Open Access

Comparative Performance of Logistic Regression, XGBoost, and Multilayer Perceptron for Mortality and Treatment Strategy Prediction in Blunt Thoracic Aortic Injury

PHILLIP D. JENKINS, MD[1*], SHELBY WILLIS, MD[1], VICTOR ANDUJO, MD[1], TIFFANY LIAN, MD[1], RUCHI THANAWALA, MD, MS[1], CASTIGLIANO BHAMIDIPATI, DO, PHD[1], JUSTIN REGNER, MD[1], JULIE DOBERNE, MD PHD[1], MICHAEL R. KOLESNIKOV, MSN, PHD[1]

[1] *Oregon Health & Science University, Department of Surgery, Portland, OR, USA*
[*] *Corresponding author: jenkinph@ohsu.edu*

Manuscript Received: 27 January 2026
Manuscript Accepted: 22 February 2026
Date of Publication: 2 March 2026
This article is available from: https://pnwjs.org/index.php/journal/article/view/15
doi: https://doi.org/10.70422/n2panf23

INTRODUCTION: Blunt thoracic aortic injury (BTAI) is a rare but lethal trauma mechanism. Mortality prediction has traditionally relied on logistic regression (LR), yet machine learning (ML) methods may better capture complex nonlinear relationships.

METHODS: Using the Aortic Trauma Foundation registry, we compared LR, extreme gradient boosting (XGB), and multilayer perceptron (MLP) for two tasks: (1) in-hospital mortality prediction and (2) classification of management strategy, open surgical repair (OSR), thoracic endovascular aortic repair (TEVAR), or medical management alone (MMA). Explanatory LR models identified independent predictors; predictive models targeted ~80% sensitivity in test sets. Performance metrics included area under the receiver operating characteristic curve (AUC), average precision, accuracy, F1 score, precision, recall, and Brier score.

RESULTS: Explanatory LR for mortality achieved excellent discrimination (AUC = 0.961), with higher Injury Severity Score (ISS), age, and ICU length of stay predicting increased mortality, and conversely, higher Glasgow Coma Scale, pH, and hospital stay protective. Predictive LR and XGB performed similarly (AUCs = 0.849 and 0.855; Brier = 0.083 and 0.101), both outperforming MLP (AUC = 0.757). For management strategy, explanatory LR identified limited predictors (e.g., ISS for OSR; age, heart rate, platelet count, and ISS for TEVAR). Predictive performance was modest: XGB (accuracy = 0.690; macro-AUC = 0.648) slightly outperformed MLP (accuracy = 0.667; macro-AUC = 0.565).

CONCLUSION: Well-specified LR matched XGB for mortality prediction in BTAI, offering interpretability with high accuracy. Management strategy prediction remained limited across models, likely reflecting unmeasured institutional and procedural factors. Enriching datasets with imaging and practice-pattern variables may improve operative decision modeling.

Keywords: *Aortic Injury, Blunt Thoracic Aortic Injury, Logistic Regression, Machine Learning, Multilayer Perceptron, Trauma Surgery, Predictive Model, XGBoost.*

INTRODUCTION

Blunt thoracic aortic injury (BTAI) is an uncommon but devastating consequence of high-energy blunt trauma, accounting for approximately 1% of trauma admissions yet remaining the second most lethal injury mechanism after traumatic brain injury.[1] Mortality rates approach 80% in the prehospital setting, and in-hospital mortality can range from 20–46% depending on injury severity and associated trauma burden.[2] Advances in prehospital care, imaging, and endovascular repair have improved survival;[3,4] however, the optimal timing and selection of management strategies, including medical therapy, urgent TEVAR, delayed TEVAR, and open surgical repair (OSR), remain the subject of ongoing debate.[5,6]

Risk prediction in BTAI is challenging due to its rarity, heterogeneous presentation, and the confounding influence of associated injuries.[5,7] Traditional risk factors such as age, hemodynamic instability, aortic injury grade, and Injury Severity Score (ISS)[8] are well-established but may not fully capture complex nonlinear interactions between physiologic and injury variables.[9] Logistic regression has long been the mainstay of clinical prediction modeling due to its interpretability,[10] but machine learning (ML) methods offer the potential to uncover complex relationships and improve discrimination.[11]

The Aortic Trauma Foundation (ATF) is a nonprofit, international, multidisciplinary initiative established in 2014, dedicated to improving outcomes for patients with traumatic aortic injuries, particularly blunt thoracic aortic injury (BTAI), through education, collaborative research, consensus-building, and the creation of a prospective, non-industry-driven registry.[3,12] In 2024, Lu et al. published the first study applying an automated ML pipeline "STREAMLINE" (A Simple, Transparent, End-To-End Automated Machine Learning Pipeline Facilitating Data Analysis and Algorithm Comparison) to the ATF registry to predict in-hospital mortality in BTAI.[12] Their work demonstrated that ML could achieve high discrimination comparable to logistic regression, while identifying novel predictors such as injury location along the lesser curvature.[12] However, their analysis was limited to a single automated pipeline, without comparison to other ML architectures.[12]

In the present study, we expand on this work by applying both logistic regression and two distinct ML models, multilayer perceptron (MLP) and extreme gradient boosting (XGBoost), to the ATF dataset. Our objectives were to (1) compare the predictive performance of logistic regression with MLP and XGBoost for in-hospital mortality, (2) examine model-specific variable prioritization, and (3) evaluate whether ML offers incremental benefit over traditional models while maintaining clinical interpretability. This study was conducted under IRB #24081 and with permission of the Aortic Trauma Foundation.

