements like curves, dividers, and flyovers that directly impact motorbike handling and safety.

Following these, ‘Pavement condition’ and ‘Traffic movement’ are also influential, suggesting riders value smooth road surfaces and organized vehicle flow. Other features such as ‘Perceived Endangerment’, ‘Motorbike condition’, ‘Pedestrian activity’, ‘Bikers' driving behavior’, and ‘Weather environment’ show progressively lower influence. Overall, the feature importance distribution shows that multiple factors affect how riders perceive traffic control and law enforcement, and addressing any of these can improve road safety for motorcyclists.

In contrast, feature importance with respect to motorcycle users (passengers/general road users) shows that ‘Traffic movement’ is the most important feature influencing their perception of ‘Traffic control & law’. This includes right turns, merging, cut-in/cut-out behavior, and two-way traffic, all of which affect safety from a passenger's point of view. These users rely on the rider's ability to navigate these conditions safely and interpret proper traffic flow as a sign of regulatory control. The second most important feature for users is ‘Sign marking & lighting’, indicating that visible infrastructure remains a key element in perceived traffic order.

Other influential features for users include ‘Driving Environment’, ‘Pavement condition’, and ‘Weather environment’, though with lesser weights than in the rider model. Features like ‘Pedestrian activity’, ‘Bikers' driving behavior’, ‘Perceived Endangerment’, and ‘Motorbike condition’ have lower importance in the user model.

Feature importance for the whole dataset was also developed. It reveals that traffic movement holds the highest significance in this prediction, followed by sign marking and lighting. Additionally, driving environment, pavement condition, bikers driving behavior, weather environment, pedestrian activity, and motorbike condition also impact traffic control and law progressively lower. Overall, the feature importance values emphasize the collective influence of these factors on traffic control and law, underscoring the potential for improving road safety by addressing any of these contributing elements.

A chi-square test was used to examine the significance of the association between Traffic Control and Law and other contributing variables. The test was conducted using contingency tables based on the frequency of responses, with calculations performed using Python. The test results, presented in Table 1, confirm that all variables are statistically significant (p-value < 0.05), supporting the appropriateness of the selected attributes.

Table 1: Results of Chi-square Tests

Variable	Chi-square Value	P-value
Traffic Movement	895.31	5.84×10^{-184}
Sign Marking & Lighting	581.27	1.06×10^{-116}
Driving Environment	580.02	1.95×10^{-116}
Pavement Condition	489.55	3.71×10^{-97}
Bikers Driving Behavior	393.12	1.08×10^{-76}
Pedestrian Activity	147.34	1.96×10^{-25}
Perceived Endangerment	136.45	3.12×10^{-23}
Motorbike Condition	95.76	3.75×10^{-15}
Weather Environment	82.57	1.33×10^{-12}

According to the violin plots of figure 5, the majority of respondents, including nonusers, users, and riders, believe that traffic laws and legislation play a significant role in causing motorcycle accidents, scoring this component between 4 and 5 though there are contrasting feature importance on perceived risk from these two different road users.

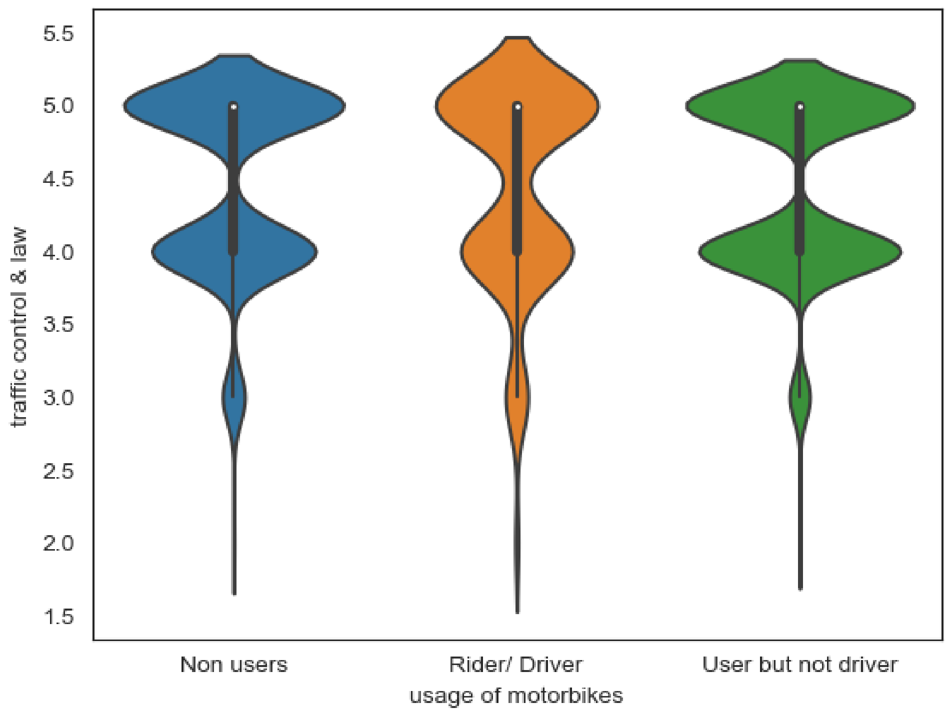


Figure 5 Violin plots for rating of 'Traffic control and Law' in terms of three type of road users

Figure 6 displays kernel density plots illustrating the relationship between traffic control & law and nine key influencing factors. The x-axis represents the rating of traffic control & law, while the y-axis shows each predictor. Darker blue areas indicate where responses are more concentrated.

