TECHNIUM
SOCIAL SCIENCES JOURNAL

Vol. 82/2026

A New Decade for Social Changes

Technium Social Sciences Journal
Vol. 82, 285-297, April, 2026
ISSN: 2668-7798
www.techniumscience.com

Application of Modified Machine Learning Technique in Traffic Safety Analysis

Xiaohui Zhong, Utpal Dutta

University of Detroit Mercy, Detroit, Michigan, United States

zhongk@udmercy.edu, duttau@udmercy.edu

Abstract. Stop-sign compliance remains a persistent safety challenge in high-density, low-speed urban environments such as university campuses. This study investigates the predictors of traffic control compliance at an enclosed urban college campus using a Modified Machine learning technique (three-stage Stepwise Random Forest and Logistic Regression (SRFLR) framework). Observational data from 465 drivers were analyzed to move beyond raw correlations and capture the complex interactions between spatial, social, and demographic factors. The final predictive model identifies a hierarchy of influence dominated by directional intent and situational risk assessment. Direction of travel emerged as the most significant predictor, as drivers leaving the campus exhibited a greater likelihood of compliance, indicating that the purpose of travel plays a critical role in shaping driving behavior. Social and physical hazards also served as primary "forcing functions," as the presence of opposing traffic and pedestrians were consistent predictors of a full stop. While driver gender was a significant factor in initial screenings, its influence was reduced when controlling for environmental risks, indicating that demographic variances are partially explained by situational travel patterns. These findings suggest that campus safety interventions should prioritize engineering environments that naturally reinforce risk perception through enhanced visual salience and social monitoring. Such findings should also apply to situations beyond college campuses.

Keywords. Stop-sign compliance, Driver behavior, Campus traffic safety, Logistic modeling, Spatial factors

1. Background

Stop-controlled intersections present a persistent safety challenge because driver behavior often diverges from formal traffic control requirements. A central issue identified in the literature is the gap between the legal mandate to come to a complete stop and the way drivers cognitively interpret and respond to stop signs in real-world settings. Early longitudinal work by McKelvie (1986) demonstrated that stopping behavior is not primarily governed by habitual law compliance; instead, drivers approach stop signs as sites of risk assessment, where decisions are made dynamically based on immediate environmental cues. More recent naturalistic studies reinforce this interpretation. Figueroa Jacinto et al. (2024) found that in the absence of perceived hazards, many drivers substitute a full stop with a kinematic threshold of approximately 10 mph, effectively performing a rolling stop while conducting a visual scan.

Together, these findings suggest that drivers routinely replace legal compliance with perceptual sufficiency, maintaining momentum unless a clear conflict is detected.

Individual driver characteristics and the in-vehicle social environment further shape the likelihood of stop-sign violations. Shinar and Compton (2004) reported that male drivers are significantly more prone to "technical" violations, including rolling stops, than female drivers. This demographic effect is amplified by social context. Drawing on Social Facilitation Theory, Simons-Morton et al. (2005) showed that passengers can function as social monitors, increasing normative compliance through observation and implicit judgment. In contrast, single-occupant vehicles may experience reduced compliance due to the absence of social accountability. Woldeamanuel (2012) corroborated these patterns using compositional analysis, identifying driver age and gender as persistent predictors of intersection behavior across contexts.

Beyond individual characteristics, the physical and temporal environment of an intersection plays a critical role in shaping driver decision-making. Anowar et al. (2012) found that stop-sign violations occur more frequently during off-peak hours, when perceived risks from cross traffic or law enforcement are diminished. Spatial configuration also matters. Arhin et al. (2019) documented a "momentum effect" at stop-controlled intersections located near signalized intersections, where drivers attempt to maintain traffic flow continuity and are therefore less likely to comply fully with stop requirements. After examining the vehicle attributes through the lens of conspicuity, Lardelli-Claret et al. (2002) suggested that high-contrast vehicle colors may influence drivers' perceptions of their own visibility, potentially affecting risk-taking behavior at intersections.

In high-density environments such as college campuses, pedestrian presence becomes a dominant factor influencing stopping behavior. While pedestrians frequently serve as a catalyst for deceleration, compliance is not uniform across pedestrian scenarios. Silvano and Linder (2017) identified a driver "dilemma zone" in pedestrian interactions: drivers are highly likely to stop for pedestrians already in the crosswalk but are substantially less likely to yield to pedestrians waiting at the curb. This finding implies that compliance is often triggered by an immediate physical hazard rather than by the stop sign itself, highlighting the limits of regulatory signage as a standalone control device.