METHODS

Study Population and Variables

We retrospectively analyzed patients from the ATF BTAI registry (November 2016-July 2025) with complete data on in-hospital mortality and management strategy. Predictor variables included demographic, initial physiologic, laboratory, and injury characteristics recorded at admission. Patients were classified into one of three management strategies: open surgical repair (OSR), thoracic endovascular aortic repair (TEVAR), or medical management (e.g. anti-impulse therapy) alone (MMA).

Continuous variables were summarized using medians and interquartile ranges and compared using the Kruskal–Wallis test for multiple group comparisons or the Wilcoxon rank-sum test for two-group comparisons. Categorical variables were compared using Pearson's chi-squared test or Fisher's exact test, as appropriate. A two-sided p-value < 0.05 was considered statistically significant.

Logistic Regression Analysis of Management Strategy

To explore associations between patient characteristics and management strategy, separate multivariable logistic regression models were constructed for each treatment type. Variables for inclusion in each model were first identified through univariate screening, with variables considered significant if the p-value was <0.1. For each treatment-specific model, individual logistic regressions were then iteratively fitted, removing variables sequentially until only those with p-values <0.05 remained. Based on this process, the OSR model included ISS only; the TEVAR model included age, heart rate, platelet count, and ISS; and the MMA model included systolic blood pressure and international normalized ratio (INR). These models were intended to identify independent associations rather than to generate predictive estimates. Model fit was evaluated using deviance and the Akaike Information Criterion (AIC).[13]

Explanatory Analysis of Mortality

An explanatory logistic regression was fit using age, Glasgow Coma Scale (GCS), pH, ISS, intensive care unit (ICU) length of stay, and hospital length of stay to identify factors independently associated with in-hospital mortality.

Predictive Modeling of Mortality: Unconstrained Evaluation

In a secondary analysis, models were evaluated without a fixed sensitivity constraint. Logistic regression (LR), extreme gradient boosting (XGB), and multilayer perceptron (MLP) were trained using a single randomized 70:30 train–test split with stratification by outcome and a fixed random seed (random_state = 42). Performance was assessed on the held-out test set across the full ROC curve and at default probability thresholds. Evaluation metrics included area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, F1 score, and Brier score. Unlike the threshold-constrained approach, this analysis reflects each model's maximum discriminative potential rather than performance at a clinically targeted sensitivity. For the XGBoost model, feature importance was calculated using gain-based importance metrics, reflecting the average reduction in loss attributable to splits on each variable across all trees.

Predictive Modeling of Mortality: Sensitivity Constrained Evaluation

A predictive modeling approach targeted approximately 80% sensitivity for mortality detection in the test set, comparing logistic regression (LR), extreme gradient boosting (XGB), and multilayer perceptron (MLP). The dataset was split using the same randomized 70:30 stratified train–test split (random_state = 42), and model performance was evaluated on the held-out test set using the area under the receiver operating characteristic curve (AUC), average precision (AP), accuracy, precision, recall, F1 score, Brier score, and confusion matrices.

Hyperparameter Sensitivity Analysis

In a secondary sensitivity analysis, XGBoost performance was further evaluated using randomized hyperparameter optimization within the training set. A randomized search over 40 candidate parameter combinations was conducted using five-fold stratified cross-validation (200 total model fits). Hyperparameters evaluated included number of estimators, maximum tree depth, learning rate, subsampling rates, γ (minimum loss reduction), and L1 and L2 regularization terms. The optimal configuration identified through cross-validation was then evaluated on the held-out test set.

Multiclass Prediction of Management Strategy

Multiclass prediction models were developed for management strategy classification using MLP and XGB. The outcome variable was the management category, encoded as OSR, TEVAR, or MMA. Predictor variables included age, ISS, systolic blood pressure, INR, heart rate, and platelet count. The dataset was cleaned to remove rows with missing predictor or outcome values, and the target variable was label-encoded. Data were split into training and test sets with a 70:30 ratio. Features were standardized for the MLP model but used in their raw scale for XGB. The MLP classifier was configured with one hidden layer of 50 neurons, ReLU activation, the Adam optimizer, early stopping, and a maximum of 1,000 iterations. The XGB classifier was configured with 100 trees, learning rate 0.1, maximum depth 5, and a multiclass softmax objective. Both models were evaluated on the test set using accuracy and multiclass AUC (macro-averaged one-vs-rest).

RESULTS

Study Population and Variables

Across treatment groups (**Table 1**), patients undergoing open surgical repair were older (median 51 years) compared with those receiving medical management alone (44 years) or endovascular repair (38 years, $p = 0.013$). Several physiologic parameters differed significantly between groups: systolic and mean arterial blood pressures were lowest in the open surgical repair cohort, while heart rate was higher in the endovascular group. Severe physiologic derangements were more pronounced among open surgical repair patients, including lower GCS scores, higher base deficit, and elevated coagulation times. Laboratory values also showed group differences, with lower platelet counts and hemoglobin levels and higher INR in the open surgical group. In-hospital mortality was substantially higher for open surgical repair (50%) compared with medical management (10%) and endovascular repair (9.7%) ($p < 0.001$).

