SEATBELT NON-COMPLIANCE BEHAVIORAL TYPOLOGY: WHAT DOES THE DATA TELL US?

Samia Sahid [*1], Md Asif Raihan [2]

[1] *Graduate Student, Department of Civil Engineering, Bangladesh University of Engineering and Technology (BUET), Bangladesh,*
e-mail: samiasahid14@gmail.com

[2] *Associate Professor, Accident Research Institute (ARI), Associate Director, Directorate of Students' Welfare (DSW), Bangladesh University of Engineering and Technology (BUET), Bangladesh,*
e-mail: raihan@ari.buet.ac.bd

***Corresponding Author**

ABSTRACT

Seatbelt use is a simple yet profoundly effective intervention to reduce road crash injuries and fatalities, a pressing challenge in developing countries like Bangladesh where compliance remains critically low. Despite clear evidence that wearing seat belts reduces injury risk by up to 50%, complex behavioral, attitudinal, and contextual factors drive widespread non-compliance, which remain insufficiently understood. This study addresses these gaps by leveraging a unique dataset of 400 survey responses from Bangladeshi drivers to unravel the multifaceted dynamics influencing seatbelt use. A two-stage machine learning framework was applied to explore the determinants of seatbelt non-use and to classify different patterns of non-compliance. In the first stage, a Random Forest model was developed using demographic variables (age, gender, income, education), behavioral factors (attitudes toward seatbelt use, risky driving practices), and contextual attributes (trip length, weather, time of day). Model interpretability was achieved through SHapley Additive exPlanations (SHAP), which identified key drivers of non-compliance such as younger age, complacency among experienced drivers, and situational effects like hot weather or short trips. The second stage uses hierarchical clustering to divide non-compliant drivers into different behavioral typologies according to SHAP attributes that have a high influence. For instance, "Young Risk-Takers" who value convenience over safety and "Situational Non-Compliers" who use seatbelts sporadically depending on the circumstances of the trip are examples of defined categories. A better understanding of how demographic and attitudinal characteristics interact to influence risk behaviors is made possible by these data-driven profiles. Beyond raising general awareness, our findings give policymakers and road safety authorities practical insights that are suited to the Bangladeshi environment, enabling them to create focused, evidence-based measures. Through the identification of particular high-risk groups, this research enables customized approaches, such targeted education campaigns for younger drivers and increased enforcement in high-risk situations, to promote seatbelt use and ultimately save lives.

Keywords: *Unsupervised Learning, Predictive Analysis, Behavioural Segmentation, SHAP Explainability, Traffic Safety.*

1. INTRODUCTION

Motor vehicle crashes are one of the significant causes of morbidity and mortality worldwide. Seat belt use is a simple and cost-effective safety measure that greatly reduces injuries and deaths in vehicle crashes. Synthesizing over 200 studies, an extensive crash database analysis estimated average fatality risk reduction of 56-61% and serious injury reduction of 51-55% across vehicle types when seat belts are used (Dissanayake & Ratnayake, 2007). Recent studies showed that seatbelts reduce fatal injuries by 45% and serious injuries by 50%, specifically lowering the prevalence and severity of traumatic brain injuries and cuts death risk for drivers and front passengers by 45–60% and cuts serious injury risk by around 50% in U.S. crash statistics (Ganti et al., 2021, CDC "Facts About Seat Belt Use", 2025). The use of seatbelt varies among different countries of the world and depends to some extent on the regulations of the place. Geographically, the lowest usage rates among drivers and passengers were observed in Asia, the Middle East, and Africa, while women drivers exhibited significantly higher compliance (51.47%) compared to men drivers (38.27%) (Kargar et al., 2023). In Bangladesh a mixed-method study of 363 road traffic injury victims found that severe injuries frequently involved the head (34.44% of highest severity ratings), correlating strongly with extremely low seatbelt use (10.24%) (Rozars et al., 2025)

Seatbelt use is shaped not only by legislation but also by behaviour, perception, and social context. Decisions to buckle up reflect demographic traits, cognitive judgments, and personal attitudes, with factors such as age, gender, education, and driving experience influencing risk-taking and safety habits (Özkan et al., 2006). Socio-cognitive factors-perceived accident risk, fatalistic beliefs, and attitudes toward law enforcement-also play key roles (Sheveland et al., 2020). In settings like Bangladesh, cultural norms, heat-related discomfort, and inconsistent enforcement further contribute to variability in seatbelt use.