From the plots, it is evident that respondents who rated traffic control & law highly (4 or 5) also tended to give high ratings to most other factors. This clustering suggests that these elements play a significant role in shaping how people perceive traffic regulation. On the other hand, lower ratings (1 or 2) show more scattered patterns, indicating less consensus among respondents when they perceive weaker traffic control.

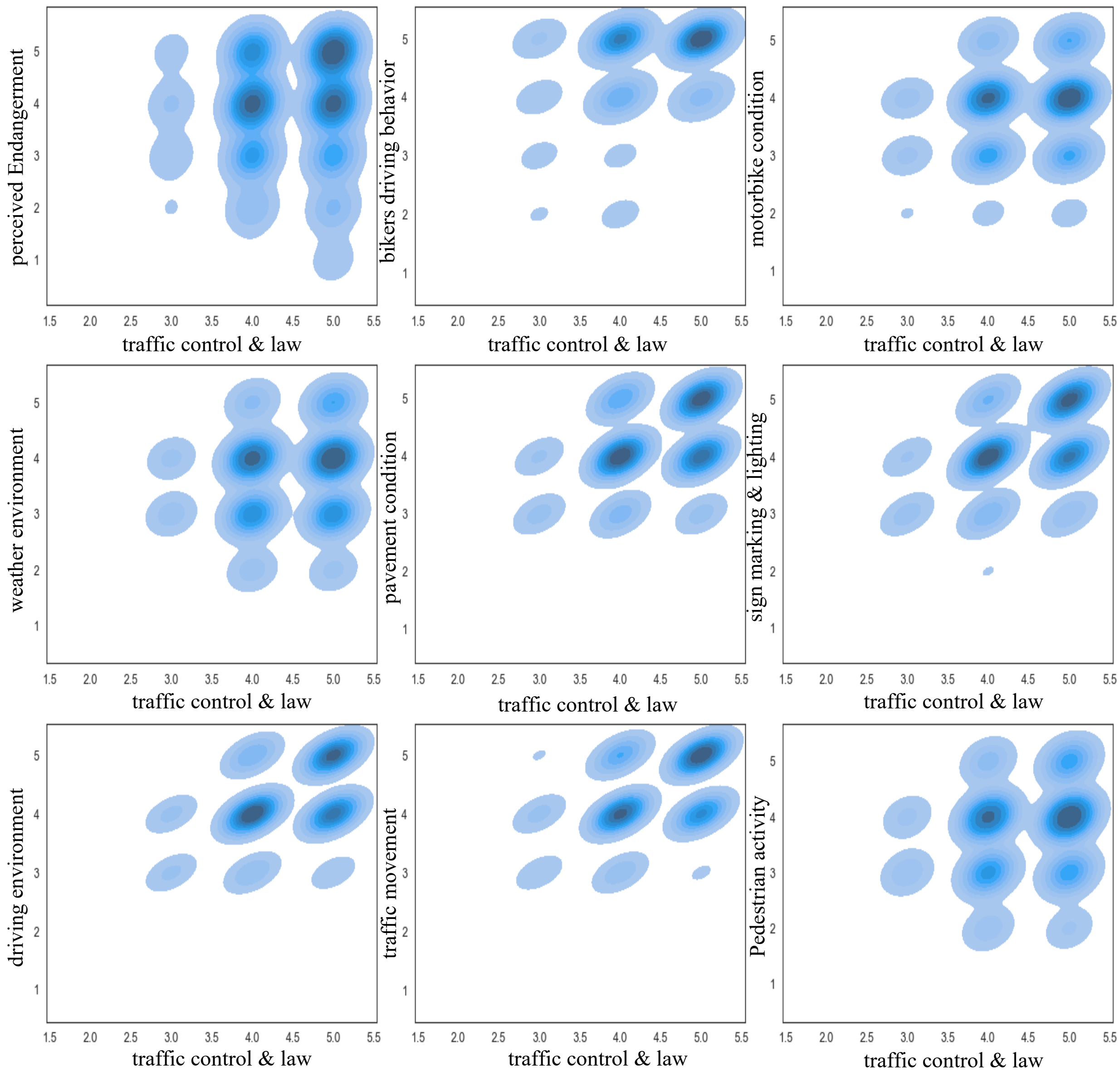


Figure 6 Kernal Density Estimation Plot

The self-reported accident probability of motorbikes due to traffic control and law was analyzed with scatterplots of figure 7. These scatter plots show the relationship between the top 3 predictors of traffic control and law and the traffic control and law ratings of 4 and 5. Each circle in the plots reflects the likelihood of reporting a rating of 5 for traffic control and law enforcement in the left subplots and a rating of 4 in the right subplots. The data points for ‘sign marking and light’, ‘traffic movement’ and ‘driving environment’ with a rating of 5, tend to cluster predominantly in the upper part of the plots (left subplots). This clustering suggests that people are more inclined to give a rating of 5 for traffic control and law under these conditions. A similar pattern is observed for a traffic control and law rating of 4. If someone assigns a rating of 4 to all those attributes, the likelihood of giving a rating of 4 for traffic control and law is also notably high. Conversely, for ratings of 1 and 2 of predictors, very few data points are observed, and they appear to be clustered in the lower probability region. Therefore, these ratings provide a lower probability of accident risk due to traffic control and law because of reporting a rating between 1 to 3 for traffic control and law.