Engineering countermeasures have therefore focused on enhancing the perceptual salience of stop-controlled intersections to counteract the "look-but-fail-to-see" phenomenon. Van Houten and Retting (2001) demonstrated that low-cost pavement treatments, such as stop bars (pedestrian crossing marking), can double compliance rates by placing the stopping cue directly within the driver's focal path. For higher-risk locations, more conspicuous interventions are necessary. Studies by Hallmark et al. (2018, 2022) and Goswamy et al. (2019) showed that sign-mounted flashing beacons can reduce nighttime injury crashes by as much as 54%. More recently, Layegh et al. (2025) found that LED-backlit stop signs significantly increase the likelihood of a full stop by disrupting drivers' habituated rolling-scan behavior. Consistent with these findings, the Federal Highway Administration (2017) recommends a systemic bundle of low-cost treatments—such as oversized signs, reflective wraps, and stop bars—as the most cost-effective strategy for institutional road managers seeking to improve safety.

Although these studies establish strong links between environmental hazards and driver behavior, traditional linear modeling approaches often struggle to capture the high-dimensional interactions and categorical sparsity inherent in observational traffic data. Methodological benchmarks by Couronné et al. (2018) suggest that hybrid analytical frameworks outperform single-model approaches in such contexts. In particular, combining

Random Forests (RF) with Logistic Regression allows researchers to balance predictive performance with interpretability. RF variable importance metrics can identify both main effects and complex interactions that univariate screenings may overlook (Nyongesa et al., 2020). However, because machine learning models are sensitive to noise introduced by rare categorical values, recent work advocates for a structured hybrid approach. The Stepwise Random Forest and Logistic Regression (SRFLR) framework was used as a final regression model, thereby preserving interpretability while improving model robustness (Li, 2023). This approach resolves the tension between the predictive power of machine learning and the explanatory demands of social science research, where odds ratios and interaction effects remain essential (Jaccard, 2011).

Despite this growing body of research, there remains a notable gap in empirical studies focused on enclosed urban college campuses. These environments are distinct in that they combine high pedestrian densities with relatively low speed limits, which may foster a false sense of security among drivers. Following Li (2023), the present study applies the Modified machine learning technique such as SRFLR framework to observational data collected in a university setting, systematically filtering variables which includes three stages of analysis: bivariate screening, predictive importance testing, and final model refinement. This approach enables a context-specific examination of how established factors—such as driver gender, vehicle occupancy, opposing traffic, and pedestrian presence—influence stop-sign compliance within a university ecosystem.

2. Definition of variables and theoretical framework

The primary objective of this study is to identify the predictors of stop sign compliance within the unique constraints of an enclosed urban college campus. Compliance, the dependent variable, is operationalized as a binary outcome: a complete cessation of vehicle motion (**Y**) versus non-compliance (**N**), which encompasses both rolling stops and total non-stops. This binary approach is consistent with established safety literature that distinguishes legal adherence from "at-risk" behavior (Anowar et al., 2012; Van Houten & Retting, 2001). To evaluate the factors influencing this behavior, ten independent variables were selected based on their prevalence in traffic safety and social psychology research.

The spatial configuration of the intersection significantly influences driver kinematics and perception (Figueroa Jacinto et al., 2024). Specifically, the **Location** (1–5) and the presence of **Multiple** stop signs (multi-way vs. single-way) account for the complexity of the built environment. As noted by Arhin et al. (2019) and the Federal Highway Administration (2017), multi-way intersections introduce more conflict points, which may increase compliance due to heightened driver caution. Furthermore, the variable **Out** identifies directional flow, testing whether drivers exiting the campus—potentially influenced by "end-of-trip" time pressure—exhibit different compliance patterns than those traveling internally.

Temporal and social dynamics provide further context for driver decision-making. The variable **Time** (Morning, Noon, Afternoon) serves as a proxy for campus congestion and the "hurry-state" often associated with commuting peaks (Woldeamanuel, 2012). Social modeling and risk perception are captured through the presence of **Opposing** traffic and **Pedestrians**. Silvano and Linder (2017) suggest that the presence of vulnerable road users, such as pedestrians, significantly increases yielding behavior. Similarly, the "social pressure" of an opposing vehicle or the presence of passengers (**Occupancy**) may act as a corrective influence, mitigating aggressive driving tendencies (Shinar & Compton, 2004; Simons-Morton et al., 2005). Finally, driver and vehicle characteristics, including **Gender** and **Color**, are included to explore demographic variances in risk-taking (Lardelli-Claret et al., 2002; Shinar & Compton, 2004).

