

Machine Learning Prediction of Length of Stay in Hand Injury Patients Undergoing Emergency Surgery

Yuxiong Weng (✉ yxweng1218@163.com)

Wuhan Union Hospital

Lizhao Yan

Wuhan Union Hospital

Kan Wang

Wuhan Union Hospital

Fangxing Ai

Wuhan Union Hospital

Yu Kang

Wuhan Union Hospital

Nan Gao

Wuhan Union Hospital

Research Article

Keywords: Hand injury, Emergency surgery, Length of stay, Machine Learning, Random forest

Posted Date: February 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1358631/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The In-hospital length of stay (LOS) is expected to increase as the complexity of hand injuries increases. This will impact healthcare systems, especially given the current situation of reduced bed capacity and rising costs. Therefore, accurate prediction of LOS would positively impact healthcare interventions. This research aims to develop a machine learning-based model to predict the length of hospital stay in patients with hand injuries.

Methods: Patients who underwent emergency surgery with hand injuries were selected from the hand surgery department of Wuhan Union Hospital. Prolonged LOS was defined as a LOS of more than or equal to 9 days. Data were analyzed using Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGboost) in R software with glm, caret, and random forest packages. Prediction accuracy and area under the curve (AUC) were calculated.

Results: A prospective cohort with 3858 patients with hand injuries were identified from 2016 to 2020. The area under the curve (AUC) , accuracy in the test set were LR(0.805,0.760), NB(0.786,0.715), RF(0.883,0.809), SVM(0.821,0.752), and XGboost(0.851,0.763), respectively. Multivariable regression indicated that longer operation time, lower hemoglobin (HGB), vascular injury, bone injury, general anesthesia, payment (commercial insurance, self-paying), and tourniquet use were linked to longer LOS. The random forest model performed best among these models and was developed into an encrypted web-based interface: <https://union.shinyapps.io/PredictLOS/>.

Conclusions: We showed that machine learning methods accurately predict LOS for hand injury patients. The model we developed can be used in clinical bed management and resources allocation.

Background

According to the US National Center for Health Statistics¹, injuries to the hands and wrists are common in emergency departments, accounting for 14.4 percent of 29.9 million injury-related emergency department visits nationwide in 2016. Due to the volume of injuries sustained annually, hand and wrist injuries rank first in the order of most expensive injury types, before knee and lower limb fractures, hip fractures, and skull-brain injury². The number of hand injured patients also shows high volatility due to various factors, including weather³, calendar data⁴, etc. Coupled with the limited number of beds, the Department of Hand Surgery must operate at full capacity during certain admission seasons, resulting in great difficulty in allocating hospital resources.

LOS is a simple and commonly used metric for measuring resource allocation and healthcare costs in health care. Accurate prediction of LOS is critical for capacity management, resource planning, and staffing levels. Several factors have been studied as predictors or risk factors for LOS in trauma, including trauma severity factors^{5,6} and inflammatory response factors⁷⁻⁹, etc. In addition, a variety of different statistical algorithms have also been used to develop optimal LOS prediction models, as shown in a

review¹⁰ that summarized 74 published studies of LOS in hand surgery^{11, 12}, thermal burns¹³, neonatal units^{11, 14}, and intensive care units(ICU)¹⁵, etc. Regression has been the most popular prediction method among these models, but it is gradually being supplanted by more sophisticated techniques such as machine learning¹⁰.

Machine learning is a branch of artificial intelligence and is widely used in healthcare due to its advantages in discovering knowledge from high-dimensional data. In this report, we compare the performance of regression model and four commonly used machine learning models in predicting LOS. We also developed an online web application based on the best models to support clinical decision making. To our knowledge, This is the first study to apply machine learning to LOS prediction in patients with hand injuries. The results will provide clinicians and policy makers with needed insights.

Methods

Study Population and Data Source

A total of 3858 patients with hand injuries who underwent emergency surgery at the Department of Hand Surgery, Wuhan Union Hospital between January 2016 and January 2020 were analysed. Patient data and attributes were extracted from electronic medical records (EMRs) by experienced physicians and trained nurses. The inclusion criteria were: (1) patients older than 18 years and younger than 90 years, (2) patients who underwent emergency surgery for hand and wrist injuries, (3) patients with complete clinical records, and (4) patients discharged after standard treatment procedures.

Variable definition and collection

A prolonged LOS is considered more than or equal to nine days. Hospitalizations that were shorter than nine days are considered short LOS. Factors that may be relevant to LOS were in this study, including age, gender, history of alcohol abuse, history of smoking, diabetes mellitus, hypertension, insurance status, intraoperative discovery, time from injury to surgery (time to surgery), and preoperative blood test results. To avoid pain, patients with hand injuries are generally not evaluated further before surgery after determining the need for surgery. Therefore, intraoperative anatomic findings were selected as predictors, including bone injury, muscle injury, nerve injury, tendon injury, and vascular injury. Insurance status was categorized into three groups: commercial insurance, national medical insurance, self-paying. Finally, white blood cell (WBC), red blood cell (RBC), neutrophil count (NE#), and hemoglobin (HGB) were selected from the preoperative blood test to assess the degree of blood loss and inflammatory response of the body.

Feature Selection

Feature selection can reduce the dimension of the feature space and the complexity of modeling¹⁶. In this paper, we first excluded significantly collinear variables based on the correlation coefficient between each variable. Next, we propose an embedded feature selection scheme that ranks features based on the feature importance values derived from the Random Forest. Then, the lower-ranked features are removed one by one, with 10-fold cross-validation of the performance of the selected feature group in each round. In addition, we also performed the least absolute shrinkage and selection operator method (LASSO) with 10-fold cross-validation to further test the relationship between the selected features and the performance of the linear model.