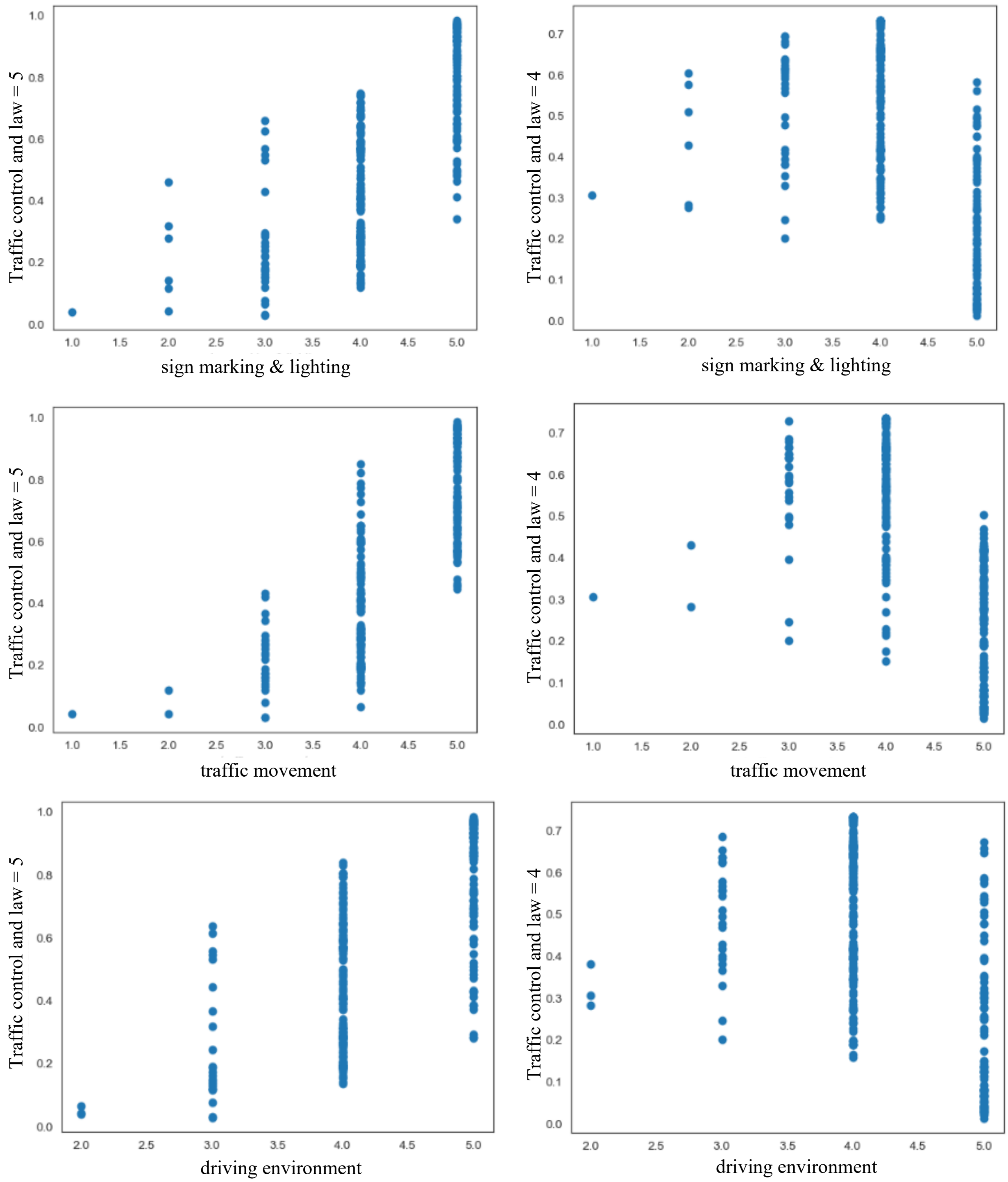


Figure 7 Scatter plots

A heatmap in figure 8, displaying the linear association between several attributes and the attribute 'traffic control and law' was evaluated as part of the pairwise correlation analysis. With a value of 0.49, the attribute with the highest correlation is ‘traffic movement’. Additionally, with a correlation coefficient of 0.47, the ‘driving environment’, ‘sign marking and lighting’ show a strong positive linear association. With a correlation coefficient of 0.42, the third most important association is found between 'pavement condition' and 'traffic control and law'. 'Perceived endangerment', on the other hand, is the feature least associated with the 'traffic control and laws'.