Understanding driver behaviour is central to road safety, as behavioural factors strongly influence crash risk and compliance. Traditional methods-logistic regression, probit models, and SEM-have long been used to examine links between demographic, attitudinal, and behavioural variables (Lajunen & Özkan, 2011), but their linear assumptions limit their ability to model complex interactions. In contrast, machine learning has gained prominence for its capacity to handle large, nonlinear datasets. Recent studies show that ensemble models such as Random Forest and XGBoost outperform traditional regression approaches in predicting crash severity and driver risk by better capturing interactions among environmental, human, and contextual factors (Yan & Shen, 2022; Sun et al., 2023). In addition to supervised learning, unsupervised learning methods-especially cluster analysis-have also advanced understanding of driver behaviour by revealing latent groups with distinct risk and compliance patterns. Early psychological studies showed that low seatbelt use clusters with sensation seeking and rule-violating tendencies (Taubman-Ben-Ari et al., 2004). Later work using K-means cluster analysis on DBQ and DSI data identified subgroups such as "unsafe and offensive," "safe and skilful," and "unskilled but relatively safe" drivers (Shirmohammadi et al., 2019). Recent hierarchical clustering of young drivers uncovered contrasting profiles-for example, individuals who view texting-while-driving as risky yet still engage in it-highlighting the need for targeted interventions (Hayashi et al., 2023).

As ML models become central to safety and behavioral research, explainable AI (XAI) methods have gained importance for ensuring transparent decision-making. SHAP, introduced by Lundberg and Lee (2017), offers a unified approach to interpreting model predictions by quantifying each feature's contribution. Recent transportation studies have used SHAP with models such as Random Forest to reveal how environmental and driver-related factors interact nonlinearly to influence crash severity and risky behaviour (Sun et al., 2023). Broader reviews show SHAP's value for summarizing causal patterns in crash data and communicating complex models to policymakers (Abdulrashid et al., 2024).In Bangladesh, seatbelt use is mandated under the Road Transport Act 2018, yet compliance remains low due to weak enforcement, minimal fines, and the lack of automated monitoring (Islam et al., 2023; Shahriar et al., 2024). Socio-cognitive barriers-such as lower education levels and scepticism about belt effectiveness-further reduce use (Kargar et al., 2023). These challenges parallel trends in many developing countries where limited enforcement and awareness impede compliance. Evidence from Saudi Arabia shows that automated detection systems can significantly raise seatbelt use (Alghamdi et al., 2018). Overall, improving compliance in Bangladesh requires a coordinated approach combining stronger enforcement, public education, and infrastructural upgrades informed by regional best practices.

In Bangladesh, existing studies on seatbelt use rely mostly on simple statistics or regression models, which inadequately capture the complex interactions among demographic, attitudinal, and contextual factors. Explainable machine learning and behavioural segmentation are rarely applied, leaving heterogeneity in non-compliant drivers unexplored. To address these gaps, this study uses a two-stage framework combining Random Forest modelling, SHAP for feature interpretability, and hierarchical clustering to identify distinct behavioural typologies. Specifically, the objectives of this study are to (i) identify the key factors associated with seatbelt non-compliance using a Random Forest model, (ii) interpret the relative and directional influence of these factors through SHAP-based explainability, and (iii) segment non-compliant drivers into meaningful behavioural groups using hierarchical clustering. Demographic and attitudinal factors interact with situational variables such as trip length, weather, and time of day to shape non-compliance, producing actionable profiles like "Young Risk-Takers" and "Situational Non-Compliers." These data-driven typologies enable targeted interventions, including awareness campaigns, enforcement strategies, and infrastructure improvements, providing practical insights for evidence-based road safety measures in Bangladesh.

2. DATA COLLECTION AND PREPARATION

2.1 Data Collection

This analysis uses the data of a questionnaire survey to better understand drivers' attitudes and behaviours toward seatbelt use in Dhaka, Bangladesh. This survey is administered by the Accident Research Institute (ARI) of Bangladesh University of Engineering and Technology (BUET) as part of a research project titled "National and Dhaka City Road Safety Summary and Observational Study of Risk Factors in Bangladesh" funded by the World Bank (WB). One of the main purposes of this research is to understand drivers' response and assess their awareness level towards using seatbelt.

Participation of this study involved completing a 48 items anonymous survey. Most items required to provide response on a scale of strongly agree to strongly disagree. The questionnaire consisted of 48 structured questions divided into demographic, attitudinal, and behavioural sections (Table-1). These variables were later used for predictive modelling and clustering analysis.