Linear Regression Model

Linear Regression (LR) is a standard statistical generalized linear model method used in data mining, automatic disease diagnosis, economic forecasting, and other broad applications¹⁷. The algorithm is essentially a conventional two-category model, with the object's category determined by inputting the object's attribute sequence. To classify the data, the model assumes that the data follows the Bernoulli distribution and employs the method of maximizing the likelihood function to solve the parameters with gradient descent. In our study, a multivariable LR model was built using the function `glm` of the R package `stats`. Odds ratios were calculated for each risk factor by exponentiating the LR coefficients. Odds ratios less than 1.0 indicate a decreased risk, while odds ratios greater than 1.0 indicate an increased risk, and the p-value of <0.05 was considered significant.

Machine Learning Models

Four classical machine learning models with five-fold cross-validation were developed to predict LOS, namely Naive Bayes (NB), Random Forest (RF), Support Vector Machines (SVM), and XGboost. All machine learning models were built using the `randomforest`, `caret`, and `XGboost` packages in the R programming language (version 3.3.1). As an extra precaution, the models were trained and tested by two team members (L.Z.Y. and K.W.) to ensure that the models were not inadvertently trained with withheld test data.

RF

The random forest method is a machine learning technique that mixes many decision trees to create a single classification model. The random forest approach generates a forest of multiple decision trees by selecting various dividing characteristics and training samples. When predicting unknown samples, each tree in the forest is trained to make decisions, significantly increasing prediction accuracy compared to a single decision tree. After statistically assessing the decision outcomes, the classification with the most votes is recognized as the official classification result.

NB

NB classifier is a highly scalable supervised learning technique. Bayesian reasoning is based on probability to derive conclusions about the ideal decision's probability distribution. It has been effectively implemented in various scientific areas and consistently performs well even when a few variables are considered.

SVM

SVM is a machine learning approach that is based on statistical learning theory. A support vector machine aims to minimize generalization error by creating a hyperplane in a high-dimensional space and utilizing a maximum margin to separate feature vectors belonging to distinct classes. When a support vector machine is used for linear classification, an $n-1$ dimensional hyperplane is used, where n is the dimension of the data.

XGboost

XGboost is one of the most extensively used machine learning classifiers in bioinformatics. It is based on a tree model that classifies using a boosting method. Regularization elements are added to the cost function to minimize the model's complexity and prevent overfitting. Additionally, the parallel computing function is enabled by the XGboost algorithm, which significantly accelerates calculation.

Model evaluation and validation

We evaluated the predictive performance of the models in two ways: calibration plots, which represent the agreement between the predicted likelihood and the actual likelihood; AUC and ROC curves, which evaluate the classification performance of each specific model.

Results

Basic Characteristics

A total of 3858 patients with hand injuries who had undergone emergency surgery were included in the study. A total of 20 features were also extracted from the EMRs. The baseline information of the patients at the time of enrollment is shown in Table 1. Their mean age was 43.85 ± 12.48 years, and 18.7% were female. The LOS classes included 40.6% and 59.4% for short and long, respectively. Of note, there was a significant difference between patients with prolonged LOS and short LOS in terms of age ($p = 0.048$), operation time ($p < 0.001$), and time to surgery ($p < 0.001$). At the same time, the proportions of low HGB, low RBC, diabetes mellitus, hypertension, general anesthesia, tourniquet use, and complicated injuries

(bone, muscle, nerve, and vascular) were higher in the prolong-LOS group than in the short-LOS group (all $p < 0.05$). These patients were randomly divided into a training group ($n=3087$) and a test group ($n=771$) (Table 2). In the training and validation sets, the distribution of baseline characteristics was similar in both, and no significant difference was found between the two cohorts, indicating a high value of internal validation. Figure 1 illustrates the workflow to build machine learning models.

Feature Selection

We excluded RBC, NE#, and muscle injury based on the correlation coefficient, which appears to be collinear with others (Figure 2). For the remaining 17 variables, we included them in the random forest algorithm for feature selection. After five rounds of 10-fold cross-validation, it is found that the model is generally stable when the number of variables is greater than 8, and the average error rate reaches the lowest value when the number of variables is reduced to 8 (Figure 3). Therefore, the first eight variables were selected in order of importance (Figure 4). In addition, we conducted LASSO regression for feature selection, and seven coefficients were determined as the best combination according to the 1SE rule (selecting the highest lambda whose error was within one standard error of the optimum) (Figure 5, Figure 6). Given the stable performance of the two models with more variables and considering the different characteristics of linear and nonlinear models, we decided to include the union of the variables selected by the two methods. Finally, nine variables were selected for model training: operation time, age, HGB, time to surgery, vascular injury, anesthesia, payment, bone injury, tourniquet use.

Model Comparisons

The predictive performance of the machine learning and LR models is shown in Table 3. All models performed well in terms of AUC, of which the RF model had the highest AUC of 0.883 (Figure 7). Notably, the RF model also had good clinical applicability with an accuracy of 80.93%, sensitivity of 84.03%, specificity of 78.82%, positive predictive value (PPV) of 73.06%, and negative predictive value (NPV) of 87.83% (Table 3). NB (AUC-ROC=0.786) significantly underperformed as compared to other algorithms.

The calibration plots (Figure 8) showed that the predicted probabilities compared to the observed LOS showed excellent agreement for the RF model, followed by the XGboost and SVM models. The calibration of the NB model tended to overestimate the LOS over the entire range of prediction probabilities compared to the other models. The LR model appeared to have a good calibration except for the lower left, which under-predicted the probabilities.