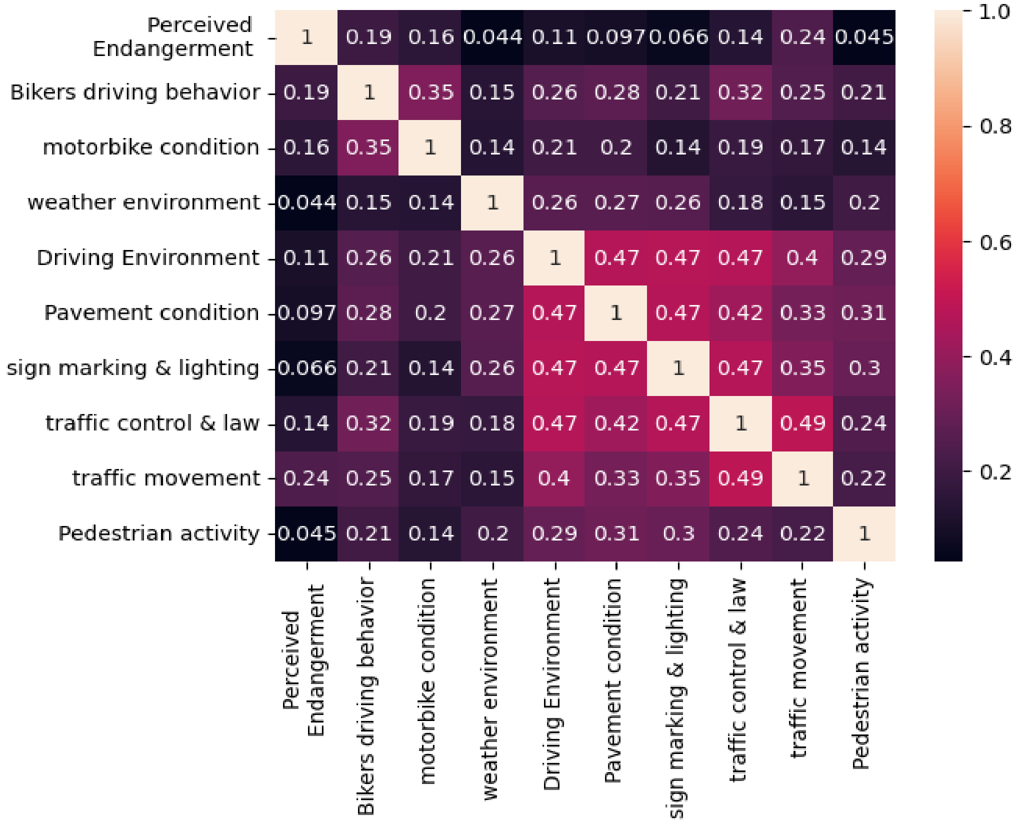


Figure 8 Correlation Matrix Heatmap

The SHAP plot illustrates the influence of different attributes on ‘traffic control and law’ rating 5. It reveals that ‘traffic movement’, ‘sign marking and lighting’, ‘driving environment’, and ‘pavement condition’ have a more pronounced impact on the rating of 5 compared to other attributes. When the ratings of these attributes increase, the model becomes more accurate in predicting ‘traffic control and law’ rating 5.

SHAP values for both male and female dataset from figure 9(b) and figure 9(c) reflect that the top 5 attributes (Traffic movement, sign marking & lighting, driving environment, pavement condition, and bikers' driving behavior) have similar significance in predicting traffic control and law. An important result from the female SHAP plot is the chance of providing a rating of 5 in the ‘traffic control and law’ rating increases with the ‘perceived endangerment’ rating, demonstrating that women view motorcycle accidents due to traffic control and law as riskier than males.

