Characterizing ILKATBO trigger detection in Volvo Trucks Using XGBoost and SHAPLEY Values

Dag Palmstierna,
Hampus Stålhandske

Bachelor's thesis
2026:K15

Lund Institute of Technology
Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

CENTRUM SCIENTIARUM MATHEMATICARUM

Abstract

This thesis examines how ILKATBO trigger frequency varies across the Volvo truck fleet and which operating conditions and vehicle characteristics are associated with elevated trigger behavior. Using logged vehicle data, trigger rate was analyzed at fleet level and within different driving-condition subsets. The results show that trigger behavior is systematic rather than random. At fleet level, road condition is the most important factor, with rougher driving environments associated with higher trigger frequency, while altitude and temperature also show clear associations. Some vehicle-configuration effects appear consistently across conditions, particularly for front axle load, while others seem more closely tied to regional or operational clusters. Overall, the study shows that ILKATBO trigger frequency is concentrated in identifiable parts of the fleet and that both operating environment and vehicle setup contribute to that variation. Results describe vehicles at the extremes of the trigger rate distribution and do not necessarily characterize the full fleet.

Contents

1 Introduction

A sudden front tire blowout at speed pulls a truck sharply toward the damaged tire. The pressure loss happens in a fraction of a second, leaving the driver very little time to respond before the vehicle deviates from its lane.

Volvo Dynamic Steering (VDS) counters this pull by helping the driver hold their intended steering position, which reduces the sudden wheel rotation caused by the puncture. Some lateral pull remains, however, because the deflated tire generate extra drag on the affected side. Improved Lane Keeping at Tire Blowout (ILKATBO) extends this by detecting a front tire blowout, applying additional corrective steering, and, if lane markings are available, activating lane keeping to help the truck stay within its lane (R. Rutström, 2026).

This thesis examines how often ILKATBO triggers and what drives those rates. No distinction is made between confirmed blowouts and false triggers, all recorded events are treated equally. The central question is whether certain vehicle configurations, usage patterns, or operating conditions are associated with systematically higher or lower trigger rates.

The analysis focuses on vehicles at the extremes of the distribution, those that trigger very frequently versus those that almost never do. The results help Volvo evaluate whether the detection logic behaves as expected and whether it may need adjustment.

To address these questions, three separate XGBoost classifiers are trained on logged fleet data, one for overall operating conditions and two for vehicles operating on rough and smooth road surfaces respectively. The models identify the vehicle characteristics and operating conditions associated with high and low trigger rates. SHAP-based sensitivity analysis is then used to quantify feature contributions and determine the factors driving these differences. A central methodological challenge is that vehicle configuration and operating environment often covary, making it difficult to attribute trigger rate differences to configuration rather than environmental factors.

1.1 Research question 1

How does the trigger rate differ under different driving conditions?

1.1.1 Rationale

Since ILKATBO is expected to react more often in certain demanding situations, the analysis investigates which driving conditions are associated with higher or lower trigger frequency. This provides a first understanding of how the function behaves in relation to the vehicle's operating environment.

1.2 Research question 2

How is vehicle configuration associated with trigger frequency under different driving conditions?

1.2.1 Rationale

Different vehicle configurations may affect how sensitive the trigger functionality is. If so, similar driving conditions could produce different trigger rates depending on the vehicle.

Configurations with unusually high or low rates would indicate the function is not behaving consistently across vehicle types.

2 Background

This section covers VDS, the ILKATBO functionality within it, and the XGBoost method used to analyze ILKATBO trigger behavior.

2.1 Volvo Dynamic Steering

VDS combines a conventional hydraulic power steering system with an electric motor on the steering gear. The system reads signals from various sensors to determine the truck's direction and the driver's intent, adjusting the motor 2,000 times per second to correct unwanted steering movements and supply additional torque where needed. It also includes active safety functions. If it detects that the truck is beginning to lose traction, it can immediately apply a small corrective steering input, often before the driver has time to react (Volvo Trucks, n.d. -b).

2.2 ILKATBO

ILKATBO is a dedicated VDS function designed to detect front tire blowouts and apply additional counter-steering in order to reduce the remaining path deviation beyond what standard VDS alone can compensate for. For this to work reliably, the detection logic must perform consistently across the full range of vehicle configurations and operating conditions (R. Rutström, 2026).

The ILKATBO functionality listens to three different sensors. Information from these sensors are input to four different parameters. Each of these parameters contributes to a weight, and if enough requirements are fulfilled, a tire blowout is triggered (R. Rutström, 2026).

2.3 Logged Data Analytics within Volvo

Volvo's logged data analytics (LDA) platform consists of logged vehicle data including parameters, global transport application (GTA) data, and vehicle specifications (VDA) (Lindroth, 2025). GTA framework was created within Volvo Group to describe driving conditions, behavior, and vehicle usage through a set of standardized parameters (Volvo Trucks, 2025). VDA typically contains variant information for a vehicle, these are vehicle specifications set by the customer at time of purchase.

2.4 XGBoost

XGBoost (Extreme Gradient Boosting) is an ensemble learning method based on gradient boosted decision trees, introduced by Chen & Guestrin (2016). It is widely used in applied machine learning for its predictive performance, computational efficiency, and built-in handling of missing data. The approach has been applied to commercial truck data for purposes such as crash factor identification (Yang et al., 2021) and fleet-scale driver risk prediction from telematics (Mehdizadeh et al., 2021). Here, XGBoost is used to classify vehicles into High and Low trigger-rate groups and to identify which features contribute most to this classification

2.4.1 Decision Trees

A decision tree partitions the feature space by recursively splitting on input variables. Each leaf node assigns a constant predicted value to the observations that fall within the corresponding region. Formally, a tree with T leaf nodes defines:

$$f(\mathbf{x}) = w_{q(\mathbf{x})}, \quad w \in \mathbb{R}^T,\ q : \mathbb{R}^d \to \{1, \ldots, T\} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^d$ is the input feature vector, $q(\mathbf{x})$ maps each observation to a leaf index, and w is the vector of leaf weights (Chen & Guestrin, 2016).

2.4.2 Gradient Boosting

Rather than fitting a single tree, gradient boosting builds an ensemble of K trees sequentially. The prediction for observation i after K rounds is:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F} \tag{2}$$

where $\mathcal{F}$ is the space of all regression trees. At each round t, a new tree f_t is added to minimize the regularized objective (Chen & Guestrin, 2016):

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\Big(y_i,\ \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\Big) + \Omega(f_t) \tag{3}$$

where l is a differentiable loss function and $\Omega(f_t)$ penalizes model complexity:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{4}$$

γ penalizes the number of leaves T and λ penalizes the magnitude of the leaf weights, both acting to limit overfitting (Chen & Guestrin, 2016).

2.4.3 Second-Order Approximation

To make the optimization tractable, XGBoost applies a second-order Taylor expansion of the loss around $\hat{y}_i^{(t-1)}$:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t(\mathbf{x}_i)^2 \right] + \Omega(f_t) \tag{5}$$

where g_i and h_i are the first and second-order gradients of the loss. For a fixed tree structure q, the optimal leaf weight for leaf j is:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{6}$$

where $I_j = \{i \mid q(\mathbf{x}_i) = j\}$ is the set of observations assigned to leaf j. The corresponding optimal objective value is:

$$\mathcal{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{7}$$

This score is the splitting criterion when evaluating candidate tree structures (Chen & Guestrin, 2016).

2.5 SHAP

While XGBoost provides built-in feature importance measures such as gain, these only indicate how frequently and effectively a feature is used for splitting across all trees. They do not reveal the direction of a feature's effect or how it contributes to individual predictions. To address this, SHAP (SHapley Additive exPlanations) is used to interpret model predictions (Lundberg et al., 2020).

SHAP decomposes each prediction into additive contributions from individual features. For a prediction $\hat{y}_i$:

$$\hat{y}_i = \phi_0 + \sum_{j=1}^{p} \phi_{ij} \tag{8}$$

where ϕ_0 is the base value (the mean model output over the training set), p is the number of features, and ϕ_{ij} is the SHAP value for feature j and observation i, representing its contribution to the deviation from ϕ_0 (Lundberg et al., 2020).

Overall feature importance is assessed by computing the mean absolute SHAP value per feature across all observations:

$$\overline{|\phi_j|} = \frac{1}{n} \sum_{i=1}^{n} |\phi_{ij}| \tag{9}$$

This ranks features by how much they contribute to the model's predictions on average, regardless of direction.

To capture the direction of a feature's contribution, the SHAP value can be computed per feature value:

$$\bar{\phi}_{j,k} = \frac{1}{n_k} \sum_{i \in k} \phi_{ij} \tag{10}$$

where k denotes a specific feature value and n_k is the number of observations with that value. A positive $\bar{\phi}_{j,k}$ indicates that observations with value k are on average pushed toward a higher model output, a negative value indicates the opposite. Together, these measures make it possible to examine not only which features matter most but also how specific feature values shift individual predictions.

3 Method

This section presents the data, preprocessing steps, modelling procedure, and interpretation approach used in the study.

3.1 Data Gathering

The data used in this study was retrieved from LDA. It consists of GTA and VDA parameters, where the target variable is computed using logged trigger events and distance driven for each vehicle.

3.2 Data Filtering and Feature Engineering

To ensure consistency in the logged data, several filtering steps were applied before constructing the target variable. Only vehicles running on a specific software version as well as vehicles logging trigger events were included in the dataset.

