Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 11, No. 04, April, 2025

Optimized Feature Selection for Predicting the Number of Casualties in Traffic Crashes

Muamer Abuzwidah [1*], Ahmed Elawady [1], Jaeyoung Jay Lee [2, 3, 4], Ghazi G. Al-Khateeb [1], Salah Haridy [5, 6], Waleed Zeiada [1, 7]

[1] *Department of Civil and Environmental Engineering, College of Engineering, University of Sharjah, Sharjah, United Arab Emirates.*

[2] *School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, China.*

[3] *Department of Civil, Environmental & Construction Engineering, University of Central Florida, United States.*

[4] *Queensland University of Technology, School of Civil and Environmental Engineering, Australia.*

[5] *Department of Industrial Engineering and Engineering Management, University of Sharjah, Sharjah, United Arab Emirates.*

[6] *Benha Faculty of Engineering, Benha University, Benha, Egypt.*

[7] *Department of Public Works Engineering, Mansoura University, Mansoura 35516, Egypt.*

Received 28 January 2025; Revised 01 March 2025; Accepted 13 March 2025; Published 01 April 2025

Abstract

Traffic crash prediction remains a critical challenge in transportation safety management, with increasing emphasis on leveraging machine learning techniques for accurate casualty prediction. This study aims to develop an optimized feature selection framework for traffic crash casualty prediction by comparing six selection techniques: Design of Experiments (DOE), Forward and Backward Sequential Feature Selection, Information Gain, Lasso Regularization, and Random Forest (RF) Feature Importance, with subsequent integration using the Borda count method. By analyzing 517,000 UK traffic crash records (2019-2023), 25 machine learning models (linear models, decision trees, ensemble methods, and neural networks) were evaluated across 12 critical attributes. Results demonstrate eXtreme Gradient Boosting (XGBoost)'s superior performance with a Root Mean Square Error (RMSE) of 0.671 and Mean Absolute Error (MAE) of 0.372 using the proposed Borda count integration method while maintaining efficient computation time (11.3 minutes compared to the baseline's 17 minutes). Five factors consistently emerged as the most influential predictors across all selection methods: number of vehicles involved, speed limit, police officer attendance, day of the week, and urban/rural classification, while environmental factors showed lower importance than traditionally assumed. The novel integration of multiple feature selection techniques through Borda count provides a more robust feature subset than any individual method, offering an optimal balance between computational efficiency and prediction accuracy. The framework enables transportation safety authorities to implement more efficient crash prediction systems while providing actionable insights about key risk factors for targeted interventions, especially to support the Highway Safety Manual development.

Keywords: Traffic Crash Analysis; Feature Selection; Machine Learning; Traffic Safety; Predictive Analytics.

1. Introduction

Traffic crashes continue to pose a severe global health issue, resulting in significant human suffering and economic impacts. The World Health Organization reports that road traffic crashes claim about 1.19 million lives each year, with countless others incurring injuries or long-term disabilities [1]. Economically, the global cost of road traffic crashes amounts to about 3% of national GDPs, and this figure rises to 5% in low- and middle-income nations [2, 3]. For

* Corresponding author: mabuzwidah@sharjah.ac.ae

http://dx.doi.org/10.28991/CEJ-2025-011-04-01

example, in the United States, the National Highway Traffic Safety Administration (NHTSA) noted that motor vehicle crashes had an economic impact of $417 billion in 2024 alone [4]. The economic impact of road crashes encompasses not only immediate healthcare and property repair costs but also broader implications such as reduced productivity, psychological distress, and significant rehabilitation requirements. Additional indirect costs include legal fees, emergency services, insurance processing, traffic congestion, and loss of workplace productivity [5-7].

This highlights the urgent need for effective crash prevention strategies, where predictive modeling plays a crucial role in identifying high-risk zones and implementing successful safety measures. Particularly vulnerable areas within transportation networks, such as high-speed freeways and intersections, are noted for their elevated risk levels [8], with intersections alone being the site of around half of all traffic incidents [9]. Such locations frequently see serious injuries and broadly impact road users.

There has been significant evolution in crash prediction techniques, moving from traditional statistical approaches to sophisticated machine learning models. This evolution reflects the growing complexity of predictive challenges and the increased availability of diverse datasets. The initial focus on basic machine learning algorithms with elementary feature selection has evolved, establishing essential benchmarks for both performance and methodological approaches. One notable study underscored the effectiveness of ensemble methods, where XGBoost achieved a 93% accuracy rate, surpassing older methods like Decision Trees (88%) and Logistic Regression (63%) [10]. Another research initiative developed a system to evaluate driving behavior and risk prediction using feature extraction and selection with XGBoost, which successfully identified crucial risk indicators, achieving 89% accuracy using NGSIM data and demonstrating the model's precision [11]. Moreover, a study highlights the critical role of advanced feature selection in improving the outcomes of models handling extensive crash data. It introduced the Weighted Fusion-Based Feature Selection (WFFS) method, which significantly enhanced the accuracy of UK traffic crash predictions to 97.28% using a random tree-based bagging technique, vital for strengthening road safety initiatives [12]. Additionally, another research utilized XGBoost and SHAP in analyzing high-risk lane changes, markedly enhancing the capabilities of Advanced Driving Assistance Systems (ADAS) through better predictive accuracy [13]. An examination of more than 91,000 crash records indicated that neural networks outperformed conventional models like Support Vector Machines in predicting casualties [14]. This continued with studies into the importance of features within ensemble methods, where Random Forest achieved an 81.45% accuracy in predicting crash severity [15]. The research also applied GIS-based spatial clustering techniques together with machine learning models to accurately identify crash hotspots and predict incidents, underscoring the importance of data-driven approaches in traffic safety management [16]. Concurrently, the study demonstrated Gaussian Process Regression's effectiveness in prediction tasks, achieving notable accuracy with an R^2 value of 0.63, highlighting the significance of choosing the right model for specific scenarios [17].

Further investigations analyzed 117,000 crash records, revealing that nonlinear SVM particularly excelled, attaining 78.32% accuracy in severity prediction and underscoring the trade-offs between model accuracy and computational efficiency [18]. More recent work improved on these results using Random Forest to achieve a 91.72% F1 score for three-category severity classification, showing the benefits of refined feature selection strategies [19]. This shift towards more sophisticated prediction methods is characterized by a growing focus on optimizing feature selection. Recent investigations have explored diverse feature selection techniques, each bringing unique benefits. For instance, a study implemented a deep learning framework integrated with CatBoost feature selection, achieving a weighted average precision of 0.80 and an F-measure of 0.82 [20]. This effort proved crucial for selecting spatially correlated features to improve prediction accuracy. The field progressed with a multi-task learning strategy that enhanced feature sharing across related tasks, achieving a 13.93% average improvement in accuracy compared to traditional methods [21]. Another study reaffirmed the superiority of advanced machine learning techniques over traditional models, achieving an R^2 of 0.80 through meticulous feature selection and model optimization, marking significant advancements in predictive methodologies [22]. Additionally, parallel research indicated that technological advances could reduce weather-related crash risks in traffic safety modeling, thereby enhancing prediction accuracy [23].

These developments underscore the essential role of feature selection in contemporary crash prediction systems, though challenges persist in fine-tuning these methods for optimal performance. Analysis via machine learning has shown that in developing regions, infrastructure elements have a more significant impact on traffic safety than human factors, diverging from conventional findings. Applying Multiple Linear Regression (MLR) has proven effective for crash analysis in urban environments [24]. Utilizing target speed modeling, behavioral analysis, and the APW technique, findings indicate human-related causes contribute to 66.8% of highway collisions, underscoring the urgency for predictive strategies and location-specific traffic safety improvements [25, 26]. Additionally, employing machine learning to pinpoint critical safety factors substantially improves traffic management efficacy [27]. Another study explores advanced feature selection for accident classification using deep learning, significantly enhancing road safety and aligning with prior research by boosting predictive accuracy and strategic traffic management interventions [28]. Further investigations have adopted a hybrid feature selection-based machine learning classification to assess injury severity in road incidents, using Random Forests and the Boruta Algorithm to identify key attributes evaluated by various classifiers, with XGBoost showing high accuracy in both single- and multi-vehicle accidents [29]. Additionally, a study has introduced a univariate feature selection method to address liability in vehicle-cyclist collisions in China, employing SVM, LDA, and ANN to determine cyclists' pre-collision activities, with SVM achieving 81.84% accuracy [30]. Another study found that behavioral compliance and machine learning jointly enhance traffic safety through predictive insights and targeted interventions [31]. Research builds on previous findings by simplifying the analysis of injury

severity in Taiwanese traffic incidents, categorizing twenty-eight factors based on their importance [32]. Utilizing a decade of UK traffic data, another study applied models like Random Forest and Logistic Regression, achieving an 87% accuracy rate in predicting traffic accident severity [33]. Additional analysis using Poisson and Negative Binomial models revealed critical roadway factors, highlighting speed's negative correlation and urgent safety enhancements [34].

Overall, the high dimensionality of crash data, encompassing numerous environmental, temporal, and infrastructural factors, necessitates efficient feature selection strategies. Current approaches often rely on single feature selection techniques, potentially missing important feature interactions or overlooking crucial predictive factors. The comparative effectiveness of different feature selection methods in the context of crash prediction remains understudied, particularly regarding their impact on model performance, training efficiency, and generalization capabilities. Furthermore, while individual feature selection methods have shown promise, the potential benefits of combining multiple techniques have not been thoroughly explored.

This study addresses these gaps by conducting a comprehensive comparison of six feature selection techniques: Design of Experiments (DOE), Forward Sequential Feature Selection (FSFS), Backward Sequential Feature Selection (BSFS), Information Gain (IG), Lasso Regularization, and Random Forest Feature Importance. Each method brings unique strengths to feature selection: DOE provides systematic factor screening, FSFS and BSFS offer iterative feature optimization, IG captures information-theoretic importance, Lasso enables sparse feature selection through regularization, and Random Forest importance leverages ensemble learning for feature ranking. By comparing and merging these approaches, we aim to develop a more robust framework for feature selection in crash prediction. Given these research opportunities, this study's objectives are to:

- Analyze the impact of different feature selection approaches on machine learning models performance.
- Develop a framework for integrating multiple feature selection methods to identify the most robust and predictive feature subset.
- Compare the computational efficiency and performance across different scenarios.

2. Data Pre-Processing

This section presents the methodology employed for data preprocessing and the subsequent exploratory analysis of the crash dataset. The preprocessing phase encompassed data cleaning, feature selection, and preliminary statistical analysis, while the exploratory analysis focused on understanding the distributions and relationships among the selected variables. The initial dataset comprised 520,000 traffic crash records from the United Kingdom, spanning from 2019 to 2023, with 35 distinct attributes. Through systematic analysis and domain knowledge, 12 critical features were identified as potential predictors for crash severity and casualty numbers. Table 1 presents these selected features, their respective data types, and descriptions. The features encompass both numeric attributes (number of vehicles, number of casualties, speed limit, and time of day) and nominal attributes (day of the week, road type, junction detail, first road class, light conditions, weather conditions, road surface conditions, urban or rural area, and police officer attendance). The dataset underwent a cleaning process to handle missing values, resulting in a final dataset of 517,000 records (99.4% retention rate). This high retention rate suggests robust initial data collection procedures and minimal impact of the cleaning process on the dataset's representativeness. Additionally, the crash severity variable, originally categorized into three classes (slight, serious, and fatal), was transformed into a binary classification problem by merging the 'serious' and 'fatal' categories into a single 'severe' class while maintaining 'slight' as 'not severe'. This transformation was implemented to enhance model robustness and address class imbalance issues, as fatal crashes represented a relatively small proportion of the dataset.

Table 1. Description of Selected Features for the Study

Attribute Name	Data Type	Description
Number of Vehicles	Numeric	Count of vehicles involved in the crash
Number of Casualties	Numeric	Count of people injured or killed in the crash
Speed Limit	Numeric	Posted speed limit at the crash location (mph)
Time of the Day	Numeric	Time when the crash occurred (0-23 hours)
Crash Severity	Nominal	Severity level of the crash
Day of Week	Nominal	Day of the week when the crash occurred
Road Type	Nominal	Categorization of road: Roundabout, One-way Street, Dual carriageway, Single carriageway, Slip Road
Junction Detail	Nominal	Type of junction where the crash occurred
First Road Class	Nominal	Classification of the first road (A, B, C, Motorway, or Unclassified)
Light Conditions	Nominal	Lighting conditions of the road at the time of the crash
Weather Conditions	Nominal	Weather conditions at the time of crash
Road Surface Conditions	Nominal	Whether the surface of the road was dry, wet, snowy, or oily
Urban or Rural Area	Nominal	The type of area where the crash occurred
Did Police Officer Attend	Nominal	Whether a police officer attended the scene of the crash, or it was self-reported

Descriptive statistics for the numerical variables are presented in Table 2. The analysis reveals that crashes typically involved a mean of 1.833 vehicles (σ = 0.69 vehicles, range: 1-17 vehicles) and resulted in 1.28 casualties on average (σ = 0.71 casualties, range: 1-70 casualties). Speed limits at crash locations exhibited considerable variation (μ = 36 mph, σ = 14.14 mph, range: 20-70 mph).

Table 2. Descriptive statistics of numerical variables in the dataset

Factor	Mean	σ	Min	Max
Number of Vehicles	1.833	0.69	1	17
Number of Casualties	1.28	0.71	1	70
Speed Limit	36	14.14	20	70
Time of crash	13.7	5.15	0	23

The temporal distribution of crashes showed a mean occurrence time of 13.7 hours (σ = 5.15 hours). The distribution analysis of the categorical variables is shown in Figure 1.

