safety

MDPI

Article

Explainable Boosting Machine: A Contemporary Glass-Box Model to Analyze Work Zone-Related Road Traffic Crashes

Raed Alahmadi [1], Hamad Almujibah [2,*], Saleh Alotaibi [3], Ali. E. A. Elshekh [2], Mohammad Alsharif [4] and Mudthir Bakri [5]

[1] Department of Civil Engineering, College of Engineering, Al-Baha University, Al-Baha P.O. Box 1988, Saudi Arabia; rnaif@bu.edu.sa
[2] Department of Civil Engineering, College of Engineering, Taif University, Taif P.O. Box 11099, Saudi Arabia; a.elheber@tu.edu.sa
[3] Department of Civil and Environmental Engineering, Faculty of Engineering—Rabigh Branch, King Abdulaziz University, Jeddah P.O. Box 21589, Saudi Arabia; salnufiae@kau.edu.sa
[4] Department Architecture, College of Engineering, Al-Baha University, Al-Baha P.O. Box 1988, Saudi Arabia; malsharif@bu.edu.sa
[5] Department of Civil Engineering, College of Engineering, Qassim University, Unaizah P.O. Box 56452, Saudi Arabia
* Correspondence: hmujibah@tu.edu.sa

Abstract: Examining the factors contributing to work zone crashes and implementing measures to reduce their occurrence can significantly improve road safety. In this research, we utilized the explainable boosting machine (EBM), a modern glass-box machine learning (ML) model, to categorize and predict work zone-related crashes and to interpret the various contributing factors. The issue of data imbalance was also addressed by utilizing work zone crash data from the state of New Jersey, comprising data collected over the course of two years (2017 and 2018) and applying data augmentation strategies such synthetic minority over-sampling technique (SMOTE), borderline-SMOTE, and SVM-SMOTE. The EBM model was trained using augmented data and Bayesian optimization for hyperparameter tuning. The performance of the EBM model was evaluated and compared to black-box ML models such as combined kernel and tree boosting (KTBoost, python 3.7.1 and KTboost package version 0.2.2), light gradient boosting machine (LightGBM version 3.2.1), and extreme gradient boosting (XGBoost version 1.7.6). The EBM model, using borderline-SMOTE-treated data, demonstrated greater efficacy with respect to precision (81.37%), recall (82.53%), geometric mean (75.39%), and Matthews correlation coefficient (0.43). The EBM model also allows for an in-depth evaluation of single and pairwise factor interactions in predicting work zone-related crash severity. It examines both global and local perspectives, and assists in assessing the influence of various factors.

Keywords: traffic safety; work zones crashes; explainable boosting machine

check for updates

Citation: Alahmadi, R.; Almujibah, H.; Alotaibi, S.; Elshekh, A.E.A.; Alsharif, M.; Bakri, M. Explainable Boosting Machine: A Contemporary Glass-Box Model to Analyze Work Zone-Related Road Traffic Crashes. *Safety* **2023**, *9*, 83. https://doi.org/10.3390/safety9040083

Academic Editor: Raphael Grzebieta

Received: 12 October 2023
Revised: 14 November 2023
Accepted: 22 November 2023
Published: 28 November 2023

1. Introduction

The execution of construction activities, corrective and preventative repair initiatives, as well as rehabilitation efforts within work zones play a pivotal role in the maintenance and enhancement of road infrastructure on a global scale. The escalating need for the refurbishment and rebuilding of deteriorating infrastructure for transportation in the United States has led to the initiation of numerous roadway construction projects across the nation. Over recent years, there has been a notable rise in the quantity of work zones, primarily attributed to the expansion of highway renovations within the State of New Jersey. Despite the potential disruption to traffic and increased risk of crashes, this occurrence is an inescapable reality that cannot be disregarded. Work zones are susceptible to a higher likelihood of crashes due to the presence of fluctuating traffic patterns, decreased right-of-way, and ongoing roadwork [1]. The Federal Highway Administration (FHWA) reported that a total of 27,037 individuals, with an annual average of 773 fatalities, lost their lives in work zone accidents in the United States between the years 1982 and 2017. The likelihood of crashes occurring in work zones is a matter of grave concern for drivers, traffic enforcement agencies, as well as road traffic safety experts [2]. A precise prediction of the severity of crashes related to work zones and the evaluation of factors that contribute to these crashes are of considerable significance. Machine Learning (ML) techniques demonstrate a notable level of adaptability and superior performance. Presently, there is a growing interest in the utilization of ML models. Despite the high accuracy of prediction in various models based on machine learning (ML), a notable limitation is their inherent transparency, commonly referred to as the "black-box" nature. Consequently, the utilization of post hoc explanation techniques becomes necessary to facilitate further interpretation. Therefore, in contrast to black-box models, our aim is to develop a "glass-box" model that can effectively predict work zone-related crashes while also providing interpretation of the various contributing factors.

2. Literature Review

Researchers have analyzed various crash risk factors in work zones by studying work zone crash records. The vulnerability of severe injury in work zones is influenced by various factors, including the characteristics of individual vehicles and workers, as well as the behavior of drivers within work zones. These factors include the preferred speed of drivers, their braking actions, the paths they choose to navigate through the work zone, etc. In a study conducted by Morgan et al. [3], driving simulator tests were utilized to examine the relationship between work zone crashes and taper lengths. The findings indicated that crashes were more likely to occur in work zones with shorter taper lengths, particularly when drivers' ability to anticipate hazards was hindered by reduced viewing distances. Weng et al. [4] examined the influence of different variables on the risks of injury and fatality for drivers in work zones, considering both short-term and long-term scenarios. They established that various factors, including light conditions, day of the week, gender, age, airbag availability, restraint use, and vehicle age, played a significant role in contributing to fatalities in both types of work zones. Concerns have been raised about how exceeding speed limits and driving swiftly in work zones may jeopardize safety. Debnath et al. [5] formulated a model with the objective of evaluating the probability and magnitude of nonadherence to speed limits in various work zone settings. The findings indicated that during the late afternoon and early morning periods, characterized by increased traffic volumes and a greater proportion of noncompliant vehicles in the vicinity, leaders of platoons with larger front gaps exhibit a greater propensity for speeding, both in terms of likelihood and magnitude. Light vehicles and their adherents demonstrated a greater inclination towards exceeding speed limits in comparison to other types of vehicles.

Numerous investigations have been undertaken by researchers to investigate the impact of truck accidents within work zones. Research has indicated that there is a notable propensity for work zone crashes involving trucks to exhibit a considerably higher degree of severity [6,7] The study conducted by Osman et al. [8] examined the correlation between the severity of injuries and different risk factors associated with significant truck accidents that took place in Minnesota between 2003 and 2012. Significant variations in the severity of injuries in truck-related crashes were observed, contingent upon the functional road classes of rural and urban areas. Bai et al. [9] utilized portable changeable message signs (PCMSs) in work zones to investigate the optimal placement of a PCMS within a work zone. They measured the alterations in the speed profiles of trucks and passenger cars to determine the effectiveness of different PCMS locations. The disparity in speed fluctuations between trucks and passenger cars has been identified as a significant contributing factor to work zone crashes that involve trucks.

The majority of these studies have employed statistical methods. Statistical models possess well-defined functional forms, albeit contingent upon several assumptions. Assumptions that underlie statistical models can lead to inaccurate or biased estimates.