Furthermore, risk factors for LOS used to fit an LR model are summarized in Table 3. Factors such as longer operation time, lower HGB, vascular injury, bone injury, general anesthesia, payment (commercial insurance, self-paying), and tourniquet use increased the chances of prolonged LOS.

Predictive modeling interface

RF is the best-performing algorithm for predicting LOS and was developed into an encrypted web-based interface that can be accessed at <https://union.shinyapps.io/PredictLOS/>.

Discussion

Prolonged LOS correlates with overcrowding in inpatient units and, as a result, lower quality of care and medical staff satisfaction¹⁸. Models that improve patient care while increasing efficiency for providers are urgently needed. This study investigated LOS factors and developed the first ML model to predict LOS in hand injury patients undergoing emergency surgery.

Using computers to guide the treatment of critically ill patients is not a new concept. In a 1977 published study, computerized systems have been proposed to monitor ICU patients¹⁹. Compared to traditional methods, machine learning techniques could process multidimensional parameters simultaneously and were not constrained by data distribution²⁰.

Presently, there have been many studies on surgery using machine learning²¹⁻²³. Still, only a few studies applied machine learning models to predict the length of hospital stay in trauma patients. Some studies exploited the artificial neural network to predict individual time to colorectal cancer surgery, and the highest AUC was up to 0.865²². Also, Houthoof et al.²⁴ trained a support vector machine model to forecast patient survival and length of stay using data from 14,480 patients. The model's area under the curve (AUC) for predicting a prolonged length of stay was 0.82. However, they did not compare multiple machine learning models' performances in predicting ability.

Our study successfully provided a novel strategy to predict the length of hospital stay in hand injury patients by machine learning models. Based on the results of ROC plots for four machine learning methods and LR methods, the RF model best predicted the LOS. And the three most important factors affecting RF's diagnostic performance were operation time, time to surgery, and age. In our study, longer operation time and older patients were associated with the prolonged LOS, indicating that the patient was in a poor physical condition, consistent with the previous study^{25,26}. All the models were validated by the internal data and showed good performance. The results of ROC plots might contribute to better managing efficiency in the hand surgery department, including the cost of patients and the waiting time for newly-clinical admissions.

There were some limitations in this study. First, because it was a retrospective study, it might result in a potential selection bias. Second, our prediction models were not validated in the external cohort to assure the model's generalizability across centers. Thus, a multi-center much larger sample size prospective clinical research was required in the future.

Conclusions

The findings of this study reveal that machine learning algorithms, specifically RF, are excellent at predicting the length of stay for patients undergoing emergency surgery for hand injuries. These algorithms have significant clinical implications, as they offer patients and clinicians with a means of predicting the predicted length of stay following surgery. Specific efforts can be made using this information to reduce hospitalization costs, hence lowering expenditures for patients and contributing to capacity management, resource planning, and staffing levels.

Abbreviations

LOS: Length of stay; LR: Logistic regression; NB: Naive bayes; RF: Random forest ; SVM: Support Vector Machine; XGboost: eXtreme Gradient Boosting; AUC: The area under the curve; HGB: Hemoglobin; ICU: Intensive care units; WBC:White blood cell; RBC:Red blood cell; NE#: Neutrophil count; LASSO: The least absolute shrinkage and selection operator method.

Declarations

Ethics approval and consent to participate

The ethics committee approved the study protocol of Wuhan Union Hospital of Huazhong University of Science and Technology.

Consent for publication

Informed consent was obtained from all participants.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

Yuxiong Weng, Lizhao Yan, Kan Wang designed the work. Lizhao Yan and Kan Wang analyzed datasets. Fangxing Ai, Yu Kang and Nan Gao wrote this paper. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Rui P KK, Ashman JJ. National Hospital Ambulatory Medical Care Survey: 2016 emergency department summary tables. Available at: https://www.cdc.gov/nchs/data/nhamcs/web_tables/2016_ed_web_tables.pdf. (Accessed November 12, 2021). 2016.
2. de Putter CE, Selles RW, Polinder S, Panneman MJ, Hovius SE, van Beeck EF. Economic impact of hand and wrist injuries: health-care costs and productivity costs in a population-based study. *J Bone Joint Surg Am*. 2012 May 2;94(9):e56. Epub 2012/05/04.
3. Marcilio I, Hajat S, Gouveia NJAem. Forecasting daily emergency department visits using calendar variables and ambient temperature readings. 2013;20(8):769-77.
4. Wargon M, Guidet B, Hoang T, Hejblum GJEMJ. A systematic review of models for forecasting the number of emergency department visits. 2009;26(6):395-9.
5. Salehi O, Dezfuli SAT, Namazi SS, Khalili MD, Saeedi MJTm. A new injury severity score for predicting the length of hospital stay in multiple trauma patients. 2016;21(1).
6. El-Abbassy I, Mohamed W, El-Hariri HM, El-Setouhy M, Hirshon JM, El-Shinawi MJAJoEM. Delay in hospital discharge of trauma patients in a University Hospital in Egypt: A prospective observational study. 2021;11(4):459-63.
7. Aneja A, Landy DC, Mittwede PN, Albano AY, Teasdall RJ, Isla A, et al. Inflammatory cytokines associated with outcomes in orthopedic trauma patients independent of New Injury Severity score: A pilot prospective cohort study. *J Orthop Res*. 2021 Nov 2. Epub 2021/11/04.
8. Namas RA, Almahmoud K, Mi Q, Ghuma A, Namas R, Zaaqoq A, et al. Individual-specific principal component analysis of circulating inflammatory mediators predicts early organ dysfunction in trauma patients. *J Crit Care*. 2016 Dec;36:146-53. Epub 2016/11/05.
9. Dekker AB, Krijnen P, Schipper IB. Predictive value of cytokines for developing complications after polytrauma. *World J Crit Care Med*. 2016 Aug 4;5(3):187-200. Epub 2016/09/22.
10. Lequertier V, Wang T, Fondrevelle J, Augusto V, Duclos A. Hospital Length of Stay Prediction Methods A Systematic Review. *Medical Care*. 2021 Oct;59(10):929-38.
11. Atashi A, Verburg IW, Karim H, Miri M, Abu-Hanna A, de Jonge E, et al. Models to predict length of stay in the Intensive Care Unit after coronary artery bypass grafting: a systematic review. *Journal of Cardiovascular Surgery*. 2018 Jun;59(3):471-82.