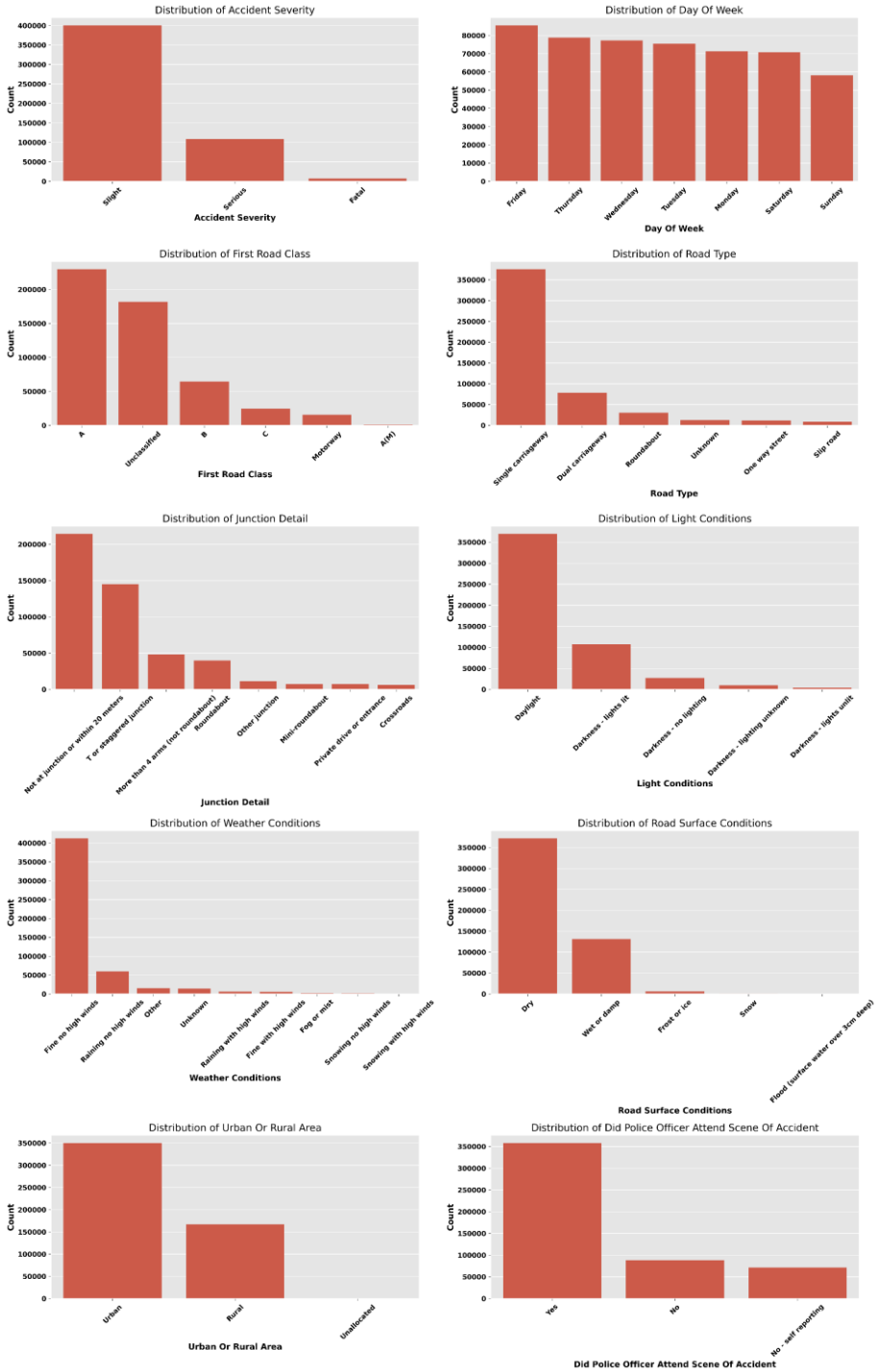


Figure 1. Distribution plots of categorical variables

Figure 2 presents the temporal distribution of crashes from 2019 to 2023, highlighting a significant reduction in crash numbers during the COVID-19 period in 2020, followed by a gradual return to pre-pandemic levels. Weekly patterns, depicted in Figure 3, demonstrate systematic variation in crash frequencies, with peak occurrences on Fridays (n ≈ 85,000) and minimum occurrences on Sundays (n ≈ 58,000). This analysis is extended by disaggregating the data by crash severity, revealing relatively consistent severity proportions across weekdays despite varying total frequencies. Daily patterns, presented in Figure 4, exhibit a characteristic bimodal distribution with prominent peaks during morning (08:00-09:00) and afternoon (16:00-17:00) rush hours, and a minimum between 03:00 and 04:00. The severity distribution across hours indicates an elevated proportion of severe crashes during night-time hours, despite lower overall frequencies.

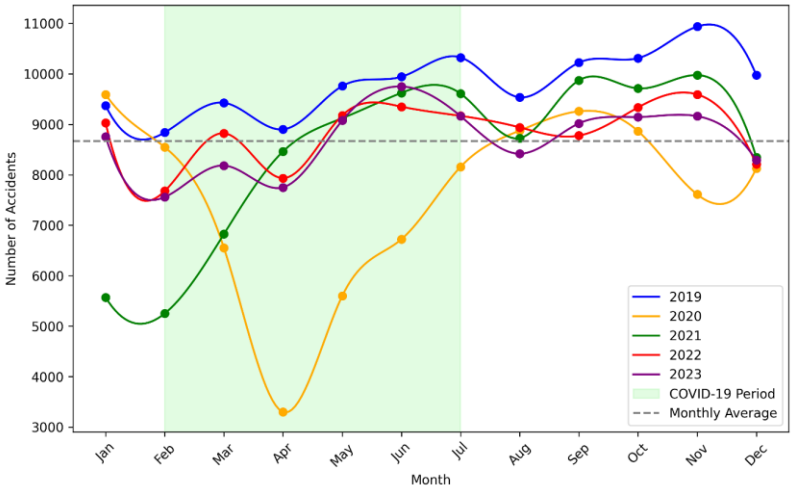


Figure 2. Monthly distribution of road crashes in the UK from 2019 to 2023

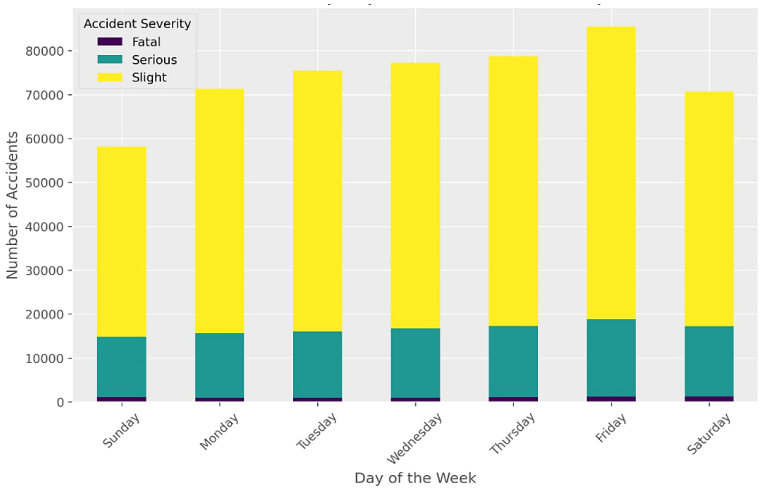


Figure 3. Stacked bar chart showing the distribution of crash severity across different days of the week

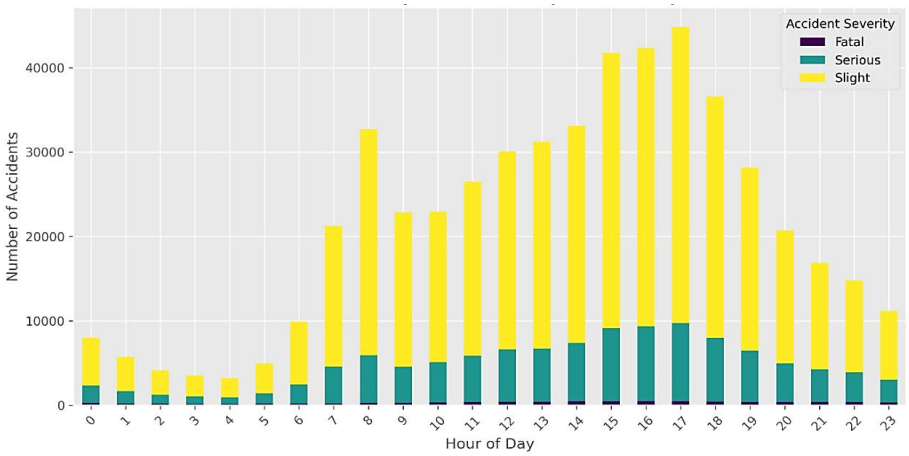


Figure 4. Stacked bar chart showing the distribution of crash severity across different hours of the day

Figure 5 presents the correlation matrix between all variables, with particular focus on their relationships with the model's target variables (crash severity and number of casualties). The analysis reveals weak correlations with crash severity, including number of vehicles (-0.069), number of casualties (-0.086), and speed limit (-0.11). For the number of casualties, weak positive correlations are observed with the number of vehicles (0.21) and urban/rural classification (0.14). The light conditions show a weak correlation with crash severity (-0.049), while weather conditions demonstrate a slight negative correlation (-0.037). Interestingly, the presence of police officers at the crash scene shows a modest positive correlation with severity (0.18), potentially indicating their more frequent attendance at more severe crashes. These correlations, while relatively modest in magnitude, provide initial insights into potential relationships between predictor variables and target outcomes that warrant further investigation through more advanced modeling techniques. It's important to note that weak linear correlations do not necessarily exclude significant non-linear relationships that may be captured by machine learning models.

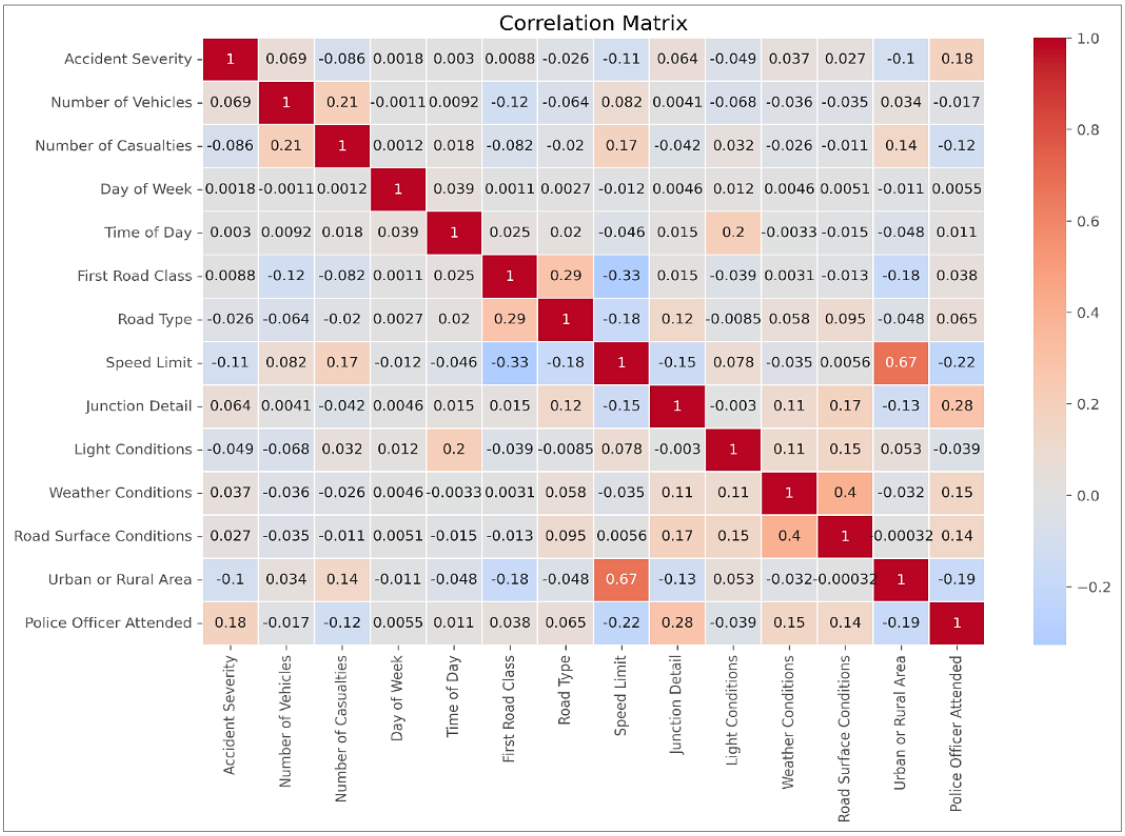


Figure 5. Correlation matrix heatmap showing the relationships between all variables

3. Research Methodology

This study presents a comprehensive framework as shown in Figure 6 for predicting crash casualties through systematic comparison of feature selection techniques and machine learning models. The framework consists of three main components: (1) feature selection through multiple techniques, (2) machine learning model implementation, and (3) performance evaluation and comparison. The methodology enables both individual assessment of each feature selection technique and evaluation of their combined effectiveness.

3.1. Feature Selection Techniques

This study implements five distinct feature selection approaches to identify the most influential factors affecting crash casualties. The number of selected features (K) for all methods was determined by the number of significant factors identified through DOE analysis at $\alpha = 0.05$, ensuring consistent comparison across different techniques. The selection of these six feature selection techniques was deliberate, aiming to provide comprehensive coverage across different methodological paradigms: experimental design (DOE), wrapper methods (FSFS and BSFS), information theory (IG), regularization techniques (Lasso), and ensemble learning (Random Forest). This diversity enables robust comparison while leveraging the complementary strengths of each approach.