In addition, the complex and multifaceted characteristics of modern datasets pose challenges when applying traditional statistical methods for modeling. In contrast, machine learning (ML) models exhibit a high degree of flexibility and require minimal reliance on assumptions. Currently, there is a growing interest in the application of ML across various fields [10–15]. Regardless of their high precision in forecasting, a significant drawback of numerous models based on ML is their inherent "black-box" nature, necessitating the use of post hoc explanation techniques for further interpretation. Multiple approaches exist for the interpretation of ML models, including partial dependency plot (PDP), Shapley additive explanations (SHAP), local interpretable model agnostic explanations (LIME), and permutation feature importance (PFI) [16–18]. The aforementioned approach offers valuable insights into model interpretations, yet it is crucial to exercise caution in its application to avoid drawing erroneous conclusions. Several potential pitfalls can arise when employing these interpretation techniques. These include misapplying the techniques in inappropriate contexts, interpreting models that lack generalization, disregarding feature dependencies and interactions, neglecting uncertainty estimates, and failing to address issues that arise in high-dimensional settings. Additionally, making unjustified causal interpretations is another common pitfall associated with these techniques.

In contrast, ML models known as "glass-box" models have been specifically built to hold inherent interpretability, thereby indicating that the explanations they produce are both reliable and readily comprehensible to users [19]. The explainable boosting machine (EBM) is a glass-box ML model that falls under the category of tree-based, cyclic gradient-boosting, general additive models. It allows for simple comprehension and interpretation of its internal mechanisms. The utilization of automatic interaction detection is implemented in accordance with the approach described by Nori et al. [20]. With regard to its reliability, EBM has demonstrated outcomes comparable to advanced ML models such as gradient boosting, support vector machine (SVM), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), artificial neural networks (ANNs), k-nearest neighbor (KNN), and random forest (RF) [21,22]. The utilization of factor-specific shape functions in this method holds promise for yielding results that are inherently interpretable. The potential benefits of EBM over complex black-box ML models are apparent in their capacity to offer interpretable decision-making processes and final estimates, which can be perceived in both local and global contexts without requiring additional interpretation methods. EBM is being successfully applied in various fields, including health sciences, computer and communications security, the education sector, and advanced manufacturing technology [23–25]. However, the application of the EBM model to the evaluation of the severity of road traffic accidents has not yet been investigated.

3. Data and Methods

In this study, we simultaneously address the issues of data imbalance and interpretability in ML models by employing different data augmentation techniques and utilizing the EBM model. The objective of our study is to predict the severity of crashes occurring in work zones and assess the influence of different risk factors. The hyperparameters of the EBM were optimized through the implementation of a Bayesian optimization strategy [26]. This approach facilitated the automated adaptation of the hyperparameters, obviating the necessity for manual intervention. To assess the efficacy of the proposed EBM model, a comparative analysis was conducted between its predictive outcomes and those generated by a black-box ML model such as combined kernel and boosting tree (KTBoost) [27], light gradient boosting machine (LightGBM) [28], and extreme gradient boosting (XGBoost) [29]. Furthermore, the EBM model was employed to analyze the factors within the global and local context. The complete research framework is illustrated in Figure 1.

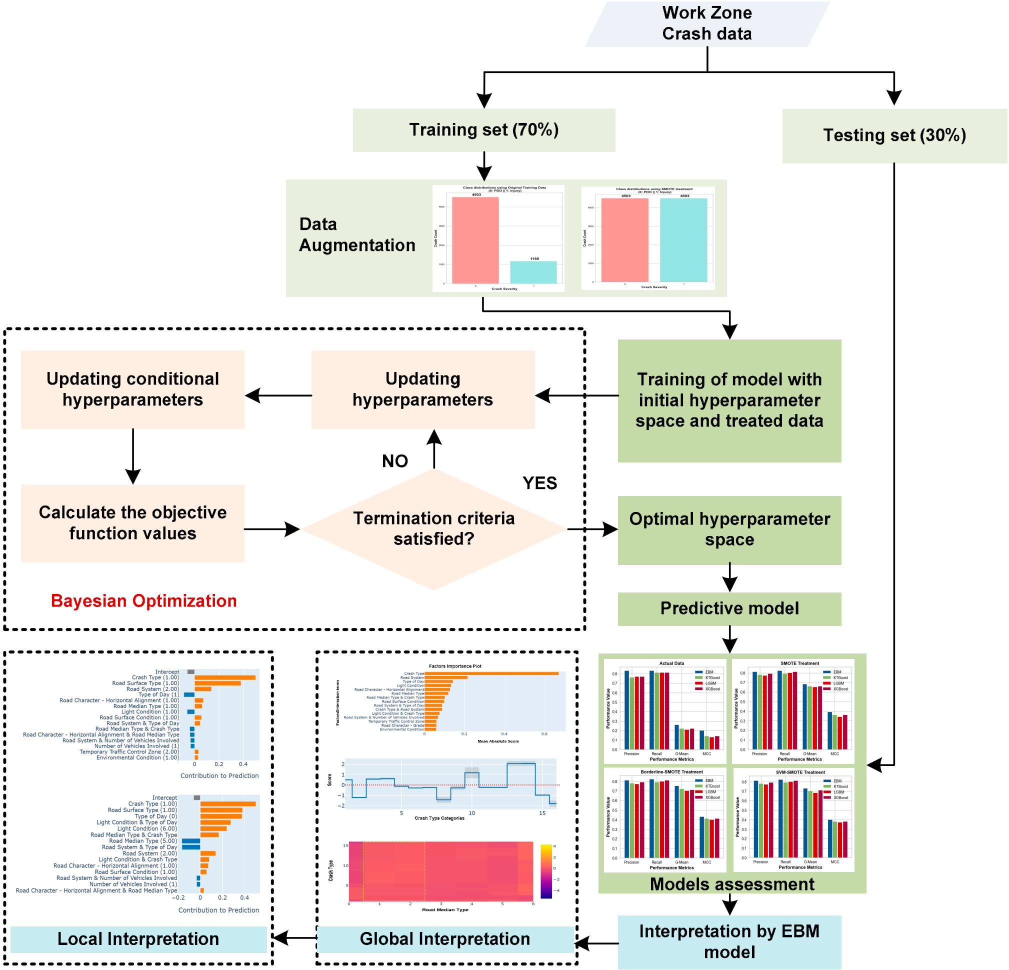

Work Zone Crash data
Training set (70%)
Testing set (30%)
Data Augmentation
Training of model with initial hyperparameter space and treated data
Updating hyperparameters
Updating conditional hyperparameters
Calculate the objective function values
Termination criteria satisfied?
NO
YES
Bayesian Optimization
Optimal hyperparameter space
Predictive model
Models assessment
Interpretation by EBM model
Global Interpretation
Local Interpretation

Figure 1. Proposed framework of the research.

3.1. Data Description

The data utilized in this study were sourced from the publicly accessible crash records database of the New Jersey Department of Transportation (NJDOT), which can be accessed through the NJDOT website. The dataset comprised a total of 8102 recorded incidents of vehicular crashes occurring within work zones in the state of New Jersey during the time frame stretching from 2017 to 2018. The crashes were categorized into two distinct classifications, specifically PDO (property damage only) and injury, according to the severity of the crashes. The injury class consisted of fatalities, severe injuries (or injuries causing debilitation), and minor injuries. Table 1 contains a listing of the crash risk factors utilized in this investigation, accompanied by their respective codes and relative frequencies.

Table 1. Coding and relative frequencies of work zone risk factors.