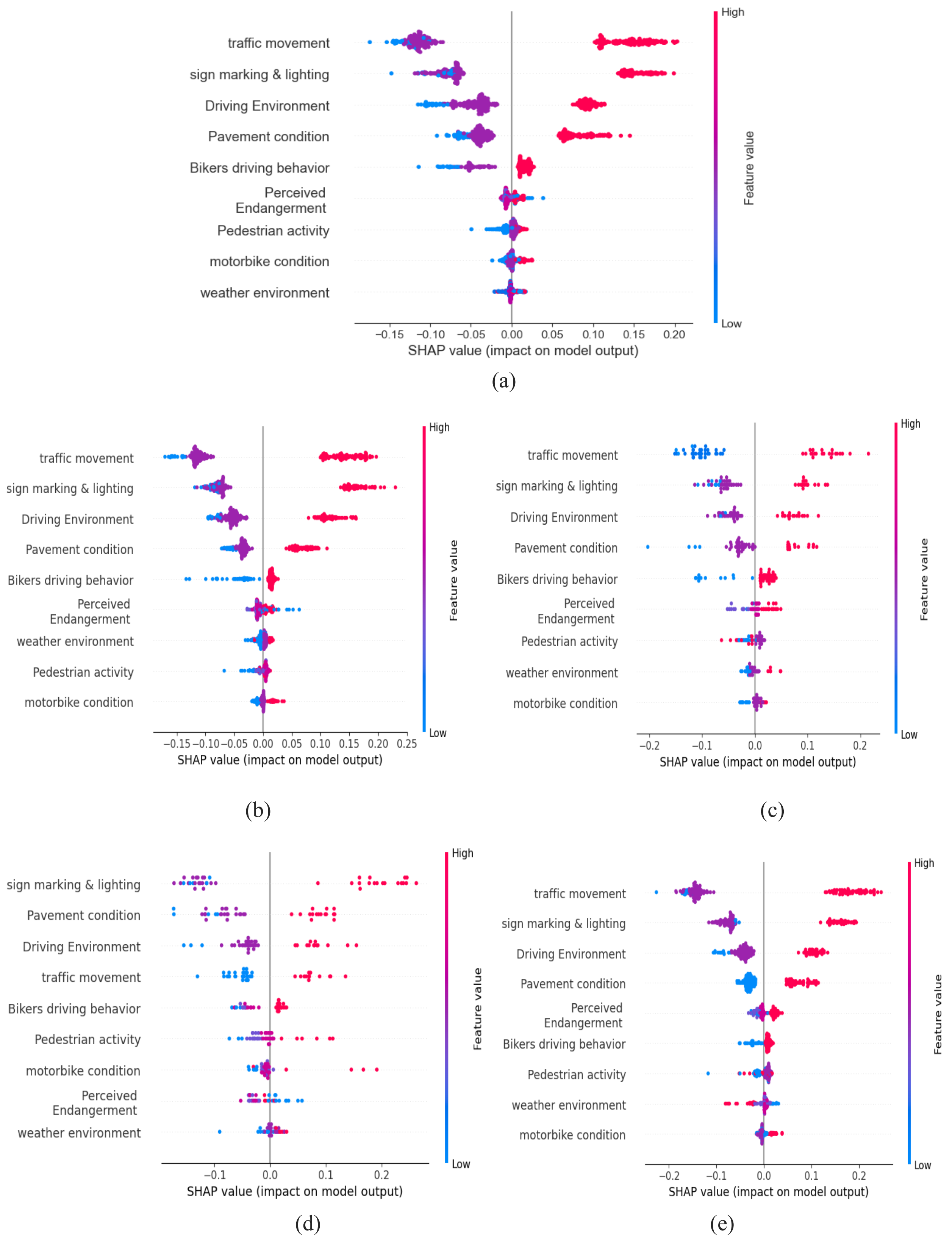


Figure 9 SHAP plots for rating 5 in 'traffic control and law' with respect to (a) driver (b) male (c) female (d) rider (e) user but not rider

The SHAP plots in figure 9(d) and figure 9(e) also highlight a notable disparity between the perceptions of risk held by motorbike users and riders. For users (but not drivers) of motorbikes, as the perceived endangerment rating increases, the probability of giving a rating of 5 also increases. This suggests that motorbike users tend to feel more concerned about potential risks and dangers associated with motorbike accidents. On the other hand, when considering drivers of motorbikes, the probability of giving a rating of 5 increases when the perceived endangerment rating decreases. This implies that drivers of motorbikes may feel less apprehensive about the risk of crashes compared to motorbike users. Again, bike users think weather environment doesn't create impact on traffic control and law violation which is depicted from SHAP plot with respect to motorbike user.

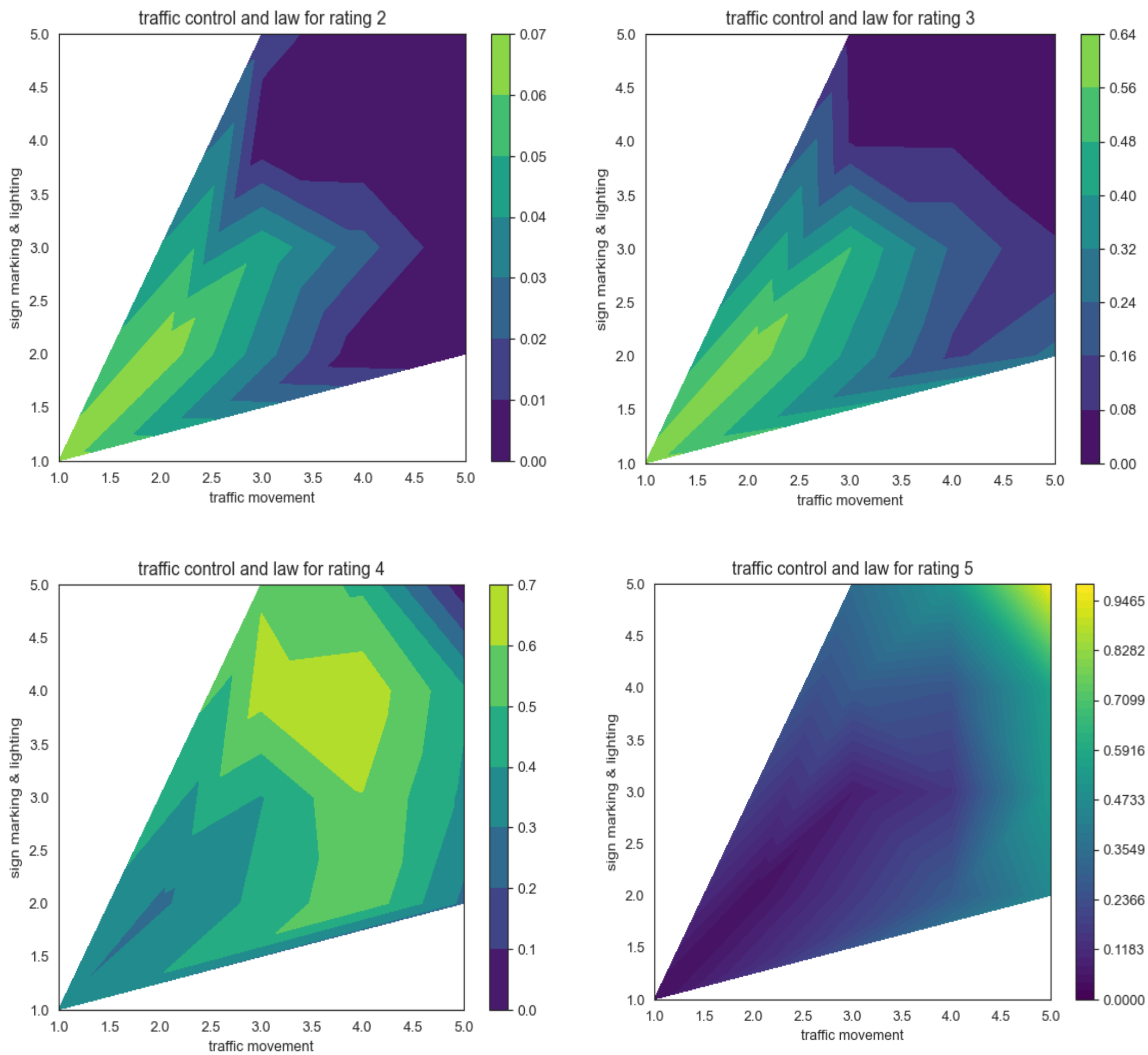


Figure 10 Contour Maps for 'Traffic Control and law' rating probability with 'traffic movement' and 'Sign marking and lighting'

From Figure 10, it is evident that when an individual assigns a rating of 5 to both 'traffic movement' and 'sign marking & lighting', there is a higher probability (nearly 0.95) of them also giving a rating of 5 in "traffic control and law". This implies a strong positive correlation between these attributes with a positive opinion of traffic movement and sign visibility being linked to a positive assessment of 'traffic control and law'. Similarly, if someone rates both ‘traffic movement’ and ‘sign marking & lighting’ with a rating of 4, there is a higher likelihood (nearly 0.7) of them assigning a rating of 4 to "traffic control and law" as well. Therefore, all figures illustrate a pattern where the evaluations of ‘traffic control and law’ align with the ratings given to ‘traffic movement’ and ‘sign marking & lighting’.