12. Almashrafi A, Elmontsri M, Aylin P. Systematic review of factors influencing length of stay in ICU after adult cardiac surgery. *Bmc Health Services Research*. 2016 Jul 29;16.
13. Hussain A, Dunn KW. Predicting length of stay in thermal burns: A systematic review of prognostic factors. *Burns*. 2013 Nov;39(7):1331-40.
14. Seaton SE, Barker L, Jenkins D, Draper ES, Abrams KR, Manktelow BN. What factors predict length of stay in a neonatal unit: a systematic review. *Bmj Open*. 2016 2016;6(10).
15. Verburg IWM, Atashi A, Eslami S, Holman R, Abu-Hanna A, de Jonge E, et al. Which Models Can I Use to Predict Adult ICU Length of Stay? A Systematic Review. *Critical Care Medicine*. 2017 Feb;45(2):E222-E31.
16. Wei HL, Billings SA. Feature subset selection and ranking for data dimensionality reduction. *IEEE Trans Pattern Anal Mach Intell*. 2007 Jan;29(1):162-6. Epub 2006/11/17.
17. Hosmer DW, Lemeshow S. *Jr. Special topics*. 2000:260-351.
18. Neuhaus K, Ho ES, Low N, Forrest CR. *JPS. Analysis of Plastic Surgery Consultations in a High-Volume Paediatric Emergency Department: A Quality Improvement Initiative*. 2021;29(4):272-9.
19. Gardner RM, Scoville DP, West BJ, Bateman B, Cundick Jr RM, Clemmer TP, editors. *Integrated computer systems for monitoring of the critically ill. Proceedings of the Annual Symposium on Computer Application in Medical Care; 1977: American Medical Informatics Association*.
20. Zhang L, Zhang H, Ai H, Hu H, Li S, Zhao J, et al. Applications of Machine Learning Methods in Drug Toxicity Prediction. *Curr Top Med Chem*. 2018;18(12):987-97. Epub 2018/07/28.
21. Park YW, Choi YS, Kim SE, Choi D, Han K, Kim H, et al. Radiomics features of hippocampal regions in magnetic resonance imaging can differentiate medial temporal lobe epilepsy patients from healthy controls. *Sci Rep*. 2020 Nov 11;10(1):19567. Epub 2020/11/13.
22. Curtis NJ, Dennison G, Salib E, Hashimoto DA, Francis NK. Artificial Neural Network Individualised Prediction of Time to Colorectal Cancer Surgery. *Gastroenterol Res Pract*. 2019;2019:1285931. Epub 2019/07/31.
23. Bissonnette V, Mirchi N, Ledwos N, Alsidieri G, Winkler-Schwartz A, Del Maestro RF, et al. Artificial Intelligence Distinguishes Surgical Training Levels in a Virtual Reality Spinal Task. *J Bone Joint Surg Am*. 2019 Dec 4;101(23):e127. Epub 2019/12/05.
24. Houthoofd R, Ruysinck J, van der Hertten J, Stijven S, Couckuyt I, Gadeyne B, et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artif Intell Med*. 2015 Mar;63(3):191-207. Epub 2015/01/13.
25. Upadhyaya A, Hu T, Meng Z, Li X, He X, Tian W, et al. Harmonic versus LigaSure hemostasis technique in thyroid surgery: A meta-analysis. *Biomed Rep*. 2016 Aug;5(2):221-7. Epub 2016/07/23.
26. Puthuchery ZA, Gensichen JS, Cakiroglu AS, Cashmore R, Edbrooke L, Heintze C, et al. Implications for post critical illness trial design: sub-phenotyping trajectories of functional recovery among sepsis survivors. 2020;24(1):1-12.

Tables

Table 1

	Level	Overall	Long LOS	Short LOS	P-value
n		3858	2293	1565	
Age (mean (SD))		43.85 (12.48)	44.17 (12.29)	43.37 (12.75)	0.048
Operation time (mean (SD))		2.65 (2.74)	3.44 (3.01)	1.50 (1.75)	<0.001
WBC (%)	<=9.5	2766 (71.7)	1656 (72.2)	1110 (70.9)	0.401
	>9.5	1092 (28.3)	637 (27.8)	455 (29.1)	
HGB (%)	<115	553 (14.3)	457 (19.9)	96 (6.1)	<0.001
	>=115	3305 (85.7)	1836 (80.1)	1469 (93.9)	
RBC (%)	<3.8	550 (14.3)	473 (20.6)	77 (4.9)	<0.001
	>=3.8	3308 (85.7)	1820 (79.4)	1488 (95.1)	
NE (%)	<=6.3	2228 (57.8)	1367 (59.6)	861 (55.0)	0.005
	>6.3	1630 (42.2)	926 (40.4)	704 (45.0)	
Gender (%)	Female	722 (18.7)	430 (18.8)	292 (18.7)	0.975
	Male	3136 (81.3)	1863 (81.2)	1273 (81.3)	
Payment (%)	Commercial insurance	490 (12.7)	315 (13.7)	175 (11.2)	0.005
	National medical insurance	2509 (65.0)	1445 (63.0)	1064 (68.0)	
	Self-paying	859 (22.3)	533 (23.2)	326 (20.8)	
Alcohol abuse (%)	No	3109 (80.6)	1870 (81.6)	1239 (79.2)	0.072
	Yes	749 (19.4)	423 (18.4)	326 (20.8)	
Smoking (%)	No	2893 (75.0)	1737 (75.8)	1156 (73.9)	0.197
	Yes	965 (25.0)	556 (24.2)	409 (26.1)	
Diabetes mellitus (%)	No	3712 (96.2)	2174 (94.8)	1538 (98.3)	<0.001