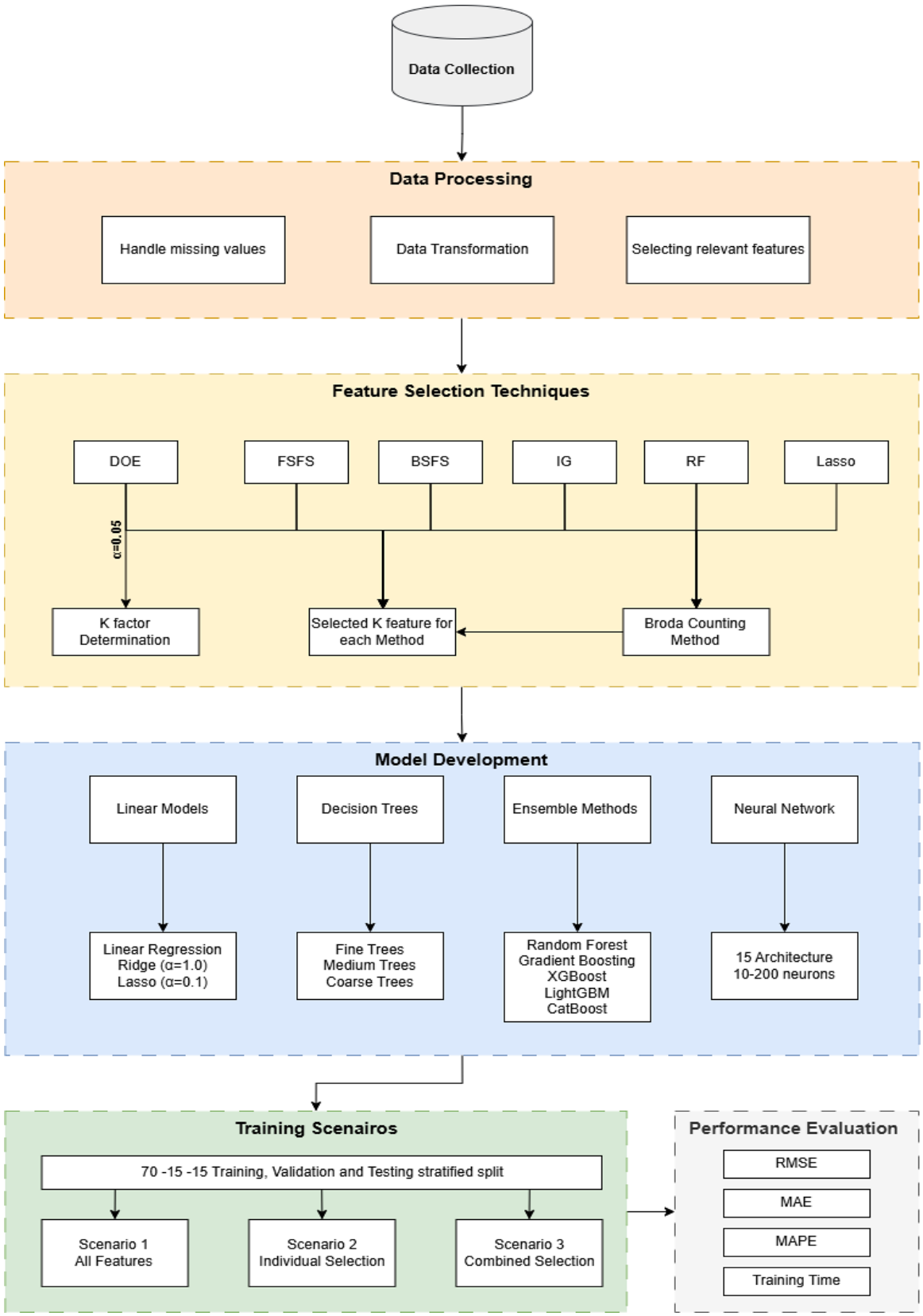


Figure 6. Methodology framework

3.1.1. Design of Experiments

The DOE approach implemented a systematic full factorial design to evaluate the impact of both categorical and numerical factors on crash casualties. This methodology enables comprehensive analysis of main effects and interaction effects between factors while maintaining statistical rigor. The experimental design incorporated twelve factors, each mapped to two levels to facilitate factorial analysis, resulting in 2^12 = 4,096 possible factor combinations. The experimental design matrix D was constructed by Equation 1:

$$D = \begin{bmatrix} +1 & +1 & \cdots & +1 \\ +1 & -1 & \cdots & +1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & -1 \end{bmatrix}_{2^{12} \times 12} \quad (1)$$

where each row represents an experimental run, and each column represents a factor level combination. For categorical variables, logical groupings based on factor characteristics and their relationship to crash severity were implemented. The mapping process followed a structured approach that preserved the underlying risk characteristics while reducing dimensionality. Temporal factors were mapped based on established traffic patterns: day of week distinguished between weekdays and weekends, acknowledging the distinct traffic patterns and risk profiles associated with each period. Infrastructure-related factors received similar treatment: first road class differentiated between major roads including motorways and A-roads versus minor roads, while road type separated major infrastructures such as roundabouts and dual carriageways from minor road configurations. Environmental and situational factors were mapped according to their risk implications. Light conditions were simplified to daylight versus darkness, reflecting the fundamental visibility distinction.

Weather conditions distinguished between fine and adverse conditions, while road surface conditions were categorized as either dry or slippery, capturing the primary friction characteristics affecting vehicle control. For numerical variables, threshold-based binary classifications were established based on domain knowledge and statistical analysis of crash patterns based on Equation 2:

$$x_{\text{binary}} = \begin{cases} 1 & \text{if } x \geq \text{threshold} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The effect of each factor was calculated using the standard DOE formula (Equation 3):

$$E_i = \frac{1}{n} \sum_{j=1}^{n} y_j x_{ij} \quad (3)$$

where E_i represents the effect of factor i, n is the number of experimental runs, y_j is the response (number of casualties) in run j, and x_{ij} is the coded level (-1 or +1) of factor i in run j. This formula calculates the average difference in response between the high level (+1) and low level (-1) runs for each factor. Table 3 showed the proposed mapping for the categorical and numerical factors.

Statistical significance was assessed through Analysis of Variance (ANOVA), with the test statistic in Equation 4:

$$F = \frac{MS_{\text{factor}}}{MS_{\text{error}}} = \frac{SS_{\text{factor}}/df_{\text{factor}}}{SS_{\text{error}}/df_{\text{error}}} \quad (4)$$

where MS represents mean squares, SS represents sum of squares, and df represents degrees of freedom. Factors were considered significant at $\alpha = 0.05$, and this significance level determined the number of features (K) selected for subsequent analysis methods. The analysis also considered two-way interaction effects between factors through the interaction term using Equation 5:

$$E_{ij} = \frac{1}{n} \sum_{k=1}^{n} y_k x_{ik} x_{jk} \quad (5)$$

where E_{ij} represents the interaction effect between factors i and j.

Table 3. Factors and their 2 levels proposed mapping

Factor	Categories/ Rule	Proposed Mapping
Day of week	Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday	Weekend (Saturday, Sunday): 1, Weekday (Monday–Friday): 0
First Road Class	Motorway, A(M) Road, A Road, B Road, C Road, Unclassified	Major Roads (Motorway, A(M), A Road): 1, Minor Roads: 0
Road type	Roundabout, One way street, Dual carriageway, Single carriageway, Slip road, Unknown	Major Roads (Roundabout, One way, Dual carriageway): 1, Minor Roads: 0
Junction detail	Not at junction, Roundabout, Mini-roundabout, T junction, Slip road, Crossroads, etc.	At Junction (e.g., Roundabout, T-junction, etc.): 1, Not at Junction: 0
Light conditions	Daylight, Darkness – lights lit, Darkness – lights unlit, Darkness – no lighting, etc.	Dark (with or without lighting): 1, Daylight: 0
Weather conditions	Fine without high winds, Raining, Snowing, Fine with high winds, Raining with high winds, etc.	Adverse Weather (Raining, Snowing, Fog, etc.): 1, Fine: 0
Road surface conditions	Dry, Wet or damp, Snow, Frost or ice, Flood, etc.	Slippery (Wet, Snow, Ice, Flood, etc.): 1, Dry : 0
Urban or Rural area	Urban, Rural	Urban: 1, Rural: 0
Police officer at the crash	Yes, No, crash self-reported	Yes: 1, No or Self-reported: 0
Speed limit	High speed (≥ 50 mph)	High Speed: 1, Low Speed: 0
Number of vehicles	Multiple vehicles (> 2)	Multiple Vehicles: 1, Single/Two Vehicles: 0
Time	Peak hours (7–9 AM, 4–6 PM)	Peak Hours: 1, Off-Peak Hours: 0

3.1.2. Forward Sequential Feature Selection

FSFS represents a wrapper-based feature selection method that builds a feature subset incrementally, starting from an empty set and sequentially adding the most beneficial features. The fundamental principle behind FSFS is to evaluate each candidate feature's contribution to the model's performance through an iterative process, selecting features that maximize predictive accuracy while maintaining model parsimony. This approach proves particularly valuable in datasets with high dimensionality, where identifying the most informative features can significantly improve model efficiency and interpretability. The selection process can be formalized through the following optimization problem:

For each step k, the algorithm selects feature x* that satisfies Equation 6:

$$x* = \arg\min_{x \in X \setminus S_k} \text{MSE}(f(S_k \cup \{x\}), Y) \quad (6)$$

where $X \setminus S_k$ denotes the set of features not yet selected. The Mean Squared Error (MSE) across the 5-fold cross-validation is computed according to Equation 7:

$$\text{MSE} = \frac{1}{5}\sum_{i=1}^{5} \frac{1}{n_i}\sum_{j=1}^{n_i} \left(y_{ij} - \hat{y}_{ij}\right)^2 \quad (7)$$

where n_i represents the number of samples in the i-th fold. y_{ij} represents the true value for the j-th sample in the i-th fold. $\hat{y}_{ij}$ represents the predicted value for the j-th sample in the i-th fold. The implementation of FSFS in this study employs a neural network with one hidden layer as the base model for feature evaluation. The process as shown in Figure 7(a) begins with an empty feature set and iteratively adds features that contribute most significantly to reducing the prediction error. The selection process incorporates a 5-fold cross-validation strategy, ensuring robust performance estimation and minimizing the risk of overfitting to particular data segments. To handle the mixed nature of variables in the traffic crash dataset, a comprehensive data preprocessing strategy was implemented. Categorical variables, including road type, weather conditions, and junction details, undergo one-hot encoding prior to the selection process. This transformation expands the initial feature space while preserving the categorical information structure. Numerical variables, such as vehicle count and speed limit, are standardized to zero mean and unit variance, ensuring equal scale consideration during the selection process.

3.1.3. Backward Sequential Feature Selection

Backward Sequential Feature Selection (BSFS) represents a wrapper-based feature selection approach that begins with the complete set of features and iteratively removes the least significant features until reaching a desired subset size. This method complements the forward selection approach by examining feature significance from the opposite direction, potentially capturing different feature interactions and dependencies. The elimination process can be formalized through the following optimization problem:

For each step k, the algorithm removes feature x* that satisfies Equation 8:

$$x* = \arg\min_{x \in S_k} \text{MSE}(f(S_k \setminus \{x\}), Y) \quad (8)$$

where $S_k \setminus \{x\}$ denotes the set of features remaining after removing feature x. The Mean Squared Error (MSE) across the 5-fold cross-validation is computed similar to FSFS. The implementation of BSFS employs the same neural network architecture with one hidden layer used in FSFS, maintaining consistency in the evaluation framework. The process begins with all twelve features and systematically eliminates features whose removal results in the smallest increase in prediction error. Like FSFS, the selection process incorporates a 5-fold cross-validation strategy to ensure robust performance estimation. The data preprocessing strategy for handling mixed variables remains consistent with the FSFS approach. Categorical variables undergo one-hot encoding, while numerical variables are standardized to zero mean and unit variance before the selection process begins. This preprocessing ensures fair comparison of feature importance regardless of their original scale or type. The process framework for FSFS implementation is shown in Figure 7(b).

(a)

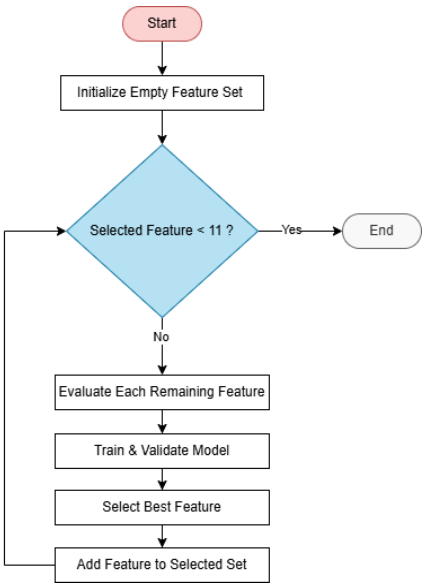


Figure 7. (a) FSFS Framework

(b)

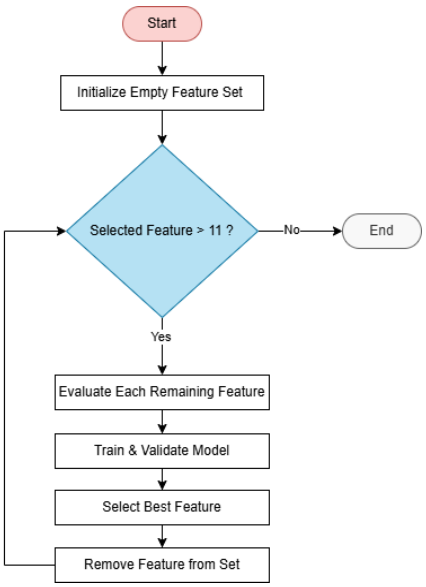


Figure 8. (b) BSFS Framework

3.1.4. Information Gain

Information Gain represents a filter-based feature selection approach adapted for regression problems, quantifying each feature's contribution to reducing uncertainty in predicting crash casualties. For a continuous target variable Y (number of casualties), the differential entropy H(Y) is calculated as shown in Equation 9:

$$H(Y) = -\int p(y)\log(p(y))dy \tag{9}$$

where p(y) is the probability density function of Y, estimated using kernel density estimation. The conditional entropy is defined by Equation 10:

$$H(Y \mid X) = -\iint p(x,y)\log(p(y \mid x))dydx \tag{10}$$

The Information Gain for each feature is subsequently calculated using Equation 11:

$$IG(Y,X) = H(Y) - H(Y \mid X) \tag{11}$$

Features are selected based on a significance threshold determined through Equation 12:

$$\tau = \mu_{IG} + a\sigma_{IG} \tag{12}$$

where μ_{IG} is the mean of all IG values, σ_{IG} is their standard deviation, and a is a scaling factor set to 1.5 to ensure selection of features with significantly higher information content. This adaptive threshold accounts for the varying scales of information gain across different datasets and feature types. To maintain consistency with the number of features (K) identified through the DOE analysis at α = 0.05, an iterative adjustment procedure was implemented for α until the number of selected features approximately matches K.

3.1.5. Lasso Regularization

Lasso (Least Absolute Shrinkage and Selection Operator) regularization represents a regression analysis method that performs both variable selection and regularization to enhance prediction accuracy and interpretability of the resulting statistical model. The Lasso estimator for the regression problem in this study is formulated as shown in Equation 13:

$$\hat{\beta}_{\text{lasso}} = \arg\min_{\beta}\left\{\frac{1}{2n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1\right\} \tag{13}$$

where y represents casualties, X is the feature matrix, β is the coefficient vector, and λ is the regularization parameter. The optimal λ is selected using 5-fold cross-validation to minimize mean squared error. Feature importance scores are calculated using Equation 14:

$$\text{Importance Score}\ (X_i) = |\hat{\beta}_i^{\text{std}}| \tag{14}$$

where $\hat{\beta}_i^{\text{std}}$ represents the standardized coefficient for feature i. The standardization ensures fair comparison across features of different scales.