For each vehicle, the first and last available logging entries were identified. Vehicles for which the first and last logging dates were identical were removed, as no valid observation interval could be established. Vehicles with an insufficient distance driven was also removed.

A consistency check was applied to detect invalid logging behavior. Vehicles with negative changes between successive readings of either trigger counts or driven distance were flagged and removed. Such negative changes indicate counter resets or logging errors and would otherwise lead to invalid rate calculations.

After these filtering steps, the total change in trigger counts and driven distance was computed for each vehicle using the difference between the first and last valid readings. The trigger rate was then defined as the ratio between the total number of trigger events and the total driven distance for each vehicle.

Feature selection started with roughly 800 feature families from the logged vehicle data. Features missing values for more than 50% of vehicles were removed first. The remaining features were then screened for relevance to tire blowout triggering, and those with no plausible connection to tire behavior or vehicle operation were dropped based on manual inspection of feature definitions and domain expert input. Approximately 30 features remained.

To construct a clean binary target, only vehicles in the outer quantiles of the trigger rate distribution were retained. Vehicles below the first quartile (Q_1) were labelled as Low and vehicles above the third quartile (Q_3) as High, while vehicles in the middle of the distribution were excluded. The quantile thresholds were computed separately for each of the three analyses:

$$\text{GTA:} \quad Q_1 = 0.5307, \quad Q_3 = 3.3732 \tag{11}$$

$$\text{SMOOTH:} \quad Q_1 = 0.4481, \quad Q_3 = 2.5711 \tag{12}$$

$$\text{ROUGH:} \quad Q_1 = 1.1462, \quad Q_3 = 7.8060 \tag{13}$$

3.3 Fitting XGBoost

The processed and labeled dataset was used to train an XGBoost classifier to separate High and Low trigger vehicles based on the engineered features. The subsequent analysis focused on which features drove the classifications, assessed through SHAP values.

3.3.1 Hyperparameter Tuning

Hyperparameters were selected through a grid search over all combinations of candidate values for `n_estimators` (number of boosting rounds), `learning_rate` (contribution of each tree), and `max_depth` (maximum tree depth). The full set of candidate values is provided in Section A.

To reduce the risk of overfitting, 20% of the labeled vehicles were set aside as a stratified validation set. A grid search with 5-fold cross-validation was run on the remaining 80%, and the best configuration was evaluated against the held-out 20%. The selected hyperparameters, shown in Table 1, were identical across all three models.

Table 1: Final XGBoost hyperparameters selected via grid search.

Hyperparameter	**GTA**	**SMOOTH**	**ROUGH**
`n_estimators`	300	300	300
`learning_rate`	0.1	0.1	0.1
`max_depth`	4	4	4

Table 2: Final XGBoost hyperparameters selected via grid search.

3.4 Evaluation of Performance

The evaluation had two stages. First, the classifier's predictive performance was checked: if the model had not learned meaningful patterns in the data, the SHAP interpretation would be unreliable. Second, a SHAP analysis identified the main drivers of ILKATBO trigger frequency.

3.4.1 Classification Performance

With the selected hyperparameters fixed, predictive performance was estimated using a stratified 5-fold cross-validation. In each fold, the model was trained on 80% of the vehicles and evaluated on the remaining 20%. Performance is reported as the mean and standard deviation of accuracy and ROC-AUC across the five test folds.

The final model fitted on the full dataset was then used for SHAP-based interpretation.

3.4.2 SHAP Analysis

To interpret the fitted models and identify what drives ILKATBO trigger frequency, SHAP values were computed for each feature on the trained classifiers, following the framework described in Section 2.5. The analysis was carried out only after confirming classification performance well above chance, since the SHAP interpretation depends on the model having learned to separate the two classes reliably.

The analysis proceeded in two parts. First, overall feature importance was assessed by computing the mean absolute SHAP value per feature across all vehicles. This ranks features by how consistently they distinguish High-trigger from Low-trigger vehicles, regardless of direction.

Second, the directionality and magnitude of each feature's contribution were examined using two complementary approaches. For numerical features and other features with a natural internal ordering, beeswarm plots were used to visualise the relationship between feature values and SHAP values at the observation level. These plots provide insight into whether relatively high or low values of a feature tend to push predictions toward the High-trigger class.

Furthermore, mean SHAP values were computed for each feature-value category. Positive mean SHAP values indicate that vehicles belonging to that category are, on average, pushed toward High-trigger classification, whereas negative mean SHAP values indicate the reverse. For numerical features with more than 20 distinct values, values were binned prior to the category-level analysis. This was done to improve interpretability and statistical robustness. By grouping values into intervals, the analysis captures broader and more reliable patterns in how different value ranges affect the prediction. These binned numerical categories were also included in the top 20 and bottom 20 feature value bar plots.

Feature values with fewer than 30 observations were excluded from this comparison, as mean SHAP values derived from small groups are dominated by individual vehicle effects and cannot be interpreted as representative of that configuration.

3.4.3 Exploratory Data Analysis

To better check whether the SHAP results from the vehicle configuration models reflect genuine hardware effects or regional clustering, the ROUGH and SMOOTH models were re-run with the GTA parameters included as additional features. `GTA_Road_Condition_1year` was excluded since it defines the subgroup split, and `Country` was added to directly capture regional concentration.

After fitting the extended models, the SHAP feature-value results for vehicle configuration variables were examined separately, while GTA variables and Country were excluded from the interpretation step. The aim was not to assess whether GTA or regional variables were important in themselves, but rather to evaluate whether the importance of configuration-related feature values remained stable once such contextual information was made available to the model. If a vehicle configuration feature retains its SHAP importance even when the model has access to operating environment variables, the signal is more likely to reflect a genuine configuration effect.

4 Results

This section presents the results of the three models. First, the classification performance of each model is reported. The SHAP analysis results are then presented, grouped by plot type. Finally, key observations from the exploratory data analysis are discussed before a brief summary concludes the section.

4.1 Classification Performance

The predictive performance of the XGBoost classifier was evaluated using stratified 5-fold cross-validation for the GTA, ROUGH, and SMOOTH analyses. Table 3 summarizes the test performance across folds, reported as mean ± standard deviation.

Table 3: Cross-validated classification performance.

Metric	GTA	ROUGH	SMOOTH
Accuracy	0.7737 ± 0.0080	0.8190 ± 0.0110	0.8627 ± 0.0056
ROC-AUC	0.8574 ± 0.0067	0.8968 ± 0.0064	0.9308 ± 0.0071

The results show that all three models achieved stable predictive performance across the cross-validation folds. Among the three analyses, the SMOOTH model performed best, followed by the ROUGH model and then the full GTA model. The ROC-AUC values indicate that the classifier was able to separate High- and Low-trigger vehicles consistently.

4.2 SHAP Analysis

The SHAP analysis is presented in four parts. First, mean absolute SHAP values rank features by overall importance for each model. Second and third, the top and bottom feature values by mean SHAP value reveal which specific values push toward High- or Low-trigger classification. Fourth, beeswarm plots show how numerical feature values relate to the model output.

4.2.1 Feature Importance

Figures 1, 2, and 3 show the mean absolute SHAP values for the GTA, ROUGH, and SMOOTH models respectively. These values summarize how strongly each feature contributed to the classification on average.

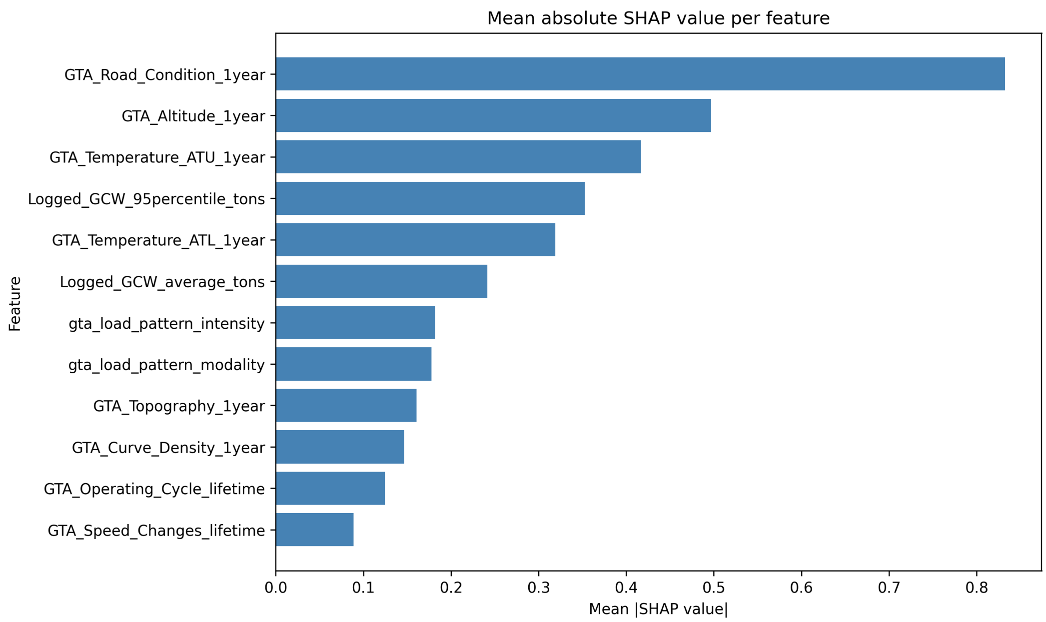

Figure 1: Mean absolute SHAP values for the full GTA model, ranked by importance.