	Level	Overall	Long LOS	Short LOS	P-value
	Yes	146 (3.8)	119 (5.2)	27 (1.7)	
Hypertension (%)	No	3425 (88.8)	2060 (89.8)	1365 (87.2)	0.013
	Yes	433 (11.2)	233 (10.2)	200 (12.8)	
Anesthesia (%)	General anesthesia	383 (9.9)	368 (16.0)	15 (1.0)	<0.001
	Local anesthesia	3475 (90.1)	1925 (84.0)	1550 (99.0)	
Bone injury (%)	No	1457 (37.8)	684 (29.8)	773 (49.4)	<0.001
	Yes	2401 (62.2)	1609 (70.2)	792 (50.6)	
Muscle injury (%)	No	1943 (50.4)	1106 (48.2)	837 (53.5)	0.002
	Yes	1915 (49.6)	1187 (51.8)	728 (46.5)	
Nerve injury (%)	No	2450 (63.5)	1340 (58.4)	1110 (70.9)	<0.001
	Yes	1408 (36.5)	953 (41.6)	455 (29.1)	
Tendon injury (%)	No	2118 (54.9)	1240 (54.1)	878 (56.1)	0.227
	Yes	1740 (45.1)	1053 (45.9)	687 (43.9)	
Vascular injury (%)	No	2087 (54.1)	987 (43.0)	1100 (70.3)	<0.001
	Yes	1771 (45.9)	1306 (57.0)	465 (29.7)	
Tourniquet use (%)	No	2464 (63.9)	1258 (54.9)	1206 (77.1)	<0.001
	Yes	1394 (36.1)	1035 (45.1)	359 (22.9)	
Time to surgery (mean (SD))		6.75 (2.86)	6.90 (2.83)	6.53 (2.90)	<0.001

Table 2

	Level	Overall	Train data	Test data	P-value
n		3858	3087	771	
Age (mean (SD))		43.85 (12.48)	43.84 (12.41)	43.88 (12.78)	0.934
Operation time (mean (SD))		2.65 (2.74)	2.65 (2.72)	2.68 (2.84)	0.806
WBC (%)	<=9.5	2766 (71.7)	2210 (71.6)	556 (72.1)	0.807
	>9.5	1092 (28.3)	877 (28.4)	215 (27.9)	
HGB (%)	<115	553 (14.3)	429 (13.9)	124 (16.1)	0.136
	>=115	3305 (85.7)	2658 (86.1)	647 (83.9)	
RBC (%)	<3.8	550 (14.3)	433 (14.0)	117 (15.2)	0.448
	>=3.8	3308 (85.7)	2654 (86.0)	654 (84.8)	
NE (%)	<=6.3	2228 (57.8)	1775 (57.5)	453 (58.8)	0.555
	>6.3	1630 (42.2)	1312 (42.5)	318 (41.2)	
Gender (%)	Female	722 (18.7)	576 (18.7)	146 (18.9)	0.900
	Male	3136 (81.3)	2511 (81.3)	625 (81.1)	
Payment (%)	Commercial insurance	490 (12.7)	401 (13.0)	89 (11.5)	0.547
	National medical insurance	2509 (65.0)	2003 (64.9)	506 (65.6)	
	Self-paying	859 (22.3)	683 (22.1)	176 (22.8)	
Alcohol abuse (%)	No	3109 (80.6)	2487 (80.6)	622 (80.7)	0.985
	Yes	749 (19.4)	600 (19.4)	149 (19.3)	
Smoking (%)	No	2893 (75.0)	2294 (74.3)	599 (77.7)	0.059
	Yes	965 (25.0)	793 (25.7)	172 (22.3)	
Diabetes mellitus (%)	No	3712 (96.2)	2979 (96.5)	733 (95.1)	0.079

	Level	Overall	Train data	Test data	P-value
	Yes	146 (3.8)	108 (3.5)	38 (4.9)	
Hypertension (%)	No	3425 (88.8)	2732 (88.5)	693 (89.9)	0.306
	Yes	433 (11.2)	355 (11.5)	78 (10.1)	
Anesthesia (%)	General anesthesia	383 (9.9)	308 (10.0)	75 (9.7)	0.889
	Local anesthesia	3475 (90.1)	2779 (90.0)	696 (90.3)	
Bone injury (%)	No	1457 (37.8)	1171 (37.9)	286 (37.1)	0.698
	Yes	2401 (62.2)	1916 (62.1)	485 (62.9)	
Muscle injury (%)	No	1943 (50.4)	1537 (49.8)	406 (52.7)	0.166
	Yes	1915 (49.6)	1550 (50.2)	365 (47.3)	
Nerve injury (%)	No	2450 (63.5)	1956 (63.4)	494 (64.1)	0.746
	Yes	1408 (36.5)	1131 (36.6)	277 (35.9)	
Tendon injury (%)	No	2118 (54.9)	1682 (54.5)	436 (56.5)	0.322
	Yes	1740 (45.1)	1405 (45.5)	335 (43.5)	
Vascular injury (%)	No	2087 (54.1)	1664 (53.9)	423 (54.9)	0.661
	Yes	1771 (45.9)	1423 (46.1)	348 (45.1)	
Tourniquet use (%)	No	2464 (63.9)	1963 (63.6)	501 (65.0)	0.498
	Yes	1394 (36.1)	1124 (36.4)	270 (35.0)	
Time to surgery (mean (SD))		6.75 (2.86)	6.78 (2.87)	6.65 (2.84)	0.261
LOS (%)	Long LOS	2293 (59.4)	1835 (59.4)	458 (59.4)	1.000