Table-1: Summary of the Questionnaire Form

Focus Area	Variables	Response
Target Variable	Use of Seatbelt (q15-for Passengers, q16-for Drivers)	Always, Very often-most of the time, Sometimes, Rarely, Never, Not Applicable
Demography of the Participants	q1. Gender	Male, Female, Prefer not to comment
	q2. Age	<18, 18-24, 24-30, 30-36, 36-42, 42-48, 48-54, 54+
	q3. Education	No formal Education, Primary, Secondary, Higher Secondary, Graduate, Post Graduate
	q4. Monthly income (in BDT)	<20k, 20-40k, 40-60k, 60-80k, 80-100k, 100k+
	q5. Current Profession	Govt. Employee, Private Employee, Self-employed, Student, Unemployed, Others
	q6. Type of private car driving	Professional, Non-Professional
	q7. Normally you drive	Rented Car, Own Car
	q8. Way of Private Car riding	As a professional driver, As a non-professional driver, As a passenger
	q9. If Non-professional, then what is the major purpose of your trip	Job, Shopping, Education, Drop-off and pickup, Recreation/exercise/prayer
	q10. How much time do you spend on private	<1hr, 1-3hr, 3-6hr, 6-9hr, 9-11hr, >11hr

	car daily		
	q11. Private car driving experience		1-2yr, 3-4yr, 5-6yr, 7-8yr, 9-10yr, >10yr
	q12. Years of getting the driving license		1-2yr, 3-4yr, 5-6yr, 7-8yr, 9-10yr, >10yr, No license
	q13. As a private car driver, how many times you fell in crash/injured within last 1 year		Never, 1-2 times, 3-5 times, >5 times
	q14. Where do you drive private car generally		Inside Dhaka city, Outside Dhaka city
	q17. *What are the main reasons for not wearing seat belt? - Multiple answers but with ranking*	q18. *When do you not wear Seatbelt? – Multiple answers but with ranking*	Yes, No
	Uncomfortable	Riding for a short trip	
	Causes movement restriction	During the hot weather	
	Causes neck pain	Don't anticipate meeting a policeman	
	Feel hot	During the daytime	
	Self-complacency	During the night	
	Long experience	During weekdays	
	Riding in low-speed local road	During weekend	
	Head or hair damage	Ride on a local road	
	Others	Ride on highways	
Attitudes towards Seatbelt Use	q19. It is acceptable to drive private car or sit in front seat without a seatbelt		Strongly Agree, Agree, Neither, Disagree, Strongly Disagree [Scale: 1-5]
	q20. Passengers don't have to wear seatbelts		
	q21. Wearing a seatbelt does not reduce the severity of a crash-related injury		
	q22. Seatbelts are not needed when riding at low speed		
	q23. A child doesn't have to wear a seatbelt		
	q24. A seatbelt is not needed for a short trip (less than 2km).		
	q25. Seatbelts are not necessary in hot weather.		
	q26. Seat-belt use is more necessary during the daytime than night		
	q27. Wearing seatbelt is important just to avoid police fines		
	q28. There is no need for the experienced driver to wear seatbelt		
Attitudes related to driving behavior	q29. Driving a little faster is acceptable for a good driver.		Strongly Agree, Agree, Neither Disagree, Strongly Disagree [Scale: 1-5]
	q30. When the road is clear, there is no need to reduce speed at intersection		
	q31. If the front vehicle goes in wrong side or violating traffic rules, it is acceptable to follow him		

	q32. It is acceptable to receive any important phone call during riding/driving car	
	q33. It is not necessary always to reduce speed or yield at pedestrian when they are crossing the road.	
	q34. Punishments for speeding or rule violation should be more severe.	
Drivers' behavior (Practice as a driver)	q35. Do not follow road sign and marking while driving	Never, Rarely, Often, Most of the time, Always [Scale: 1-5]
	q36. Receive mobile calls while driving the private car	
	q37. Carry front seat passenger without seat belt	
	q38. Drive while not being physically fit or not having sound mind	
	q39. Drive private car with mechanical problems	
	q40. Overtake slow moving vehicles, like, rickshaw, bicycle from left side to save time	
	q41. Speeding to cross intersection during red signal or when police signals/stops	
	q42. Drive with expired documents	
	q43. Do not pay attention to the pedestrians when they walk or try to cross the road or intersection	
	q44. Use minimum path (short-cut) while turning right or left	
	q45. Drive on the wrong-side of the road to avoid traffic congestion at 3 or 4-leg intersections or even to overtake other vehicles	
	q46. Forget to look at the side mirrors or to put on indicators while lane changing or turning	
	q47. Try to overtake the vehicle in front (ahead of me) without paying attention to his indicators	
	q48. Use horn frequently while riding/driving so that vehicles or pedestrians in front move aside	

2.2 Data Preparation

The raw dataset obtained from the questionnaire survey was first imported using pandas and cleaned to ensure consistency. All column headers were standardized by removing extra spaces, punctuation, and converting them to lowercase. Responses with missing or invalid seatbelt data (code "6" or blank values) were excluded. Seatbelt-use questions (q15 for passengers and q16 for drivers) were reversed so that higher values indicated safer behavior. Binary target variables were then created, assigning 1 (compliant) for responses “Always” or “Most of the time” and 0 (non-compliant) otherwise. Multi-select situational items (q18) were converted into binary form, and the total number of selected reasons was computed as a new variable (q18_count). Behavioral questions (q35–q48) were reversed to align their direction with safety-oriented interpretation.