In the GTA model (Figure 1), `GTA_Road_Condition_1year` has the highest mean absolute SHAP value, followed by `GTA_Altitude_1year`, `GTA_Temperature_ATU_1year`, `GTA_Temperature_ATL_1year`, and `Logged_GCW_95percentile_tons`. The remaining features contribute less individually but still retain non-negligible importance.

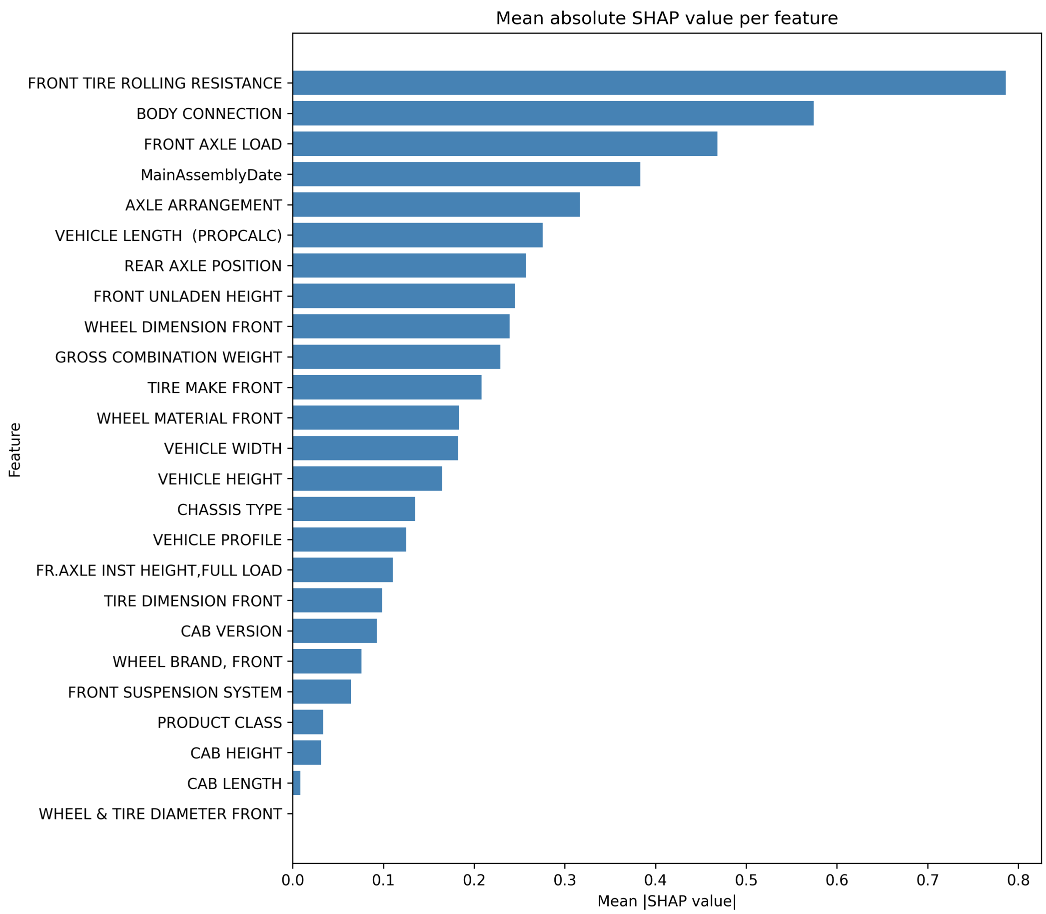

Figure 2: Mean absolute SHAP values for ROUGH model, ranked by importance.

In the ROUGH model (Figure 2), `FRONT TIRE ROLLING RESISTANCE` and `BODY CONNECTION` have the highest mean absolute SHAP values. `FRONT AXLE LOAD` ranks third, followed by `MainAssemblyDate` and `AXLE ARRANGEMENT`. A second tier of moderately important features includes `VEHICLE LENGTH (PROPCALC)`, `REAR AXLE POSITION`, `FRONT UNLDEN HEIGHT`, `TIRE MAKE FRONT`, `GROSS COMBINATION WEIGHT`, and `WHEEL MATERIAL FRONT`.

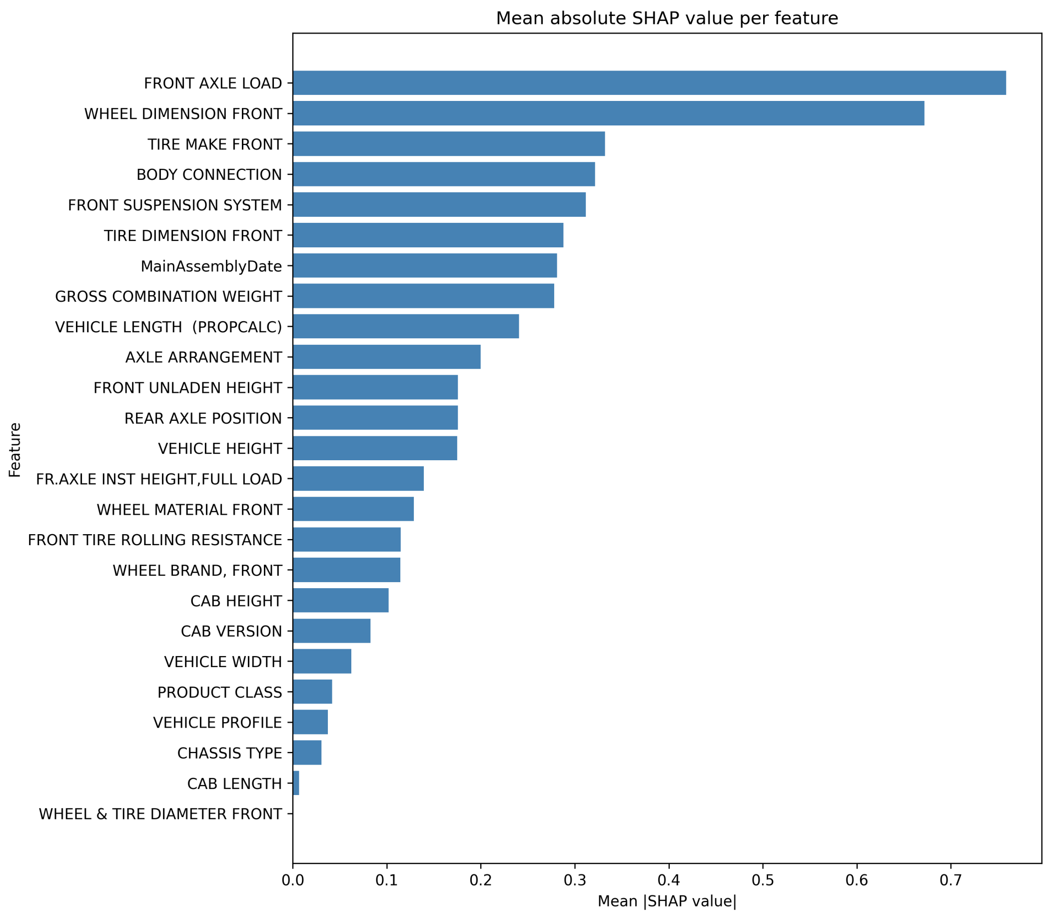

Figure 3: Mean absolute SHAP values for SMOOTH model, ranked by importance.

In the SMOOTH model (Figure 3), `FRONT AXLE LOAD` and `WHEEL DIMENSION FRONT` have the highest mean absolute SHAP values and `TIRE MAKE FRONT` ranks third, followed by `BODY CONNECTION`, `FRONT SUSPENSION SYSTEM`, and `TIRE DIMENSION FRONT`. Notably, `FRONT TIRE ROLLING RESISTANCE` and `BODY CONNECTION`, which ranked first and second under ROUGH conditions, are less dominant under SMOOTH conditions, ranking fifteenth and fourth respectively.

4.2.2 Top Feature Values

Figures 4, 5, and 6 show the feature values with the highest mean SHAP values for the GTA, ROUGH, and smooth models respectively.

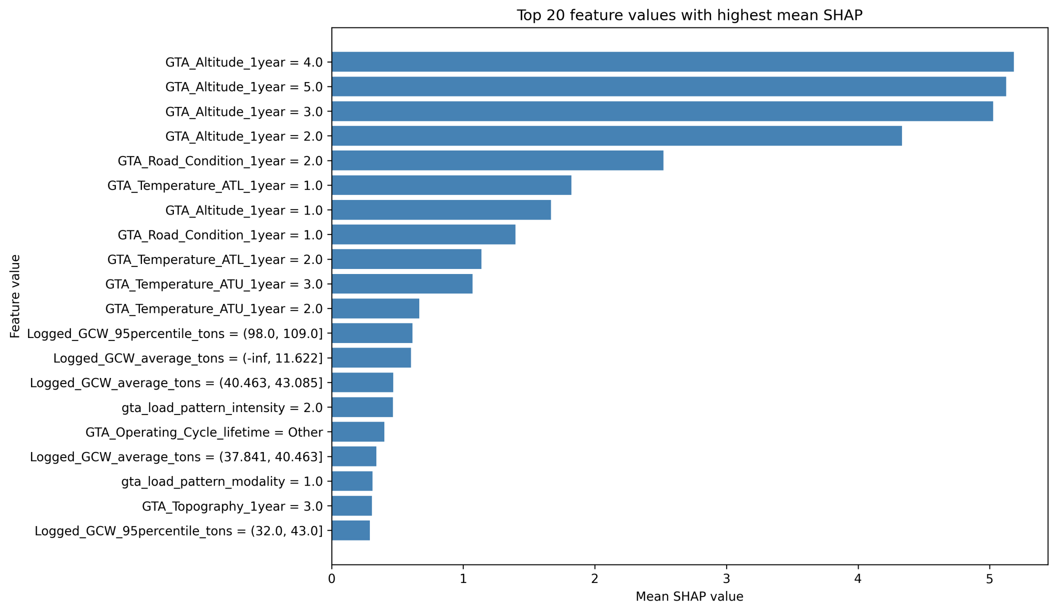

Figure 4: Top 20 feature values with the highest mean SHAP values in the GTA model.