Level	Overall	Train data	Test data	P-value
Short LOS	1565 (40.6)	1252 (40.6)	313 (40.6)	

Table 3

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC-ROC
LR	0.7601	0.7732	0.7511	0.6798	0.8289	0.8053
RF	0.8093	0.8403	0.7882	0.7306	0.8783	0.8833
NB	0.7147	0.8371	0.6310	0.6079	0.8500	0.7862
SVM	0.7523	0.7700	0.7402	0.6694	0.8248	0.8206
XGboost	0.7629	0.7590	0.7655	0.6872	0.8240	0.8509

Table 4

Characteristic	OR ¹	95% CI ¹	p-value
Operation.time	1.49	1.40, 1.59	<0.001
HGB			<0.001
<i>>=115</i>	–	–	
<i><115</i>	1.77	1.31, 2.41	
Time to surgery	1.01	0.98, 1.04	0.48
Vessel injury			<0.001
<i>No</i>	–	–	
<i>Yes</i>	1.83	1.54, 2.19	
Anesthesia			<0.001
<i>Local anesthesia</i>	–	–	
<i>General anesthesia</i>	13.7	7.68, 27.2	
Payment			0.041
<i>National medical insurance</i>	–	–	
<i>Commercial insurance</i>	1.36	1.06, 1.75	
<i>Self-paying</i>	1.14	0.93, 1.40	
Bone injury			0.007
<i>No</i>	–	–	
<i>Yes</i>	1.27	1.07, 1.51	
Tourniquet use			<0.001
<i>No</i>	–	–	
<i>Yes</i>	1.43	1.19, 1.73	

¹OR = Odds Ratio, CI = Confidence Interval

Figures

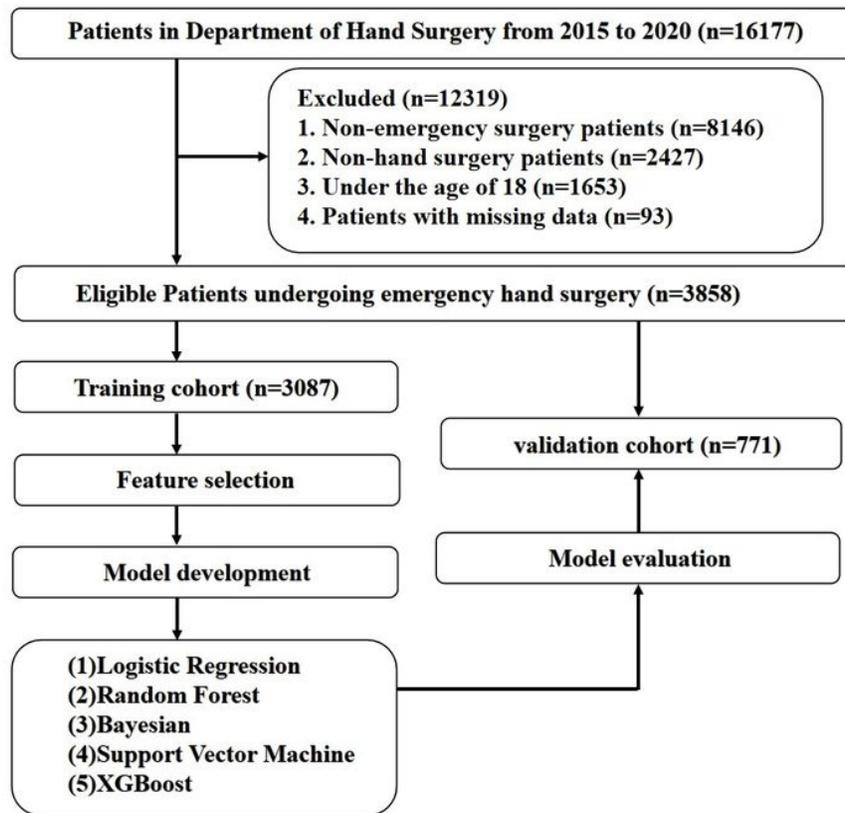


Figure 1

Study profile and analysis pipeline.