When stratified by mortality status (**Table 2**), non-survivors were significantly older (median 51 vs. 39 years, $p < 0.001$) and more likely to have undergone open surgical repair (6.9% vs. 1.4%, $p < 0.001$), while survivors more frequently received endovascular repair (61% vs. 35%, $p < 0.001$). Non-survivors demonstrated greater physiologic instability, with markedly lower systolic and mean arterial pressures, lower GCS, higher lactate and creatinine levels, lower hemoglobin and platelet counts, prolonged coagulation times, lower pH, and higher base deficit values (all $p < 0.001$). Injury severity scores were significantly higher in non-survivors (median 45 vs. 30), and both ICU and hospital lengths of stay were shorter, consistent with early mortality. These findings highlight distinct physiologic and injury severity profiles by treatment type and their association with survival.

Logistic Regression Analysis of Management Strategy

OSR was significantly associated only with higher ISS ($\beta = 0.0408$, $p = 0.0017$; null deviance 220.41; residual deviance 211.22; AIC 215.22). TEVAR was associated with younger age ($\beta = -0.0120$, $p = 0.0032$), higher heart rate ($\beta = 0.00647$, $p = 0.0269$), lower platelet count ($\beta = -0.00325$, $p < 0.0001$), and lower ISS ($\beta = -0.0113$, $p = 0.0330$; null deviance 1219.1; residual deviance 1192.1; AIC 1202.1). MMA was associated with higher systolic blood pressure ($\beta = 0.00691$, $p = 0.0145$) and lower INR ($\beta = -0.9339$, $p = 0.0094$; null deviance 929.04; residual deviance 909.39; AIC 915.39).

Explanatory Analysis of Mortality

The initial multivariable logistic regression for mortality demonstrated excellent discrimination (AUC = 0.9607) and fit (null deviance 577.75; residual deviance 234.57; AIC 248.57). (**Figure 1**). Increased mortality risk was independently associated with higher age ($\beta = 0.049$, $p < 0.0001$), higher ISS ($\beta = 0.070$, $p < 0.0001$), and longer ICU length of stay ($\beta = 0.453$, $p < 0.0001$). Protective factors included higher GCS ($\beta = -0.125$, $p < 0.0001$), higher pH ($\beta = -5.835$, $p < 0.0001$), and longer hospital length of stay ($\beta = -0.528$, $p < 0.0001$). Management strategy was not an independent predictor of mortality after adjustment.

When evaluated across the full ROC curve and at default thresholds, LR achieved the highest performance with an AUC of 0.936 and accuracy of 0.896, followed by XGB (AUC 0.915, accuracy 0.891). The MLP model again underperformed with an AUC of 0.610 and lower accuracy (**Figure 2**). Calibration was slightly better for LR than XGB, consistent with the threshold-constrained analysis. These results highlight that LR provides the strongest overall separation between survivors and non-survivors when evaluated without sensitivity constraints (**Table 3**).

Figure 1

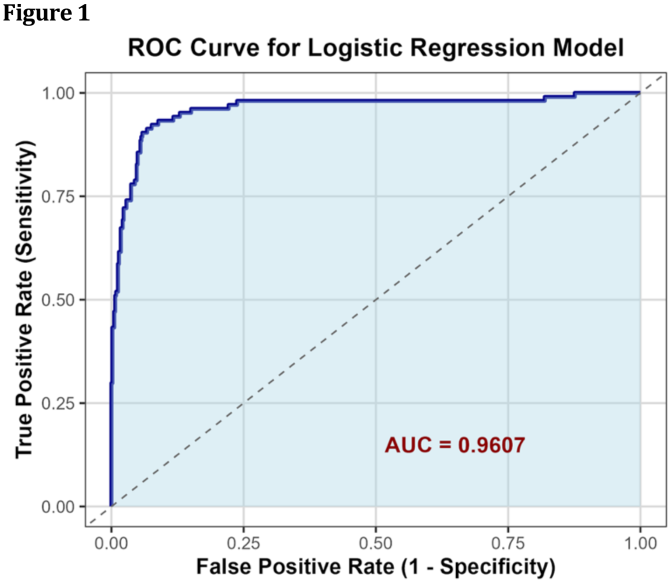

Figure 1. ROC Curve for in-hospital mortality logistic regression model.

In the hyperparameter sensitivity analysis, the optimized XGBoost configuration included 400 estimators, a maximum tree depth of 6, a learning rate of 0.1, subsample and column sampling rates of 0.8, $\gamma = 0.5$, L1 regularization ($\alpha = 0.01$), and L2 regularization ($\lambda = 1$). This tuned model achieved a cross-validated AUC of 0.958 within the training data and an AUC of 0.915 on the held-out test set. Although hyperparameter optimization modestly improved internal cross-validation performance, improvement in test-set discrimination was limited, suggesting that baseline model performance was already strong and that additional tuning may yield incremental rather than substantial gains.