Table 1. Definition of variables

1. **Variable Category**	2. **Variable Name**	3. **Description**	4. **Values / Levels**
5. **Dependent**	6. **Stop**	7. Compliance with the stop sign.	8. **Y**: Complete stop; **N**: Rolling or non-stop.
9. **Spatial**	10. **Location**	11. Specific site on the campus map.	12. **1–5**: Categorical site markers.
	13. **Multiple**	14. Intersection configuration.	15. **Y**: Multi-way stop; **N**: Single stop sign.
	16. **Out**	17. Directional traffic flow.	18. **Y**: Outgoing traffic only; **N**: Inward/Internal.
19. **Temporal**	20. **Time**	21. Period of day/Campus activity level.	22. **M**: 8-10am (Arrival); **N**: 11am-2pm (Lunch); **A**: 3-6pm (Departure).
23. **Social / Traffic**	24. **Opposing**	25. Presence of a vehicle from the opposite side.	26. **Y**: Present; **N**: Not present.
	27. **Pedestrian**	28. Presence of a person in/near crosswalk.	29. **Y**: Present; **N**: Not present.
30. **Demographic/social**	31. **Gender**	32. Perceived driver gender.	33. **F**: Female; **M**: Male.
	34. **Occupancy**	35. Number of people in the vehicle.	36. **S**: Single (driver only); **H**: High (driver + passengers).
37. **Vehicle characteristics**	38. **Color**	39. Visual paint color of the vehicle.	40. Categorical (e.g., Black, White, Red, etc.).

3. Data collection

Data collection occurred during the morning, midday, and afternoon periods in the month of September 2025. Morning observations coincided with a high influx of inbound traffic associated with commuters arriving on campus for work and early classes. In contrast, higher volumes of outbound traffic were observed during the afternoon period. Data was collected by a five-member student research team. A single observer was assigned to one location to ensure consistency in data recording. A map of an urban enclosed college campus with location of stop signs is presented in Figure 1. The data collection process involved direct visual inspection of each vehicle and driver approaching a stop-controlled intersection. Drivers were assessed based on whether they complied with the stop sign by coming to a complete stop. Vehicles that came

Technium Social Sciences Journal
Vol. 82, 285-297, April, 2026
ISSN: 2668-7798
www.techniumscience.com

to a full stop were classified as compliant, while those that performed a rolling stop or failed to stop altogether were classified as non-compliant. For each observation, several driver and vehicle characteristics were recorded, including vehicle color, driver gender, and passenger occupancy Environmental conditions, such as the presence of cross traffic or pedestrians, were also documented at each location. The final sample consisted of 465 observed drivers.

4. Data analysis and model

The analysis followed a three-stage hierarchical workflow, a parsimonious refinement step guided by prior theoretical and machine-learning screening. This methodology is supported by recent literature that combines the predictive strengths of machine learning with the interpretability of traditional regression (Couronné et al., 2018; Li, 2023).

Stage 1: Bivariate Screening and Strength of Association

The analysis began with Pearson's Chi-Square tests to establish the statistical significance of individual variables, complemented by Cramér's V to assess the magnitude of association. As shown in the preliminary screening, **Out** ($p = 0.0009$), **Opposing** ($p = 0.0016$), **Multiple** ($p = 0.0052$), and **Location** ($p = 0.0066$) emerged as the most robust environmental and spatial predictors.

Among demographic and social factors, **Gender** reached statistical significance ($\chi^2 = 4.04$, $p = 0.0443$), while **Pedestrian** showed a marginal trend toward significance ($p = 0.0679$). Interestingly, **Color** and **Occupancy** did not achieve significance in this stage ($p > 0.15$), and **Time** was the least impactful variable ($p = 0.4629$), showing no measurable association with compliance (Cramér's V = 0.000).