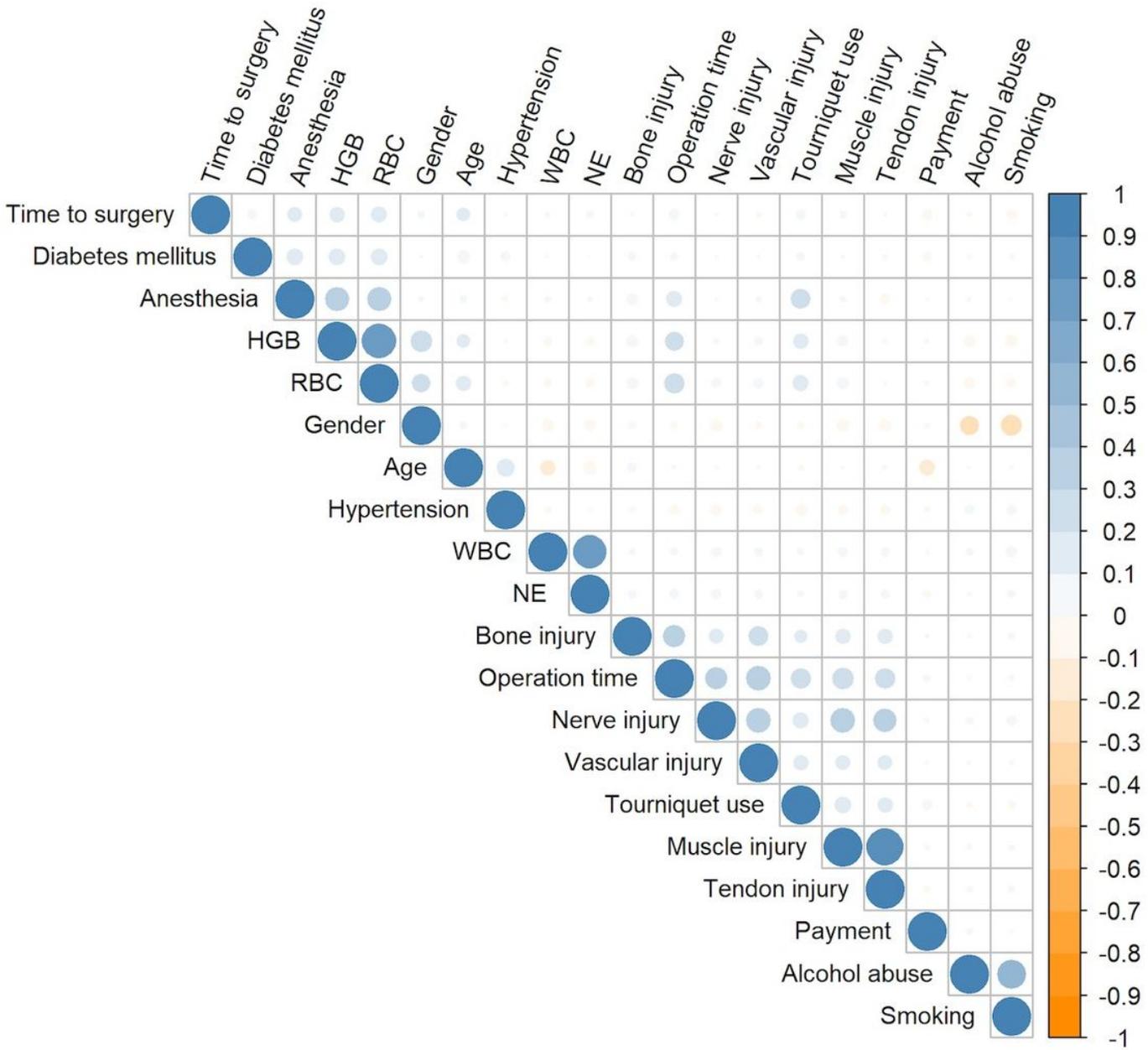


Figure 2

Correlation coefficients for each pair of variables in the data set.

The estimated correlation values are distributed within the range of -1 to 1. They are represented by color depth, with a number reaching either end value implying a stronger negative correlation and positive correlation, respectively.