In the GTA model (Figure 4), the feature values with the strongest positive SHAP contributions are high-altitude values: `GTA_Altitude_1year = ALTI4000` has the highest mean SHAP value, followed by `ALTI5500`, `ALTI2500`, and `ALTI3000`. `GTA_Road_Condition_1year = VERY ROUGH` and `ROUGH` rank fifth and seventh respectively. Low lower limit temperature values also contribute positively, with `GTA_Temperature_ATL_1year = ATL40` and `ATL25` ranking sixth and eight. The two highest upper temperature limits (`GTA_Temperature_ATU_1year = ATU-VH` and `ATU40`) also push towards higher trigger rates.

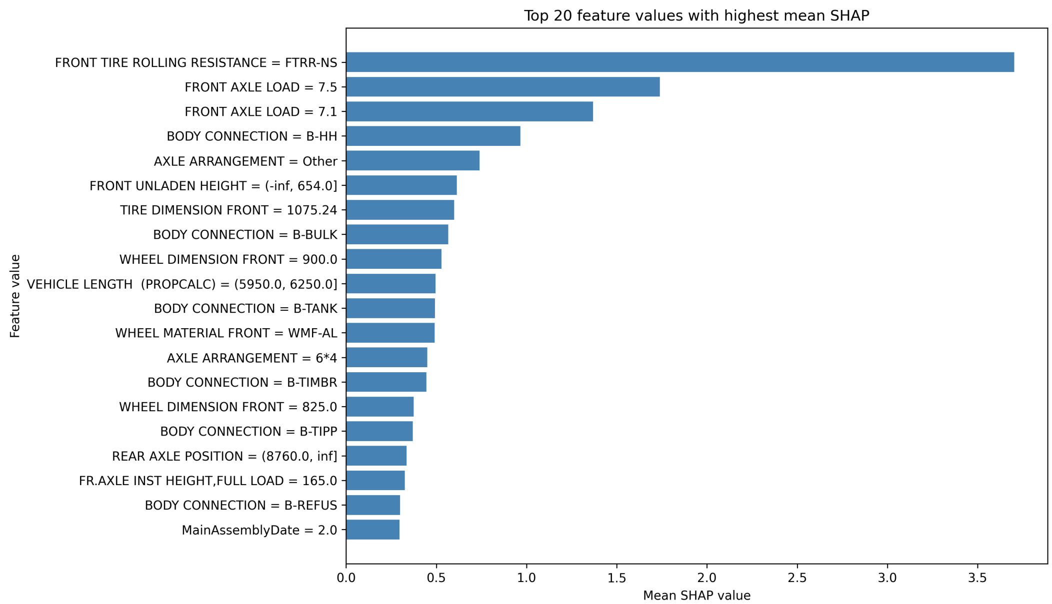

Figure 5: Top 20 feature values with the highest mean SHAP values under rough/very rough conditions.

In the ROUGH model (Figure 5), the feature value with the strongest positive SHAP contribution is `Front Tire Rolling Resistance = FTRR-NS`, with a mean SHAP value above 3.0. This is followed by `Front Axle Load = 7.5` and `Front Axle Load = 7.1`. `Body Connection = B-TIMBR` and `Body Connection = B-HH` also rank among the top positive contributors. Further entries include `Fr.Axle Inst Height, Full Load = 135.0` and `130.0`, `Axle Arrangement = Other`, `Body Connection = B-BULK`, `Gross Combination Weight = 40.0`, and `Wheel Material Front = WMF-AL`.

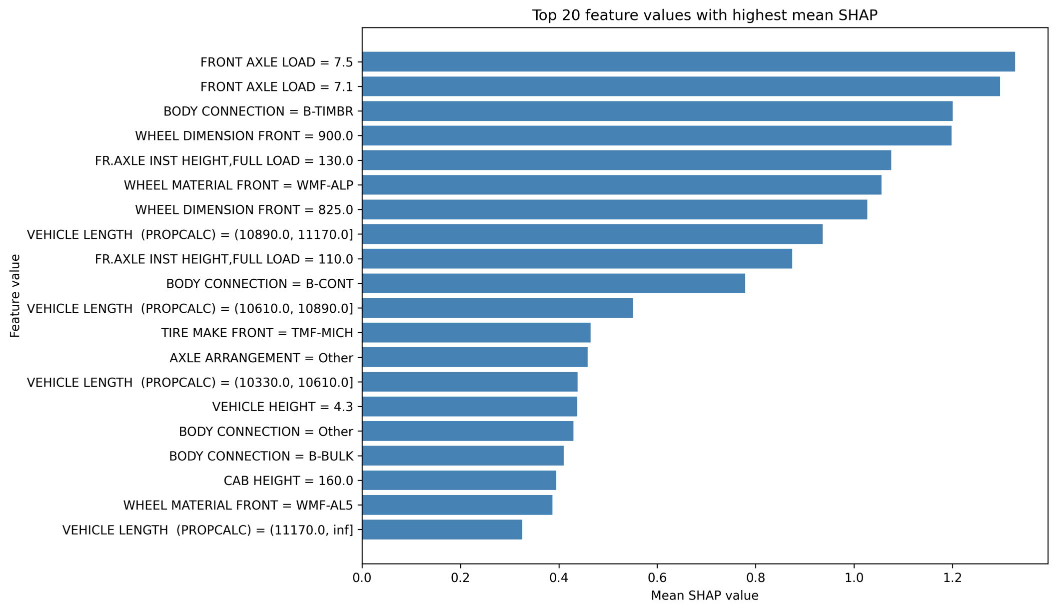

Figure 6: Top 20 feature values with the highest mean SHAP values under smooth conditions.

In the SMOOTH model (Figure 6), the feature value with the strongest positive SHAP contribution is `Fr.Axle Inst Height, Full Load = 135`, followed by `Front Axle Load = 7.5` and `Front Axle Load = 7.1`. `Body Connection = B-TIMBR` and `Wheel Dimension Front = 900.0` also rank among the top entries, along with `Fr.Axle Inst Height, Full Load = 130.0`, `Wheel Material Front = WMF-ALP`, and `Wheel Dimension Front = 825.0`.

4.2.3 Bottom Feature Values

Figures 7, 8, and 9 show the feature values with the lowest mean SHAP values for the GTA, ROUGH, and SMOOTH models respectively.

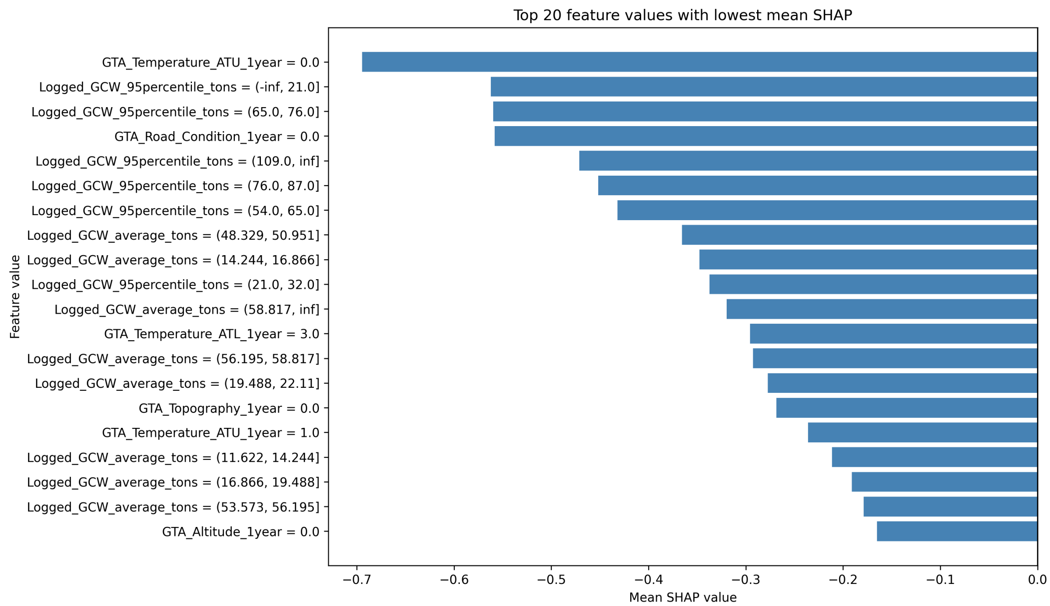

Figure 7: Top 20 feature values with the lowest mean SHAP values in the GTA model.

In the GTA model (Figure 7), the feature values most strongly associated with Low-trigger classification include `GTA_Temperatue_ATU_1year = ATU25`, `Logged_GCW_95percentile_tons = (56.0, 70.0)`, and `GTA_Road_Condition_1year = SMOOTH`. Several `Logged_GCW_95percentile_tons` intervals, including `(7.99, 24.0)` and `(24.0, 32.0)`, also appear among the top negative contributors.