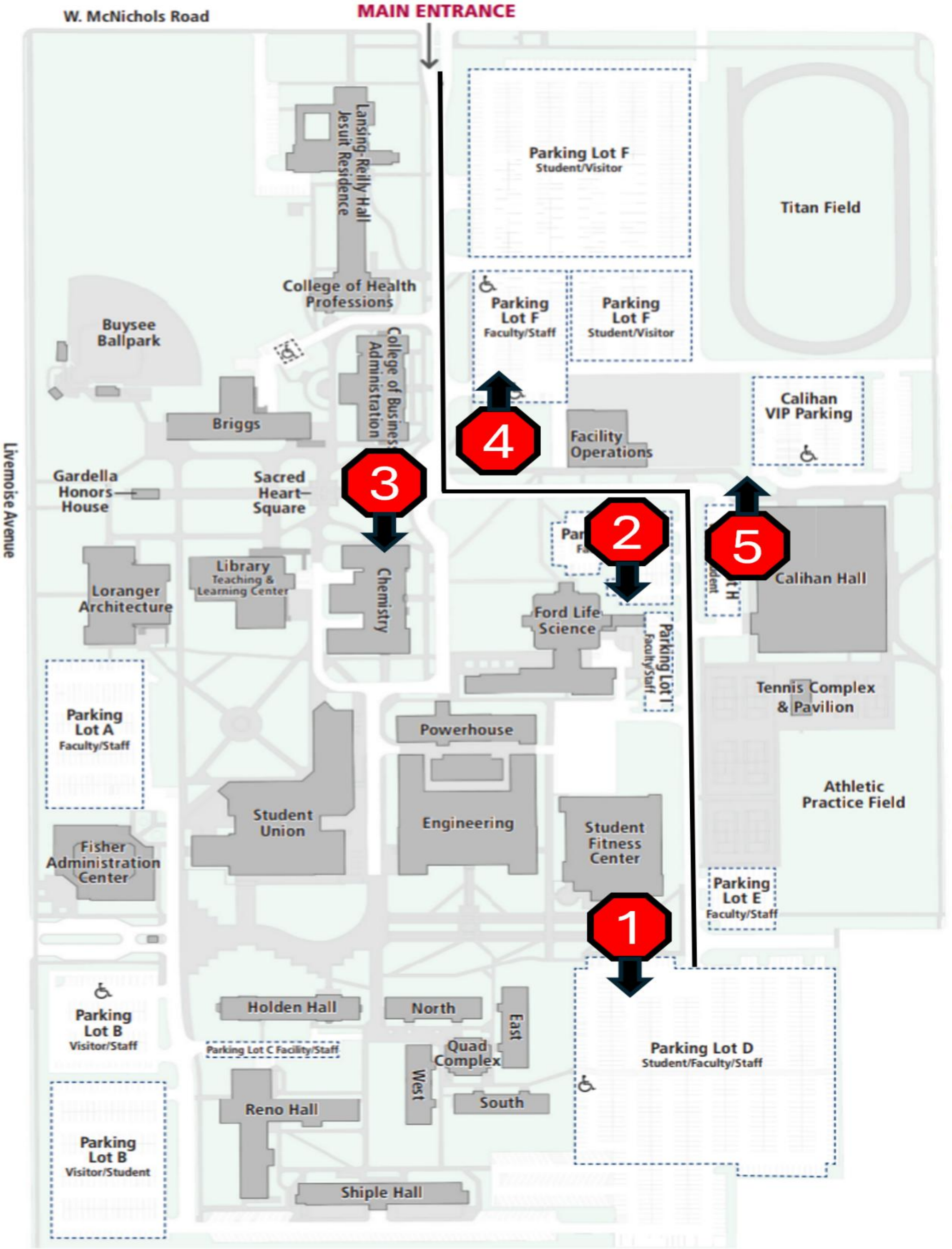

Figure 1 Locations of Data Collection Sites

TECHNIUM SOCIAL SCIENCES JOURNAL
Technium Social Sciences Journal
Vol. 82, 285-297, April, 2026
ISSN: 2668-7798
www.techniumscience.com

Stage 2: Machine Learning Sensitivity Analysis and Iterative Refinement
Following the bivariate screening, a Random Forest (RF) classifier was employed to capture both main effects and complex interactions. This stage was conducted as an iterative process to ensure the model was not skewed by categorical sparsity.

- **Round 1 (Diagnostic):** An initial RF model was run on the raw dataset. In this round, **Color** achieved an exceptionally high importance score. However, a granular diagnostic revealed that this ranking was artificially inflated by rare color categories with very low sample frequencies ($n < 10$). Because RF algorithms are sensitive to sparsity, the model was over-fitting these infrequent categories (6 Brown, 2 Gold, 2 Green, 1 Orange, 1 Tan), leading to "categorical noise."
- **Refinement:** To correct for this, these rare categories were consolidated into a single **"Other"** group. This consolidation was necessary to maintain statistical stability and ensure that the final model reflected meaningful behavioral trends rather than data anomalies (Nyongesa et al., 2020).
- **Round 2 (Final Importance):** The RF analysis was re-run using the consolidated variables. The results from this second round (as shown Table 2) provide the finalized importance hierarchy:
    - **Color (0.4306):** Even after consolidation, color remained the most influential feature, suggesting a complex: though not necessarily linear relationship with compliance.
    - **Social and Spatial Factors: Opposing (0.1099)** and **Gender (0.1098)** emerged as the next most stable high-ranking features.
    - **Secondary Predictors: Occupancy (0.0804)**, **Location (0.0690)**, and **Pedestrian (0.0669)** showed moderate predictive power, while **Multiple (0.0335)** and **Out (0.0425)** contributed less to the overall information gain despite their significance in Stage 1.
    - **Temporal Minimalist: Time** remained a less impactful variable (**0.0573**), confirming its negligible role in predicting stopping behavior on this campus.
    -