Categorical features (e.g., trip purpose, location, time, weather) were encoded using OneHotEncoder, while numeric variables were converted to numeric type and imputed with their median values to handle missing entries. Finally, all variables were standardized using StandardScaler to ensure comparability across different scales before model training.

3. MEHTHODOLOGY

3.1 Model Development

Two predictive models were developed using the Random Forest Classifier to estimate seatbelt compliance separately for drivers (q16) and passengers (q15). The preprocessed dataset was divided into training (80%) and testing (20%) subsets using stratified sampling to preserve the proportion of compliant and non-compliant cases. Each model was trained with 300 decision trees (n_estimators=300) and the parameter class_weight='balanced' to correct for any class imbalance. The Random Forest algorithm was chosen for its robustness, ability to capture nonlinear relationships, and inherent feature importance estimation. Model performance was evaluated using precision, recall, F1-score, and ROC-AUC on the test data, ensuring reliable and generalizable predictive capability. The passenger model did not give high accuracy value. So, the result of driver model is used for interpretability analysis through SHAP values.

3.2 Framework

An outline of the modelling workflow of this study is depicted in Figure 1. The data mentioned above in Table 1 have been used for the complete and smooth execution of the modelling works in the workflow.

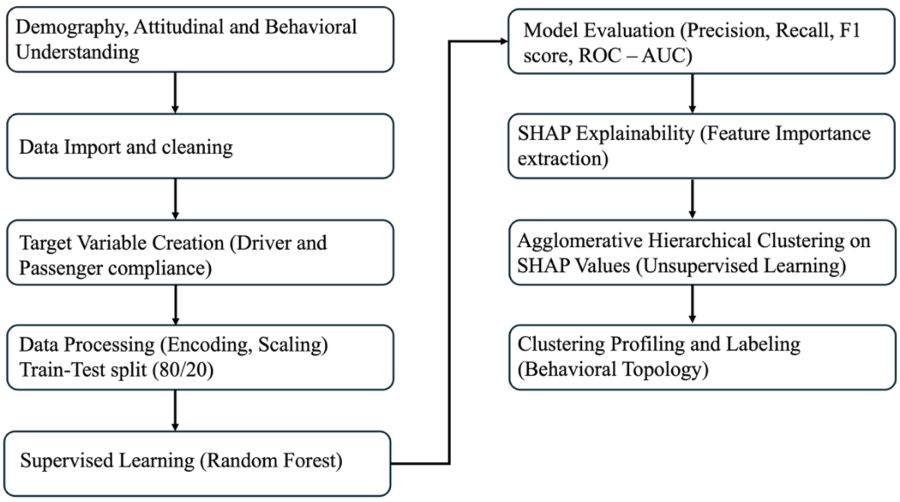

Figure 1: Modelling workflow

3.3 Supervised Machine learning model

Random Forest (RF)

Random Forest regression model was employed in this study to effectively account for non-linear relationships and intricate interactions among the variables. RF is very commonly used in traffic safety, especially in works that involve classification and prediction of crash severity and driver risk (Zhang et al., 2022; Sun et al., 2023). The RF algorithm, first proposed by Breiman (2001), constructs a collection of decision trees by utilizing two core techniques: bootstrap aggregation (bagging) and random feature selection. A RF model is built by training multiple Decision Tree (DT) classifiers. This is done using bagging (bootstrap aggregation) meaning the algorithm generates several subsets of the original dataset by resampling the training data with replacement. For each DT, node splitting is restricted to considering only a randomly selected subset of the available explanatory factors (features).After training, the final classification is determined by an ensemble vote: each trained DT casts a vote for a possible outcome, and the result is the outcome that receives most of the votes from the ensemble of classifiers (Wen et al., 2021).

For this study, two separate predictive models were developed using the RF Classifier from scikit-learn: one for driver seatbelt compliance (Q16) and another for passenger compliance (Q15). Each model used 300 decision trees (n_estimators=300) with class_weight='balanced' to address class imbalance.

This approach ensures robust performance estimation and prevents overfitting by evaluating the model on unseen data.

3.4 Evaluation Matrix

Table 2 presents the matrix used for calculating performance metrics. True positives (TP) refer to the number of samples that truly belong to the positive class ("use seatbelt") and are correctly predicted as such. False negatives (FN) represent the positive samples that were incorrectly classified. True negatives (TN) denote the correctly identified instances of the negative class ("don't use seatbelt"), while false positives (FP) indicate the negative samples that were mistakenly classified as positive.