3.1.6. Random Forest Feature Importance

Random Forest Feature Importance leverages the ensemble nature of random forests to provide a robust measure of feature significance in predicting crash casualties. This method combines the predictive power of multiple decision trees with built-in feature evaluation mechanisms, making it particularly suitable for this mixed-type dataset containing both categorical and numerical variables. The importance score for each feature is calculated using the Mean Decrease in Impurity (MDI), specifically the decrease in variance for this regression task as of Equation 15:

$$\text{Importance}(X_i) = \frac{1}{N_T}\sum_{T=1}^{N_T} \sum_{n \in N_T} \Delta I(n, X_i) \quad (15)$$

where N_T is the number of trees in the forest, $\Delta I(n, X_i)$ is the impurity decrease at node n where feature X_i is used for splitting calculated as of Equation 16:

$$\Delta I(n, X_i) = w_n \sigma^2(n) - w_{l(n)} \sigma^2(l(n)) - w_{r(n)} \sigma^2(r(n)) \quad (16)$$

3.1.7. Integrated Feature Selection Ranking

The Combined Feature Selection Ranking methodology integrates the results from all five feature selection techniques using the Borda Count method, a positional voting system that accounts for the complete ranking information from each selection approach. This method provides a comprehensive aggregation of feature rankings while being robust to scale differences between different selection techniques. For a dataset with n features, each selection method contributes to the final ranking by assigning points based on position. For each feature X_i, the Borda count score is calculated as of Equation 17:

$$\text{Score}_{\text{Borda}}(X_i) = \sum_{m=1}^{6} (n - r_m(X_i) + 1) \quad (17)$$

where n is the total number of features (12 in this study), $r_m(X_i)$ is the rank of feature X_i in method m, and m represents each feature selection method (DOE, FSFS, BSFS, IG, Lasso, and RF). The final selection of features (K) is based on the ordered Borda scores.

3.2. Machine Learning Models

This study implemented twenty-five machine learning models across four categories to predict crash casualties, with configurations optimized for the traffic safety domain:

3.2.1. Linear Models

Three linear regression variants were implemented: standard Linear Regression, Ridge Regression (α_{ridge} = 1.0), and Lasso Regression (α_{lasso} = 0.1). These models established baseline performance and provided interpretable coefficients for factor impact assessment.

3.2.2. Decision Tree Models

Decision trees with varying complexity levels were implemented: Fine Tree (no maximum depth, minimum split: 2 samples), Medium Tree (maximum depth: 6, minimum split: 10 samples), and Coarse Tree (maximum depth: 3, minimum split: 20 samples). These configurations balanced pattern capture capability with generalization potential.

3.2.3. Ensemble Methods

Seven ensemble methods were evaluated: Random Forest (100 trees, maximum depth: 10), Bagged Trees (10 estimators, 50% sample bootstrapping), Gradient Boosting Regression (50 estimators, learning rate: 0.1), AdaBoost Regression (50 estimators), XGBoost, Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost). Default parameters were used for the last three models except for early stopping set at 10 rounds.

3.2.4. Neural Network Models

Fifteen configurations of Multi-Layer Perceptron Regressors were tested, each with a single hidden layer but varying neuron counts (10-200 neurons in increments of 10 up to 100, then 150 and 200). All networks used ReLU activation for hidden layers, linear activation for output, and a maximum of 500 training iterations with early stopping (patience: 10). The range of neuron counts (was selected through manual experimentation to explore the relationship between model complexity and performance in crash casualty prediction. For comprehensive coverage of different architectural complexities, a systematic approach was implemented with 10-neuron increments in the lower range (10-90) to capture fine-grained performance differences in simpler architectures. After 100 neurons, we increased the increment to 50 neurons (150, 200) to efficiently explore more complex architectures while managing computational resources. This approach allowed to thoroughly investigate the performance-complexity tradeoff across the spectrum from simple to highly complex neural networks while maintaining computational feasibility.

3.3. Training Configurations

The model training process was structured into three distinct scenarios to comprehensively evaluate the effectiveness of different feature selection approaches. In Scenario 1 (baseline), all models were trained using the complete set of 12 factors. Scenario 2 involved training models with top K features identified by each feature selection technique separately, where K was determined through the Design of Experiments (DOE) analysis at $\alpha = 0.05$. This approach allowed for direct comparison of the effectiveness of different feature selection methods while maintaining consistent dimensionality across experiments. In Scenario 3, models were trained using a combined feature set determined through the Borda counting method, which integrated rankings from all six feature selection techniques to create a robust, consensus-based feature subset. Model performance was evaluated using multiple metrics to provide a comprehensive assessment: Root Mean Square Error (RMSE) for overall prediction accuracy, Mean Absolute Error (MAE) for average magnitude of errors, Mean Absolute Percentage Error (MAPE) for scale-independent error measurement, and training time to assess computational efficiency. The dataset was split into 70% training, 15% validation and 15% testing sets using Stratified Shuffle Split, maintaining the same split across all experiments to ensure fair comparison.

4. Results and Discussion

4.1. Feature Selection Results

4.1.1. Design of Experiments

The Design of Experiments (DOE) analysis, conducted at $\alpha = 0.05$, revealed nine significant factors influencing crash casualties, as shown in the Pareto chart and Normal plot of standardized effects (Figures 9 and 10).

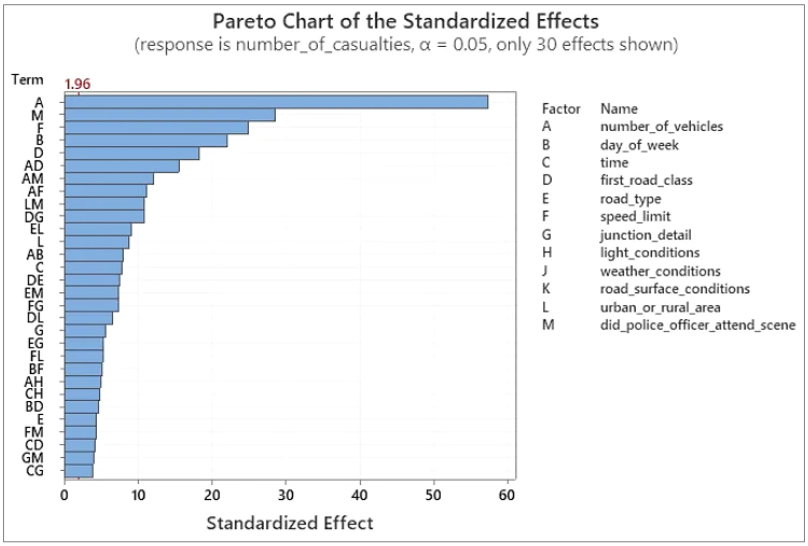
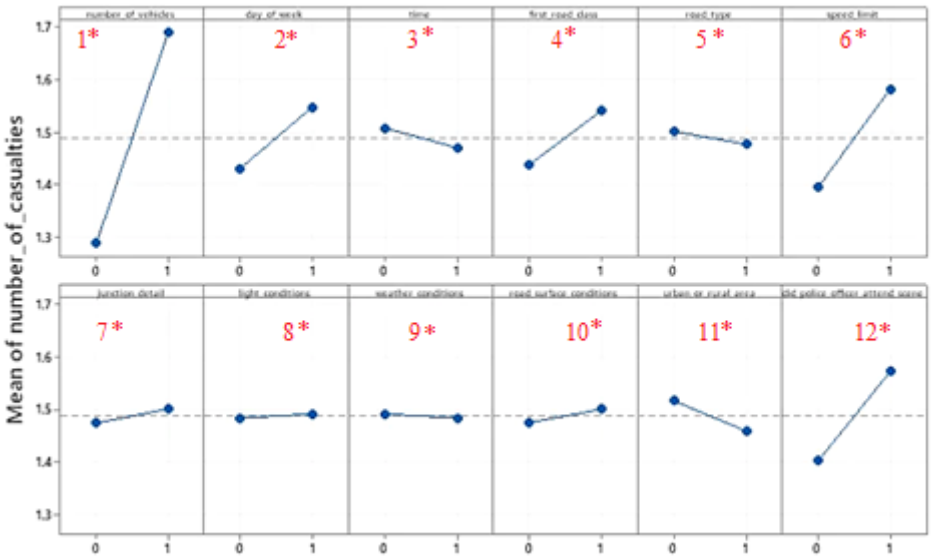


Figure 9. Pareto Chart at α = 0.05



1*= Number of Vehicles, 2*= Day of Week, 3*= time, 4*= First road Class, 5*= Road type, 6*= Speed limit, 7*= Junction Detail, 8*= light condition, 9*= Weather Condition, 10*= Road Surface Condition, 11*= Urban or rural area, 12*= Did police officer attend

Figure 10. Main effects plot for number of casualties fitted means

The number of vehicles involved emerged as the most influential factor with the highest standardized effect (approximately 60), followed by the day of the week and the time of the crash. The main effects plot demonstrates that the number of vehicles has the strongest positive correlation with casualty numbers, showing a marked increase from single- to multiple-vehicle crashes.

4.1.2. Forward Sequential Feature Selection

The FSFS analysis identified feature importance through an iterative process, as shown in Figure 11. The number of vehicles emerged as the most significant predictor with an importance score of 0.095, substantially higher than other features.

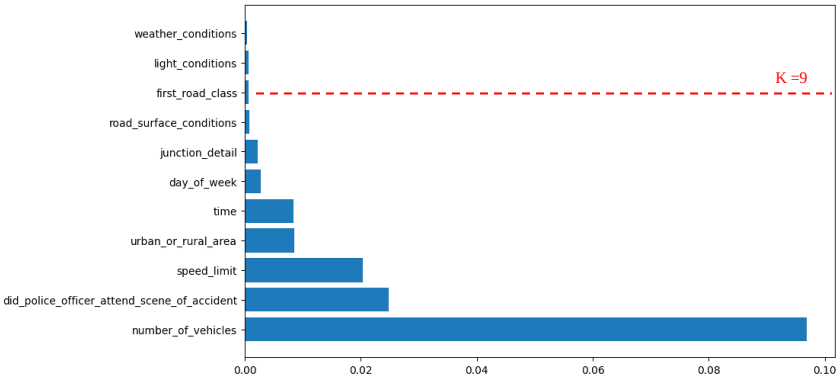


Figure 11. FSFS feature importance results

Police officer attendance and speed limit followed as the second and third most important features, with importance scores of 0.028 and 0.023 respectively. Temporal and spatial factors showed moderate importance, with time and urban/rural classification having scores around 0.01. Environmental conditions such as weather, light conditions, and road surface conditions demonstrated minimal predictive power with importance scores below 0.005, suggesting their relatively minor role in predicting crash casualties.

4.1.3. Backward Sequential Feature Selection

The BSFS, initialized with all features and retaining the top 11 variables, revealed a similar importance pattern to FSFS but with some notable differences, as illustrated in Figure 12. The number of vehicles maintained its position as the most crucial predictor with a dominant importance score of 0.085, more than twice the importance of the next feature. Police officer attendance and speed limit were again identified as significant predictors, with importance scores of 0.025 and 0.023 respectively. Time and urban/rural classification showed moderate predictive power, with importance scores around 0.01-0.015. An interesting distinction from FSFS was the inclusion of road type in the selected feature set, though with relatively lower importance (0.005). Environmental factors again demonstrated minimal impact, with weather conditions showing the lowest importance score among all features.

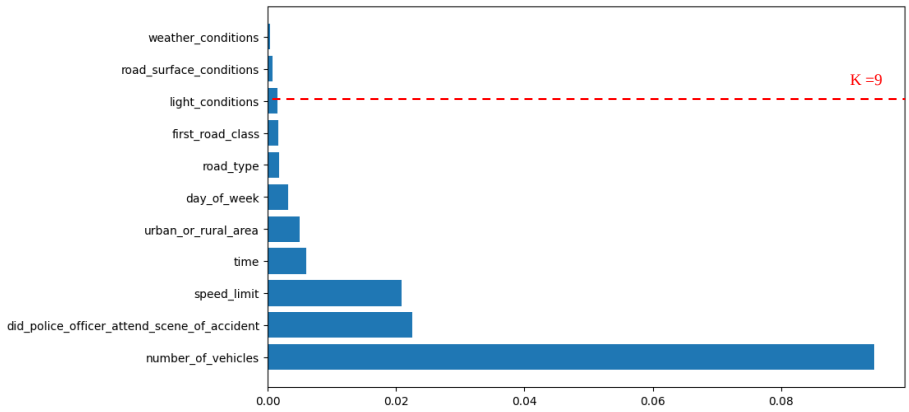


Figure 12. BSFS feature importance results

4.1.4. Information Gain (IG) Analysis

The IG analysis provided a different perspective on feature importance, quantifying each feature's contribution to reducing uncertainty in predicting crash casualties. As shown in Figure 13, the number of vehicles again emerged as the most informative feature with a standardized importance score of 0.25, followed closely by the speed limit (0.22).

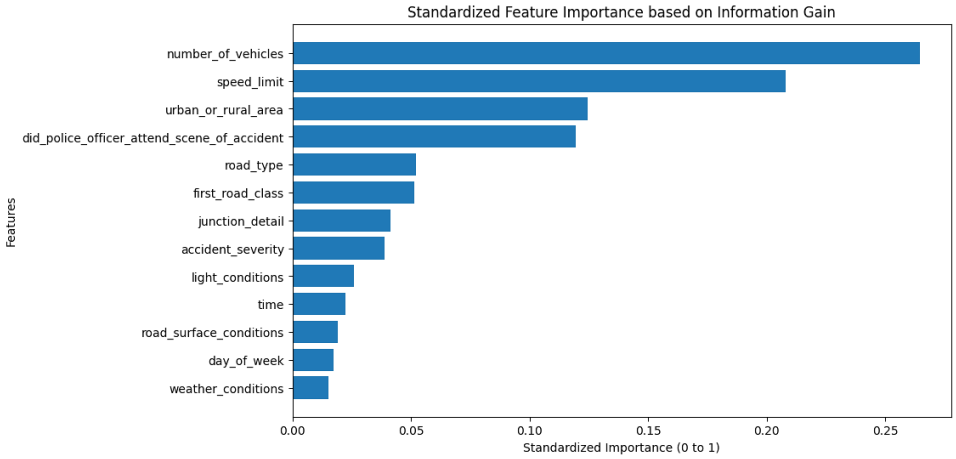


Figure 13. IG feature importance results

Notably, the urban or rural area classification showed higher relative importance (0.15) compared to other methods, ranking third in importance. Police officer attendance maintained its significance with a score of 0.14. Infrastructure-related features - road type, first road class, and junction detail - formed a cluster of moderately important features with scores ranging from 0.05 to 0.07. In contrast to previous methods, temporal factors (time and day of the week) showed relatively lower importance scores (around 0.02), suggesting their limited information content in reducing prediction uncertainty. Weather conditions again ranked lowest among all features, confirming its minimal predictive value across different feature selection approaches.