Risk Factors	Codes and Description
Road Character—Road Horizontal Alignment	0: 'Unknown (1.45%)', 1: 'Straight (90.45%)', 2: 'Curved Left (3.48%)', 3: 'Curved Right (4.62%)'
Road Character—Road Grade	0: 'Unknown (1.33%)', 1: 'Level (86.46%)', 2: 'Down Hill (3.33%)', 3: 'Up Hill (3.76%)', 4: 'Hill Crest (2.26%)', 5: 'Sag (2.86%)'
Road Surface Type	0: 'Asphalt (87.43%)', 1: 'Concrete (12.57%)'
Road Surface Condition	0: 'Dry (85.37%)', 1: 'Wet (8.64%)', 2: 'Snowy (3.54%)', 3: 'Icy (2.45%)'

Table 1. *Cont.*

Risk Factors	Codes and Description
Light Condition	0: 'Day light (78.47%)', 1: 'Dark (Spot Street Lights) (7.38%)', 2: 'Dark (No Street Light) (4.68%)', 3: 'Dark (Continuous Street Light) (3.07%)', 4: 'Dusk (2.87%)', 5: 'Dawn (1.95%)', 6: 'Dark (OFF Street Light) (1.58%)'
Environmental Condition	0: 'Clear (84.67%)', 1: 'Rain (5.94%)', 2: 'Overcast (4.77%)', 3: 'Snow (1.31%)', 4: 'Fog (1.27%)', 5: 'Snow (1.21%)', 6: 'Fog (0.83%)'
Road Median Type	0: 'Painted Median (18.72%)', 1: 'Barrier Median (38.62%)', 2: 'Curbed Median (12.41%)', 3: 'Grass Median (7.46%)', 4: 'Others (0.64%)', 5: 'None/Absence of Median (22.15%)'
Temporary Traffic Control Zone	0: 'Construction Zone (88.47%)', 1: 'Maintenance Zone (7.54%)', 2: 'Utility Zone (3.99%)'
Crash Type	0: 'Non-Fixed Object (4.56%); 1: 'Rear End-Same Direction (8.11%)', 2: 'Side Swipe-Same Direction (2.51%)', 3: 'Right Angle (11.68%)', 4: 'Head On-Opposite Direction (21.17%)', 5: 'Side Swipe-Opposite Direction (2.51%)', 6: 'Struck Parked Vehicle (1.87%)', 7: 'Left Turn/U Turn (0.12%)', 8: 'Backing (3.85%)', 9: 'Encroachment (0.11%)', 10: 'Overturn (8.61%)', 11: 'Fixed Object (1.15%)', 12: 'Animal (0.56%)', 13: 'Pedestrian (17.31%)', 14: 'Pedal cyclist (14.16%)', 15: 'Others (1.72%)'
Road System	0: 'Private Property (0.11%)', 1: 'Interstate (7.84%)', 2: 'State Highway (23.67%)', 3: 'State/Interstate Authority (17.53%)', 4: 'State Park or Institution (12.57%)', 5: 'County (10.16%)', 6: 'Co Auth, Park or Inst (12.96%)', 7: 'Municipal (8.05%)', 8: 'Mun Aith, Park or Inst (5.64%)', 9: 'US Govt Property 1.46%)'
Type of Day	0: 'Weekend (36.12%)', 1: 'Weekday (63.88%)'
Number of Vehicles Involved	0: 'Single Vehicle (22.38%)', 1: 'Multiple Vehicles (77.62%)'

3.2. Data Augmentation Strategies

In scenarios of work zone-related crashes, there exists an imbalance in the crash records, whereby the number of instances for each class (property damage only and injury) is not equitably represented. The aforementioned disparity poses a significant detriment to both ML and statistical classification algorithms, resulting in a substantial decline in accuracy. There are several primary factors contributing to this phenomenon. (1) In case of imbalanced data, ML and statistical classification algorithms disregard minority class instances as extraneous data points and produce a rudimentary classifier that estimates all samples as belonging to the majority class. (2) They often exhibit bias towards a class that is numerically dominant due to their optimization objective of maximizing classification accuracy. This approach treats classification errors for all classes equally, which is not suitable for imbalanced datasets.

In this study, various oversampling-based data augmentation techniques were utilized to address the issue of class imbalance in work zone crashes. These techniques included SMOTE (synthetic minority over-sampling technique) [30], borderline-SMOTE [31], and SVM-SMOTE [32]. Those with a keen interest in obtaining comprehensive information regarding these strategies are suggested to refer to the corresponding reference for further details.

3.3. Explainable Boosting Machine: A Glass-Box ML Model

The EBM model is built upon generalized additive models (GAMs), which are widely recognized as the benchmark to demonstrate a high level of comprehensibility. Given that $\Delta = (x_r, y_r)$ is a training dataset with a length of R, the input vectors that $(x_1, x_2, \ldots, x_R)$ with 'ϕ' attributes, and y_r is the target factor, then the GAMs takes the form as shown in Equation (1).

$$\Theta(E[y]) = \beta_o + \sum \Gamma_r(x_r) \tag{1}$$

where x_r represents the r^{th} factor within the attribute set, while Θ denotes the link function that aligns the generalized additive model (GAM) with either regression (e.g., Θ = identity) or classification (e.g., Θ = logistic), and Γ_r refers to the attribute function.

When compared to conventional GAMs, EBM, which employs bagging and gradient boosting, offers a number of important improvements. The training process involves focusing on individual attributes sequentially, utilizing a significantly small learning rate. Round-robin boosting is employed to disregard the order of the attributes. In order to mitigate the effects of collinearity, EBM iterates through the attributes, aiming to identify the most influential attribute function Γ_r for each attribute. Subsequently, it incorporates the information from each attribute into the prediction process. Specifically, each function Γ_r is utilized as a lookup table in which the term contribution is added and transmitted via the link function Θ to generate individual predictions. The feasibility of determining the attribute with the greatest influence on an individual prediction can be attributed to the concepts of additivity and modularity. This allows for the ordering and visualization of contributions. One additional benefit of the EBM is its capacity to enhance precision through the integration of pairwise interactions into conventional GAMs, resulting in the formation of GA2Ms, as represented by Equation (2).

$$\Theta(E[y]) = \beta_o + \sum \Gamma_r(x_r) + \sum \Gamma_{rs}(x_r, x_s) \tag{2}$$

Here, the representation of the 2D interactions $\Gamma_{rs}(x_r, x_s)$ can be visualized as a heatmap on the 2D $x_r - x_s$ plane, maintaining a notably high level of comprehensibility. The GA2M algorithm first constructs an optimal GAM, and then examines the residuals in order to identify and rank all potential interaction combinations based on their significance.

3.4. Hyperparameter Tuning of ML Models

The optimization of hyperparameters is a crucial step that must be undertaken prior to the learning of ML models in order to minimize over-fitting and reduce the model's complexity. Various approaches to hyperparameter tuning have been extensively examined that encompass grid search cross-validation (GS-CV), random search cross-validation (RS-CV), and the Bayesian optimization approach [33,34]. However, it has been observed that GS-CV and RS-CV are methods that systematically investigate the complete spectrum of potential hyperparameter values without considering previous outcomes, which enhances the computation time to reach the optimal values. In contrast, Bayesian optimization, in the choice of subsequent hyperparameters, takes into account previous evaluations. This approach enables the determination of the most suitable hyperparameter values while minimizing the number of iterations required [33]. Therefore, we employed Bayesian optimization for hyperparameter tuning of EBM and other models in our research. We employed the G-mean performance metric to facilitate hyperparameter adjustment.