Table 2: Confusion Matrix

	Predicted Lebel	
Actual Label	Use Seatbelt	Don't use seatbelt
Use Seatbelt	True Positive (TP)	False Negative (FN)
Don't use seatbelt	False Positive (FP)	True Negative (TN)

In this study using the four numbers in Table 2, model performance was evaluated on the test set using accuracy, F1-score, and the Area Under the ROC Curve (AUC).

3.4.1 Accuracy & F1 - score

Accuracy is the ratio of correct predictions to the total number of predictions. It measures the overall correctness of the model (Hastie, 2009). Which is expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Because traffic safety data are highly imbalanced, accuracy alone is insufficient for evaluating classifier performance. Therefore, precision, recall, and F1-score are used to assess the multiclass model (Wen et al., 2021). The F1-score, the harmonic mean of precision and recall, ranges from 0 (poor performance) to 1 (perfect performance) and is computed as follows:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

3.4.2 ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve is generated by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). This curve visually represents the prediction accuracy of the dataset as the classification probability threshold is varied. To quantify the model's overall performance, the Area Under the Curve (AUC) value can be calculated. The quality of the model is directly proportional to its AUC value; an AUC that is closer to signifies superior performance.

$$TPR = \frac{TP}{TP+FN} \tag{5}$$

$$FPR = \frac{FP}{FP+TN} \tag{6}$$

$$AUC = [x\ axis{:}\ FPR, y\ axis{:}\ TPR] \tag{7}$$

3.5 Model Interpretation

Shapley Additive exPlanations (SHAP)

Although machine learning offers strong predictive accuracy, understanding how input variables shape model outcomes remains difficult. This study examines the interpretability of tree-based ensemble models to support strategies for improving helmet-use safety. To evaluate variable importance, it applies SHAP (SHapley Additive exPlanations), a game-theoretic method introduced by (Lundberg and Lee., 2017) that explains model outputs by computing each feature's contribution through Shapley values. Given the value of an explanatory variable, i, the Shapley value is calculated as follows:

$$\varphi_i = \sum_{s \subseteq F-i} \frac{|S|!(|F|-|S|-1)!}{|S|!} \left[f_{SU\{i\}}\left(x_{SU\{i\}}\right) - f_s(x_s) \right] \tag{8}$$

Here, $|F|$ denotes the total number of explanatory variables, while S represents any subset of these variables excluding the i-th one, and $|S|$ indicates the number of elements in that subset. The term $f_{SU\{i\}}\left(x_{SU\{i\}}\right)$ refers to the model trained with the inclusion of the i-th variable, whereas $f_s(x_s)$ represents the model trained without it.

In this study, to interpret the model predictions, SHAP (SHapley Additive exPlanations) was applied to the driver model. The TreeExplainer algorithm was used to compute each feature's contribution to the predicted compliance probability for every respondent. The global SHAP summary ranked variables by their average absolute influence, identifying the most important behavioral, attitudinal, and situational predictors of seatbelt use.

A beeswarm plot was generated to visualize the direction and magnitude of these feature effects.

3.6 Unsupervised Learning

Cluster analysis and behavioural segmentation

To explore patterns of behavioral similarity, SHAP value vectors were standardized and grouped using Agglomerative Hierarchical Clustering (Ward linkage) which follows a "bottom-up" approach Agglomerative Hierarchical Clustering with Ward Linkage (often referred to as Ward's Method) is a highly effective unsupervised learning technique frequently used in transportation engineering due to its tendency to form compact, spherical, and relatively equal-sized clusters (Li et al., 2017).

The optimal cluster number (four) was selected based on interpretability and meaningful subgroup separation.

Each cluster was then profiled by its mean compliance rate, situational factors (q18_count), and dominant SHAP feature influences. Clusters were qualitatively labeled as Disciplined Safety-Followers, Routine Rule-Adherers, Situational Non-Compliers and Young Risk-Takers.

These typologies represent distinct behavioral profiles explaining variations in seatbelt compliance.

4. RESULTS

4.1 Sample Overview

After initial data cleaning, 380 valid responses remained, and 377 cases contained complete information required for modelling. The demographic and driving-profile distribution closely followed the questionnaire structure, including age, education, income, profession, trip purpose, driving experience, licensing years, crash history, and driving location.

4.2 Model Performance

Two Random Forest classifiers were trained to predict seatbelt compliance for drivers (q16) and passengers (q15).

Table 3: Distinction between two RF Classifiers

Driver Model	Passenger Model
• Accuracy: 0.92 • ROC AUC: 0.768 • Class imbalance observed: only 5 non-compliers in test set • The model identified compliant drivers very well (Recall = 0.99) but had limited ability to detect the small number of non-compliant drivers.	• Accuracy: 0.75 • ROC AUC: 0.746 • Better balance: 24 non-compliers, 52 compliers • Reasonable discrimination, with stronger recall for compliant passengers (0.92).