ROC curve (receiver operating characteristic curve) was also plotted (figure 11 and figure 12) for both random forest model and CNN model. CNN performs better to predict each class except class 4 for which AUC (Area Under the ROC Curve) value is less than Random Forest model’s AUC value.

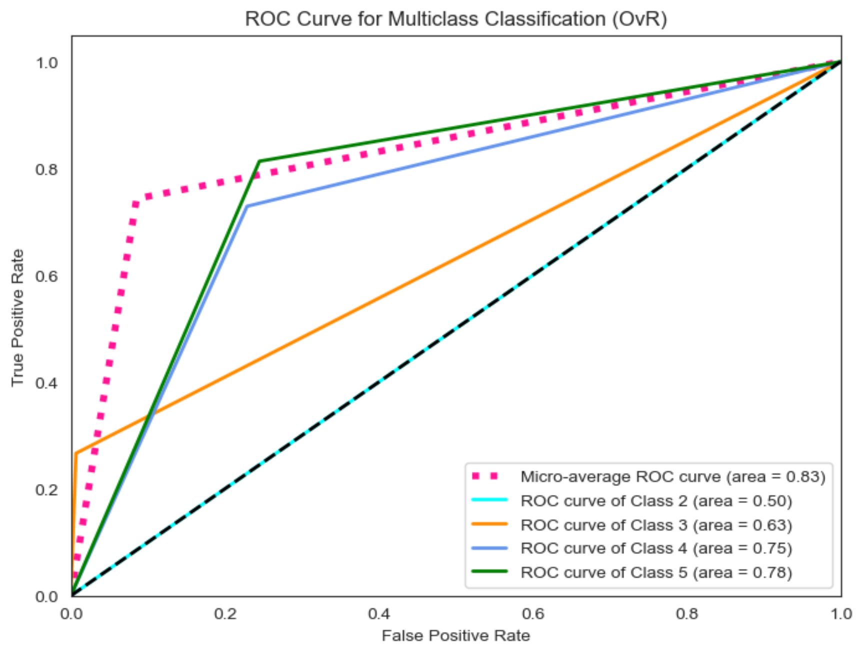


Figure 11 ROC curve for Random Forest model

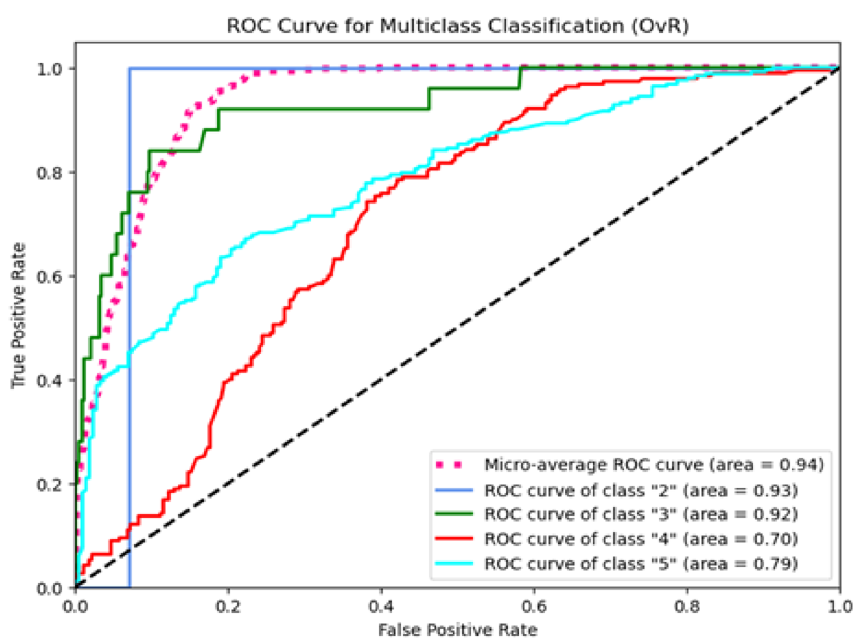


Figure 12 ROC curve for Deep Learning

CONCLUSIONS & POLICY IMPLICATIONS

With increasing short trip demand and usage of ride sharing motorcycle services, motorcycle safety impacts much on transport sector of urban areas of developing country like Bangladesh. In Bangladesh, risk issues regarding motorcycle are to be handled cautiously as both crash rate and death rate are rising rapidly. As a developing country Bangladesh are facing problems of inadequate structure, overcrowded roads, unsafe vehicles and road designs, limited law enforcement, lack of public transport facilities, inadequate training, and licensing system, corruption, lack of awareness and education on accident and emergency response, political unwillingness, and limited research and development projects. To make the transportation sector sustainable, this research trained model with data from questionnaire and developed prediction models to scale perceived risk of motorcycle due to traffic control and law.