3.5. Performance Evaluation of ML Models

The evaluation of the proposed EBM and other black-box ML models can be conducted using various metrics that are typically derived from the contingency or confusion matrix of the model, as illustrated in Figure 2. In the confusion matrix, the true positives are defined as the instances in which a classifier correctly forecasts the positive class for a given set of outcomes. Similarly, true negatives refer to the outcomes in which the classifier accurately predicts the absence of a specific class. When a predictive model erroneously classifies an instance as belonging to the positive class when it actually does not, it is referred to as a false positive (∇^{ρ}). Similarly, false negatives refer to the instances in which the classification model produces an erroneous prediction concerning the negative class. In order to calculate the classification accuracy (CA), a commonly employed metric to assess the performance of ML models, we divide the overall count of accurate predictions by the total count of estimations produced. The utilization of this metric may result in inaccurate findings when applied to imbalanced datasets, as it assigns greater weight to the class that is more prevalent. In the given circumstances, the utilization of classification accuracy as an efficiency metric would not be feasible. In order to address this issue, several performance metrics are employed alongside accuracy, such as precision (P), recall (R), Matthews' correlation coefficient (MCC), and geometric mean (G-mean). The calculations for each metric can be found in Equations (3) through (7). The G-mean is a metric used to evaluate the efficiency of ML models. The evaluation measures the balance between the accuracy of classifying minority and majority instances. The utilization of the MCC can provide valuable insights into the effectiveness of a classification algorithm that has been trained on imbalanced data. Its value should ideally lie within the interval of −1 to 1. Values that approach +1 indicate an elevated degree of performance, while values that deviate from +1 indicate a lower level of performance.

$$\mathrm{CA} = \frac{\Delta^{\rho}+\Delta^{\eta}}{\Delta^{\rho}+\Delta^{\eta} + \nabla^{\eta} + \nabla^{\rho}} \tag{3}$$

$$\mathrm{P} = \frac{\Delta^{\rho}}{\Delta^{\rho} + \nabla^{\rho}} \tag{4}$$

$$\mathrm{R} = \frac{\Delta^{\rho}}{\Delta^{\rho} + \nabla^{\eta}} \tag{5}$$

$$\text{G-Mean} = \sqrt{\left(\frac{\Delta^{\rho}}{\Delta^{\rho} + \nabla^{\eta}}\right)\left(\frac{\Delta^{\eta}}{\nabla^{\rho}+\Delta^{\eta}}\right)} \tag{6}$$

$$\mathrm{MCC} = \frac{\Delta^{\rho}\ ?\ \Delta^{\eta} - \nabla^{\rho}\ ?\ \nabla^{\eta}}{\sqrt{(\Delta^{\rho} + \nabla^{\rho})(\Delta^{\rho} + \nabla^{\eta})\left(\Delta^{\eta} + \nabla^{\rho}\right)\left(\Delta^{\eta} + \nabla^{\eta}\right)}} \tag{7}$$

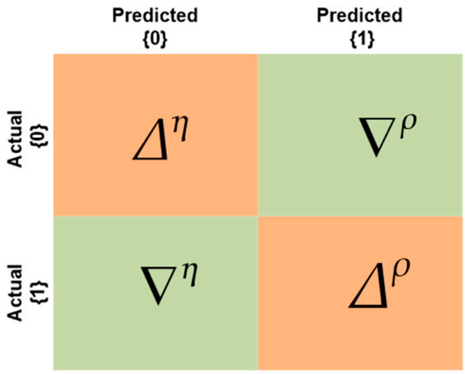


Figure 2. Confusion matrix plot for classification.

4. Results and Discussion

The NJDOT work zone crash records utilized in this research comprise a total of 8102 work zone crashes that occurred within the time frame of 1 January 2017 to 31 December 2018. The dataset consisted of various outcomes, which encompassed 10 instances of fatalities, constituting approximately 0.1% of the overall count. Additionally, there were 1609 incidents of crashes resulting in nonfatal injuries, accounting for approximately 20% of the total count. The remaining 6473 crashes were categorized as property damage only (PDO) events. The low number of observations for fatal, major, and minor crashes necessitated the combination of fatal, major, and minor crashes with injury crashes for the estimation of the model. The idea of binary classification into injury and PDO classes has been taken from similar studies [35–37].

The complete dataset, consisting of 8102 work zone crash records, was divided into two subsets in order to facilitate model training and validation, as well as for the purpose of evaluation. The training–validation set comprises 70% of the complete dataset, consisting of 5671 reports. This set includes 4503 cases of PDO (property damage only) and 1168 cases of injuries, with a balancing ratio of 0.25. The remaining 30% of the data, containing 2433 crash records, was set aside for testing purposes.

4.1. Data Treatment and Hyperparameter Tuning

Data-level augmentation techniques were implemented on the training data before applying the EBM model and conducting hyperparameter tuning. After implementing multiple data augmentation techniques, the training datasets that have undergone resampling are depicted in Figure 3. Subsequently, Bayesian optimization is employed to ascertain the optimal values for the hyperparameters of ML models. Utilizing both the original and the augmented data, the goal was to maximize the G-mean metric. The optimal hyperparameters for EBM and other competitive glass-box and black-box ML models with different data augmentation strategies are demonstrated in Table A1 in Appendix A.