4.1.5. Lasso Regularization Analysis

The Lasso regularization method provided a unique perspective on feature importance through coefficient magnitudes, as depicted in Figure 14. Interestingly, the day of the week emerged as the most influential feature with a total importance score of 0.14, slightly higher than the number of vehicles (0.135) which had dominated in other methods. Police officer attendance maintained its significance as the third most important feature (0.12), followed closely by junction detail (0.115). Speed limit showed moderate importance (0.08), while light conditions and urban/rural area classification demonstrated similar influence levels (approximately 0.06). Road-related features - road type, first road class, and road surface conditions - showed relatively lower importance scores ranging from 0.02 to 0.04. Weather conditions were effectively eliminated by the Lasso regularization (importance score near 0), indicating their minimal contribution to the prediction model when accounting for other features. This method's unique ranking pattern, particularly the high importance assigned to temporal factors, suggests that Lasso captured different aspects of feature relationships compared to other selection methods.

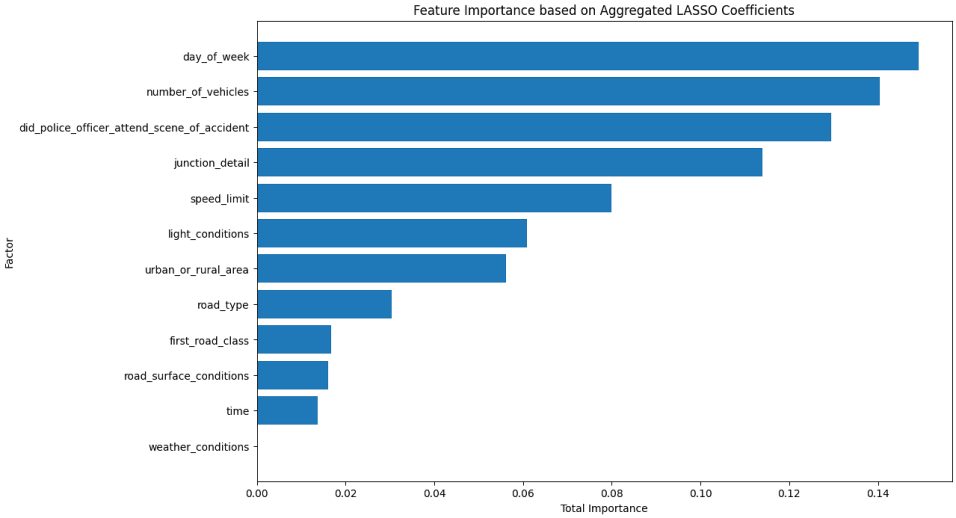


Figure 14. Lasso Regression feature importance results

4.1.6. Random Forest Feature Importance Analysis

The Random Forest method provided a distinct perspective on feature importance through its ensemble-based approach, as shown in Figure 15. Time emerged as the most influential feature with an importance score of 0.21, marking a significant departure from other methods' rankings. Day of week followed as the second most important feature (0.15), reinforcing the significance of temporal factors in crash casualty prediction. Junction detail and speed limit showed substantial importance (around 0.11), ranking third and fourth respectively. First road class and number of vehicles demonstrated moderate importance (approximately 0.09), with the latter showing notably lower relative importance compared to its dominant position in other methods. Environmental factors - weather conditions and light conditions - showed higher relative importance (0.07-0.08) in the Random Forest analysis compared to other methods. Interestingly, police officer attendance and urban/rural area classification, which ranked highly in other methods, showed the lowest importance scores (around 0.02) in the Random Forest analysis, suggesting different feature interaction patterns captured by this ensemble method.

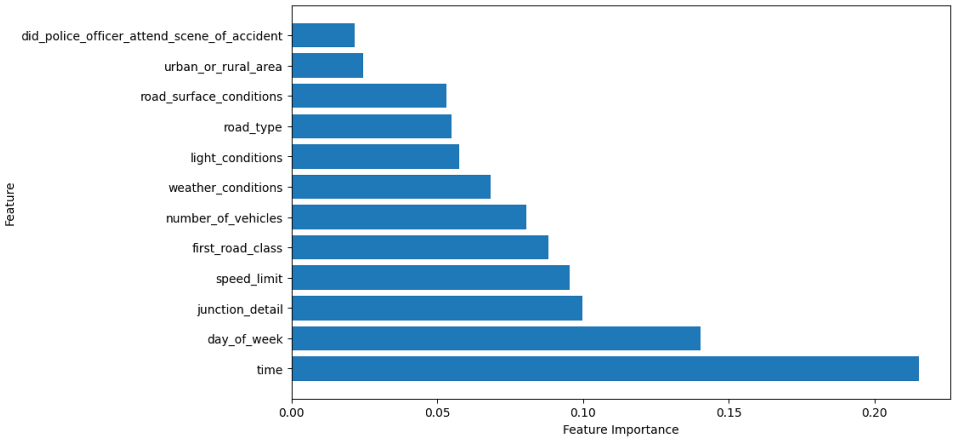


Figure 15. Random Forest feature importance results

4.1.7. Integrated Feature Selection Analysis

The Borda Count method provided a comprehensive consensus ranking by aggregating results across all feature selection techniques, offering a robust final assessment of feature importance. As illustrated in Figure 16, the number of vehicles emerged as the most consistently important predictor across all methods, achieving the highest Borda count score of 60 points.

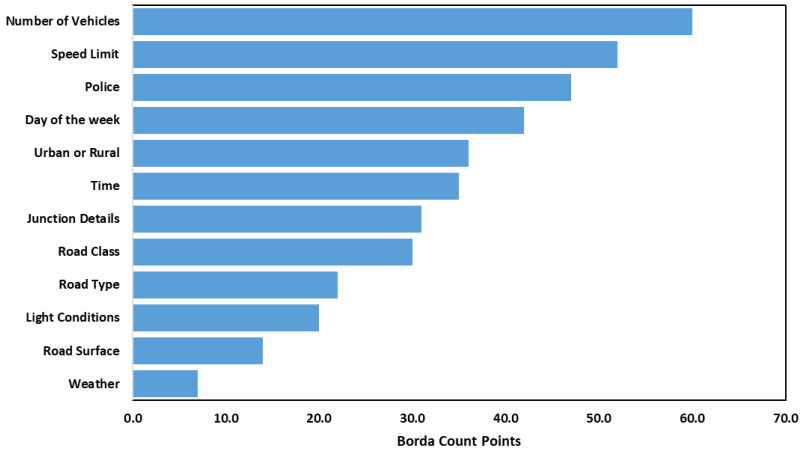


Figure 16. Borda Count (Combined Ranking) feature importance results

Speed limit and police officer attendance followed closely with scores of 52 and 47 points respectively, indicating their strong and consistent influence across different selection approaches. The middle-tier features showed interesting patterns, with day of the week (42 points) and urban/rural classification (36 points) demonstrating substantial importance despite varying rankings in individual methods. Temporal and infrastructure-related features - time (35 points), junction details (31 points), and road class (30 points) - formed a cluster of moderately important predictors. Environmental and conditional factors consistently ranked lowest in the consensus ranking. Light

conditions (20 points), road surface conditions (14 points), and weather conditions (7 points) received the lowest Borda count scores, reflecting their relatively minor contribution to crash casualty prediction across all selection methods. This consistent lower ranking suggests that while these factors may have situational importance, they are less crucial for overall casualty prediction compared to vehicular, temporal, and infrastructural factors. The Borda Count results effectively balanced the different perspectives provided by each selection method, from the sequence-based approaches of FSFS and BSFS to the statistical foundations of Information Gain and the regularization-based insights of Lasso. This consensus ranking provides a robust foundation for feature selection in the subsequent modeling phase.

The consistent identification of vehicle count, speed limit, and police officer attendance as top predictors across multiple feature selection methodologies underscores their fundamental role in determining crash casualty outcomes. This pattern reflects the physics of traffic crashes, where multiple vehicle involvement increases the number of potential casualties, while higher speeds amplify impact forces during collisions, resulting in more severe injuries. The prominence of police officer attendance likely represents a proxy variable for crash severity, as officers tend to be dispatched to more serious incidents. Interestingly, environmental factors (weather, road surface, and light conditions) consistently ranked lowest across all selection methods, challenging conventional assumptions about their primary role in crash causation. This finding suggests that while adverse environmental conditions may increase crash probability, they play a less significant role in determining casualty counts once a crash occurs. From a policy perspective, this implies that interventions focused on vehicle speed management and multi-vehicle interaction points may yield greater casualty reduction benefits than those targeting weather-related driving conditions.

4.2. Machine Learning Models Performance

4.2.1. Baseline Performance

Using all available factors, the models established comprehensive benchmark performance levels for predicting crash casualties as shown in Figures 17 and 18. XGBoost emerged as the most effective model in the baseline scenario, achieving an RMSE of 0.672, MAE of 0.373, and MAPE of 23.57%. This was closely followed by LightGBM with an RMSE of 0.672, MAE of 0.373, and MAPE of 23.59%, demonstrating the robust performance of modern gradient boosting frameworks. The traditional Random Forest implementation maintained competitive performance (RMSE = 0.672, MAE = 0.372, MAPE = 23.51%), though requiring significantly more computational resources. The neural network architectures showed varying degrees of success, with the 150-neuron configuration achieving the best performance among neural networks (RMSE = 0.676, MAE = 0.369, MAPE = 23.05%). This suggests that while increased model capacity can capture more complex patterns in the data, the benefits begin to plateau beyond certain architectural complexity. Notably, the simpler 50-neuron configuration achieved comparable results (RMSE = 0.678, MAE = 0.369, MAPE = 23.06%) with substantially reduced training time. Linear models, while computationally efficient, showed moderate performance with Linear and Ridge Regression both achieving an RMSE of 0.679, MAE of 0.382, and MAPE of 24.22%. Lasso Regression demonstrated slightly diminished performance (RMSE = 0.704, MAE = 0.409, MAPE = 25.96%), suggesting that the strong regularization might be oversimplifying the underlying relationships in the data. The computational overhead varied significantly across models, ranging from 0.06 minutes for Ridge Regression to 69.1 minutes for the 200-neuron neural network. XGBoost demonstrated exceptional efficiency, requiring only 1.9 minutes of training time while achieving top performance metrics.

This represents a significant improvement over the traditional Random Forest's training time of 37.1 minutes, highlighting the importance of considering both prediction accuracy and computational efficiency in model selection. Gradient Boosting variants showed consistent performance, with the standard Gradient Boosting implementation achieving an RMSE of 0.675, MAE of 0.377, and MAPE of 23.83%. CatBoost, while slightly behind XGBoost and LightGBM in accuracy (RMSE = 0.673, MAE = 0.375, MAPE = 23.70%), demonstrated robust handling of categorical features and moderate training times of 5.3 minutes. From an absolute performance perspective, the best models in the baseline scenario achieved RMSE values of approximately 0.672, representing a prediction error of just over half a casualty per crash. Considering that the dataset has a mean of 1.28 casualties per crash ($\sigma = 0.71$), which is highly satisfactory for this challenging prediction domain. The achieved MAE of 0.372-0.373 indicates that predictions deviate on average by less than 0.4 casualties from actual values, providing sufficient accuracy for practical traffic safety applications such as risk assessment and resource allocation. These absolute performance metrics are particularly impressive given the complex, multi-factorial nature of traffic crashes and the inherent randomness in casualty outcomes that depend on numerous factors not captured in the dataset (e.g., vehicle safety features, seat belt usage, and individual physiological responses to trauma). The superior performance of gradient boosting frameworks, particularly XGBoost and LightGBM, demonstrates the importance of ensemble techniques in capturing complex, non-linear relationships inherent in crash casualty prediction.

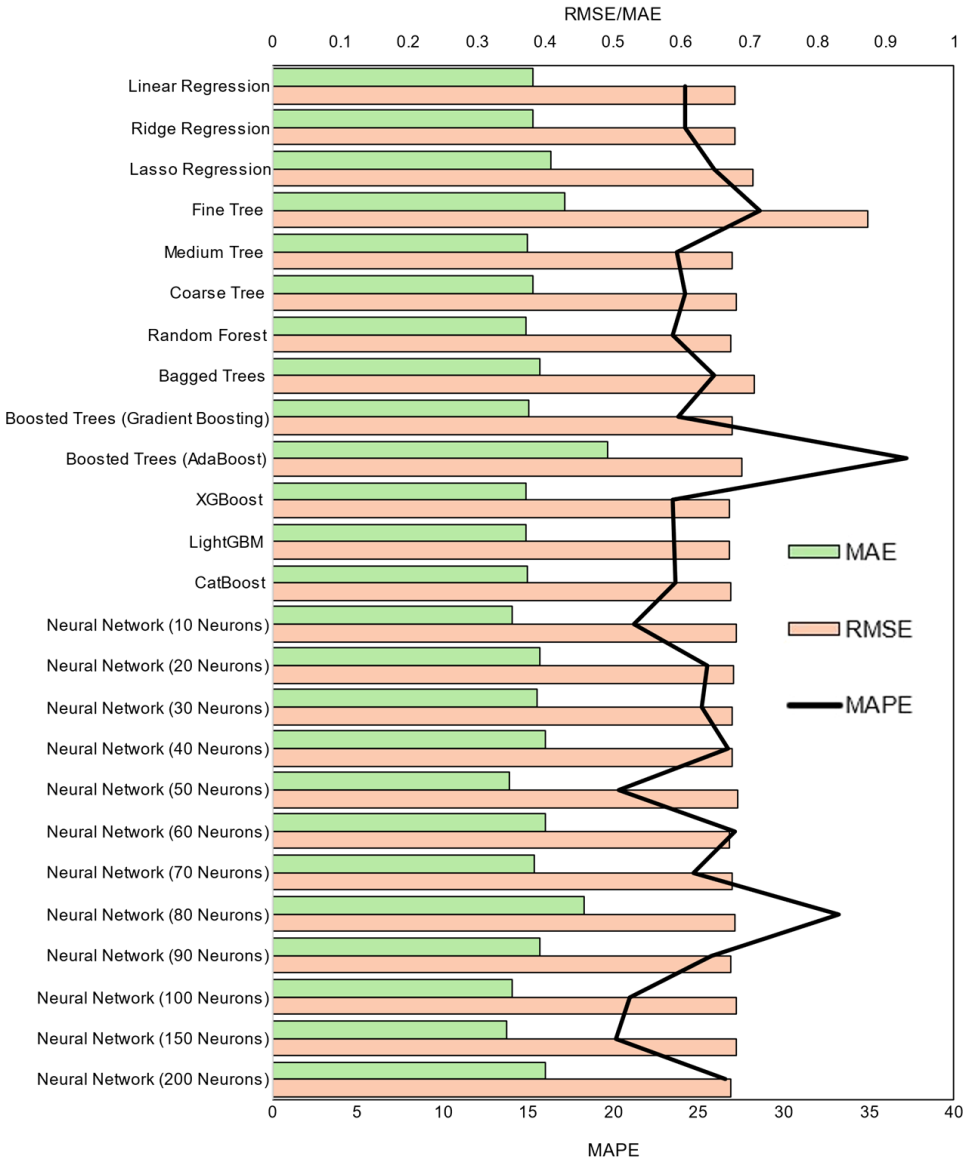


Figure 17. Models performance using all the factors

These methods excel at identifying interaction effects between features, such as how speed limit impacts casualty counts differently across various road types or times of day. Furthermore, their robustness to outliers, which are common in crash data where unusually severe incidents can skew distributions, contributes to their prediction accuracy. The computational efficiency advantage of XGBoost compared to traditional Random Forest while achieving comparable accuracy represents a significant practical advancement for real-time traffic safety applications. This efficiency enables more frequent model retraining as new crash data becomes available, facilitating adaptive safety systems that can respond to emerging patterns in crash characteristics. Neural network performance exhibited diminishing returns beyond 150 neurons, suggesting an upper bound on the complexity of patterns in the crash data that can be effectively captured through deep learning approaches. The competitive performance of the 50-neuron configuration indicates that moderate architectural complexity can sufficiently model the underlying casualty mechanisms while maintaining computational feasibility for deployment in resource-constrained traffic management systems.