Overall, the models showed moderate predictive power, appropriate for behavioural survey data, with performance limited by class imbalance in the driver dataset.

4.3 SHAP-Based Feature Importance

SHAP analysis provided global explanations of the driver model, as the accuracy of this model is higher. The features contributing most to predictions were dominated by unsafe driving practices**,** situational excuses for not wearing seatbelts**,** and attitudinal beliefs about seatbelt necessity**.** Unsafe driving behaviors (q35–q48) were the strongest predictors of non-compliance. Not checking mirrors (q46), speeding at intersections (q41), improper overtaking (q47), and driving while unfit (q38) were the top predictors of lower seatbelt use. Situational excuses (q18_count) had strong negative influence. Respondents reporting many reasons like short trip, hot weather, time of the trip or riding on highways for not wearing a seatbelt were consistently non-compliant. Attitudes justifying non-use (q24- Short trip less than 2km, q28-Seatbelt not necessary for experienced drivers, q20-Seatbelt not necessary for passengers) increased the likelihood of non-compliance. Driving experience (q11-Time period of driving experience, q12-Time period of holding a license) had moderate but meaningful influence. Vehicle maintenance behaviors (q39-Driving car with mechanical problems) and road-rule adherence (q35- Not following road sign and marking, q43-Not paying attention to the pedestrians, q45-Driving on the wrong side to avoid congestion and even to overtake other vehicles) were also significant.

A SHAP beeswarm (Figure 3) and feature importance bar plot (Figure 2) illustrate these insights.

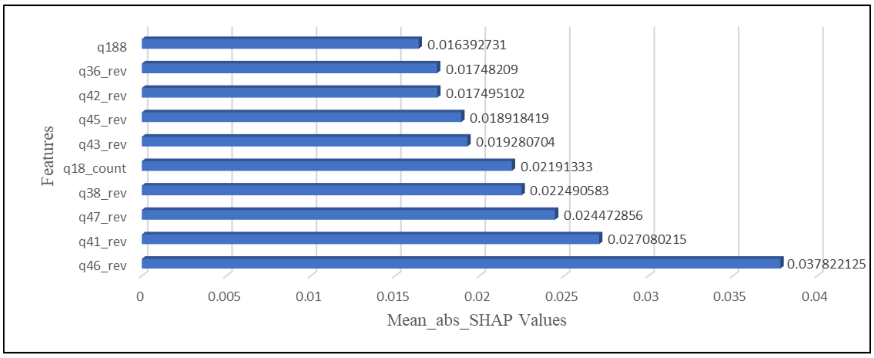

Figure 2: Top 10 SHAP_Feature_importance Bar Chart (Driver Model)

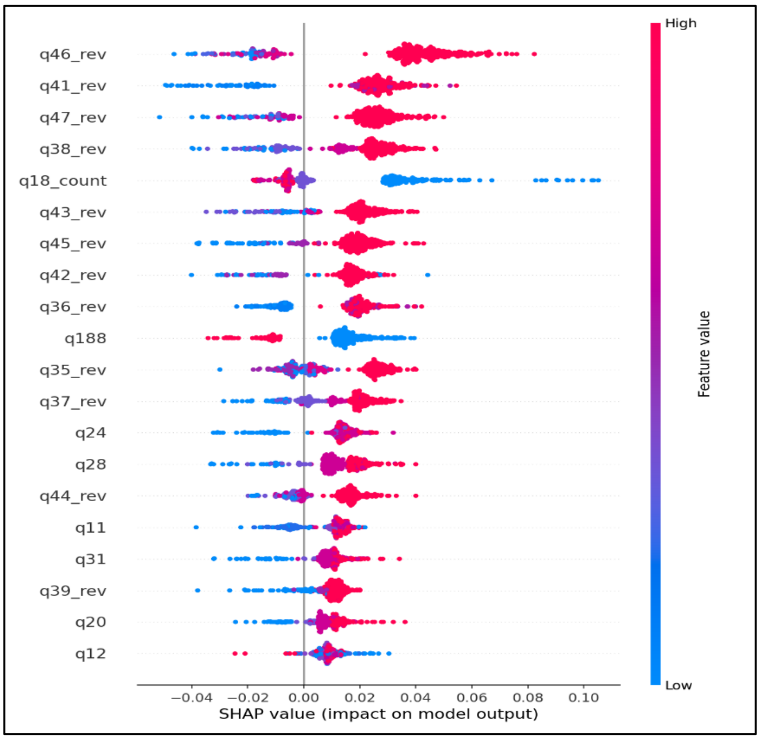

Figure 3: SHAP Beeswarm

4.4 Clustering of Behavioral Profiles

Hierarchical clustering on SHAP vectors revealed four distinct behavioral typologies of seatbelt users is shown in Table 4. Cluster profiles are shown graphically in Figure 4, from which we can get the attitudinal and behavioral characteristics that are influencing noncompliance behavior. Converging evidence from SHAP and clustering suggest that seatbelt compliance is highly consistent with general safe driving behaviour. Situational beliefs (e.g., "short trip", "no police around") meaningfully drive non-compliance. Attitudes that downplay the necessity of seatbelts (e.g., q24, q28) create clear psychological risk profiles. Experience alone does not guarantee compliance, though more experienced drivers tend to be slightly safer.