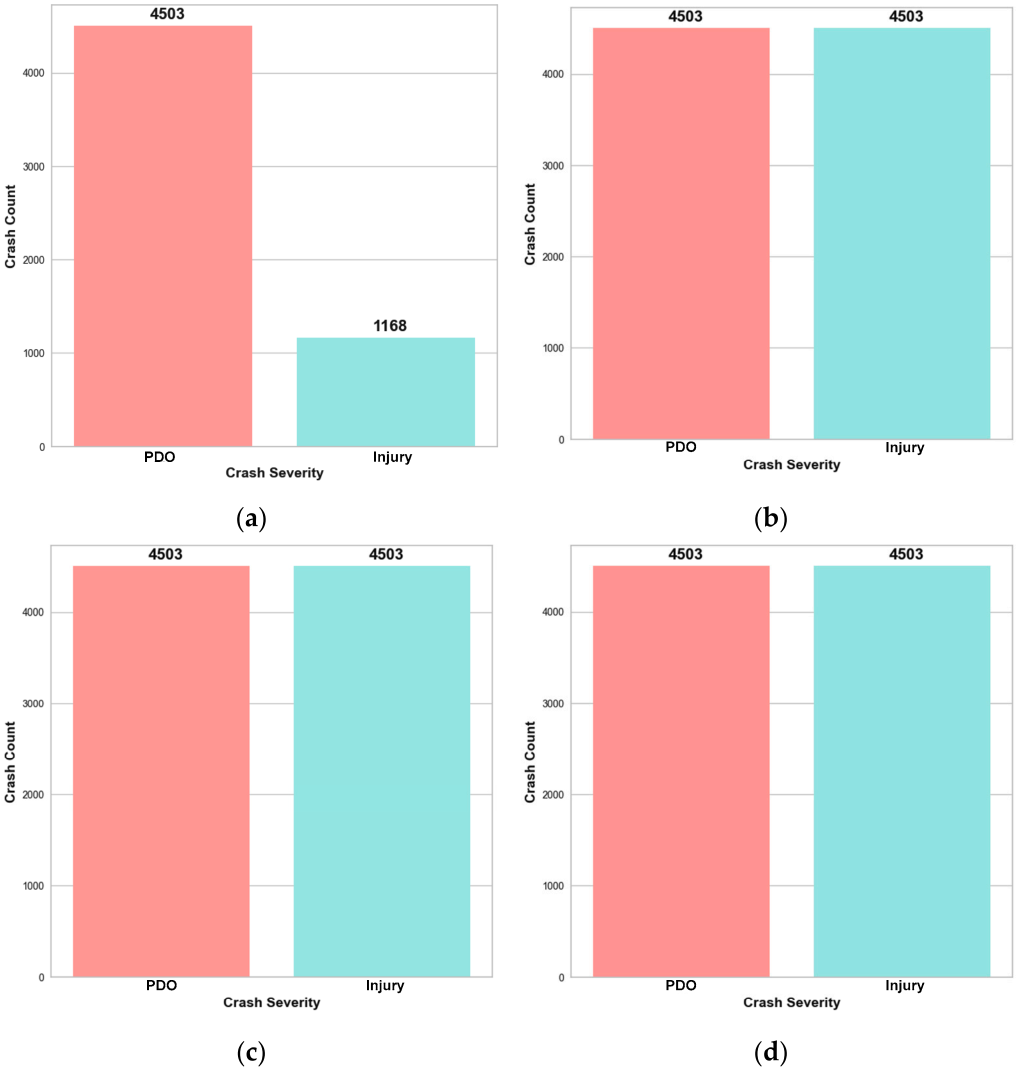


Figure 3. Resampling by different data augmentation strategies; (**a**) original data; (**b**) SMOTE treatment; (**c**) borderline-SMOTE treatment; (**d**) SVM-SMOTE treatment.

4.2. Performance Comparison of ML Models

In this study, injury and PDO were designated as the positive and negative classes, respectively, for the purpose of developing models and assessments. The efficiency of ML models was assessed by analyzing various performance metrics, including recall value, precision value, G-mean, and MCC. In the initial analysis of the untreated data (Figure 4a), it was observed that the EBM model exhibited a G-mean of 0.26 and an MCC of 0.20. These values were found to be slightly higher compared to the other black-box ML models. However, the values of these indicators are quite low due to the imbalanced nature of the crash data. Consequently, we are compelled to utilize a balanced dataset. Following the implementation of various data augmentation techniques, it has been observed that the G-mean and MCC values have exhibited an increase across all models. The application of the SMOTE to the data led to improved performance metrics in the EBM model, specifically a higher G-mean of 0.68 and a higher MCC of 0.39. The KTBoost and XGBoost models also demonstrated improved performance, albeit to a lesser extent, with a G-mean of 0.66 and an MCC of 0.36 (Figure 4b). When considering data that have been treated with borderline-SMOTE, the EBM demonstrated superior performance in terms of G-mean (0.75) and MCC (0.43). Subsequently, the KTBoost yielded a G-mean value of 0.72 and an MCC value of 0.41 (Figure 4c). In the data that underwent the SVM-SMOTE treatment, EBM exhibited improved performance in terms of G-mean (0.73) and MCC (0.40), followed by XGBoost with G-mean (0.71) and MCC (0.38) (Figure 4d). Based on these outcomes, the EBM model with borderline-SMOTE-treated data can be used for further interpretation of factors from both global and local perspectives.

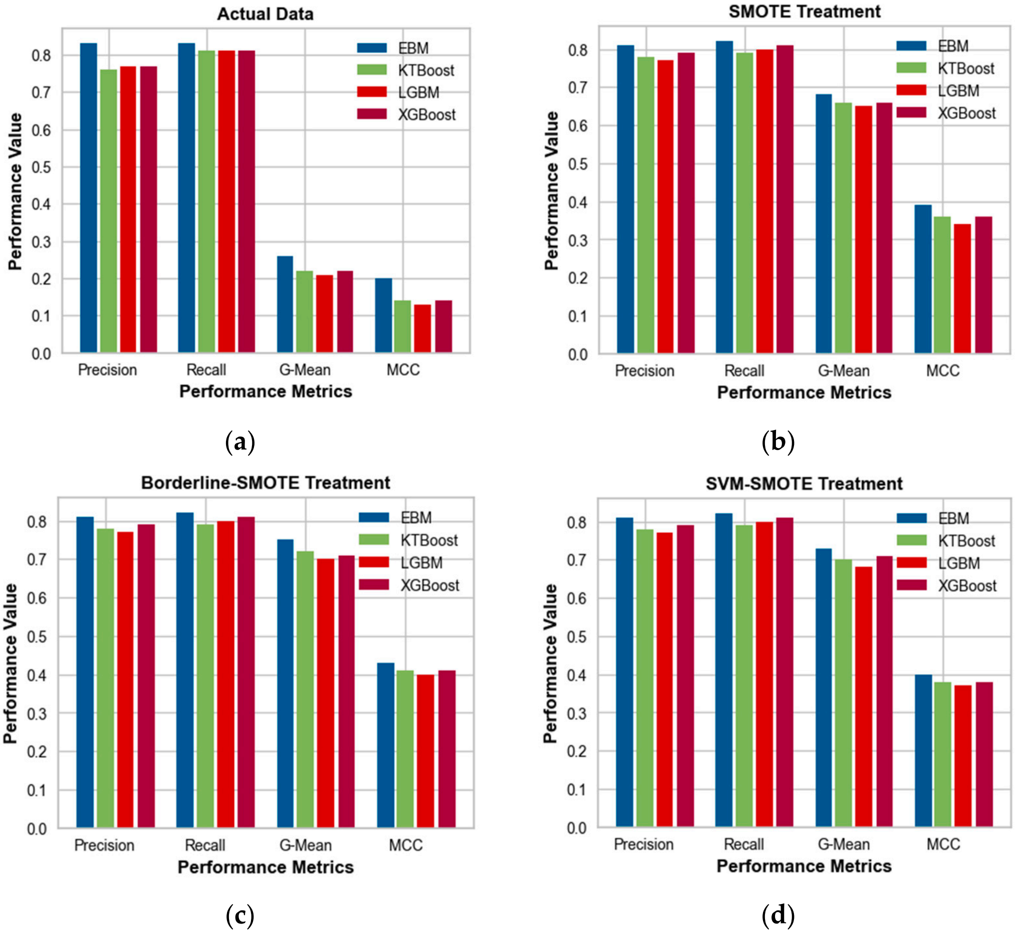


Figure 4. Performance measures of different models using various data augmentation strategies; (**a**) Original Data; (**b**) SMOTE treatment; (**c**) borderline-SMOTE treatment; (**d**) SVM-SMOTE treatment.

4.3. EBM-Based Factor Importance and Pairwise Interaction

In addition to prediction performance, this study also provides a comprehensive analysis of the EBM model, incorporating borderline-SMOTE-treated data. This strategy facilitates an in-depth investigation of the impact of factors and combinations of factors on the severity of work zone-related crashes. Figure 5 presents a comprehensive overview of the importance of each individual factor, as well as their interactions in pairs. When considering individual factors, it becomes evident that both the crash type and the road system play crucial roles in influencing the probability of injuries in work zone crashes. However, in terms of pairwise interaction, the combined effect of road median type and crash type significantly increased the probability of injuries occurring in work zone-related crashes, followed by the combination of road system and day of the week.