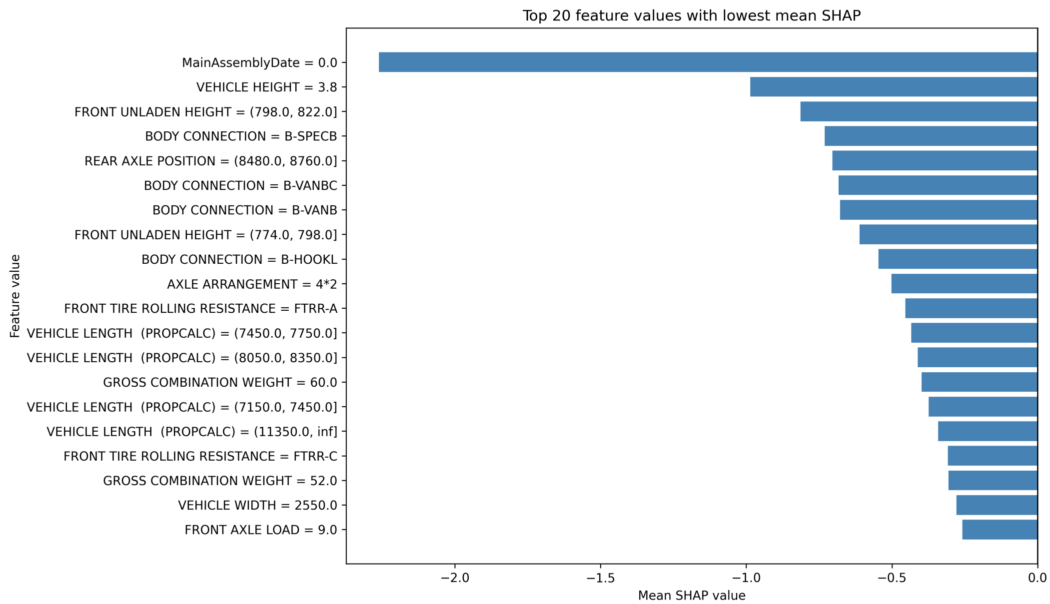

Figure 8: Top 20 feature values with the lowest mean SHAP values under rough/very rough conditions.

In the ROUGH model (Figure 8), the feature values most strongly associated with Low-trigger classification are `MainAssemblyDate = 0.0`, `VEHICE HEIGHT = 3.8`, and several `BODY CONNECTION` values: `B-VANBC`, `B-VANB`, `B-SPECB`, and `B-HOOKL`. `FRONT TIRE ROLLING RESISTANCE = FTRR-A` also contributes negatively, as do several `MainAssemblyDate` values (`8.0`, `13.0`, `10.0`, `11.0`, `7.0`) and `FRONT UNLADEN HEIGHT = (760.0, 900.0)`.

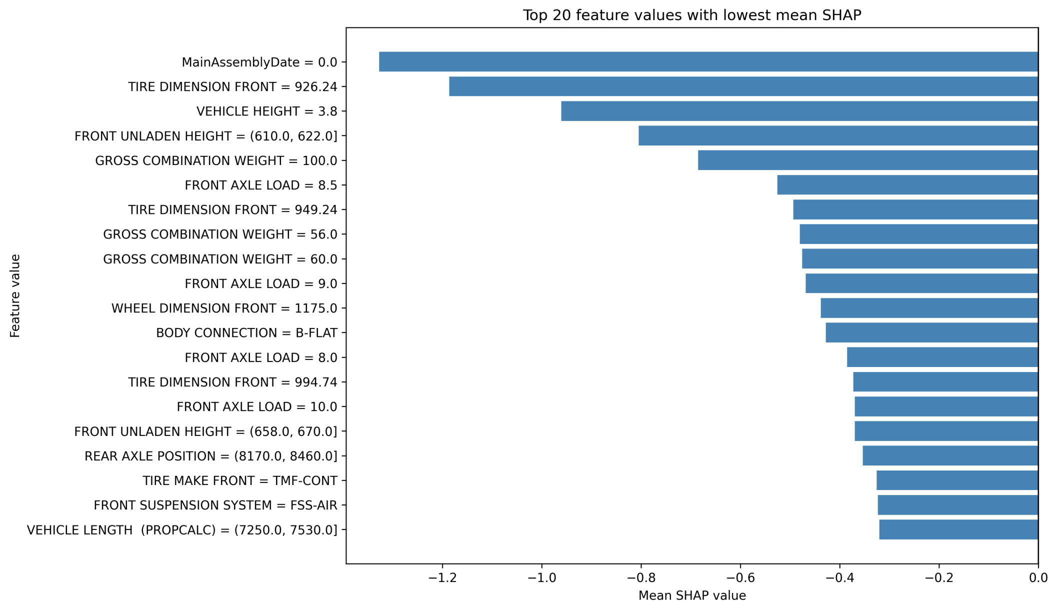

Figure 9: Top 20 feature values with the lowest mean SHAP values under smooth conditions.

In the SMOOTH model (Figure 9), the feature values most strongly associated with Low-trigger classification are `MainAssemblyDate = 0.0`, `Vehicle Height = 3.8`, and `TIRE DIMENSION FRONT = 926.24`. These are followed by `MainAssemblyDate = 9.0`, `BODY CONNECTION = B-FLAT`, and `WHEEL DIMENSION FRONT = 1175.0`. High `FRONT AXLE LOAD` values (`8.5`, `9.0`, `10.0`, `8.0`) appear repeatedly among the negative contributors, as do several `GROSS COMBINATION WEIGHT` values (`120`, `100.0`, `56.0`, `60.0`) and `FRONT SUSPENSION SYSTEM = FSS-AIR`.

4.2.4 Beeswarm Plots

Figures 10, 11, and 12 show the SHAP beeswarm plots for numerical and ordinal features for the GTA, ROUGH, and SMOOTH models respectively. Each point represents a vehicle, and the horizontal position shows the SHAP value of that feature for that vehicle. The colour indicates the relative feature value within that feature, where red corresponds to higher values and blue to lower values. Grey points represent vehicles for which no information was available for that feature. The features are ordered from top to bottom by overall SHAP feature importance, such that the most influential features appear at the top and the less influential ones lower down.

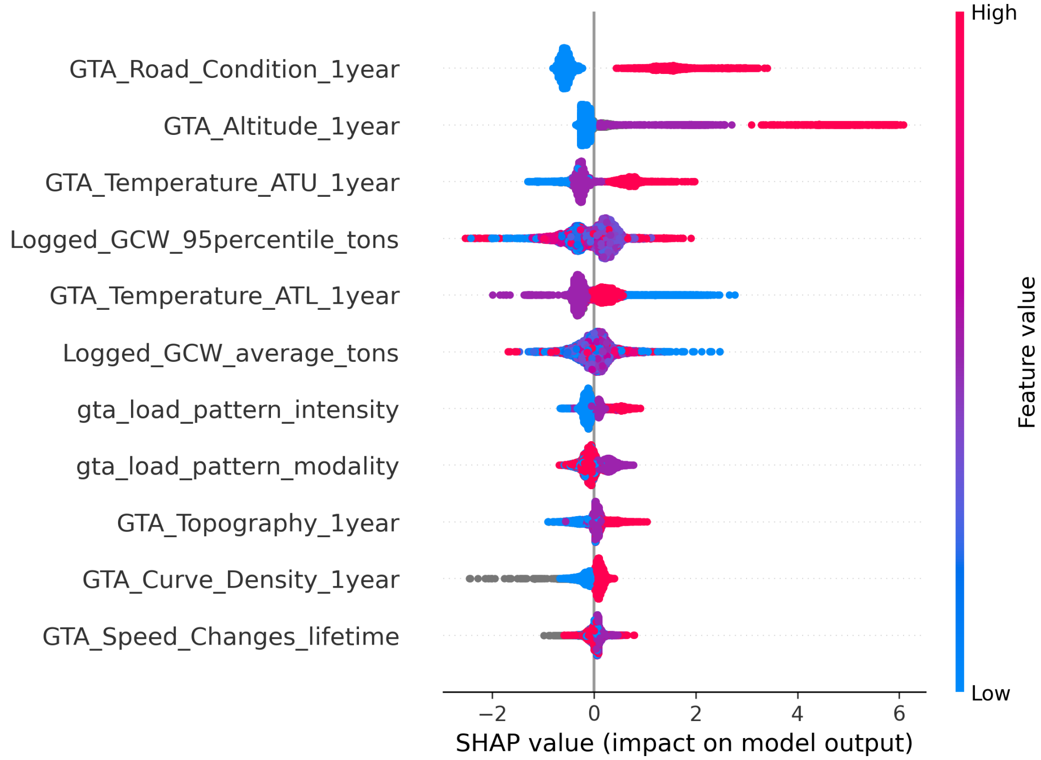

Figure 10: SHAP beeswarm plot for numerical features in the GTA model.

The top-ranked features in the GTA beeswarm plot (Figure 10) all show clear colour gradients along the SHAP axis. `GTA_Road_Condition_1year` separates cleanly, with blue (smoother conditions) clustering on the negative side while red (rougher conditions) fans out across positive SHAP values. `GTA_Altitude_1year` follows a similar pattern, with high-altitude vehicles producing some of the largest positive contributions in the plot.

`GTA_Temperature_ATU_1year` has a three-band structure where low temperatures sit on the negative side, intermediate values near zero, and high temperatures shift toward positive SHAP. `GTA_Temperature_ATL_1year` is broadly similar, though the lowest values sit near the centre rather than the negative side, with intermediate values pulling slightly negative and high values pushing positive.