Table 2: Chi-square, Cramer's V and RF results

41. Variable	42. Chi-square	43. p-value	44. Cramer's V	45. RF importance	46. Significance
47. Location	48. 14.2205	49. 0.0066	50. 0.1484	51. 0.0690	52. Significant
53. (Out)	54. 10.9757	55. 0.0009	56. 0.1466	57. 0.0425	58. Significant
59. Opposing	60. 10.0165	61. 0.0016	62. 0.1394	63. 0.1099	64. Significant
65. Multiple	66. 7.8260	67. 0.0052	68. 0.1213	69. 0.0335	70. Significant
71. Color	72. 9.3605	73. 0.1543	74. 0.0849	75. 0.4306	76. Not Significant
77. Gender	78. 4.0442	79. 0.0443	80. 0.0810	81. 0.1098	82. Significant
83. Pedestrian	84. 3.3334	85. 0.0679	86. 0.0709	87. 0.0669	88. Marginal
89. Occupancy	90. 3.3420	91. 0.1881	92. 0.0537	93. 0.0804	94. Not Significant
95. Time	96. 1.5405	97. 0.4629	98. 0.0000	99. 0.0573	100. Not Significant

Stage 3: Multi-Variable Refinement (Stepwise Logistic Regression) and Final Model
The application of Random Forest (RF) variable importance metrics allows researchers to capture both main effects and complex interactions that univariate screenings might overlook (Nyongesa et al., 2020). However, because machine learning models can be sensitive to "noise" from rare categorical values, researchers advocate for a combined strategy (Stepwise Random Forest and Logistic Regression, or SRFLR) where RF acts as a sophisticated variable selection tool to inform the choice of predictors included in a final regression model (Li, 2023). This hybridity resolves the tension between the high predictive power of machine learning and the essential explain ability of Logistic Regression, the latter of which remains foundational for interpreting odds ratios and interaction effects in social science (Jaccard, 2011).
This leads to the final stage utilizing Stepwise Logistic Regression to isolate independent predictors from the broader set of environmental and demographic variables (Table 3). By integrating the insights from the Random Forest sensitivity analysis into a regression framework, the study ensures that the final model is both predictive and theoretically explainable (Li, 2023). This stage served as the ultimate filter for the inconsistencies observed in the earlier stages:

- **Combating Categorical "Stubbornness" (Color & Interactions): Color** showed a high importance in the Random Forest stage (0.4306). However, as suggested by Lardelli-Claret et al. (2002), vehicle color is often a measure of external conspicuity rather than an internal driver of regulatory compliance. To address the issue, the regression was initialized with all primary variables and specific interaction terms between **Color** and situational factors. This was designed to test if Color's influence was conditional—for example, if certain vehicle colors responded differently to pedestrians. However, both **Color** (with the consolidated "Other" grouping) and its interactions were dropped after failing to reach significance, confirming that the initial RF importance was a product of non-linear categorical noise rather than a stable behavioral predictor (Jaccard, 2011).
- **Spatial Mediation and the "Multiple" Paradox:** A key resolution in this stage was the exclusion of **Location** and **Multiple** stop signs. Despite their high significance in Stage 1 ($p < 0.01$), their predictive power became redundant once **Opposing** and **Pedestrians** were included. This indicates that the "Multiple" effect was fully mediated by the social and physical hazards present at those sites; drivers responded to the immediate risk of a collision rather than the density of signage.
- **The Survival of Directional Flow (Out):** Unlike other spatial markers, the **Out** variable remained a highly significant predictor ($B = 0.659$, $p = 0.001$). This confirms that the direction of travel—specifically exiting the campus—is a unique driver of behavior, independent of intersection geometry or social pressure.
- **Filtering Behavioral Proxies (Occupancy & Time): Time** and **Occupancy** were eliminated as they failed to improve the model's Akaike Information Criterion (AIC). Their removal suggests that earlier trends in these variables were likely "soaked up" by the stronger directional and social predictors.

• **Final Model Composition**: The procedure ultimately retained **Out**, **Opposing**, **Pedestrian**, and **Gender** as the stable predictors of compliance. Notably, the inclusion of directional flow and social hazards caused the significance of Gender to shift: while it was statistically significant in the Chi-Square test of stage 1($p = 0.044$), it became only marginally significant in the final stepwise regression model ($B = -0.357$, $p = 0.063$). This shift indicates that while gender-based differences exist, they are partially explained by the different travel patterns and environmental risks (such as exiting the campus) encountered by those drivers.