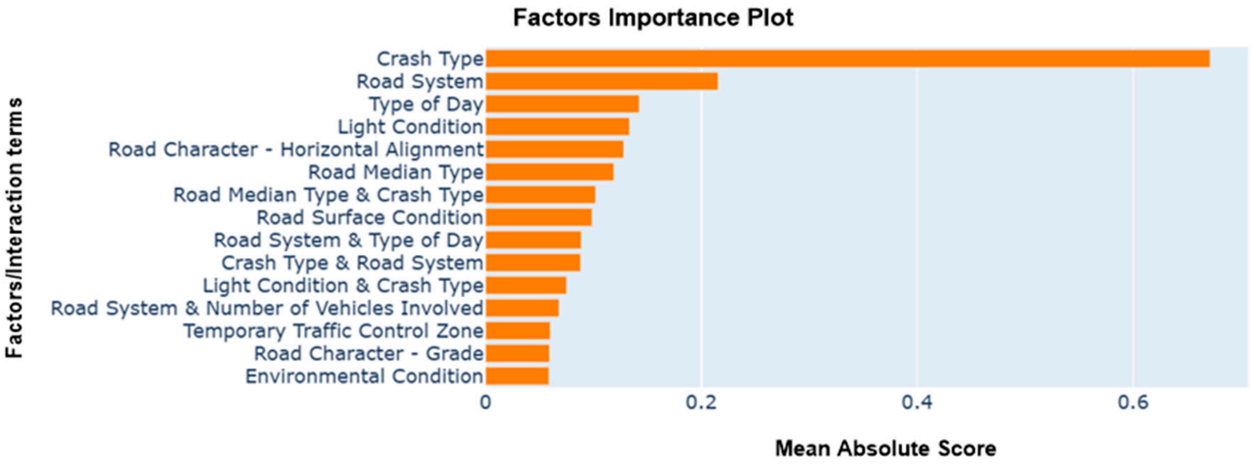


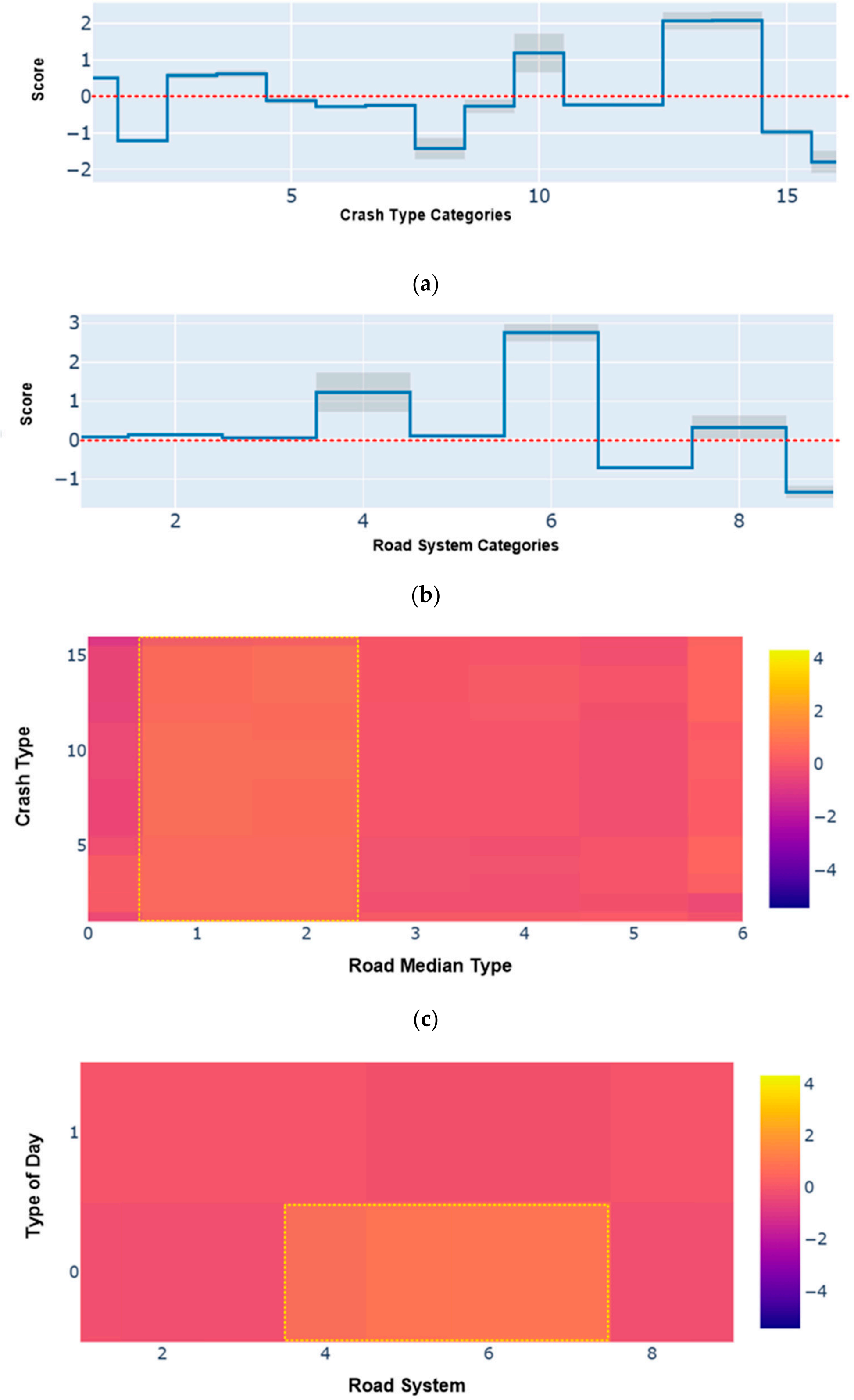
Figure 5. Importance of factors/pairwise interaction terms.

Additional outputs generated by the proposed EBM model using borderline-SMOTE-treated data for comprehension of the global outcomes regarding the severity of work zone-related crashes are illustrated in Figure 6, encompassed EBM shape functions and heatmaps. The shape function of the EBM is a unidimensional function that represents the relationship between predictor factor values and the corresponding scores. The predictor factor values are plotted along the horizontal axis, while the scores, which indicate the influence of the independent factors on the predicted logits, are plotted along the vertical axis. A score greater than zero signifies a statistically significant relationship between the independent factors and the response factor.

Figure 6a,b presents the shape functions derived from the EBM approach for the two prominent factors, namely crash type and road system. The crash types, coded as 3 (Right Angle), 4 (Head On Opposite Direction), 10 (Overturn), 13 (Pedestrian), and 14 (Pedal cyclist), exhibited score values exceeding zero, indicating a higher propensity for causing injuries in work zone crashes. Similarly, the road systems coded as 2 (State Highway), 4 (State Park or Institution), 6 (Co Auth, Park or Inst) and 8 (Mun Aith, Park or Inst) were more likely to cause injuries. Figure 6c,d depicts the heatmaps of pairwise interaction of importance pairs of factors. The area in yellow shows the higher score value, which means likelihood of injuries. It has been observed that in all types of crashes, the median type coded as 1 (Barrier Median) and 2 (Curbed Median) contributed more towards the injuries in work zone crashes. Similarly, on weekends, the road system type coded as 4 (State Park or Institution), 5 (County), 6 (Co Auth, Park or Inst), and 7 ('Municipal') were highly prone to the occurrence of injuries in work zones.