`Logged_GCW_95percentile_tons` and `Logged_GCW_average_tons` do not follow this pattern. Both show red and blue points scattered across both sides of zero, with no clear monotonic relationship between weight and SHAP contribution. Gross combination weight is not a simple linear predictor in the GTA model.

Among features with smaller SHAP magnitudes, several still carry a consistent direction. `gta_load_pattern_intensity`, `GTA_Topography_1year`, and `GTA_Curve_Density_1year` each show a colour gradient from blue on the left to red on the right. The contributions are smaller, but the direction is clear.

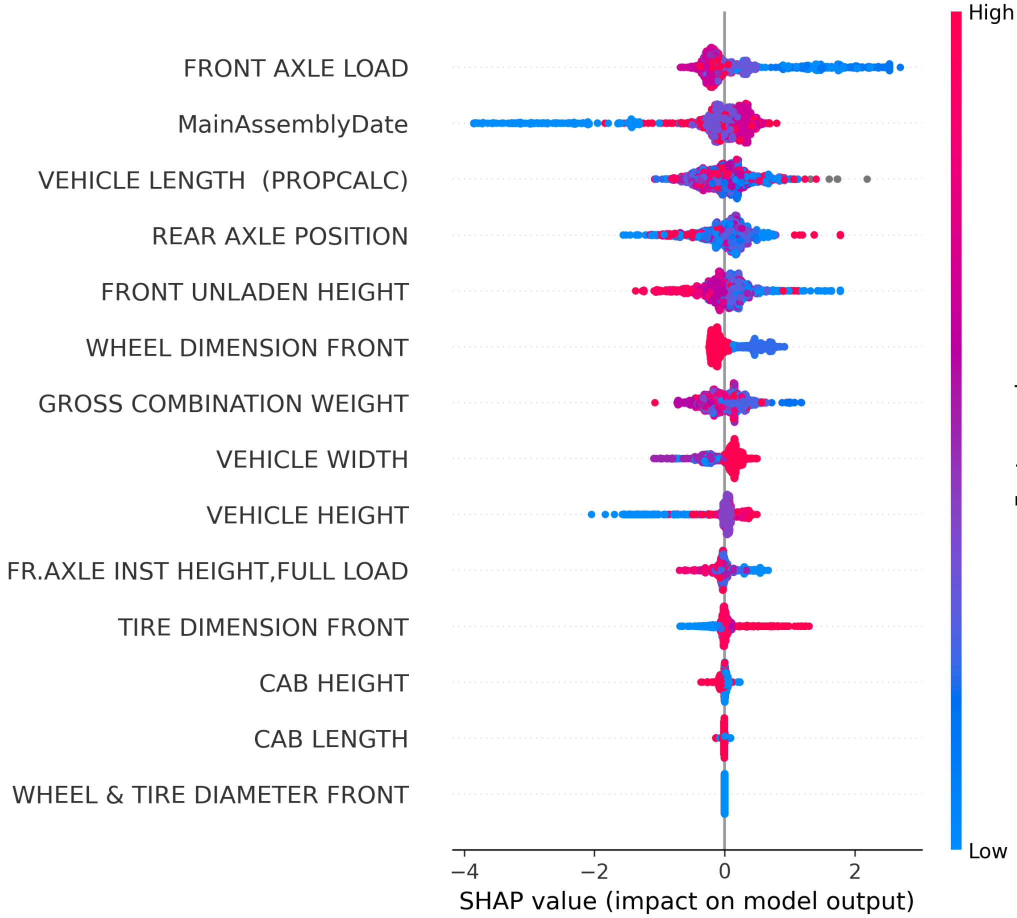

Figure 11: SHAP beeswarm plot for numerical features under ROUGH conditions.

Under rough conditions (Figure 11), `FRONT AXLE LOAD` dominates the beeswarm. Blue points (lower axle loads) extend well into positive SHAP territory, while red points (higher loads) cluster on the negative side with a narrower spread. `MainAssemblyDate` also separates by colour, with older vehicles (blue) receiving more negative SHAP contributions and newer ones shifting toward zero or slightly positive.

`VEHICLE LENGTH (PROPCALC)`, `GROSS COMBINATION WEIGHT`, and `REAR AXLE POSITION` show no consistent colour separation. The swarm for each is centred around zero with red and blue intermixed on both sides.

Several lower-ranked features still show directional patterns. `FRONT UNLDEN HEIGHT` and `WHEEL DIMENSION FRONT` both show an inverse pattern where smaller values appear more often on the positive SHAP side and larger values on the negative side. `TIRE DIMENSION FRONT` reverses this, with larger tires appearing more often on the positive side. `VEHICLE WIDTH` and `VEHICLE HEIGHT` show weaker but visible tendencies, with higher values leaning toward positive SHAP.

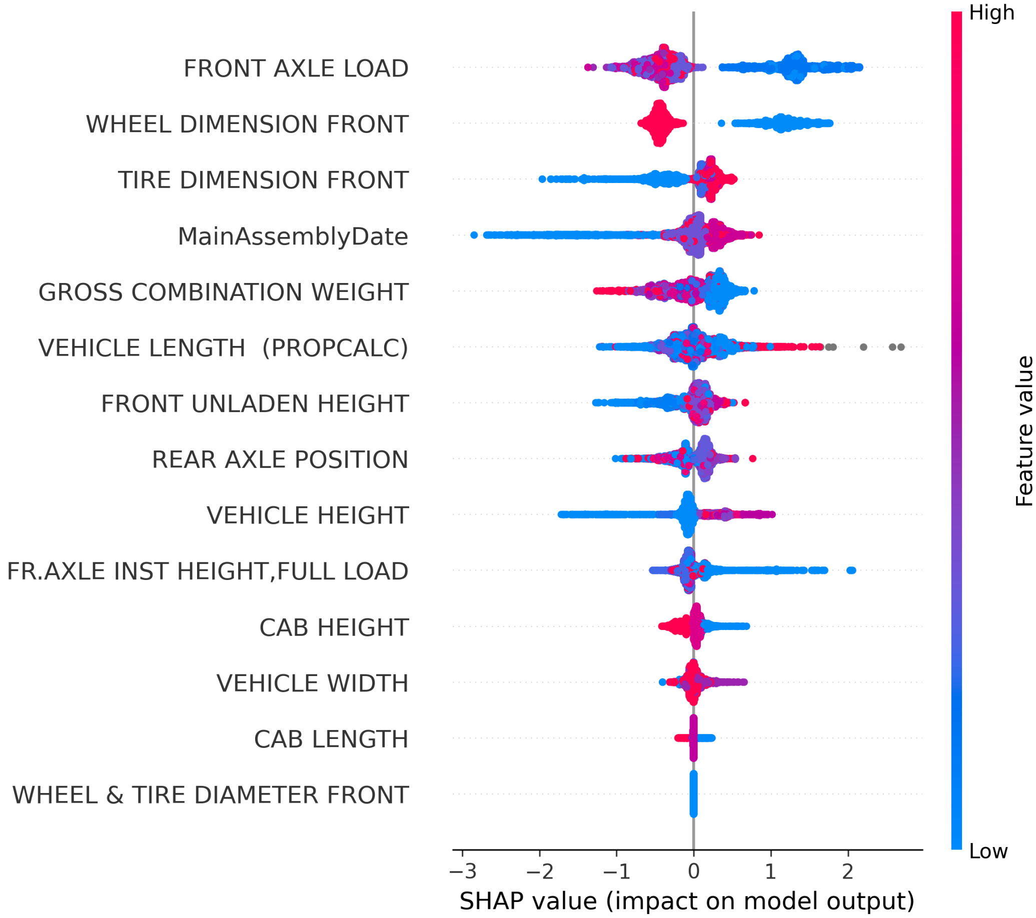

Figure 12: SHAP beeswarm plot for numerical features under smooth conditions.

In the SMOOTH beeswarm (Figure 12), `FRONT AXLE LOAD` and `WHEEL DIMENSION FRONT` both rank at the top and show sharp colour separation. For `FRONT AXLE LOAD`, blue points (lower loads) stretch further into positive SHAP territory, while red and mid-range values are compressed on the negative side. `WHEEL DIMENSION FRONT` mirrors this, with smaller dimensions on the positive SHAP side and larger ones on the negative side. `TIRE DIMENSION FRONT` shows the opposite, with larger tire dimensions on the positive side and smaller ones on the negative side. `MainAssemblyDate` shows a gradient where older assembly dates (blue) appear more often on the positive SHAP side and newer dates (red) on the negative side.

`GROSS COMBINATION WEIGHT`, `VEHICLE LENGTH (PROPCALC)`, and `REAR AXLE POSITION` show no clean colour separation along the SHAP axis, with red and blue values appearing on both sides of zero.

At the lower end of the importance ranking, the patterns are subtler. `FRONT UNLADEN HEIGHT` shows lower values more often on the negative SHAP side. `VEHICLE HEIGHT` follows a gradient with taller vehicles on the positive side and shorter ones on the negative side.

`FR.AXLE INST HEIGHT,FULL LOAD` shows lower values more often on the positive SHAP side. `CAB HEIGHT`, `VEHICLE WIDTH`, `CAB LENGTH`, and `WHEEL & TIRE DIAMETER FRONT` are all tightly clustered around zero with little discernible colour structure.

4.3 Exploratory Data Analysis

Tables 4 and 5 show the top 20 feature values with the highest mean SHAP values for the extended ROUGH and SMOOTH models respectively.