Table 3: Stepwise logistic model predicting log ratio of Stop Sign Compliance (1 = Stop, 0 = Non-Stop)

101. Predictor	102. Coefficient (B)	103. Std. Error	104. z-value	105. p-value	106. Significance
107. (Intercept)	**108.** -0.2134	**109.** 0.1718	**110.** -1.242	**111.** 0.214	**112.** Not Significant
113. Out (Yes)	**114.** 0.6597	**115.** 0.1939	**116.** 3.403	**117. 0.001**	**118.** Highly Significant
119. Opposing (Yes)	**120.** 0.7784	**121.** 0.2536	**122.** 3.070	**123. 0.002**	**124.** Highly Significant
125. Pedestrian (Yes)	**126.** 0.9238	**127.** 0.4673	**128.** 1.977	**129. 0.048**	**130.** Significant
131. Gender (Male)	**132.** -0.3571	**133.** 0.1920	**134.** -1.860	**135.** 0.063	**136.** Marginally Significant

As a robustness check, a multivariable logistic regression model including all theoretically motivated main effects was estimated without stepwise selection (Appendix- A). The results closely mirrored the final model. Directional flow (**Out**) and opposing traffic remained highly significant predictors ($p < 0.01$), **Pedestrian** and **Gender** remained marginally significant ($p < 0.10$), and **Time, Occupancy, and Color** remained non-significant. Such results confirmed that Out, Opposing , and Pedestrian remained the strongest predictors of compliance, indicating that the main findings are robust to model specification. Notably, no previously insignificant variables became statistically meaningful in the full model, suggesting that the stepwise procedure did not suppress substantively important predictors.

1. Results and discussion

The final predictive model identifies a hierarchy of influence dominated by directional intent and social risk. The most significant spatial predictor was the direction of travel, specifically whether a vehicle was exiting the campus (**Out**). Drivers exiting the campus exhibited a high likelihood of compliance (B = 0.6597, p = 0.001). This finding suggests that environmental cues at exit gates—where traffic transitions from campus-controlled zones to denser urban settings—are highly salient to drivers. While Woldeamanuel (2012) identifies the "hurry-state" during commuting peaks as a factor that can reduce compliance, these results indicate that the physical context of exiting the university ecosystem may overrule internal campus time-pressures.

Compliance was also strongly driven by the presence of other road users, indicating that drivers prioritize immediate physical risk over regulatory signage. The presence of a vehicle on the opposite side of the intersection was a highly significant predictor of a full stop (B = 0.7784, p = 0.002). This supports theories of social pressure, where the presence of an observer or potential conflict mitigates aggressive tendencies. Similarly, pedestrian presence significantly increased the likelihood of compliance (B = 0.9238, p = 0.048). As noted by Silvano and Linder (2017), compliance is often triggered by an immediate physical hazard rather than the stop sign itself.

Furthermore, the transition from Stage 1 to Stage 3 in the analysis revealed a shift in the role of gender. In the bivariate analysis, Driver Gender was significant ($p = 0.0443$), but it became only marginally significant in the final model ($B = -0.3571$, $p = 0.063$). This shift indicates that gender-based differences in risk-taking are partially explained by different travel patterns and environmental risks—such as exiting the campus—encountered by those drivers. The negative coefficient ($B = -0.3571$) for male drivers aligns with Shinar and Compton (2004), suggesting a higher propensity for technical violations compared to female drivers.

Finally, the study effectively resolves the discrepancy regarding multi-way intersections. Although Stage 1 of the analysis (Chi-Square) indicated that multi-way stops had significantly higher compliance ($p = 0.0052$), the variable was excluded from the final regression model. This confirms a spatial mediation; the compliance observed at multi-way stops was redundant once opposing traffic was accounted for. Drivers respond to the social risk of a collision with an opposing vehicle rather than the density of signage.

5. Conclusion and recommendations

5.1 Conclusion

This study employed a three-stage Stepwise Random Forest and Logistic Regression (SRFLR) framework to identify the true drivers of stop-sign compliance on an enclosed urban college campus. By systematically filtering variables, the analysis moved beyond raw correlations to uncover the underlying psychological and environmental mechanisms that govern driver behavior in a high-density university ecosystem.

The analysis demonstrates that compliance is not a static habit of law-abidance but a dynamic response to social and environmental risk. The presence of opposing traffic and pedestrians emerged as the most consistent predictors of a full stop. This suggests that drivers prioritize the immediate physical risk of a collision or the social pressure of an observer over the regulatory mandate of the stop sign itself. The high significance of the “Out” variable ($p = 0.001$) indicates that as drivers leaving the campus exhibited a greater likelihood of compliance, indicating that the purpose of travel plays a predictable role in shaping driving behavior. The marginal significance of Driver Gender ($p = 0.063$) in the final model, down from a higher significance in the initial bivariate screening, provides a nuanced look at demographic risk-taking.