Table 1

Characteristic	n=	Endovascular Repair (n= 725)	Medical Management Alone (n = 297)	Open Surgical Repair (n = 27)	p - value
Age	1,049	38.0 (27.0, 55.0)	44.0 (31.0, 57.0)	51.00 (34.00, 59.00)	**0.013**
Gender	1,048				0.054
Female		160 / 724 (22%)	86 / 297 (29%)	8 / 27 (30%)	
Male		564 / 724 (78%)	211 / 297 (71%)	19 / 27 (70%)	
Mechanism	1,045				0.40
MVC		446 / 723 (62%)	176 / 295 (60%)	16 / 27 (59%)	
Motorcycle Accident		108 / 723 (15%)	45 / 295 (15%)	2 / 27 (7.4%)	
Auto vs. Pedestrian		73 / 723 (10%)	35 / 295 (12%)	5 / 27 (19%)	
Fall		62 / 723 (8.6%)	22 / 295 (7.5%)	2 / 27 (7.4%)	
Other		31 / 723 (4.3%)	16 / 295 (5.4%)	1 / 27 (3.7%)	
Work Accident		3 / 723 (0.4%)	1 / 295 (0.3%)	1 / 27 (3.7%)	
Systolic BP	1,014	115.0 (96.0, 134.0)	120.50 (105.0, 141.0)	114.0 (92.0, 124.0)	**<0.001**
Admission MAP	827	85.00 (72.0, 99.0)	90.00 (76.0, 106.0)	80.00 (57.0, 92.0)	**<0.001**
Heart rate	961	100.00 (85.0, 118.0)	96.00 (82.0, 109.0)	90.00 (79.0, 104.0)	**0.014**
GCS	1,011	14.00 (6.0, 15.0)	15.00 (10.00, 15.0)	7.00 (3.00, 15.0)	**<0.001**
Temperature	825	36.50 (36.10, 36.89)	36.67 (36.28, 36.83)	36.55 (35.95, 36.85)	0.058
Lactate	764	3.65 (2.30, 4.99)	3.26 (2.30, 4.40)	5.20 (2.40, 7.70)	0.055
Creatinine	918	1.16 (0.97, 1.40)	1.10 (0.90, 1.35)	1.20 (0.88, 1.37)	0.12
Hemoglobin	931	12.8 (11.4, 14.1)	13.4 (11.9, 14.5)	12.6 (10.7, 14.3)	**0.001**
Platelet count	927	232.0 (176.0, 287.0)	259.0 (212.0, 315.0)	183.0 (147.0, 265.0)	**<0.001**
PTT	617	27.3 (24.5, 31.0)	25.80 (23.9, 28.6)	30.0 (27.2, 34.1)	**<0.001**
PT	604	14.0 (12.4, 15.7)	13.5 (12.2, 15.0)	15.7 (13.2, 17.3)	**0.016**
INR	693	1.13 (1.03, 1.30)	1.10 (1.00, 1.20)	1.30 (1.12, 1.40)	**<0.001**
pH	717	7.29 (7.21, 7.34)	7.31 (7.25, 7.35)	7.18 (7.05, 7.32)	**0.002**
Base deficit value	672	5.1 (3.0, 9.0)	4.0 (2.0, 7.0)	9.7 (0.6, 16.0)	**<0.001**
ISS	892	34.0 (24.0, 41.0)	29.00 (24.0, 41.0)	36.00 (29.0, 66.0)	**0.006**
In-Hospital Mortality	1,007	67 / 692 (9.7%)	29 / 289 (10%)	13 / 26 (50%)	**<0.001**

Table 1. Study population and variables across treatment groups. Median (Q1, Q3); n / N (%); Kruskal-Wallis rank sum test; Pearson's Chi-squared test; Fisher's exact test. Motor vehicle collision, MVC; Blood pressure, BP; Mean arterial pressure, MAP; Glasgow Coma Scale, GCS; Partial thromboplastin time, PTT; Prothrombin time, PT; International normalized ratio, INR; Injury severity score, ISS.

Table 2

Characteristic	n=	Status: Survivor	Status: Non-Survivor	p -value
Medical Management Alone	297	266 (90%)	31 (10%)	**<0.001**
Open surgical repair	27	14 (52%)	13 (48%)	**<0.001**
Endovascular repair	725	625 (86%)	100 (14%)	**<0.001**
Age	1,219	39.0 (28.00, 54.0)	51.00 (34.0, 65.0)	**<0.001**
Gender	1,218			0.4
Female	291	250 (86%)	41 (14%)	
Male	927	779 (84%)	148 (16%)	
Mechanism	1,202			**<0.001**
MVC	716	628 (88%)	88 (12%)	
Motorcycle accident	181	154 (85%)	27 (15%)	
Auto vs. Pedestrian	143	100 (70%)	43 (30%)	
Fall	104	85 (82%)	19 (18%)	
Other	53	45 (85%)	8 (15%)	
Work-related accident	5	4 (80%)	1 (20%)	
Systolic BP	1,150	118.0 (100.0, 136.0)	98.50 (71.0, 128.0)	**<0.001**
Admission MAP	954	87.0 (73.0, 101.0)	70.00 (50.0, 97.0)	**<0.001**
Heart rate	1,087	97.0 (84.0, 114.0)	100.00 (78.0, 120.0)	0.8
GCS	1,145	15.0 (9.0, 15.0)	3.0 (3.0, 12.0)	**<0.001**
Temperature	915	36.60 (36.17, 36.89)	36.50 (35.90, 36.90)	0.15
Lactate	826	3.30 (2.20, 4.60)	5.00 (3.40, 8.70)	**<0.001**
Creatinine	1,035	1.10 (0.91, 1.37)	1.30 (1.09, 1.60)	**<0.001**
Hemoglobin	1,053	13.10 (11.7, 14.4)	11.70 (9.8, 13.3)	**<0.001**
Platelet count	1,042	244.0 (191.0, 298.0)	191.00 (137.0, 249.0)	**<0.001**
PTT	713	26.65 (24.2, 30.0)	33.00 (28.0, 45.4)	**<0.001**
PT	699	13.80 (12.25, 15.30)	15.50 (13.50, 18.00)	**<0.001**
INR	795	1.10 (1.00, 1.21)	1.30 (1.14, 1.45)	**<0.001**
pH	805	7.30 (7.23, 7.35)	7.19 (7.08, 7.30)	**<0.001**
Base deficit value	761	4.9 (2.2, 8.0)	8.3 (5.1, 13.0)	**<0.001**
ISS	1,006	30.0 (24.0, 41.0)	45.0 (34.0, 59.0)	**<0.001**
ICU length of stay	1,119	6.0 (3.0, 14.0)	3.0 (0.0, 9.0)	**<0.001**
Hospital length of stay	1,147	14.0 (8.0, 26.0)	2.0 (1.0, 8.0)	**<0.001**
Ventilator days	1,032	2.0 (0.0, 7.0)	2.0 (1.0, 7.0)	**0.036**