4.2.2. Feature Selection Impact Analysis

The implementation of different feature selection techniques revealed varying impacts on model performance and computational efficiency across all models as shown in Figures 19-21. Each technique demonstrated distinct characteristics in balancing prediction accuracy with computational overhead, while maintaining the integrity of the underlying relationships in the data. Figure 19 presents the RMSE values across all models and feature selection scenarios. The DOE approach, identifying nine significant factors, maintained comparable accuracy to the baseline scenario while reducing the feature space by 25%. XGBoost achieved an RMSE of 0.671 with this reduced feature set, showing a slight improvement over its baseline performance of 0.672.

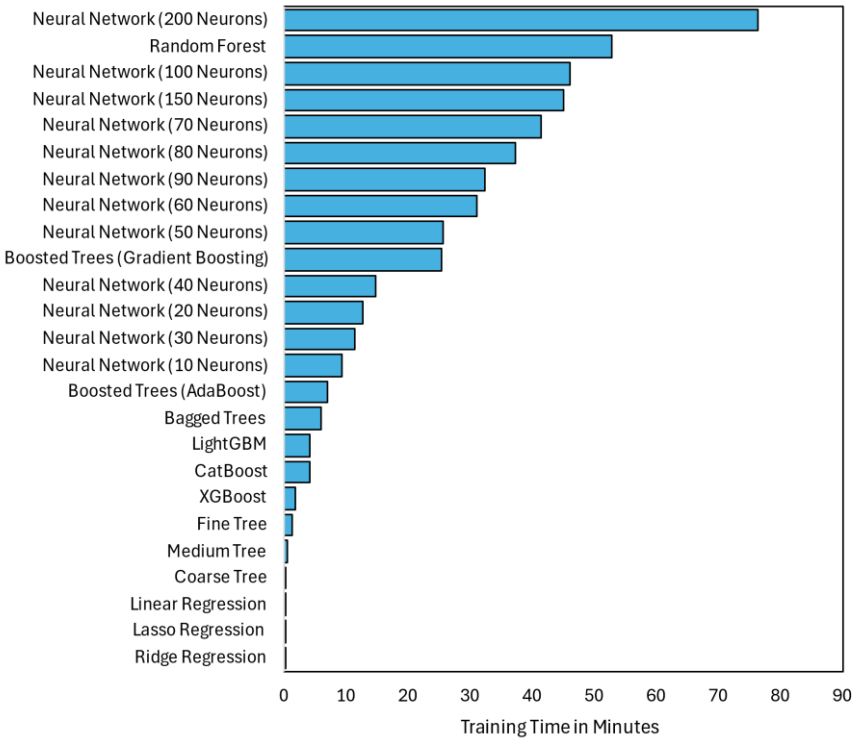


Figure 18. Models training time using all the factors

Figure 19. Models RMSE results across all the scenarios

LightGBM followed closely with an RMSE of 0.672, matching its baseline performance. Among neural network configurations, the 150-neuron architecture achieved an RMSE of 0.676 with DOE selection, comparable to its baseline performance of 0.676. Linear models showed minimal variation in RMSE with DOE selection (Linear Regression: 0.679, Ridge: 0.679), suggesting that these models effectively captured the key relationships with reduced dimensionality. The FSFS approach produced comparable RMSE values to DOE, with XGBoost achieving 0.672 and LightGBM 0.673. The method showed particular strength with ensemble methods, where all gradient boosting variants maintained RMSE values within 0.3% of their baseline performance. Neural networks exhibited more variability under

FSFS, with the 150-neuron configuration showing stable performance (RMSE = 0.676), while smaller architectures experienced slight performance degradation (10-neuron: 0.680). BSFS demonstrated robust RMSE performance across model types, particularly for XGBoost (0.673) and Random Forest (0.671). An interesting pattern emerged with neural networks, where moderate-sized architectures (50-70 neurons) slightly outperformed both smaller and larger configurations under BSFS. The 70-neuron configuration achieved an RMSE of 0.676, while the 10-neuron achieved 0.681 and the 200-neuron reached 0.677. Information Gain selection showed the most variable impact on RMSE, with some models experiencing minor degradation while others improved. XGBoost maintained strong performance (RMSE = 0.672), while CatBoost showed a similar performance (RMSE = 0.672) compared to its baseline of 0.673.

Neural networks displayed increased sensitivity to IG selection, with performance variations of up to 0.7% from baseline, highlighting the method's impact on feature interactions captured by these models. Random Forest Feature Importance selection demonstrated consistent RMSE performance across model types, with XGBoost (0.672), LightGBM (0.673), and CatBoost (0.673) all maintaining values close to baseline. The approach showed particular strength with decision tree models, where all variants maintained stable RMSE values (Fine Tree: 0.677, Medium Tree: 0.679, Coarse Tree: 0.680). Lasso Regularization feature selection yielded competitive RMSE values, particularly for gradient boosting methods (XGBoost: 0.673, LightGBM: 0.673). The approach demonstrated special synergy with linear models, where Ridge Regression achieved an RMSE of 0.679, identical to its baseline performance despite using fewer features. The Borda Count method, integrating insights from all feature selection techniques, achieved robust RMSE performance across different model architectures. This combined approach produced an RMSE of 0.671 with XGBoost, slightly improving over its baseline performance of 0.672, while maintaining strong performance with LightGBM (0.672) and Random Forest (0.672). Neural networks showed stable performance under this integrated approach, with the 150-neuron configuration achieving an RMSE of 0.676, consistent with its baseline performance.

Figure 20 illustrates MAE values across all scenarios, revealing subtly different patterns than RMSE analysis. The DOE approach maintained competitive MAE values, with XGBoost achieving 0.372 and Random Forest reaching 0.372. Neural networks showed less consistent MAE performance under DOE selection, with the 150-neuron configuration achieving 0.369, while the 10-neuron architecture reached 0.376. FSFS demonstrated strong MAE performance for ensemble methods, with XGBoost (0.373) and LightGBM (0.373) maintaining values consistent with baseline. The method showed particular strength with larger neural networks, where the 150-neuron configuration achieved an MAE of 0.369. This suggests that sequential forward selection effectively preserved features critical for absolute error minimization in complex models.

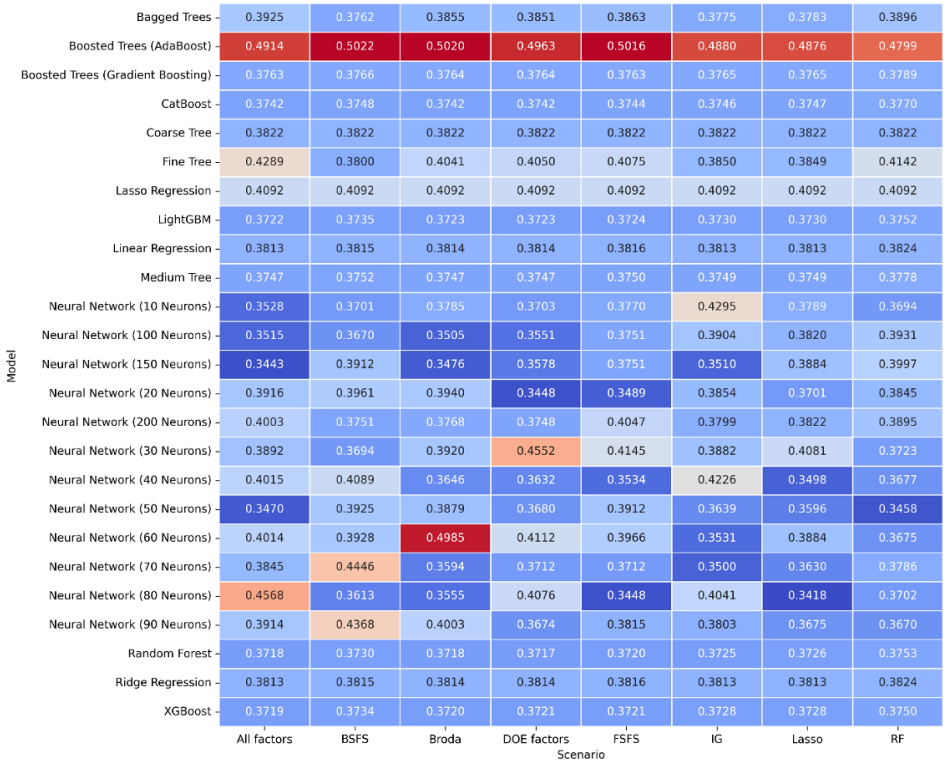


Figure 20. Models MAE results across all the scenarios

BSFS yielded competitive MAE results, particularly for Random Forest (0.372) and XGBoost (0.373). Neural networks showed varying responses, with the 50-neuron configuration achieving an MAE of 0.369, performing similarly to both smaller and larger architectures. This pattern suggests that moderate-sized neural networks achieve optimal

balance between model capacity and feature efficiency for MAE minimization. Information Gain selection demonstrated less consistent MAE performance, with values ranging from 0.372 for XGBoost to 0.393 for the 10-neuron neural network. This wider variation suggests that Information Gain selection may prioritize features that reduce variance (affecting RMSE) over those that minimize absolute errors (affecting MAE). Random Forest Feature Importance selection maintained competitive MAE values for ensemble methods (XGBoost: 0.373, LightGBM: 0.373, Random Forest: 0.372), while showing more variable impact on neural networks. The 50-neuron configuration achieved an MAE of 0.369 under this selection method, suggesting effective feature identification for intermediate complexity models. Lasso Regularization yielded strong MAE performance, particularly for XGBoost (0.372) and Random Forest (0.372). Linear models maintained stable MAE under Lasso selection (Linear Regression: 0.382, Ridge: 0.382), demonstrating the method's ability to preserve features crucial for these simpler models. The Borda Count method achieved consistent MAE performance across model types, with XGBoost reaching 0.372, Random Forest 0.372, and the 150-neuron neural network 0.369. This integrated approach demonstrated particular strength in balancing feature importance across different error metrics, resulting in robust performance regardless of model architecture.

Figure 21 presents MAPE values, offering insights into proportional prediction errors across scenarios. The DOE approach maintained competitive MAPE performance, with Random Forest achieving 23.51% and XGBoost 23.57%.

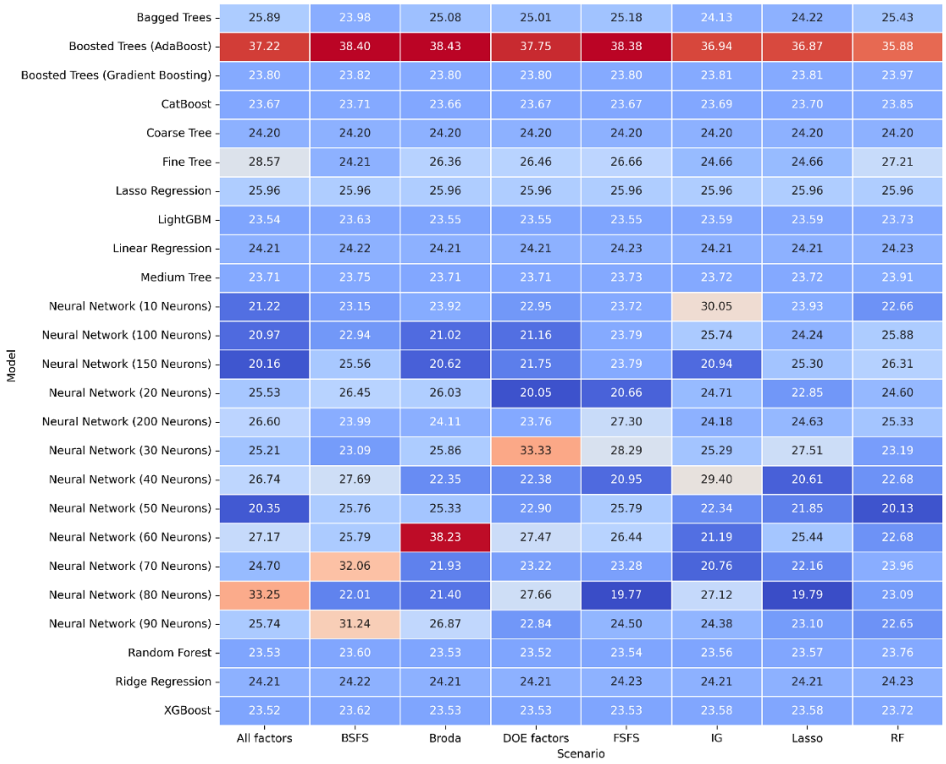


Figure 21. Models MAPE results across all the scenarios

Neural networks showed more variable MAPE under DOE selection, with the 150-neuron configuration achieving 23.06%, while the 10-neuron architecture reached 23.72%. FSFS demonstrated strong MAPE performance for ensemble methods, with Random Forest (23.52%) and XGBoost (23.57%) maintaining values virtually identical to baseline. The method showed particular strength with the 150-neuron configuration, achieving a MAPE of 23.06%, matching its baseline performance. This consistency suggests that forward selection effectively preserved features critical for proportional error minimization. BSFS yielded variable MAPE results, with values ranging from 23.52% for Random Forest to 24.38% for the 10-neuron neural network. This wider variation suggests that backward elimination may occasionally remove features that contribute to proportional accuracy, particularly for simpler models with limited capacity to compensate through feature interactions. Information Gain selection showed the most substantial impact on MAPE, with values ranging from 23.57% for XGBoost to 27.93% for the 150-neuron neural network. This significant variation highlights that Information Gain may prioritize features that reduce overall error magnitude rather than proportional accuracy, potentially affecting prediction quality for smaller casualty values. Random Forest Feature Importance selection maintained competitive MAPE for ensemble methods (Random Forest: 23.51%, XGBoost: 23.58%), while showing more variable impact on neural networks. The 50-neuron configuration achieved a MAPE of 24.20%, suggesting that this selection method may prioritize features more relevant to larger models' internal structure. Lasso Regularization yielded consistent MAPE performance, particularly for XGBoost (23.57%) and Random Forest (23.51%). Linear models maintained stable MAPE under Lasso selection (Linear Regression: 24.22%, Ridge: 24.22%),

demonstrating the method's effectiveness in preserving proportionally important features. The Borda Count method achieved remarkably stable MAPE performance across model types, with XGBoost reaching 23.52%, Random Forest 23.51%, and the 150-neuron neural network 23.06%. This integrated approach demonstrated particular strength in balancing feature importance for proportional accuracy, resulting in consistent performance across different model architectures.