Table 4: Top 20 feature values for ROUGH model with delta rank included

feature	feature_group	rank_table_1	rank_change
FRONT TIRE ROLLING RESISTANCE	FTRR-NS	1	0
FRONT AXLE LOAD	7.5	2	0
FRONT AXLE LOAD	7.1	3	0
BODY CONNECTION	B-HH	4	-5
AXLE ARRANGEMENT	Other	5	1
FRONT UNLADEN HEIGHT	(-inf, 654.0]	6	-1
TIRE DIMENSION FRONT	1075.24	7	2
BODY CONNECTION	B-BULK	8	-13
WHEEL DIMENSION FRONT	900.0	9	-8
VEHICLE LENGTH (PROPCALC)	(5950.0, 6250.0]	10	-5
BODY CONNECTION	B-TANK	11	1
WHEEL MATERIAL FRONT	WMF-AL	12	4
AXLE ARRANGEMENT	6*4	13	-30
BODY CONNECTION	B-TIMBR	14	0
WHEEL DIMENSION FRONT	825.0	15	-9
BODY CONNECTION	B-TIPP	16	5
REAR AXLE POSITION	(8760.0, inf]	17	-67
FR.AXLE INST HEIGHT,FULL LOAD	165.0	18	-8
BODY CONNECTION	B-REFUS	19	3
MainAssemblyDate	2.0	20	-17

The rough extended model keep the top 3 feature values unchanged. And kept 15 out of the original 20 top ranked feature values as the most important among vehicle configurations.

Table 5: Top 20 feature values for SMOOTH model with delta rank included

feature	feature_group	rank_table_1	rank_change
FRONT AXLE LOAD	7.5	1	-1
FRONT AXLE LOAD	7.1	2	-2
BODY CONNECTION	B-TIMBR	3	2
WHEEL DIMENSION FRONT	900.0	4	1
FR.AXLE INST HEIGHT,FULL LOAD	130.0	5	-2
WHEEL MATERIAL FRONT	WMF-ALP	6	-13
WHEEL DIMENSION FRONT	825.0	7	2
VEHICLE LENGTH (PROPCALC)	(10890.0, 11170.0]	8	-3
FR.AXLE INST HEIGHT,FULL LOAD	110.0	9	3
BODY CONNECTION	B-CONT	10	-2
VEHICLE LENGTH (PROPCALC)	(10610.0, 10890.0]	11	-34
TIRE MAKE FRONT	TMF-MICH	12	-11
AXLE ARRANGEMENT	Other	13	-19
VEHICLE LENGTH (PROPCALC)	(10330.0, 10610.0]	14	-6
VEHICLE HEIGHT	4.3	15	-50
BODY CONNECTION	Other	16	1
BODY CONNECTION	B-BULK	17	8
CAB HEIGHT	160.0	18	4
WHEEL MATERIAL FRONT	WMF-AL5	19	-85
VEHICLE LENGTH (PROPCALC)	(11170.0, inf]	20	-27

For the SMOOTH extended model the top three feature values stayed, and out of the 20 original top ranked features 15 remained.

4.4 Summary

All three XGBoost models achieved high predictive performance, indicating that all three models learned meaningful patterns that could be interpreted further with SHAP.

For Research Question 1, `GTA_Road_Condition_1year` was the most important feature. When looking at feature-value importance, high-altitude categories and `ROUGH` and `VERY ROUGH` road conditions were among the strongest positive contributors to High-trigger classification. On the Low-trigger side, `SMOOTH` road condition, lower gross combination weight intervals and lower altitude appeared among the strongest negative contributors. The beeswarm analysis further showed that road condition and altitude displayed clear directional patterns.

For Research Question 2, the most important vehicle configuration features differed between driving conditions. Under ROUGH conditions, `FRONT TIRE ROLLING RESISTANCE` and `BODY CONNECTION` ranked highest. Under SMOOTH conditions, `FRONT AXLE LOAD` and `WHEEL DIMENSION FRONT` ranked highest. Across both subgroup models, lower `FRONT AXLE LOAD` and `WHEEL DIMENSION FRONT` values were consistently associated with higher trigger frequency.

5 Discussion

The results show that ILKATBO trigger behavior is not random, but follows clear patterns related to both operating environment and vehicle setup. What remains to be established is which of those patterns carry genuine interpretive weight, and which are artefacts of the data structure, the labelling strategy, or regional fleet clustering. The discussion proceeds in three parts, model validation, operating environment addressing Research question 1 (Q1), and vehicle configuration addressing research question 2 (Q2).

5.1 Key findings

5.1.1 Model Validation

All three models perform well. The full GTA model reaches 77.4% accuracy and a ROC-AUC of 85.7% while the ROUGH and SMOOTH models reach 81.9% and 86.3% with ROC-AUC values of 89.7% and 93.1% respectively (Table 3).

The use of stratified 5-fold cross-validation reduces the risk of optimistic estimates from a lucky train-test split. Every vehicle contributes exactly once to the evaluation, and the splits preserve class balance in each fold. Performance is stable across folds, with low standard deviations (Table 3), rather than driven by one particularly favorable split. This suggests the models have identified real patterns rather than noise.

The high accuracy may partly reflect the labelling strategy. Since only the outer quartiles of the trigger rate distribution are used, the two classes are easier to separate than they would be if the full distribution were included. This likely inflates the reported accuracy compared to if borderline vehicles would have been included. That said, the goal of the study is to understand what distinguishes vehicles that trigger very frequently from those that almost never do, and for that purpose the labelling strategy is appropriate.

5.1.2 Operating Environment

The results of the GTA model show a clear effect of road condition. Road condition stands out clearly from the other features, where trigger frequency differs substantially across driving conditions, and rougher environments produce more triggers. This is consistent with ILKATBO's intended behavior. Rougher roads should produce more of the disturbances the detection logic is designed to respond to.

More noteworthy is the role of altitude. Low altitude values cluster around zero or slightly negative SHAP values in the beeswarm plot (Figure 10), while high altitude values produce strongly positive contributions, some reaching above +6. The top-values plot (Figure 4) reinforces this, the four feature values with the highest positive mean SHAP contributions are all high-altitude values. Topography and curve density, while lower in overall SHAP importance, also both show clear directional patterns in the beeswarm plot. Higher values push toward positive SHAP, which is consistent with high-altitude roads tending to be more hilly and winding. Although one might therefore expect altitude to matter only indirectly through these correlated road characteristics, the fact that it remains the second most important feature in the model suggests that it captures something that road roughness, topography, and curve density alone do not.

Temperature features also show clear patterns. Extreme temperatures in both upper and lower limits produce more positive SHAP contributions and lower values push towards low. Very high upper temperatures appear among the strongest positive contributors (Figure 4), while lower limits rank as the strongest negative contributors (Figure 7). The lower-bound equivalent shows that lower extreme temperatures push towards a higher trigger label (Figure 10), with especially lower limits appearing as the top positive contributors (Figure 4). This is broadly expected, as regions with extreme temperatures often coincide with generally harsher operating conditions through poorer road infrastructure, more demanding terrain, or less predictable driving environments. As with altitude, the fact that temperature features retain high SHAP importance along with road condition again suggests that it captures aspects of the operating environment that road condition alone does not.

The total weight of the combination, represented by the logged average and the 95th percentile GCW, is highly ranked in mean absolute SHAP importance, but does not show a clean directional pattern in the beeswarm plot (Figure 10). Both high and low weight values appear on both sides of zero with no monotonic structure.

5.1.3 Vehicle Configurations

The variant models reveal that a handful of vehicle configuration features consistently distinguish high- from low-triggering vehicles. The exploratory analysis supports the credibility of these signals. When GTA parameters are added to the models, the top-ranked configuration features largely retain their positions, reducing the risk that they are merely byproducts of regional operating patterns.

To address RQ2, the discussion examines how vehicle configuration is associated with trigger frequency within the ROUGH and SMOOTH subgroups. The discussion begins with configuration signals shared across both conditions. The next step highlights condition-specific patterns, with emphasis on features whose importance or direction differs between the ROUGH and SMOOTH models. These observations are then interpreted using the exploratory analysis with GTA variables included, in order to assess whether the signals appear robust or instead reflect operating-environment or regional clustering. The section concludes by summarising the main implications for the interpretation of vehicle configuration effects.

Shared Signals Across ROUGH and SMOOTH Conditions
Front axle load provides the most robust signal. Lower ratings (7.5 and 7.1 tons) rank among the top positive contributors in both the ROUGH and SMOOTH models, and the directional pattern is consistent across both beeswarm plots. The relationship is not strictly monotonic, however, as higher front axle load ratings do not consistently push toward lower SHAP values. It is also worth noting that front axle load is not a directly controllable parameter but rather a consequence of other configuration choices, and that it only specifies the maximum allowed load, not the actual load during driving.

Since front axle load partly depends on wheel dimensions and front suspension system, these hardware-related configurations are also expected to be reflected in the results. Smaller wheel dimensions show a similar pattern (Figure 11) for both ROUGH and SMOOTH models. For the front suspension system, the SHAP differences between air and leaf suspension are less clear, although leaf suspension shows a positive SHAP contribution for both, indicating that the model assigns some importance to this configuration as well.

Wheel and tire dimensions show consistent directional patterns across both models. In both beeswarm plots, narrower wheels push toward positive SHAP while larger values push negative, and tire dimension front follows the opposite pattern. However, the strength and stability of these signals differ between conditions, which is discussed further under condition-specific patterns. One possible interpretation is that a higher tire profile, resulting from a smaller wheel combined with a larger tire, produces a different dynamic response that the detection logic is more sensitive to.