Table 2. Study population stratified by mortality status. n / N (%); Median (Q1, Q3); Pearson's Chi-squared test; Wilcoxon rank sum test. Motor vehicle collision, MVC; Blood pressure, BP; Mean arterial pressure, MAP; Glasgow Coma Scale, GCS; Partial thromboplastin time, PTT; Prothrombin time, PT; International normalized ratio, INR; Injury severity score, ISS.

Table 3

Outcome	Model	AUC	Average Precision	Accuracy	F1	Precision	Recall	Brier	Threshold	TN	FP	FN	TP
In-Hospital Mortality	LR	0.9362	0.8648	0.8955	0.6866	0.9200	0.5476	0.0740	0.5	157	2	19	23
In-Hospital Mortality	XGB	0.915	0.8000	0.8905	0.6667	0.9167	0.5238	0.0921	0.5	157	2	20	22
In-Hospital Mortality	MLP	0.610	0.5026	0.8010	0.3333	0.5556	0.2381	0.1623	0.5	151	8	32	10

Table 3. Unconstrained Model Results. LR, Logistic Regression; XGB, XGBoost; MLP, Multilayer Perceptron; AUC, Area Under the Receiver Operating Characteristic Curve; F1, F1 Score (harmonic mean of precision and recall); Brier, Brier Score; TN, True Negative; FP, False Positive; FN, False Negative; TP, True Positive.

Table 4

Outcome	Model	AUC	Average Precision	Accuracy	F1	Precision	Recall	Brier	Threshold	TN	FP	FN	TP
In-Hospital Mortality	LR	0.8492	0.5701	0.7345	0.4593	0.3196	0.8158	0.0829	0.114	171	66	7	31
In-Hospital Mortality	XGB	0.8552	0.5701	0.7236	0.4493	0.31	0.8158	0.1012	0.0991	168	69	7	31
In-Hospital Mortality	MLP	0.7566	0.4453	0.5418	0.3298	0.2067	0.8158	0.14	0.0	118	119	7	31

Table 4. 80% Sensitivity Results. LR, Logistic Regression; XGB, XGBoost; MLP, Multilayer Perceptron; AUC, Area Under the Receiver Operating Characteristic Curve; F1, F1 Score (harmonic mean of precision and recall); Brier, Brier Score; TN, True Negative; FP, False Positive; FN, False Negative; TP, True Positive.

XGBoost Feature Importance

Feature importance analysis from the XGBoost mortality model identified Length of stay (LOS) ISS, GCS, age, and pH as the most influential predictors (**Figure 3**). The relative importance pattern was consistent with variables independently associated with mortality in logistic regression, suggesting that mortality risk in BTAI is predominantly driven by physiologic and injury severity factors rather than complex nonlinear interactions.

Sensitivity Constrained Evaluation

In the predictive modeling analysis for mortality at approximately 80% sensitivity, LR and XGB achieved similar performance with AUC values of 0.849 and 0.855, recall 0.816, and precision around 0.31. LR demonstrated slightly better calibration with a Brier score of 0.083 compared to 0.101 for XGB. The MLP model underperformed, with an AUC of 0.757 and lower accuracy and calibration metrics (**Table 4**).

Multiclass Prediction of Management Strategy

In the multiclass management strategy prediction task, the MLP achieved an accuracy of 0.667 and a macro-averaged AUC of 0.565 (**Figure 4**). The XGB model achieved an accuracy of 0.690 and a macro-averaged AUC of 0.648 (**Figure 5**). Both models demonstrated class-dependent variation in AUC, with the MLP

Figure 2

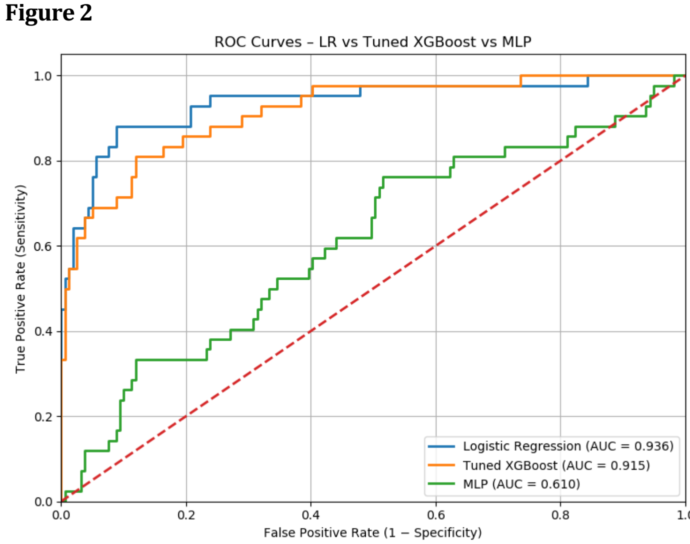

Figure 2. Receiver operating characteristic (ROC) curves for logistic regression (LR), extreme gradient boosting (XGB), and multilayer perceptron (MLP) models predicting in-hospital mortality under unconstrained evaluation. Logistic regression demonstrated the highest overall discrimination, followed closely by XGBoost, whereas MLP showed lower performance. Curves reflect model performance on the held-out 30% test set.