From model results different perceptions of bike users, drivers, females, and males were analyzed. Deployment of model in public server was accomplished so that policy makers and users can use the interface to predict 'traffic control and law' after input of rating on other features. Policy makers can take decisions about the improvements of traffic control and law enforcement for a specific route, hazardous regions.

According to the authors' knowledge, this is the first investigation into this topic in Bangladesh, despite the paucity of prior research on perceived motorcycle risk due to traffic control and law globally. Despite being in its early phases, this research represents a novel effort to improve the safety of motorcycles in developing nations by improving traffic control and law.

From the model, several key areas can help improve traffic control and road safety, especially for motorcyclists. First, both riders and users clearly believe that traffic signs and proper lighting are important for safe roads. When signs are clear and lights are working well, people feel that the area is well managed and traffic laws are being followed. Therefore, improving road signs, adding more street lights, and fixing broken ones should be a top priority. Also, smooth and well-planned traffic flow, including proper lane discipline, safe turning areas, and well-designed intersections, is seen as essential, especially by passengers who rely on the rider's ability to handle traffic safely.

In addition, road conditions and designs such as curves, flyovers, dividers, and pavement surfaces play a big role in how riders feel about safety. If the roads are damaged or poorly designed, motorcycle handling becomes risky. So, regular maintenance of roads, fixing potholes, and improving drainage systems are very important. Also, riders believe that law enforcement should check motorbikes for overloading, engine capacity, and mechanical problems to avoid crashes caused by poor vehicle condition. Meanwhile, the study shows that women and passengers feel more at risk when traffic rules are not properly enforced, especially when road conditions or signage are poor. Safety campaigns and traffic awareness programs should consider these concerns, focusing on both riders' responsibility and passenger safety.

Moreover, the study also found that riders often underestimate how their own driving behavior affects road safety, while users are more aware of this risk. This suggests that more rider training is needed, along with strict enforcement of traffic rules. Weather risks were not a major concern for users, but awareness should still be raised about how rain, fog, or heat can affect accidents. In short, making improvements in traffic movement, signs, road quality, and rider awareness, while also listening to user feedback, can help build a safer traffic system for everyone. Incorporating global strategies such as the "Vision Zero" and "Safe System" approaches, with cooperation and data sharing from the policies of developed countries, can further support efforts in developing nations to enhance traffic safety, as emphasized by the World Health Organization (28).

Lastly, these models can also be created for a particular network, region, or corridor of hazardous locations. A larger database with collaborated funding can improve interpretation of results and reliability of models. Similarly for other transportation modes this method can be applicable for improving service quality. For better accident prediction in multiple situations and implementation in decision-making and planning, advanced machine learning, deep learning, Structural Equation Modeling and artificial intelligence can be utilized with improved dataset for accident avoidance. As this is one of the first research on perceived traffic control and law risk in urban context, it may serve as the basis for more in-depth research in improving traffic control and law.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Md. Gausul Azam Noman, Md. Mushtaque Tahmid, Md Asif Raihan, Md. Shamsul Hoque; data collection: Md. Mushtaque Tahmid; analysis and interpretation of results: Md. Gausul Azam Noman, Md. Mushtaque Tahmid, Md Asif Raihan; draft manuscript preparation: Md. Gausul Azam Noman, Md. Mushtaque Tahmid, Md Asif Raihan. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Bangladesh Ride Sharing Market: With 7.5m Rides a Month, Ridesharing Services Take over Bangladesh | The Business Standard. https://www.tbsnews.net/economy/75m-rides-month-ridesharing-services-take-over-bangladesh-45453. Accessed Aug. 1, 2023.
2. Hossain, S. Bangladesh Counts Highest Death Rates from Bike Accidents. *Prothomalo*. https://en.prothomalo.com/bangladesh/accident/bangladesh-counts-highest-death-rates-from-bike-accidents. Accessed June 27, 2024.
3. Adhikary, T. S. Boom of Unlicensed Bikers. *The Daily Star*. https://www.thedailystar.net/news/bangladesh/news/boom-unlicensed-bikers-3020251. Accessed June 28, 2023.
4. Özkan, T., T. Lajunen, B. Doğruyol, Z. Yıldırım, and A. Çoymak. Motorcycle Accidents, Rider Behaviour, and Psychological Models. *Accident Analysis & Prevention*, Vol. 49, 2012, pp. 124–132.
5. Dapilah, F., B. Y. Guba, and E. Owusu-Sekyere. Motorcyclist Characteristics and Traffic Behaviour in Urban Northern Ghana: Implications for Road Traffic Accidents. *Journal of Transport & Health*, Vol. 4, 2017, pp. 237–245.
6. Sexton, B. F., C. J. Baughan, M. A. Elliott, and G. Maycock. The Accident Risk of Motorcyclists. 2004.
7. Abid, M. M., S. Bano, A. Kumar, M. Iqbal, M. L. U. R. Shahid, A. Javed, and M. A. U. R. Tariq. Trend towards Helmet Usage and the Behavior of Riders While Wearing Helmets. *Sustainability*, Vol. 14, No. 22, 2022, p. 14713. https://doi.org/10.3390/su142214713.
8. Wang, L., M. Abdel-Aty, Q. Shi, and J. Park. Real-Time Crash Prediction for Expressway Weaving Segments. *Transportation Research Part C: Emerging Technologies*, Vol. 61, 2015, pp. 1–10.
9. Huang, T., S. Wang, and A. Sharma. Highway Crash Detection and Risk Estimation Using Deep Learning. *Accident Analysis & Prevention*, Vol. 135, 2020, p. 105392.
10. Li, P., M. Abdel-Aty, and J. Yuan. Real-Time Crash Risk Prediction on Arterials Based on LSTM-CNN. *Accident Analysis & Prevention*, Vol. 135, 2020, p. 105371.
11. Thibenda, M., D. M. P. Wedagama, and D. Dissanayake. Drivers' Attitudes to Road Safety in the South East Asian Cities of Jakarta and Hanoi: Socio-Economic and Demographic Characterisation by Multiple Correspondence Analysis. *Safety science*, Vol. 155, 2022, p. 105869.
12. Hung, K. V. Education Influence in Traffic Safety: A Case Study in Vietnam. *IATSS research*, Vol. 34, No. 2, 2011, pp. 87–93.
13. Dinh, D. D., N. H. Vũ, R. C. McIlroy, K. A. Plant, and N. A. Stanton. Effect of Attitudes towards Traffic Safety and Risk Perceptions on Pedestrian Behaviours in Vietnam. *IATSS research*, Vol. 44, No. 3, 2020, pp. 238–247.
14. Zou, Y., Y. Zhang, and K. Cheng. Exploring the Impact of Climate and Extreme Weather on Fatal Traffic Accidents. *Sustainability*, Vol. 13, No. 1, 2021, p. 390. https://doi.org/10.3390/su13010390.
15. Zare Sakhvidi, M. J., J. Yang, D. Mohammadi, H. FallahZadeh, A. Mehrparvar, M. Stevenson, X. Basagaña, A. Gasparrini, and P. Dadvand. Extreme Environmental Temperatures and Motorcycle Crashes: A Time-Series Analysis. *Environmental Science and Pollution Research*, Vol. 29, No. 50, 2022, pp. 76251–76262.
16. Al-Masaeid, H. Impact of Pavement Condition on Rural Road Accidents. *Canadian Journal of Civil Engineering*, Vol. 24, 1997, pp. 523–531. https://doi.org/10.1139/cjce-24-4-523.
17. Shaaban, K., S. Gaweesh, and M. M. Ahmed. Investigating In-Vehicle Distracting Activities and Crash Risks for Young Drivers Using Structural Equation Modeling. *PLoS one*, Vol. 15, No. 7, 2020, p. e0235325.
18. Tahmid, M. M., M. A. Raihan, S. Salam, A. Muktadir, and M. S. Hoque. *Analysis of Heterogenous Motorcycle Risk Perception and Crash Exposure in Developing Country's Urban Driving Environment: Precursors and Policy Implications Using Structural 2 Equation Modeling 3*. 2024.