4.4. EBM-Based Local Factor Interpretation

The EBM-based local explanation comprises explanations that pertain to specific estimations for particular instances. A bar graph can be generated for each occurrence, visually representing a consistent intercept term in gray color; additive terms that contribute positively are displayed in orange, while additive terms that have a negative impact are represented in blue. It is important to note that the strength of factors in local interpretation may differ from that of global interpretation. Global interpretation encompasses the overall impact of instances, while local interpretation focuses on specific instances within the data.



(**a**)

(**b**)

(**c**)

(**d**)

Figure 6. EBM-based global interpretation (**a**,**b**) and EBM-based shape functions; (**c**,**d**) EBM-based pairwise interaction (heatmap).

Here, we examine a pair of randomly selected instances that have been accurately predicted and classified as work zone crash injuries. In the initial instance, a case is accurately categorized as "injury" with a probability of 0.87 (refer to Figure 7a). In this specific instance, the first three factors, namely crash type (1: Rear End Same Direction), road surface type (1: Concrete), and road system (2: State Highway), displayed a positive influence. Conversely, the type of day (1: Weekday) had a negative impact, which resulted in PDO. Figure 7b illustrates another scenario wherein an event is correctly classified as "injury" with a probability of 0.91. In this particular case, the crash type (1: Rear End, Same Direction), road surface type (1: Concrete), and type of day (0: Weekend) exhibited a significant positive impact, indicating an increased probability of causing injuries. On the other hand, the road median type (5: None or Absence of Median) and the interaction between the road system and the type of day had a negative influence, with an increase in the probability of PDO.

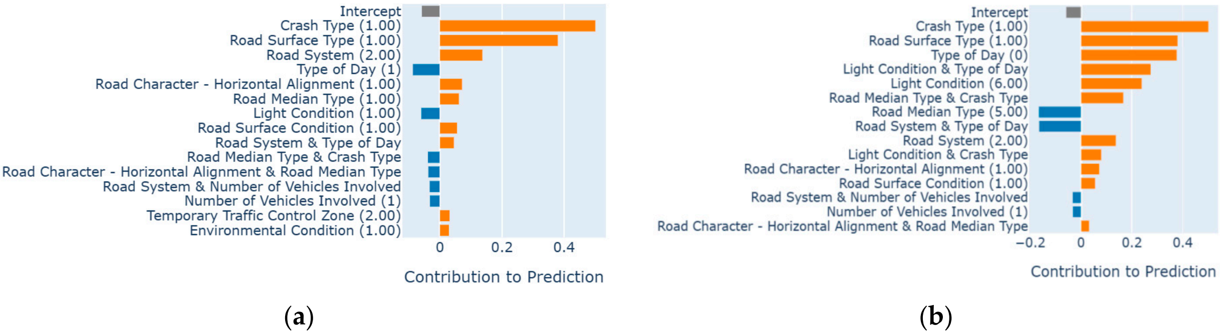


Figure 7. Local interpretation of two correctly predicted instance as "Injury"; (**a**) instance # 13 correctly classified "Injury" with probability 0.87 (**b**); instance # 22 correctly classified "Injury" with probability 0.91.

5. Conclusions and Recommendations

This research introduces a glass-box ML model known as EBM, which is utilized for predicting the severity of crashes occurring in work zones. It demonstrates a level of reliability that is comparable to that of other black-box ML models, while still maintaining the ability to be explained and understood in an intuitive manner. The present study employed a dataset comprising work zone crashes occurring on highways under the jurisdiction of the state of New Jersey between the years 2017 and 2018. The problem of imbalanced crash data was also mitigated by employing data augmentation techniques, including SMOTE, borderline-SMOTE, and SVM-SMOTE, with the objective of addressing the issue of imbalance. The EBM model was trained with Bayesian optimization using the processed data. The evaluation of the EBM model's performance was conducted by utilizing holdout data, and subsequently comparing it to black-box ML models including KTBoost, LightGBM, and XGBoost. The performance of the EBM model and other black-box ML models differed slightly, but was comparable. The finely tuned EBM model using borderline-SMOTE-treated data outperformed other black-box ML models on the testing dataset, achieving higher precision (81.37%), recall (82.53%), G-mean (75.39%), and MCC (0.43).

Subsequently, the EBM model exhibited effectiveness in the interpretation of data from both a global and local standpoint. The factor of crash type was found to be of utmost importance in determining the injuries sustained in work zone-related crashes, with road system type ranking second in significance. In the context of pairwise interaction, it was observed that the combination of road median type and crash type had the greatest impact on work zone-related crashes. The findings of this study revealed that a significant proportion of injuries in work zone-related accidents can be ascribed to distinct categories of crashes, including rear-end collisions in the same direction, right-angle collisions, head-on collisions in the opposite direction, vehicle overturning, and incidents involving pedestrians and bicyclists. Likewise, there is a higher likelihood of work zones on New Jersey state highways with concrete surfaces leading to injuries. The EBM model facilitates users in understanding the categorization of novel cases and the methodology involved in generating predictions. This fosters a greater sense of confidence in the transparent nature of the glass-box model in contrast to the black-box model.

The outcomes of this research will yield significant perspectives for the analysis of crashes associated with work zones and policymakers within the realm of traffic safety. Further research could be conducted by employing alternative sophisticated glass-box ML models. In the same way, future research efforts could be expanded by incorporating additional data augmentation techniques to address the issue of imbalanced work zone crash data. This study utilizes a dataset encompassing a two-year period from 2017 to 2018. Future studies will incorporate additional work zone crash data from the State of New Jersey and other regions.

Author Contributions: Conceptualization, R.A.; data curation, H.A.; formal analysis, R.A. and H.A.; funding acquisition, M.B.; investigation, A.E.A.E. and M.A.; project administration, A.E.A.E.; resources, H.A.; software, S.A. and M.A.; validation, M.B. and M.A.; visualization, S.A. and R.A.; writing—review and editing, A.E.A.E. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data was created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The researcher would like to acknowledge the Deanship of Scientific Research, Taif University, for funding this work.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Optimal hyperparameter values of different ML models using different data augmentation strategies.

Model	Hyperparameters	Range	Optimal Values			
			Original Data	SMOTE Treatment	Borderline-SMOTE Treatment	SVM-SMOTE Treatment
EBM	max_bins	(100, 500)	135	205	190	220
	learning_rate	(0.01, 0.2)	0.11	0.18	0.09	0.12
	max_leaves	(3, 12)	3	6	6	4
	min_samples_leaf	(1, 5)	3	3	2	3
KTBoost	number of trees	(100–3000)	300	350	550	250
	max_depth	(1–10)	2	3	2	4
	learning_rate	(0.01, 0.2)	0.14	0.13	0.11	0.17
XGBoost	n_estimators	(50–2000)	150	110	235	210
	learning_rate	(0.01, 0.2)	0.09	0.11	0.07	0.08
LightGBM	n_estimators	(50–2000)	220	90	165	110
	learning_rate	(0.01, 0.2)	0.15	0.08	0.13	0.10

References

1. FHWA Work Zone Facts and Statistics. Available online: https://ops.fhwa.dot.gov/wz/resources/facts_stats.htm#ftn2 (accessed on 15 September 2023).
2. Theofilatos, A.; Ziakopoulos, A.; Papadimitriou, E.; Yannis, G.; Diamandouros, K. Meta-analysis of the effect of road work zones on crash occurrence. *Accid. Anal. Prev.* **2017**, *108*, 1–8. [CrossRef] [PubMed]