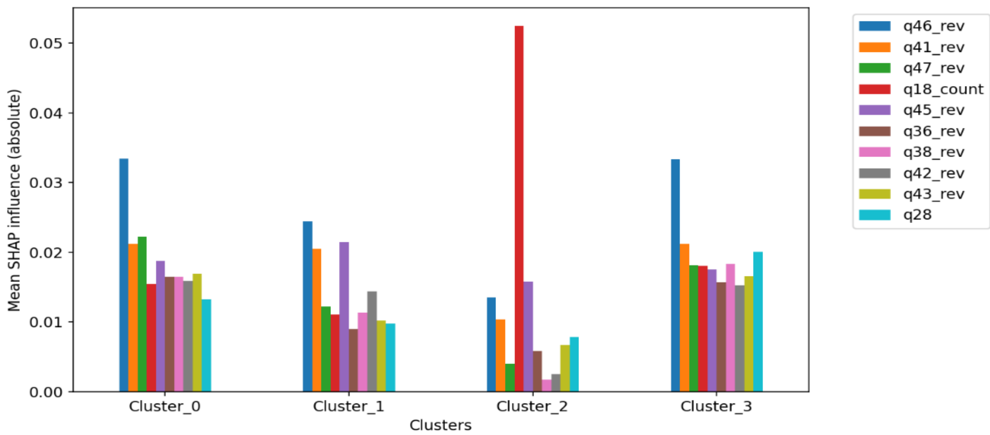

Figure 4: Top SHAP feature means by Cluster (Driver Model)

Table 4: Cluster Profiles

Cluster Segment	Cluster_0	Cluster_1	Cluster_2	Cluster_3
Description	Routine rule-adherers	Young risk-takers	Situational non-compliers	Disciplined safety-followers
Counts (n)	261	22	20	74
Compliance rate (%)	99.2	86	0	100
Situational Excuses	1.05	0.86	1.75	0.70
Characteristics	Safe driving behaviors and consistently wearing seatbelts	Prone to risky maneuvers, lower rule-following attitudes.	Does not usually wear seatbelts and reports many situational barriers.	The safest group — clean driving habits and strong attitudes toward safety.

5. DISCUSSION AND CONCLUSIONS

This study examined the determinants of seatbelt compliance among private car users by integrating survey-based behavioral variables with machine learning models. The Random Forest classifier demonstrated moderate predictive performance, with SHAP analysis highlighting that unsafe driving practices, situational excuses, and permissive attitudes toward seatbelt use were the most influential predictors of non-compliance. These findings reinforce the strong behavioral and perceptual foundations of seatbelt use, consistent with previous road safety literature.

The clustering of SHAP values revealed four distinct behavioral profiles. Two clusters-Routine Rule-Adherers and Disciplined Safety-Followers exhibited near-perfect compliance and minimal situational excuses. A smaller group, termed Young Risk-Takers, showed lower rule-following tendencies and higher propensity for risky maneuvers. The smallest but critical group, Situational Non-Compliers, reported the highest number of excuses for not wearing seatbelts and had the lowest compliance rate. These typologies illustrate that seatbelt non-use is not uniform but emerges from varying combinations of behavior, attitude, and situational reasoning. Although a separate predictive model was developed for passenger seatbelt compliance, SHAP-based interpretability and clustering were not extended to this model due to comparatively weaker class separability and the context-dependent nature of passenger behavior, which could limit the robustness of explainable insights. Overall, the results suggest that policies focusing solely on enforcement may be insufficient. Targeted behavioral interventions such as awareness campaigns addressing situational misconceptions (e.g., "short trips do not require seatbelts") and risk perception may be more effective for specific subgroups. Moreover, improving general safe driving practices may indirectly increase seatbelt use, given the strong predictive relationship observed.

In conclusion, the integration of machine learning explainability and clustering offers a valuable framework for understanding nuanced seatbelt user profiles in Bangladesh. This approach can support more tailored, data-driven strategies for improving road safety outcomes. Future work may benefit from larger and more balanced samples, as well as longitudinal data to validate behavioral patterns over time.

ACKNOWLEDGEMENTS

The authors are grateful to World Bank for issuing adequate funds, and Accident Research Institute of BUET for conducting the field surveys.