showing the highest separability for class 1 (AUC = 0.62) and the lowest for class 2 (AUC = 0.53), and the XGB model showing the highest separability for class 0 (AUC = 0.67) and the lowest for class 2 (AUC = 0.63). Discrimination was moderate overall, consistent with the logistic regression findings that management strategy is influenced by factors not fully captured in the available structured data.

Figure 3

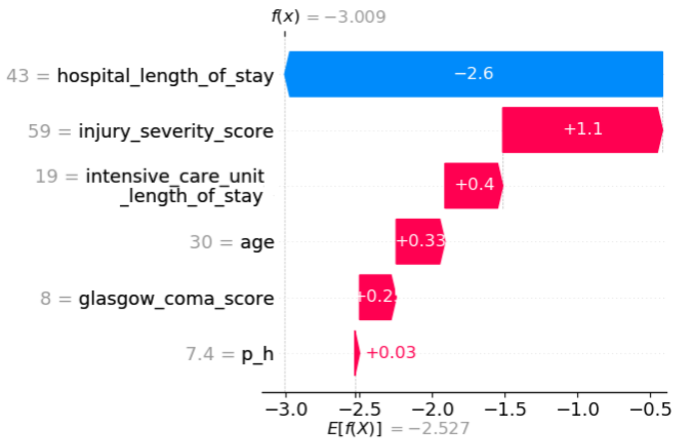

Figure 3. SHAP (SHapley Additive exPlanations) waterfall plot showing relative feature importance from the XGBoost model for in-hospital mortality prediction for a single representative observation. Gray values on the left indicate each feature's actual value for that patient. Bars reflect each variable's contribution to shifting the model output from the baseline expected value [E[f(X)] = −2.527] to the final prediction [f(x) = −3.009], with red bars increasing and blue bars decreasing the predicted log-odds of mortality. LOS, ISS, GCS, age, and pH were among the most influential predictors.

DISCUSSION

This study is the first to directly compare logistic regression with two contemporary machine learning algorithms, MLP and XGBoost, for both in-hospital mortality and management strategy prediction in BTAI using the Aortic Trauma Foundation registry. Whereas prior work by Lu et al. applied an automated ML pipeline (STREAMLINE) to mortality prediction alone,[12] our analysis incorporated multiple modeling frameworks, separated explanatory from predictive analyses, and extended the approach to include operative decision-making.

For in-hospital mortality, our explanatory logistic regression model achieved excellent discrimination (AUC = 0.96), identifying higher Injury Severity Score (ISS), lower Glasgow Coma Scale (GCS), lower pH, older age, and prolonged ICU stay as independent predictors. When tuned for ~80% sensitivity in a predictive framework, logistic regression and XGBoost achieved nearly identical performance (AUCs = 0.849 and 0.855, respectively), with multilayer perceptron underperforming (AUC = 0.757). These findings confirm and extend those of Lu et al., demonstrating that a well-specified logistic regression can match or exceed complex ML approaches for physiologically determined outcomes such as in-hospital mortality in BTAI. Importantly, XGBoost feature importance mirrored regression findings, reinforcing that established physiologic variables dominate mortality prediction in BTAI.

Figure 4

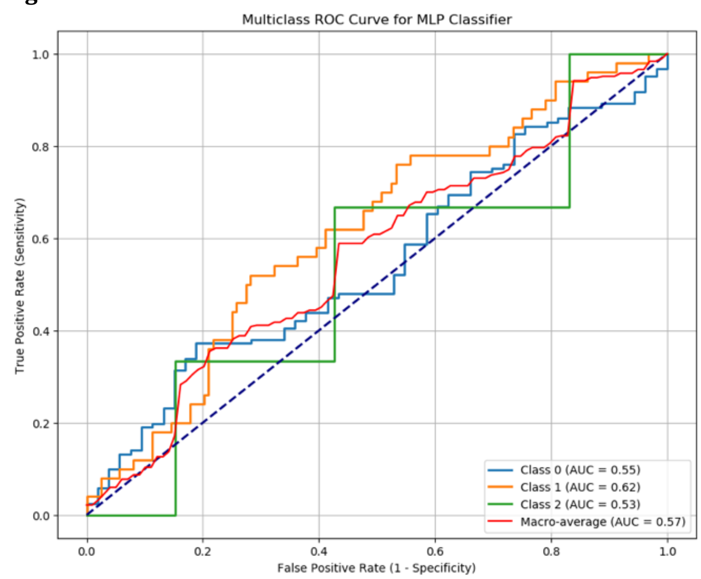

Figure 4. Multiclass ROC for MLP.