19. Tahmid, M. M., F. Al Mahmud, O. Chowdhury, and M. A. Raihan. Analysis of Mode Reliability Factors among Off-Campus Students Using Structural Equation Modeling in Dhaka City. *Engineering Proceedings*, Vol. 56, No. 1, 2023, p. 259.
20. Noman, G. A., Md. M. Tahmid, Md. A. Raihan, and Md. S. Hoque. Artificial Intelligence-Based Perceived Motorcycle Risk Prediction in Bangladesh's Urban Driving Environment. In *Proceedings of the 2nd International Conference on Advances in Civil Infrastructure and Construction Materials (CICM 2023), Volume 2* (M. S. Alam, G. M. J. Hasan, A. H. M. M. Billah, and K. Islam, eds.), Springer Nature Switzerland, Cham, pp. 379–388.
21. TAHMID, M. M., G. A. Noman, and A. Mahmud. Perceived Motorcycle Risk Prediction Using Structural Equation Modeling and Artificial Intelligence in the Urban Driving Environment of Bangladesh. 2023.
22. What Do Gender Data Reveal about the Economic Struggles of Women in Bangladesh? *World Bank Blogs*. https://blogs.worldbank.org/en/opendata/what-do-gender-data-reveal-about-economic-struggles-women-bangladesh. Accessed July 28, 2024.
23. Prajwala, T. R. A Comparative Study on Decision Tree and Random Forest Using R Tool. *International journal of advanced research in computer and communication engineering*, Vol. 4, No. 1, 2015, pp. 196–199.
24. Kabir, E., S. Guikema, and B. Kane. Statistical Modeling of Tree Failures during Storms. *Reliability Engineering & System Safety*, Vol. 177, 2018, pp. 68–79.
25. Tahmid, M. M. Regression Analysis of Superconductivity Data. *Available at SSRN 5617850*, 2025.
26. Gehl, P., F. Cavalieri, and P. Franchin. Approximate Bayesian Network Formulation for the Rapid Loss Assessment of Real-World Infrastructure Systems. *Reliability Engineering & System Safety*, Vol. 177, 2018, pp. 80–93.
27. CPSC 540: Machine Learning. https://www.cs.ubc.ca/~nando/540-2013/lectures.html. Accessed Aug. 1, 2023.
28. Wielenga, D. 073-2007: Identifying and Overcoming Common Data Mining Mistakes. 2007.
29. Xu, B., N. Wang, T. Chen, and M. Li. Empirical Evaluation of Rectified Activations in Convolutional Network. *arXiv preprint arXiv:1505.00853*, 2015.
30. Al-Sarayreh, M., M. M. Reis, W. Q. Yan, and R. Klette. Potential of Deep Learning and Snapshot Hyperspectral Imaging for Classification of Species in Meat. *Food Control*, Vol. 117, 2020, p. 107332.
31. Goodfellow, I., Y. Bengio, and A. Courville. *Deep Learning*. MIT press, 2016.