Vehicle size shows directional tendency across both models. Vehicle height of 3.8 meters is the second-strongest negative contributor in the ROUGH and third in SMOOTH (Figure 8 and Figure 9), while heights of 4.1 and above push toward positive SHAP. Longer vehicle frames appear among the top positive contributors in the SMOOTH model, with the four longest vehicle length intervals all present in the top 20 feature values (Figure 6). Hence, larger vehicles tend toward higher trigger rates across conditions.

Main assembly date of 0 is the strongest negative contributor in both models. One possible explanation is that certain configuration options associated with higher trigger rates are no longer manufactured, leading the model to associate recent assembly dates with lower trigger probability.

Body connection rank highly in both models, though individual categories vary in robustness. B-TIMBR climbs to first place in the extended SMOOTH model and remains stable in the ROUGH model, indicating that its association with higher trigger rates persists after accounting for operating conditions, making it the most credible body connection candidate for a genuine configuration-level effect. On the negative side, B-VANB and B-VANBC both push toward lower trigger predictions in the ROUGH model. These are standard body specifications typically found on vehicles designed for conventional transport, which may inherently produce fewer of the disturbances that the detection logic responds to under rough conditions. The behavior of B-BULK and B-HH is condition-dependent and is examined in the section below.

Condition Specific Patterns

Front tire rolling resistance, non specified dominates the ROUGH model but ranks only 15th under SMOOTH conditions. It survives the addition of GTA parameters, yet the inconsistency across road conditions and the difficulty of interpreting the feature physically make it hard to determine whether this reflects a genuine configuration effect or a fleet-specific deployment pattern.

In the SMOOTH model, lower front axle installation height at full load pushes predictions toward a higher trigger rate. This signal is retained in the extended EDA model, suggesting it is not driven by regional or environmental confounding. Front axle installation height at full load is a calculated parameter that depends on front unladen height and front suspension system. When examining the specific values, both 110mm and 130mm correspond to air-suspension vehicles with front axle load ratings of 7, 7.5, or 8 tonnes. The model therefore associates a specific combination of light front-end specification and air suspension with elevated trigger frequency under smooth conditions.

Although showing the same directional pattern, the wheel and tire dimension signals illustrate how the same feature can behave differently depending on road condition. Under smooth conditions, wheel dimension front ranks second in overall feature importance, and the directional pattern in the beeswarm plot is sharp. Under rough conditions, the same directional trend is present but weaker, and wheel dimension front = 900 drops eight positions in the extended EDA model, suggesting that part of its rough-road signal is absorbed by operating environment variables. This contrast indicates that the wheel dimension effect is more defensible as a genuine configuration signal under smooth conditions, where the operating environment introduces less confounding variation. A similar pattern holds for tire dimension front, which ranks higher and shows more distinct structure in the SMOOTH model.

Body connection categories B-BULK and B-HH both drop substantially when GTA parameters are added to the extended ROUGH model. B-BULK falls 13 ranks and B-HH drops from 4th to 9th (Table 4). This indicates their original importance reflected the operating environments these body types are typically deployed in rather than the body connection itself. Both should therefore be treated with caution as a credible configuration-level signal.

5.2 Limitations

An important limitation is the quality and construction of the truck data. The target variable is based on logged counters, where trigger frequency is calculated from the change in cumulative trigger counts divided by the change in cumulative driven distance. This makes the analysis practical at fleet level, but the results depend on the assumption that trucks log data in a consistent and reliable way over time. In practice, this cannot be guaranteed. During preprocessing, vehicles with clearly suspicious logging behaviour, such as negative counter changes, were removed. Even after filtering, some trucks showed unusually high trigger counters while others appeared never to trigger at all. If either the trigger counter or the distance counter is logged incorrectly, the resulting trigger rate may be misleading, and the high outliers in the data should be interpreted with caution.

There are also limitations in the explanatory variables. Some features contain broad categories such as `Other`, which may group together several technically different vehicle configurations. These categories can still be useful for prediction but are difficult to interpret physically.

The GTA variables describe usage and operating conditions over aggregated time windows rather than at the exact moment of individual triggers. Since several GTA parameters summarise the most recent year of operation, they capture long-term exposure rather than event-level conditions. They are useful for identifying broad patterns, but do not to explain individual trigger events in detail.

The feature engineering process also introduces some uncertainty. The final feature set was selected through missingness filtering, manual screening, and consultation with domain experts. This is reasonable in an applied engineering setting, but it introduces subjectivity.

A further limitation is the ecological nature of the analysis. Trigger rate is computed as a vehicle-level aggregate, triggers per kilometre driven over the vehicle's lifetime, and GTA parameters similarly describe operating conditions at aggregated annual or lifetime levels. Individual trigger events, however, are driven by instantaneous conditions such as the specific road surface at the moment of the blowout, the vehicle's speed and load at that instant. This study therefore identifies which types of vehicles tend to trigger more often in aggregate, but cannot be used to predict the conditions that produce a specific trigger event. This distinction limits the direct applicability of findings to event-level calibration decisions, for which individual-event data would be required.

Another limitation comes from how the model predictions are set. Since the trigger rate was computed and then categorized into high and low labeled vehicles based on quantiles, the model does not distinguish between a high and an extreme trigger rate. The model is not trying to predict and learn the exact trigger rates but instead learning what configurations are more apparent at high respectively low triggering vehicles.

Finally, this is an observational study based on historical fleet data. The XGBoost models and SHAP analysis identify systematic associations between features and trigger frequency, but they do not establish causality. A feature associated with elevated trigger frequency does not necessarily cause more triggers and may instead be correlated with other factors not captured in the dataset. The regional clustering discussed above is the most concrete example.

5.3 Final conclusion

Road condition is the dominant predictor of trigger frequency. Rougher roads produce more triggers, consistent with the system's design intent. This finding is expected and confirms that the detection logic behaves as intended under varying surface conditions.

Altitude is the second strongest predictor, and it retains that importance even when road condition is already included in the model. The SHAP analysis and the EDA together establish that altitude is not simply acting as a proxy for rougher terrain. The effect holds across different driving condition distributions. Curve density and topography show the same directional pattern. What the model cannot tell us is *why* altitude matters. Whether temperature, air density, braking dynamics, or driver behaviour are the operative mechanisms remains unknown. The model identifies altitude as important but cannot explain the underlying process.

At the configuration level, Front Axle Load, Wheel Dimension Front, and Tire Dimension Front show consistent directional patterns across both the ROUGH and SMOOTH models. Lower axle load ratings, smaller wheel dimensions, and larger tire dimensions are all associated with higher trigger frequency. The consistency across model splits makes these credible hardware-level effects rather than artefacts of regional clustering.

5.4 Further research

- **Snapshot-level data.** The current analysis relies on aggregated annual parameters, which limits it to describing which vehicles trigger more often rather than why individual triggers occur. Two additions would help. First, introducing a feature recording time since the previous trigger event would make it possible to distinguish isolated triggers from clusters of consecutive events, which may have very different underlying causes and carry different diagnostic value. Second, capturing snapshots in the seconds leading up to the trigger would allow direct analysis of the dynamic conditions that precede an event, rather than relying on fleet-wide averages.
- **Controlled vehicle testing.** A targeted test could be set up by selecting three vehicles with different front axle loads, for example 7.1, 8.5, and 10.0 tonnes, and running each both unloaded and at full load in a controlled environment. Comparing trigger behaviour across these six combinations would deepen the understanding of the association between front axle load and trigger frequency, and reveal whether

the detection logic should account for the ratio between actual load and specified maximum.

- **Regional validation.** Identifying vehicles with identical configurations operating in different geographical regions would help separate genuine hardware effects from regional clustering.

6 References

References

[1] R. Rutström, personal communication, 2025.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Lindroth, L. (2025). *LDA SQL and Hadoop platform*. Accessed: 13 May 2026. https://confluence.srv.volvo.com/spaces/DVG193/pages/69178854/LDA+SQL+and+Hadoop+platform

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Volvo Trucks. (2025). *GTA tables in LAT_AutoETL*. Accessed: 12 May 2026. https://confluence.srv.volvo.com/spaces/DVG193/pages/320459954/GTA+Tables?preview=/320459954/645844622/GTA%20tables%20in%20LAT_AutoETL_1.8.pdf

Volvo Trucks. (n.d.-a). *The AutoETL Info Page*. Accessed: 13 May 2026. https://volvogroup.sharepoint.com/sites/coll-support-lda/SitePages/AutoETL.aspx#overview

Volvo Trucks. (n.d.-b). *Volvo Dynamic Steering*. Accessed: 28 April 2025. https://www.volvotrucks.se/sv-se/trucks/features/volvo-dynamic-steering.html

Yang, C., Chen, M., & Yuan, Q. (2021). The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis. *Accident Analysis & Prevention*, *158*, 106153. https://doi.org/10.1016/j.aap.2021.106153

A Hyperparameter tuning

This section summarizes the hyperparameter grid used for all models and the parameter combination selected from the grid search. The selected values correspond to the configuration with the highest mean cross-validated test score.

Table 6: Grid search was performed over all combinations of the candidate values listed. All three models converged on the same configuration, suggesting this combination generalizes across the three data subsets.

Hyperparameter	Candidates tested	Selected
`learning_rate`	`0.01, 0.05, 0.1`	`0.1`
`max_depth`	`2, 3, 4`	`4`
`n_estimators`	`100, 200, 300`	`300`