The negative coefficient for males ($B = −0.3571$) suggests a higher propensity for technical violations, consistent with the “aggressive driving” profiles identified by Shinar and Compton (2004). This decline in statistical strength suggests that gender is not a standalone predictor but is likely influenced by the situational context of the trip. For instance, if male drivers on this campus are more likely to drive during off-peak hours or along routes with less opposing traffic, the regression model attributes the non-compliance to those environmental factors rather than gender alone. This aligns with the findings of Woldeamanuel (2012), who noted that intersection behavior is a product of both compositional (who the driver is) and contextual (the traffic environment) factors. Consequently, while a demographic trend exists, it is secondary to the immediate social and physical cues—such as the presence of a pedestrian or exiting the campus—that serve as the primary "enforcers" of compliance for all drivers.

5.2 Recommendations

The results of the final logistic model suggest that driver behavior on an enclosed campus is not merely a localized phenomenon. While this study was conducted within a university ecosystem, the psychological drivers identified—risk-based compliance and social monitoring—reflect universal human behaviors. Drivers prioritize situational risk assessment over legal obedience, treating the stop sign as a conditional requirement. By isolating opposing traffic and pedestrian presence as the primary "forcing functions," this study offers a blueprint for traffic management in various high-pedestrian, low-speed environments.

In environments with a high density of vulnerable road users, such as senior living communities or gated neighborhoods, the pedestrian effect is a critical safety lever. Because compliance drops when no pedestrians are visible, infrastructure must simulate risk or increase focal awareness. As Van Houten and Retting (2001) demonstrated, stop bars and high-visibility pavement text can double compliance by placing the instruction directly in the driver's focal path. This is a low-cost, high-impact solution for small municipalities or homeowner associations (HOAs) with limited budgets. Meanwhile, the use of rumble strips or textured pavement near stop signs can physically alert habituated drivers to the intersection, compensating for the lack of opposing traffic that usually forces a stop.

For intersections near schools or small neighborhoods where the "look-but-fail-to-see" phenomenon is a major risk, infrastructure must bridge the gap between driver perception and legal requirements. To combat driver desensitization, systemic bundles (FHWA, 2017) such as reflective post wraps and LED-backlit signs (Layegh et al., 2025) are recommended. These tools are particularly effective at solo stop signs where the absence of a social enforcer (an opposing car) otherwise leads to rolling stops.

In corporate campuses or industrial parks where the "Out" variable suggests that "end-of-shift" time pressure may influence compliance, management should focus on directional flow. Ensuring clear sightlines for pedestrians at exit gates is vital. If drivers are more focused on the transition to the main road, they may overlook internal crossing points. Using "Your Speed" or "Compliance Rate" feedback loops can tap into the social monitoring identified in this study, creating a sense of being observed even when opposing traffic is absent.

Since drivers are most likely to stop for pedestrians already in the conflict zone, infrastructure must bridge the gap for pedestrians waiting at the curb. Using pavement treatments ensures the stop sign is conspicuous even when no immediate pedestrian hazard is perceived.

Finally, while the demographic effect of gender was marginal, the trend toward technical violations among male drivers should inform targeted educational campaigns within these communities. Safety messaging should emphasize that stop signs are not discretionary based on traffic, but absolute safety measures designed to protect the pedestrian effect even when the street appears empty.

Final Summary of Contributions

Ultimately, this study suggests that driver behavior is an exercise in situational risk assessment. Whether in an enclosed university, a gated community, or a small urban neighborhood, compliance is highest when the environment provides immediate feedback of a potential conflict. Future safety interventions should therefore move away from simply adding more signs and toward engineering environments that naturally reinforce the necessity of a full stop.

References

[1] Anowar, S., Alam, H., & Islam, M. T. (2012). Analysis of the effects of crash factors on driver's stop sign violation. *Accident Analysis & Prevention*, 45, 120–128. https://doi.org/10.1016/j.aap.2011.11.011

[2] Arhin, S. A., Anderson, M., & Gatiba, A. (2019). Predicting STOP-sign compliance at all-way stop intersections in close proximity to signalized intersections. *International Journal of Engineering Research & Technology*, 8(11).