In contrast, prediction of management strategy, categorizing patients as undergoing open surgical repair, TEVAR, or medical management alone, proved substantially more challenging. Even with the same modeling frameworks, the best model (XGBoost) achieved a macro-average AUC of only 0.648, with MLP performing slightly worse at 0.565. This represents an absolute drop of more than 0.30 in discriminative performance compared to mortality prediction. The limited accuracy likely reflects the multifactorial nature of management decisions, which depend not only on patient physiology and injury severity but also on institutional protocols, surgeon preference, device availability, and timing of presentation, factors not captured in the current registry. Future work will focus on the role of aortic injury grade on these models.

Notably, while our logistic regression management models identified a small set of significant variables (e.g., ISS for OSR; age, heart rate, platelet count, and ISS for TEVAR; and systolic blood

Figure 5

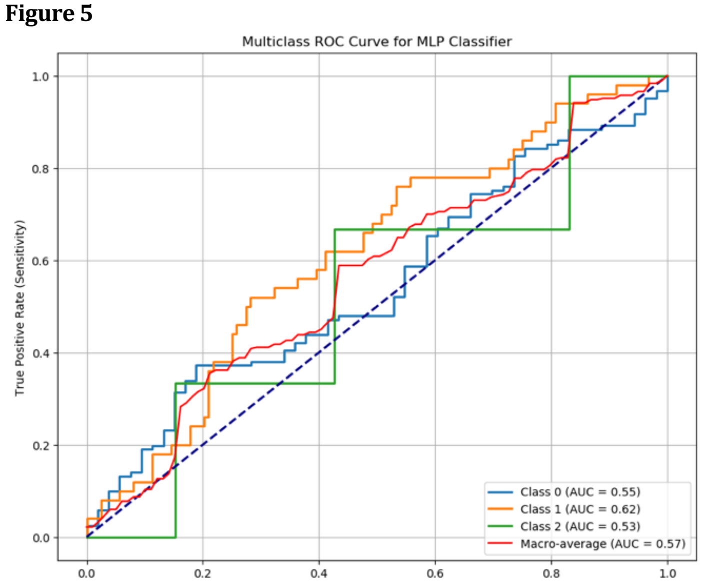

Figure 5. Multiclass ROC for XGBoost.

pressure and INR for MMA), the ML models highlighted additional predictors such as lactate and base deficit. These variables may influence decision-making in subtle, nonlinear ways not retained in traditional regression after multivariable adjustment. However, their inclusion did not yield clinically meaningful gains in discrimination, underscoring that the primary limitation is missing explanatory data rather than model choice.

Clinically, these findings suggest a clear role for logistic regression as the preferred tool for mortality risk stratification in BTAI, combining high accuracy, interpretability, and ease of implementation in bedside calculators and triage tools. For complex decisions such as operative strategy, ML approaches may offer greater potential once additional data streams, such as imaging features, institutional practice patterns, and device inventory, are integrated into prediction models. Until such data are available, model performance for management strategy will remain constrained, regardless of algorithmic sophistication.

This study has several limitations. First, as a retrospective analysis of the Aortic Trauma Foundation registry, it is subject to the inherent constraints of observational data, including potential selection bias, incomplete capture of confounding variables, and missing data. Second, the registry lacks key determinants of management strategy such as detailed imaging characteristics, institutional protocols, surgeon experience, device availability, and timing of presentation, which likely limited model performance for operative decision-making. Third, while we evaluated multiple modeling approaches, our analysis was restricted to structured clinical variables available at admission; incorporating physiologic waveform data, imaging-derived features, and longitudinal clinical information could further improve prediction. Fourth, although we performed internal validation using train–test splits, external validation in independent datasets is necessary before clinical implementation. Finally, the relatively small number of events for certain outcomes, particularly OSR, may have reduced statistical power and increased the risk of model overfitting despite regularization and cross-validation efforts.

Future work should focus on three areas: first, expanding the registry to capture unmeasured determinants of management choice; second, validating both explanatory and predictive models across independent cohorts; and third, exploring hybrid approaches that combine the transparency of regression with the flexibility of ML. Ensemble learning approaches, such as soft voting or stacking methods that combine predictions from logistic regression, XGBoost, and neural network models, may leverage complementary model structures to enhance discrimination and calibration. While not explored in the current study, future investigations may evaluate whether ensemble strategies provide incremental benefit beyond individual models in trauma mortality prediction. By pairing interpretable, high-performing mortality prediction with enriched decision-modeling frameworks, the potential exists to enhance both prognostication and operative planning for patients with BTAI.

Competing Interests: The authors attest to having no financial or commercial associations that may create conflicts of interest with the information presented in the manuscript.

Authors' Contributions: All authors provided substantial contributions to the work, and were in agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. **PJ**: Conceptualization, study design, data analysis, data interpretation, writing – original draft, critical revision. **SW**: Data interpretation, writing – review & editing. **VA**: Data interpretation, writing – review & editing. **TL**: Literature review, writing – review & editing. **CB**: Methodology, writing – review & editing. **JR**: Data interpretation, writing – review & editing. **JD**: Conceptualization, writing – review & editing. **MK**: Data curation, supervision, methodology, writing – review & editing.

Media Summary: Future work should focus on three areas: first, expanding the registry to capture unmeasured determinants of management choice; second, validating both explanatory and predictive models across independent cohorts; and third, exploring hybrid approaches that combine the transparency of regression with the flexibility of ML. Ensemble learning approaches, such as soft voting or stacking methods that combine predictions from logistic regression, XGBoost, and neural network models, may leverage complementary model structures to enhance discrimination and calibration. While not explored in the current study, future investigations may evaluate whether ensemble strategies provide incremental benefit beyond individual models in trauma mortality prediction. By pairing interpretable, high-performing mortality prediction with enriched decision-modeling frameworks, the potential exists to enhance both prognostication and operative planning for patients with BTAI.