DECLARATION OF USE OF AI

AI-assisted language editing tools (Grammarly) were used solely for grammar, spelling, and language clarity editing after completion of the manuscript. No AI-based system was used for content generation, data analysis, interpretation of results, or drawing conclusions.

REFERENCES

Dissanayake, S., & Ratnayake, I. (2007). Effectiveness of seat belts in reducing injuries (No. MBTC-2080). Mack-Blackwell National Rural Transportation Study Center (US).

Ganti, L., Bodhit, A. N., Daneshvar, Y., Hatchitt, K., Kuchibhotla, S., Pulvino, C., ... & Peters, K. R. (2021). Effectiveness of seatbelts in mitigating traumatic brain injury severity. World Journal of Emergency Medicine, 12(1), 68.

Centers for Disease Control and Prevention. (2025, January 31). Seat belt use: Facts. U.S. Department of Health and Human Services

Kargar, S., Ansari-Moghaddam, A., & Ansari, H. (2023). The prevalence of seat belt use among drivers and passengers: a systematic review and meta-analysis. Journal of the Egyptian Public Health Association, 98(1), 14.

Rozars, M. F. K., Ahmed, N., Sultana, N., Ishtiak, A. S. M., Alam, M. T., Hossan, M. E., ... & Hawlader, M. D. H. (2025). Factors associated with road traffic injury in a high-risk zone of Bangladesh: a mixed-method study. Injury prevention, 31(1), 32-39.

Özkan, T., Puvanachandra, P., Lajunen, T., Hoe, C., & Hyder, A. (2012). The validity of self-reported seatbelt use in a country where levels of use are low. Accident Analysis & Prevention, 47, 75-77.

Sheveland, A. C., Luchman, J. N., Mendelson, J., Xie, J., Bleiberg, M. A., Eby, D. W., ... & Walton, B. R. (2020). Psychological constructs related to seat belt use: A nationally representative survey study. Accident Analysis & Prevention, 148, 105715.

Lajunen, T., & Özkan, T. (2011). Self-report instruments and methods. In Handbook of traffic psychology (pp. 43-59). Academic Press

Yan, M., & Shen, Y. (2022). Traffic accident severity prediction based on random forest. Sustainability, 14(3), 1729.

Sun, Z., Wang, D., Gu, X., Abdel-Aty, M., Xing, Y., Wang, J., ... & Chen, Y. (2023). A hybrid approach of random forest and random parameters logit model of injury severity modeling of vulnerable road users involved crashes. Accident Analysis & Prevention, 192, 107235.

Shirmohammadi, H., Hadadi, F., & Saeedian, M. (2019). Clustering analysis of drivers based on behavioral characteristics regarding road safety. International Journal of Civil Engineering, 17(8), 1327-1340.

Hayashi, Y., Friedel, J. E., Foreman, A. M., & Wirth, O. (2023). A hierarchical cluster analysis of young drivers based on their perceived risk and frequency of texting while driving. Journal of safety research, 85, 398-404.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Abdulrashid, I., Farahani, R. Z., Mammadov, S., Khalafalla, M., & Chiang, W. C. (2024). Explainable artificial intelligence in transport Logistics: Risk analysis for road accidents. Transportation Research Part E: Logistics and Transportation Review, 186, 103563.

Islam, B. Z., Tune, S. N. B. K., Naher, N., & Ahmed, S. M. (2023). Trauma care scenarios following road traffic crashes in Bangladesh: a scoping review. Global Health: Science and Practice, 11(2).

Shahriar, M. Z., Huq, A. S., & Afzal, L. (2024). A Quantitative Assessment of People's Compliance with Traffic Law Enforcement to Mitigate Road Crashes in Bangladesh.

Ahsan, A. H. M., Hasan, M. K., Rumi, M. H., Ahmed, T., & Aunto, T. K. (2024). Students' safety culture at tertiary level academic institutes in Bangladesh: A cross-sectional study. Heliyon, 10(22)

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Alghnam, S., Towhari, J., Alkelya, M., Binahmad, A., & Bell, T. M. (2018). The effectiveness of introducing detection cameras on compliance with mobile phone and seatbelt laws: a before-after study among drivers in Riyadh, Saudi Arabia. Injury epidemiology, 5(1), 31.

Wen, X., Xie, Y., Jiang, L., Pu, Z., & Ge, T. (2021). Applications of machine learning methods in traffic crash severity modelling: current status and future directions. Transport reviews, 41(6), 855-879.

Zhang, S., Khattak, A., Matara, C. M., Hussain, A., & Farooq, A. (2022). Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents. PLoS one, 17(2), e0262941.

Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction.

Li, Q., Wang, K. P., Eacker, M., & Zhang, Z. (2017). Clustering methods for truck traffic characterization in pavement ME design. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 3(2), F4016003.6