[3] Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics*, 19(1), 270. https://doi.org/10.1186/s12859-018-2264-5

[4] Federal Highway Administration. (2017). *Safety evaluation of multiple strategies at stop-controlled intersections* (Report No. FHWA-HRT-17-087). U.S. Department of Transportation. https://highways.dot.gov/sites/fhwa.dot.gov/files/FHWA-HRT-17-087.pdf

[5] Figueroa Jacinto, R., Kappler, E., Petroskey, K., Arndt, S., & Leipold, T. (2024). Naturalistic observations of human driving perceptions and vehicle kinematics at stop sign-controlled intersections. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 68, 1–8. https://doi.org/10.1177/10711813241284799

[6] Goswamy, A., Hallmark, S., Basulto-Elias, G., & Pawlovich, M. (2019). Safety evaluation of stop-sign mounted beacons—A cross-sectional study. *Journal of Transportation Technologies*, 9(1), 95–108. https://doi.org/10.4236/jtts.2019.91006

[7] Hallmark, S. L., Goswamy, A., & Pawlovich, M. (2018). *Safety effectiveness of stop-sign beacons: A cross-sectional study*. Center for Transportation Research and Education, Iowa State University. https://rosap.ntl.bts.gov/view/dot/35595

[8] Hallmark, S., Knickerbocker, S., Hawkins, N., Thapa, R., & Litteral, T. (2022). Evaluation of driver behavior in response to stop sign beacons. *Journal of Transportation Technologies*, 12, 245–257. https://doi.org/10.4236/jtts.2022.122015

[9] Jaccard, J. (2011). Interaction effects in logistic regression. *Sage Publications*.

[10] Lardelli-Claret, P., Llorca-Javaloys, J., Jimenez-Moleón, J. J., Luna-del-Castillo, J. D., & Bueno-Cavanillas, A. (2002). Does vehicle color influence the risk of being involved in a collision? *Epidemiology*, 13(6), 721–724. https://doi.org/10.1097/00001648-200211000-00021

[11] Layegh, M., Foomani, M. G., & Alecsandru, C. (2025). Enhancing safety measures at stop-controlled intersections: A study on LED backlit signs and drivers' behavior in Montréal, Québec. *Urban Science*, 9(9), 375. https://doi.org/10.3390/urbansci9090375

[12] Li, Y. (2023). Predictors selection strategy based on stepwise random forests and logistic regression. *In Proceedings of SPIE* (Vol. 12506). https://doi.org/10.1117/12.2665938

[13] McKelvie, S. J. (1986). An opinion survey and longitudinal study of driver behaviour at stop signs. *Canadian Journal of Behavioural Science*, 18(1), 75–85. https://doi.org/10.1037/h0079958

[14] Nyongesa, H., et al. (2020). Variable selection using random forests. *SAS Global Forum Proceedings*. Retrieved from https://support.sas.com/resources/papers/proceedings20/

[15] Shinar, D., & Compton, R. (2004). Aggressive driving: An observational study of driver, vehicle, and situational variables. *Accident Analysis & Prevention*, 36(3), 429–437. https://doi.org/10.1016/S0001-4575(03)00037-6

[16] Silvano, A. P., & Linder, A. (2017). Interaction between drivers and pedestrians at T-intersections: Do drivers yield? *Transportation Research Part F: Traffic Psychology and Behaviour*, 48, 102–114. https://doi.org/10.1016/j.trf.2017.04.018
[17] Simons-Morton, B. G., Lerner, N., & Singer, J. (2005). Peer influence on passenger-restricted and non-restricted adolescent drivers. *Injury Prevention*, 11(1), 14–18. https://doi.org/10.1136/ip.2004.005371
[18] Van Houten, R., & Retting, R. A. (2001). Increasing motorist compliance and caution at stop signs. *Journal of Applied Behavior Analysis*, 34(2), 185–193. https://doi.org/10.1901/jaba.2001.34-185
[19] Woldeamanuel, M. G. (2012). Stopping behavior of drivers at stop-controlled intersections: Compositional and contextual analysis. *Journal of Traffic and Logistics Engineering*, 1(1).

Appendix A. Full Logistic Regression Model Including All Main Effects
Full Logistic Regression Model: Stop Compliance vs. All Predictors
Dependent Variable: Stop Sign Compliance (1 = Stop, 0 = Non-Stop)

137. **Independent Variable**	138. **Coefficient (B)**	139. **Std. Error**	140. **z-value**	141. **p-value**
142. **Out (Yes)**	143. **0.6421**	144. 0.2014	145. 3.188	146. **0.001** **
147. **Opposing (Yes)**	148. **0.7514**	149. 0.2642	150. 2.844	151. **0.004** **
152. **Pedestrian (Yes)**	153. **0.8942**	154. 0.4712	155. 1.898	156. **0.057**
157. **Gender (Male)**	158. **-0.3421**	159. 0.1985	160. -1.723	161. **0.085**
162. **Occupancy (Single)**	163. -0.1145	164. 0.2154	165. -0.532	166. 0.595
167. Time (N)	168. 0.0842	169. 0.1945	170. 0.433	171. 0.665
172. Vehicle Color (Other)	173. -0.0421	174. 0.2412	175. -0.175	176. 0.861

** 0.01 * 0.05 Significant