Funding Information: There was no funding associated with this publication.

Hashtags: #TraumaSurgery #AorticInjury #MachineLearning

Abbreviations: AIC, Akaike Information Criterion; AP, Average Precision; ATF, Aortic Trauma Foundation; AUC, Area Under the Receiver Operating Characteristic Curve; BP, Blood Pressure; BTAI, Blunt Thoracic Aortic Injury; GCS, Glasgow Coma Scale; ICU, Intensive Care Unit; INR, International Normalized Ratio; ISS, Injury Severity Score; LOS, Length of Stay; LR, Logistic Regression; MAP, Mean Arterial Pressure; ML, Machine Learning; MLP, Multilayer Perceptron; MMA, Medical Management Alone; MVC, Motor Vehicle Collision; OSR, Open Surgical Repair; PT, Prothrombin Time; PTT, Partial Thromboplastin Time; ROC, Receiver Operating Characteristic; TEVAR, Thoracic Endovascular Aortic Repair; XGB, Extreme Gradient Boosting.

REFERENCES

1. Boutin L, Caballero MJ, Guarrigue D, Hammad E, Rennuit I, Delhaye N, Neuschwander A, Meyer A, Bitot V, Mathais Q, Boutonnet M, Julia P, Olaf M, Duranteau J, Hamada SR; TraumaBase Group. Blunt Traumatic Aortic Injury Management, a French TraumaBase Analytic Cohort. Eur J Vasc Endovasc Surg. 2022 Mar;63(3):401-409. doi: 10.1016/j.ejvs.2021.09.043.

2. Akhmerov A, DuBose J, Azizzadeh A. Blunt Thoracic Aortic Injury: Current Therapies, Outcomes, and Challenges. Ann Vasc Dis. 2019 Mar 25;12(1):1-5. doi: 10.3400/avd.ra.18-00139.

3. Lee WA, Matsumura JS, Mitchell RS, Farber MA, Greenberg RK, Azizzadeh A, Murad MH, Fairman RM. Endovascular repair of traumatic thoracic aortic injury: clinical practice guidelines of the Society for Vascular Surgery. J Vasc Surg. 2011 Jan;53(1):187-92. doi: 10.1016/j.jvs.2010.08.027

4. Lamarche Y, Berger FH, Nicolaou S, Bilawich AM, Louis L, Inacio JR, Janusz MT, Evans D. Vancouver simplified grading system with computed tomographic angiography for blunt aortic injury. J Thorac Cardiovasc Surg. 2012 Aug;144(2):347-54, 354.e1. doi: 10.1016/j.jtcvs.2011.10.011

5. Marcaccio CL, Dumas RP, Huang Y, Yang W, Wang GJ, Holena DN. Delayed endovascular aortic repair is associated with reduced in-hospital mortality in patients with blunt thoracic aortic injury. J Vasc Surg. 2018 Jul;68(1):64-73. doi: 10.1016/j.jvs.2017.10.084

6. Hsu TA, Huang JF, Kuo LW, Cheng CT, Kuo YC, Liao CH, Hsieh CH, Fu CY. Does Thoracic Endovascular Aortic Repair for Blunt Thoracic Aortic Injury Increase the Probability of Delayed Intra-abdominal Hemorrhage in Patients With Concomitant Blunt Abdominal Trauma? J Endovasc Ther. 2024 Jun;31(3):428-437. doi: 10.1177/15266028221128200

7. de Mestral C, Dueck AD, Gomez D, Haas B, Nathens AB. Associated injuries, management, and outcomes of blunt abdominal aortic injury. J Vasc Surg. 2012 Sep;56(3):656-60. doi: 10.1016/j.jvs.2012.02.027

8. Baker SP, O'Neill B, Haddon W Jr, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma. 1974 Mar;14(3):187-96.

9. Mai JYL, Holmes A, Frahm-Jensen G. Grade four blunt traumatic aortic injury with massive haemothorax: Resuscitation considerations during the primary survey. Trauma Case Rep. 2020 Jun 29;29:100333. doi: 10.1016/j.tcr.2020.100333. Erratum in: Trauma Case Rep. 2023 Mar 01;45:100815. doi: 10.1016/j.tcr.2023.100815

10. Boateng, E. and Abaye, D. A Review of the Logistic Regression Model with Emphasis on Medical Research. Journal of Data Analysis and Information Processing, 2019 Nov; 7(4):190-207. doi: 10.4236/jdaip.2019.74012

11. Ding N, Guo C, Li C, Zhou Y, Chai X. An Artificial Neural Networks Model for Early Predicting In-Hospital Mortality in Acute Pancreatitis in MIMIC-III. Biomed Res Int. 2021 Jan 28;2021:6638919. doi: 10.1155/2021/6638919

12. Lu E, Dubose J, Venkatesan M, Wang ZP, Starnes BW, Saqib NU, Miller CC, Azizzadeh A, Chou EL. Using machine learning to predict outcomes of patients with blunt traumatic aortic injuries. J Trauma Acute Care Surg. 2024 Aug 1;97(2):258-265. doi: 10.1097/TA.0000000000004322

13. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974 Dec; 19(6):716-723. doi: 10.1109/TAC.1974.1100705

PNWJS