3. Morgan, J.; Duley, A.; Hancock, P. Driver responses to differing urban work zone configurations. *Accid. Anal. Prev.* **2010**, *42*, 978–985. [CrossRef] [PubMed]
4. Weng, J.; Xue, S.; Yang, Y.; Yan, X.; Qu, X. In-depth analysis of drivers' merging behavior and rear-end crash risks in work zone merging areas. *Accid. Anal. Prev.* **2015**, *77*, 51–61. [CrossRef] [PubMed]
5. Debnath, A.K.; Blackman, R.; Haworth, N. A Tobit model for analyzing speed limit compliance in work zones. *Saf. Sci.* **2014**, *70*, 367–377. [CrossRef]
6. Khattak, A.J.; Targa, F. Injury severity and total harm in truck-involved work zone crashes. *Transp. Res. Rec.* **2004**, *1877*, 106–116. [CrossRef]
7. Weng, J.; Du, G.; Ma, L. Driver injury severity analysis for two work zone types. In *Proceedings of Institution of Civil Engineers-Transport*; Thomas Telford Ltd.: London, UK, 2016; pp. 97–106.
8. Osman, M.; Paleti, R.; Mishra, S.; Golias, M.M. Analysis of injury severity of large truck crashes in work zones. *Accid. Anal. Prev.* **2016**, *97*, 261–273. [CrossRef]
9. Bai, Y.; Yang, Y.; Li, Y. Determining the effective location of a portable changeable message sign on reducing the risk of truck-related crashes in work zones. *Accid. Anal. Prev.* **2015**, *83*, 197–202. [CrossRef]
10. Dong, S.; Khattak, A.; Ullah, I.; Zhou, J.; Hussain, A. Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHAPley Additive exPlanations. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2925. [CrossRef]
11. Jiang, L.; Xie, Y.; Ren, T. Modelling highly unbalanced crash injury severity data by ensemble methods and global sensitivity analysis. In Proceedings of the Transportation Research Board 98th Annual Meeting, Washington, DC, USA, 13–17 January 2019; pp. 13–17.
12. Khattak, A.; Almujibah, H.; Elamary, A.; Matara, C.M. Interpretable Dynamic Ensemble Selection Approach for the Prediction of Road Traffic Injury Severity: A Case Study of Pakistan's National Highway N-5. *Sustainability* **2022**, *14*, 12340. [CrossRef]
13. Mafi, S.; AbdelRazig, Y.; Doczy, R. Machine learning methods to analyze injury severity of drivers from different age and gender groups. *Transp. Res. Rec.* **2018**, *2672*, 171–183. [CrossRef]
14. Parsa, A.B.; Movahedi, A.; Taghipour, H.; Derrible, S.; Mohammadian, A.K. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accid. Anal. Prev.* **2020**, *136*, 105405. [CrossRef] [PubMed]
15. Zhang, S.; Khattak, A.; Matara, C.M.; Hussain, A.; Farooq, A. Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents. *PLoS ONE* **2022**, *17*, e0262941. [CrossRef] [PubMed]
16. Du, M.; Liu, N.; Hu, X. Techniques for interpretable machine learning. *Commun. ACM* **2019**, *63*, 68–77. [CrossRef]
17. Kenny, E.M.; Ford, C.; Quinn, M.; Keane, M.T. Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies. *Artif. Intell.* **2021**, *294*, 103459. [CrossRef]
18. Murdoch, W.J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Interpretable machine learning: Definitions, methods, and applications. *arXiv* **2019**, arXiv:1901.04592. [CrossRef] [PubMed]
19. Rai, A. Explainable AI: From black box to glass box. *J. Acad. Mark. Sci.* **2020**, *48*, 137–141. [CrossRef]
20. Nori, H.; Jenkins, S.; Koch, P.; Caruana, R. Interpretml: A unified framework for machine learning interpretability. *arXiv* **2019**, arXiv:1909.09223.
21. Khattak, A.; Chan, P.-W.; Chen, F.; Peng, H. Assessing wind field characteristics along the airport runway glide slope: An explainable boosting machine-assisted wind tunnel study. *Sci. Rep.* **2023**, *13*, 10939. [CrossRef]
22. Maxwell, A.E.; Sharma, M.; Donaldson, K.A. Explainable boosting machines for slope failure spatial predictive modeling. *Remote Sens.* **2021**, *13*, 4991. [CrossRef]
23. El-Mihoub, T.A.; Nolle, L.; Stahl, F. Explainable Boosting Machines for Network Intrusion Detection with Features Reduction. In Proceedings of the International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cambridge, UK, 13–15 December 2022; pp. 280–294.
24. Sarica, A.; Quattrone, A.; Quattrone, A. Explainable boosting machine for predicting Alzheimer's disease from MRI hippocampal subfields. In Proceedings of the International Conference on Brain Informatics, Virtual, 17–19 September 2021; pp. 341–350.
25. Xiaolin, L.; Qingyuan, W.; Panicker, R.C.; Cardiff, B.; John, D. Binary ECG Classification Using Explainable Boosting Machines for IoT Edge Devices. In Proceedings of the 2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Glasgow, UK, 24–26 October 2022; pp. 1–4.
26. Khattak, A.; Chan, P.-W.; Chen, F.; Peng, H. Prediction and Interpretation of Low-Level Wind Shear Criticality Based on Its Altitude above Runway Level: Application of Bayesian Optimization–Ensemble Learning Classifiers and SHapley Additive exPlanations. *Atmosphere* **2022**, *13*, 2102. [CrossRef]
27. Sigrist, F. KTBoost: Combined kernel and tree boosting. *Neural Process. Lett.* **2021**, *53*, 1147–1160. [CrossRef]
28. Li, F.; Zhang, L.; Chen, B.; Gao, D.; Cheng, Y.; Zhang, X.; Yang, Y.; Gao, K.; Huang, Z.; Peng, J. A light gradient boosting machine for remainning useful life estimation of aircraft engines. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3562–3567.
29. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
30. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]

31. Han, H.; Wang, W.-Y.; Mao, B.-H. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *Proceedings of the International Conference on Intelligent Computing*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 878–887.
32. Tang, Y.; Zhang, Y.-Q.; Chawla, N.V.; Krasser, S. SVMs modeling for highly imbalanced classification. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2008**, *39*, 281–288. [CrossRef] [PubMed]
33. Kadam, V.J.; Jadhav, S.M. Performance analysis of hyperparameter optimization methods for ensemble learning with small and medium sized medical datasets. *J. Discret. Math. Sci. Cryptogr.* **2020**, *23*, 115–123. [CrossRef]
34. Shekar, B.; Dagnew, G. Grid search-based hyperparameter tuning and classification of microarray cancer data. In Proceedings of the 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Sikkim, India, 25–28 February 2019; pp. 1–8.
35. Moomen, M.; Rezapour, M.; Raja, M.N.; Ksaibati, K. Predicting injury severity and crash frequency: Insights into the impacts of geometric variables on downgrade crashes in Wyoming. *J. Traffic Transp. Eng. (Engl. Ed.)* **2020**, *7*, 375–383. [CrossRef]
36. Xu, C.; Liu, P.; Wang, W.; Zhang, Y. Real-time identification of traffic conditions prone to injury and non-injury crashes on freeways using genetic programming. *J. Adv. Transp.* **2016**, *50*, 701–716. [CrossRef]
37. Wei, X.; Shu, X.; Huang, B.; Taylor, E.L.; Chen, H. Analyzing traffic crash severity in work zones under different light conditions. *J. Adv. Transp.* **2017**, *2017*, 5783696. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.