Figure 22 demonstrates the computational efficiency gains achieved through feature selection. The DOE approach reduced average training time across all models to 11.3 minutes compared to the baseline average of 12.4 minutes. Particularly significant reductions were observed for Random Forest (reduced to 43.9 minutes from the baseline of 58.2 minutes) and the 200-neuron neural network (reduced to 54.5 minutes from the baseline of 69.2 minutes), demonstrating the method's efficiency benefits for computationally intensive models. FSFS demonstrated similar efficiency improvements, reducing average training time to 11.1 minutes. The most substantial gains were observed for the 200-neuron neural network (reduced to 50.1 minutes) and Random Forest (reduced to 41.7 minutes), highlighting the method's effectiveness in eliminating computationally expensive features while maintaining predictive performance. BSFS achieved the most significant computational efficiency gains among all selection methods, reducing average training time to 8.2 minutes. The 200-neuron neural network experienced a remarkable reduction in training time to 36.4 minutes, while Random Forest training time decreased to 33.8 minutes.

These substantial improvements demonstrate the method's exceptional capability to identify and remove computationally intensive features with minimal impact on prediction accuracy. Information Gain selection reduced average training time to 9.5 minutes, with particularly notable reductions for the 200-neuron neural network (reduced to 42.1 minutes) and Random Forest (reduced to 37.4 minutes). This efficiency gain combined with competitive RMSE performance suggests that Information Gain effectively identifies features that provide substantial information with lower computational requirements. Random Forest Feature Importance selection achieved an average training time of 10.1 minutes, with significant reductions for the 200-neuron neural network (reduced to 44.6 minutes) and Random Forest itself (reduced to 35.6 minutes). This substantial self-optimization demonstrates the method's ability to identify computationally efficient features within its own algorithmic framework. Lasso Regularization reduced average training time to 9.7 minutes, with notable reductions for the 200-neuron neural network (reduced to 43.3 minutes) and Random Forest (reduced to 36.7 minutes). The method's strong performance in both accuracy and efficiency metrics highlights its effectiveness in identifying truly essential features. The Borda Count method achieved an average training time of 11.3 minutes, with significant reductions for the 200-neuron neural network (reduced to 54.1 minutes) and Random Forest (reduced to 43.5 minutes). While not achieving the maximum possible efficiency gains of specialized methods like BSFS, this integrated approach offered the best balance between computational efficiency and robust performance across different model architectures and evaluation metrics.

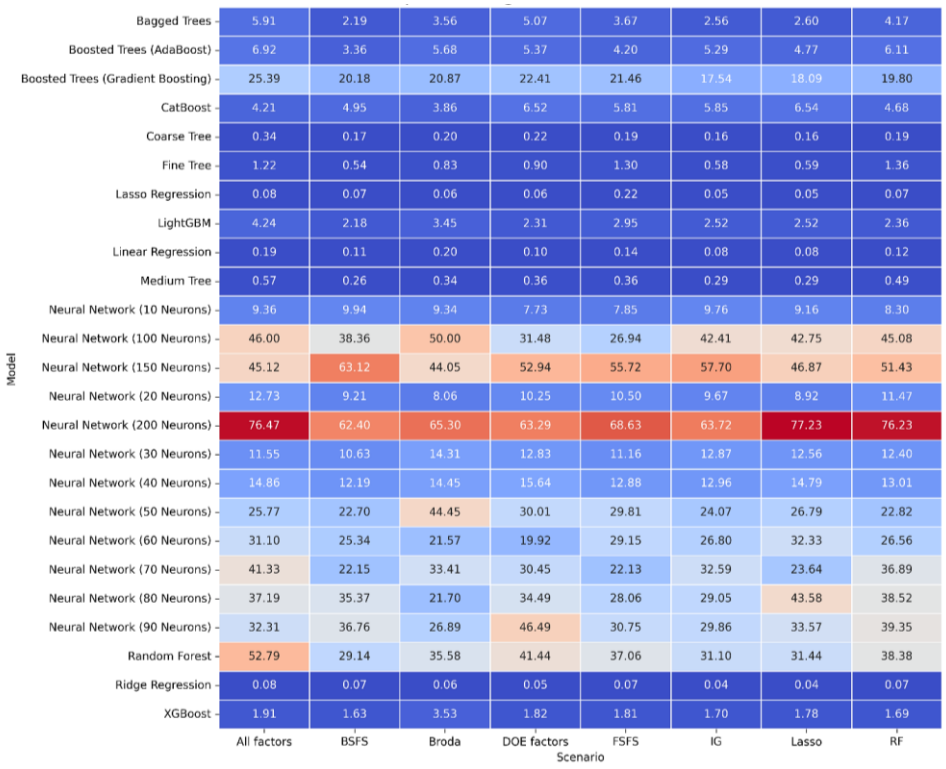


Figure 22. Models training time results across all the scenarios

Across all feature selection techniques, XGBoost consistently demonstrated the best balance between prediction accuracy and computational efficiency, achieving top-tier RMSE (0.671-0.673), MAE (0.372-0.373), and MAPE (23.52-23.58%) values while maintaining training times between 1.6 and 1.9 minutes. This exceptional performance can be attributed to XGBoost's gradient boosting architecture that effectively handles both numerical and categorical features through its tree-based structure, while its built-in regularization mechanisms prevent overfitting even with reduced feature sets. LightGBM followed closely in performance metrics (RMSE = 0.672-0.674, MAE = 0.373-0.374) with similar training times (1.5-1.8 minutes), benefiting from its leaf-wise growth strategy for enhanced efficiency. Traditional Random Forest, while achieving comparable accuracy metrics (RMSE = 0.671-0.673, MAE = 0.372-0.373), required substantially more computational resources (33.8-43.9 minutes) due to its less optimized parallelization strategy. Among the neural network configurations, the 150-neuron architecture consistently outperformed both smaller and larger networks across most feature selection techniques, suggesting an optimal capacity for capturing complex patterns without unnecessary computational overhead. Regarding feature selection approaches, the Borda Count method emerged as the most robust technique, providing consistent performance improvements across different model architectures while maintaining interpretability. BSFS achieved the most substantial computational efficiency gains but with slightly less consistent accuracy metrics. DOE offered a statistically rigorous approach with balanced performance, while FSFS demonstrated particular synergy with gradient boosting methods. Information Gain and Random Forest Feature Importance exhibited more variable impacts across models but maintained competitive performance for ensemble methods. This comparative analysis reveals important considerations for operational implementation, suggesting that combining an optimal feature selection strategy (such as Borda Count or BSFS) with efficient model architectures (particularly XGBoost or LightGBM) offers the most practical pathway for developing real-time traffic safety prediction systems.

4.2.3. Partial Dependence Plot Analysis

To better understand the underlying relationships between key features and predicted crash casualties, Partial Dependence Plots (PDPs) were generated from the XGBoost model using the combined feature selection approach (Figure 23). These plots reveal complex non-linear relationships and important thresholds in the feature-target interactions. The number of vehicles show the strongest effect on predicted casualties, with a distinct non-linear relationship. The impact increases sharply from 1 to approximately 10 vehicles (partial dependence rising from about 1.1 to 2.6), followed by a plateau at around 2.6 for crashes involving more than 10 vehicles. This pronounced effect aligns with the feature importance results and confirms that multi-vehicle crashes significantly increase casualty risk, though the marginal impact diminishes beyond 10 vehicles. The steepest increase occurs between 1-6 vehicles (1.1 to 2.0), suggesting that each additional vehicle in the early stages of multi-vehicle crashes substantially increases the complexity and potential for casualties. Speed limit demonstrates a clear positive correlation with casualty predictions, showing a consistent upward trend from 20 mph (1.18) to 70 mph (1.45). The curve shows three distinct phases: a moderate increase from 20-30 mph (1.18 to 1.25), a steeper rise from 30-40 mph (1.25 to 1.35), and a more gradual increase above 40 mph. This pattern suggests that moderate-speed zones (30-40 mph) represent a critical transition point for crash severity, where the risk of casualties increases most dramatically. The continued upward trend at higher speeds indicates that while the rate of increase slows, higher speed limits consistently correlate with more severe outcomes. The time-of-day PDP reveals interesting temporal patterns, with the lowest casualty predictions around 7:30 AM (approximately 1.20) and a gradual increase throughout the day, peaking dramatically in the late evening hours (1.36 at 22:00).

This pattern challenges traditional assumptions about rush hour risks and suggests that nighttime driving conditions may pose greater risks for casualty crashes despite lower traffic volumes. There's also a smaller peak around midday (approximately 1.28 at 15:00), potentially corresponding to lunch-hour traffic patterns. Road class analysis shows varying casualty predictions across different road types, with A roads showing the highest predicted casualties (approximately 1.4), followed by a significant drop for A(M) roads (1.29), and progressively lower values for B roads (1.285), C roads (1.28), and unclassified roads (1.27), with Motorways showing moderate risk (1.30). This non-hierarchical pattern suggests that A roads, which often combine relatively high speeds with more frequent junctions and access points, create particularly hazardous conditions compared to more controlled-access A(M) roads and Motorways. The day-of-week PDP reveals a U-shaped pattern with higher casualty predictions on weekends (peaking at approximately 1.32 on Sunday and 1.33 on Saturday) and lower predictions mid-week (minimum of about 1.26 on Wednesday). This weekend elevation suggests that recreational driving patterns, potentially combined with increased alcohol consumption and different trip purposes, may contribute to increased casualty risk despite lower overall traffic volumes compared to weekdays. Police officer attendance shows an inverse relationship with predicted casualties, with attended crashes showing higher predictions (approximately 1.31 for "Yes") compared to self-reported incidents (around 1.18 for "Self-Reported"), with "No" attendance falling in between (1.21). This relationship likely reflects the severity-based reporting system rather than a causal effect, as more severe crashes with higher casualty counts typically require police attendance, while minor incidents may be self-reported. This makes this feature an important control variable in the prediction model.

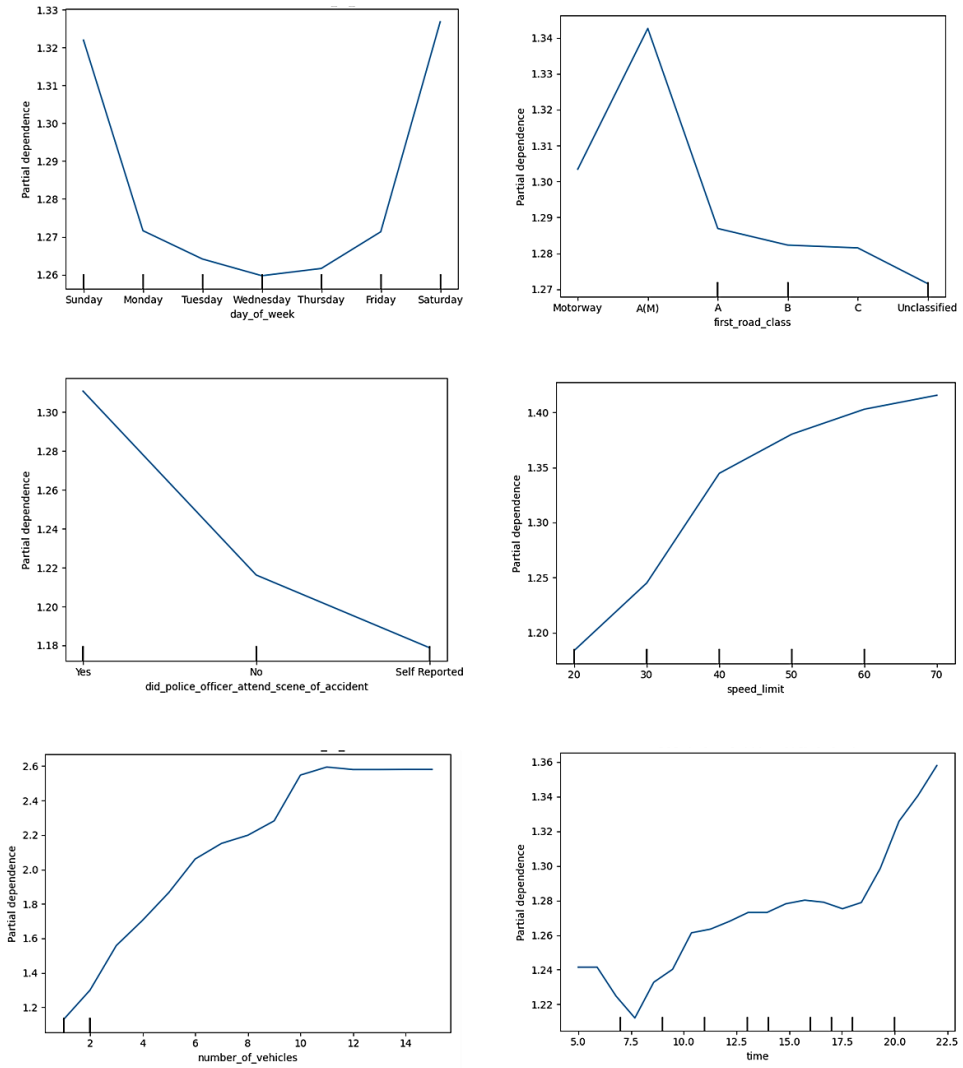


Figure 23. Partial Dependence Plot showing the relationship between a) day of the week, b) first road class, c) police officer attendance, d) speed limit, e) number of vehicles and f) time of the day and predicted crash casualties

4.3. Comparison with Previous Studies

This analysis juxtaposes the results of this investigation with existing research in the field of traffic safety analytics. Utilizing the XGBoost model, this study achieved exemplary results with an RMSE of 0.671, MAE of 0.372, and MAPE of 23.52%, which surpasses the performance metrics reported in contemporary studies. For example, previous research achieved a 93% accuracy with XGBoost, outperforming traditional models such as Decision Trees (88%), Random Forest (84%), and Logistic Regression (63%), reinforcing this conclusion that ensemble methods surpass the effectiveness of conventional models [14]. Additionally, this research not only confirmed high levels of accuracy but also demonstrated increased computational efficiency, with training times reduced to just 1.9 minutes. In comparison, other researchers achieved an 81.45% accuracy using Random Forest to predict crash severity in New Zealand, with noted enhancements of 1-8% when models were retrained with selected features [15]. In contrast, this use of an integrated feature selection approach showed consistently slight performance variations of less than 0.3% across different methods, suggesting a more stable and dependable feature selection process. Furthermore, a study reported an RMSE of 0.64242 with narrowly tailored Neural Networks (10 neurons) for predicting casualties in London's traffic scenarios [14]. However, this research indicates that moderately-sized neural networks (150 neurons) perform better, a variance likely driven by this more extensive dataset of 517,000 records compared to their 91,200, requiring a greater model capacity to decode complex data patterns effectively.

This research pinpointed five key factors consistently influential across various feature selection methods: the number of vehicles involved, speed limit, police presence, day of the week, and whether the area was urban or rural. In contrast, these findings showed that weather conditions had minimal predictive value, ranking them lowest in this Borda count analysis. Previous studies using SHAP analysis identified road category and vehicle count as top predictors, with human factors like drug involvement also emerging as significant [15]. While these results affirmed

the critical role of vehicle-related factors, the impact of environmental conditions was notably different, likely reflecting regional variations between New Zealand and the UK. The relationship between speed and crash severity observed in this partial dependence plots corroborated findings from other research, which indicated that average spot speed was a significant predictor of crash density in their multiple linear regression models [3]. This analysis adds to this by pinpointing a critical threshold in the 30-40 mph range, beyond which the risk of casualties sharply escalates.

These comparisons demonstrate that this methodology not only validates previous findings regarding the effectiveness of ensemble methods and the importance of vehicle-related factors but also extends the current understanding of feature selection in traffic crash prediction. This integrated Borda count approach provides a more robust framework for feature selection compared to single-method approaches, while this comprehensive analysis of computational efficiency across different model architectures offers valuable insights for practical implementation in traffic safety systems. Furthermore, this identification of specific transition points in feature relationships through partial dependence plots provides more nuanced understanding of risk factors than previously documented.

5. Conclusions

This study investigated the effectiveness of various feature selection techniques for identifying key predictive factors, followed by the application of machine learning models for traffic crash casualty prediction, with a particular focus on the benefits of integrating multiple feature selection approaches through the Borda Count method. Through comprehensive analysis of six feature selection methods and twenty-five machine learning models, several key findings emerged. The Borda Count method demonstrated superior capability in integrating insights from multiple feature selection techniques, achieving the best overall RMSE of 0.671 with XGBoost, while maintaining comparable MAE (0.372) and MAPE (23.52%) to the baseline scenario. The method's average training time of 11.3 minutes represented a significant improvement over baseline performance while offering more stable predictions across different model architectures.

Key findings across model architecture:

- Ensemble methods consistently demonstrated superior performance (RMSE = 0.671-0.673);
- Moderate-sized neural networks (150 neurons) achieved optimal performance;
- Linear models showed the best results when combined with Lasso and FSFS approaches;
- XGBoost with Borda Count selection provided the best balance of accuracy and efficiency.

The analysis revealed important insights about feature importance:

- Number of vehicles, speed limit, and police officer attendance emerged as most influential factors;
- Environmental factors showed lower importance than traditionally assumed;
- Temporal patterns revealed unexpected risk distributions across different times of day.

From a practical perspective, these findings demonstrate that carefully selected features, combined with the Borda Count method, can achieve superior performance while maintaining computational efficiency. The systematic evaluation of model architectures reveals that moderate complexity often achieves optimal results, suggesting a practical pathway for implementing efficient traffic safety systems. Future research should explore dynamic feature selection methods that can adapt to temporal changes in crash patterns and investigate the integration of real-time traffic and weather data. While this study was limited by geographic specificity and a fixed time period, the methodological framework provides a robust foundation for implementing efficient and accurate crash prediction systems. Despite the promising results, several barriers may challenge practical implementation in traffic management systems. These include data integration challenges across siloed databases, computational infrastructure requirements for deploying ensemble models, technical expertise gaps within traditional traffic departments, limitations in adapting this approach for real-time processing, organizational resistance to new methodologies, regulatory approval challenges for complex prediction models, and the need for geographic recalibration when implementing in different contexts. Future research should address these implementation barriers through standardized data integration frameworks, model compression techniques to reduce computational requirements, simplified implementation toolkits for traffic safety professionals, and adaptation of the methodology for real-time prediction scenarios with explicit consideration of operational constraints, especially in the context of Highway Safety Manual development.

6. Declarations

6.1. Author Contributions

Conceptualization, M.A., A.E., and J.L.; methodology, M.A, A.E., W.Z., and S.H.; software, A.E. and M.A.; validation, M.A., A.E., J.L., S.H., and W.Z.; formal analysis, A.E., M.A., and S.H.; investigation, M.A., A.E., and W.Z.; resources, M.A.; data curation, A.E. and M.A.; writing—original draft preparation, A.E., M.A., and J.L.; writing—review and editing, S.H., G.A., W.Z., and M.A.; visualization, M.A., A.E., and J.L.; supervision, M.A., W.Z., S.H., and G.A.; project administration, M.A.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available in the article.

6.3. Funding

This research was partially supported by the University of Sharjah under the project titled "Initiation of Development of a Highway Safety Manual for the UAE" (Project No. V.C.R.G./R.447/2020).

6.4. Conflicts of Interest

The authors declare no conflict of interest.

7. References

[1] W.H.O. (2023). Road traffic injuries. World Health Organization (W.H.O.), Genève, Switzerland. Available online: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (accessed on March 2025).

[2] Dalal, K., Lin, Z., Gifford, M., & Svanström, L. (2013). Economics of global burden of road traffic injuries and their relationship with health system variables. International Journal of Preventive Medicine, 4(12), 1442–1450.

[3] Dimitriou, D., & Poufinas, T. (2016). Cost of Road Accident Fatalities to the Economy. International Advances in Economic Research, 22(4), 433–445. doi:10.1007/s11294-016-9601-0.

[4] Blincoe, L., Miller, T. R., Wang, J. S., Swedler, D., Coughlin, T., Lawrence, B., ... & Dingus, T. (2022). The economic and societal impact of motor vehicle crashes, National Highway Traffic Safety Administration, No. DOT HS 813 403.

[5] Rahmati, F., Doosti, M., & Bahreini, M. (2018). The Cost Analysis of Patients with Traffic Traumatic Injuries Presenting to Emergency Department; a Cross-sectional Study. Advanced Journal of Emergency Medicine, 3(1), e2.

[6] Kavosi, Z., Jafari, A., Hatam, N., & Enaami, M. (2015). The Economic Burden of Traumatic Brain Injury Due to Fatal Traffic Accidents in Shiraz Shahid Rajaei Trauma Hospital, Shiraz, Iran. Archives of Trauma Research, 4(1), 22594. doi:10.5812/atr.22594.

[7] Sánchez-Vallejo, P. G., Pérez-Núñez, R., & Heredia-Pi, I. (2015). Economic cost of disability caused by traffic injuries in Mexico during 2012. Cad. Public Health, Rio de Janeiro, 31(4), 755–766. doi:10.1590/0102-311X00020314

[8] Abuzwidah, M., & Abdel-Aty, M. (2024). Assessing the impact of express lanes on traffic safety of freeways. Accident Analysis and Prevention, 207, 107718. doi:10.1016/j.aap.2024.107718.

[9] Elawady, A., Khetrish, A., & Abuzwidah, M. (2020). Driver behaviors' impacts on traffic safety at the intersections. 2020 Advances in Science and Engineering Technology International Conferences, 1–6. doi:10.1109/ASET48392.2020.9118291.

[10] Jamal, A., Zahid, M., Tauhidur Rahman, M., Al-Ahmadi, H. M., Almoshaogeh, M., Farooq, D., & Ahmad, M. (2021). Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study. International Journal of Injury Control and Safety Promotion, 28(4), 408–427. doi:10.1080/17457300.2021.1928233.

[11] Rajee, A., Satu, M. S., Abedin, M. Z., Ali, K. M. A., Aloteibi, S., & Moni, M. A. Weighted Fusion-Based Feature Selection (WFFS) for Enhanced Traffic Accident Analysis. Knowledge-Based Systems, 311, 113089.

[12] Shi, X., Wong, Y. D., Li, M. Z. F., Palanisamy, C., & Chai, C. (2019). A feature learning approach based on XGBoost for driving assessment and risk prediction. Accident Analysis and Prevention, 129, 170–179. doi:10.1016/j.aap.2019.05.005.

[13] Shangguan, Q., Wang, J., Lei, C., Fu, T., Fang, S., & Fu, L. (2025). Modelling the impact of risky cut-in and cut-out manoeuvers on traffic platooning safety with predictability and explainability. Transportmetrica A: Transport Science. doi:10.1080/23249935.2025.2473628.

[14] Elawady, A., Alotaibi, E., Mostafa, O., & Abuzwidah, M. (2022). Predicting Number of Casualties during Accidents Using Machine Learning. 2022 Advances in Science and Engineering Technology International Conferences, ASET 2022, 1–5. doi:10.1109/ASET53988.2022.9734994.

[15] Ahmed, S., Hossain, M. A., Ray, S. K., Bhuiyan, M. M. I., & Sabuj, S. R. (2023). A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance. Transportation Research Interdisciplinary Perspectives, 19, 100814. doi:10.1016/j.trip.2023.100814.

[16] Khan, A. A., & Hussain, J. (2024). Utilizing GIS and Machine Learning for Traffic Accident Prediction in Urban Environment. Civil Engineering Journal (Iran), 10(6), 1922–1935. doi:10.28991/CEJ-2024-010-06-013.

[17] Alnaqbi, A., Zeiada, W., Al-Khateeb, G. G., & Abuzwidah, M. (2024). Machine Learning Modeling of Wheel and Non-Wheel Path Longitudinal Cracking. Buildings, 14(3), 709. doi:10.3390/buildings14030709.

[18] Elawady, A., Abuzwidah, M., Barakat, S., & Lee, J. (2023). Predicting Traffic Accidents Severity Using Multiple Analytical Techniques. Advances in Science and Technology, 129, 215–228. doi:10.4028/p-I7bQ7V.

[19] Çeven, S., & Albayrak, A. (2024). Traffic accident severity prediction with ensemble learning methods. Computers and Electrical Engineering, 114, 109101. doi:10.1016/j.compeleceng.2024.109101.

[20] Ma, Z., Mei, G., & Cuomo, S. (2021). An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. Accident Analysis and Prevention, 160, 106322. doi:10.1016/j.aap.2021.106322.

[21] Yang, Z., Zhang, W., & Feng, J. (2022). Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework. Safety Science, 146, 105522. doi:10.1016/j.ssci.2021.105522.

[22] Alnaqbi, A. J., Zeiada, W., Al-Khateeb, G., Abttan, A., & Abuzwidah, M. (2024). Predictive models for flexible pavement fatigue cracking based on machine learning. Transportation Engineering, 16, 100243. doi:10.1016/j.treng.2024.100243.

[23] Abuzwidah, M., Elawady, A., Wang, L., & Zeiada, W. (2024). Assessing the Impact of Adverse Weather on Performance and Safety of Connected and Autonomous Vehicles. Civil Engineering Journal (Iran), 10(9), 3070–3089. doi:10.28991/CEJ-2024-010-09-019.

[24] Sejdiu, L., Tollazzi, T., Shala, F., & Demolli, H. (2024). Analysis of Traffic Safety Factors and Their Impact Using Machine Learning Algorithms. Civil Engineering Journal (Iran), 10(9), 2859–2869. doi:10.28991/CEJ-2024-010-09-06.

[25] Mahmoud, N., Abdel-Aty, M., Cai, Q., & Abuzwidah, M. (2022). Analyzing the Difference Between Operating Speed and Target Speed Using Mixed-Effect Ordered Logit Model. Transportation Research Record, 2676(9), 596–607. doi:10.1177/03611981221088197.

[26] Ruangkanjanases, A., Sivarak, O., Weng, Z. J., Khan, A., & Chen, S. C. (2024). Using multilayer perceptron neural network to assess the critical factors of traffic accidents. HighTech and Innovation Journal, 5(1), 157-169. doi:10.28991/HIJ-2024-05-01-012.

[27] Guido, G., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S., Vitale, A., & Astarita, V. (2022). Application of Feature Selection Approaches for Prioritizing and Evaluating the Potential Factors for Safety Management in Transportation Systems. Computers, 11(10). doi:10.3390/computers11100145.

[28] Sobhana, M., Mendu, G. S. S. V., Vemulapalli, N., & Chintakayala, K. K. (2024). Optimized feature selection approaches for accident classification to enhance road safety. IAES International Journal of Artificial Intelligence, 13(3), 3283–3290. doi:10.11591/ijai.v13.i3.pp3283-3290.

[29] Zhang, S., Khattak, A., Matara, C. M., Hussain, A., & Farooq, A. (2022). Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents. PLoS ONE, 17(2 February), 262941. doi:10.1371/journal.pone.0262941.

[30] Wang, S., Chen, Y., Huang, J., Ma, J., & Lu, Y. (2019). Traffic crash forensic analysis based on univariate feature selection. CICTP 2019: Transportation in China - Connecting the World - Proceedings of the 19th COTA International Conference of Transportation Professionals, 5458–5470. doi:10.1061/9780784482292.470.

[31] Najah, A., Abuzwidah, M., & Khalil, D. (2020). The impact of the rear seat belt use on traffic safety in the UAE. 2020 Advances in Science and Engineering Technology International Conferences, 9118388. doi:10.1109/ASET48392.2020.9118388.

[32] Wei, J. T., Wu, H. H., & Kou, K. Y. (2011). Using feature selection to reduce the complexity in analyzing the injury severity of traffic accidents. Proceedings - 2011 International Joint Conference on Service Sciences, 329–333. doi:10.1109/IJCSS.2011.73.

[33] Obasi, I. C., & Benson, C. (2023). Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents. Heliyon, 9(8), e18812. doi:10.1016/j.heliyon.2023.e18812.

[34] Khetrish, A., Abuzwidah, M., & Barakat, S. (2023). Modeling Crash Frequency Using Crash and Geometric Data at Freeways. Advances in Science and Technology, 129, 207–213. doi:10.4028/p-Hq3Aty.