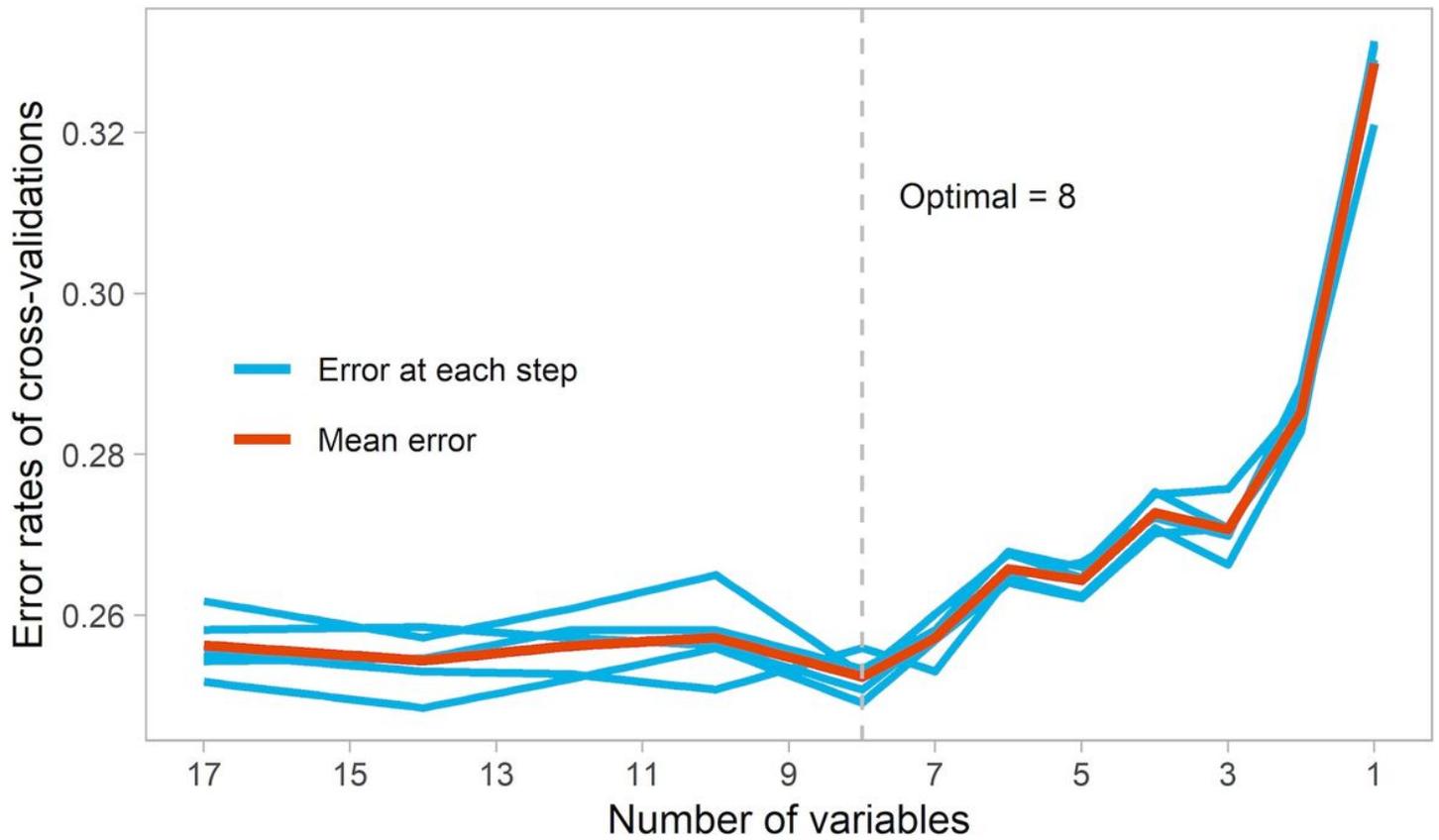


Figure 3

Ten-fold cross-validation error as a function of the number of input variables used to differentiate LOS in order of variable importance.

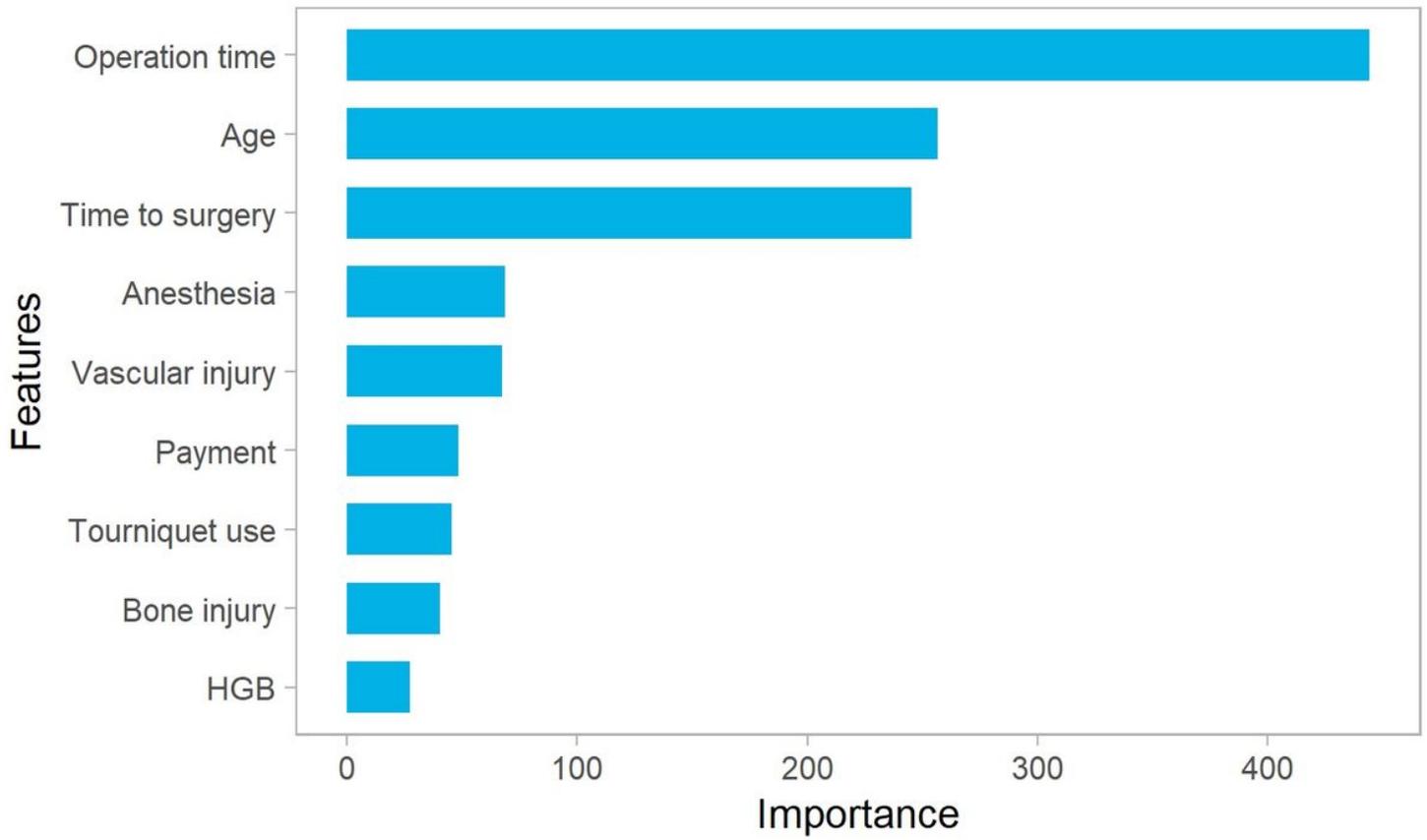


Figure 4

Variable importance of features included in random forest algorithm.

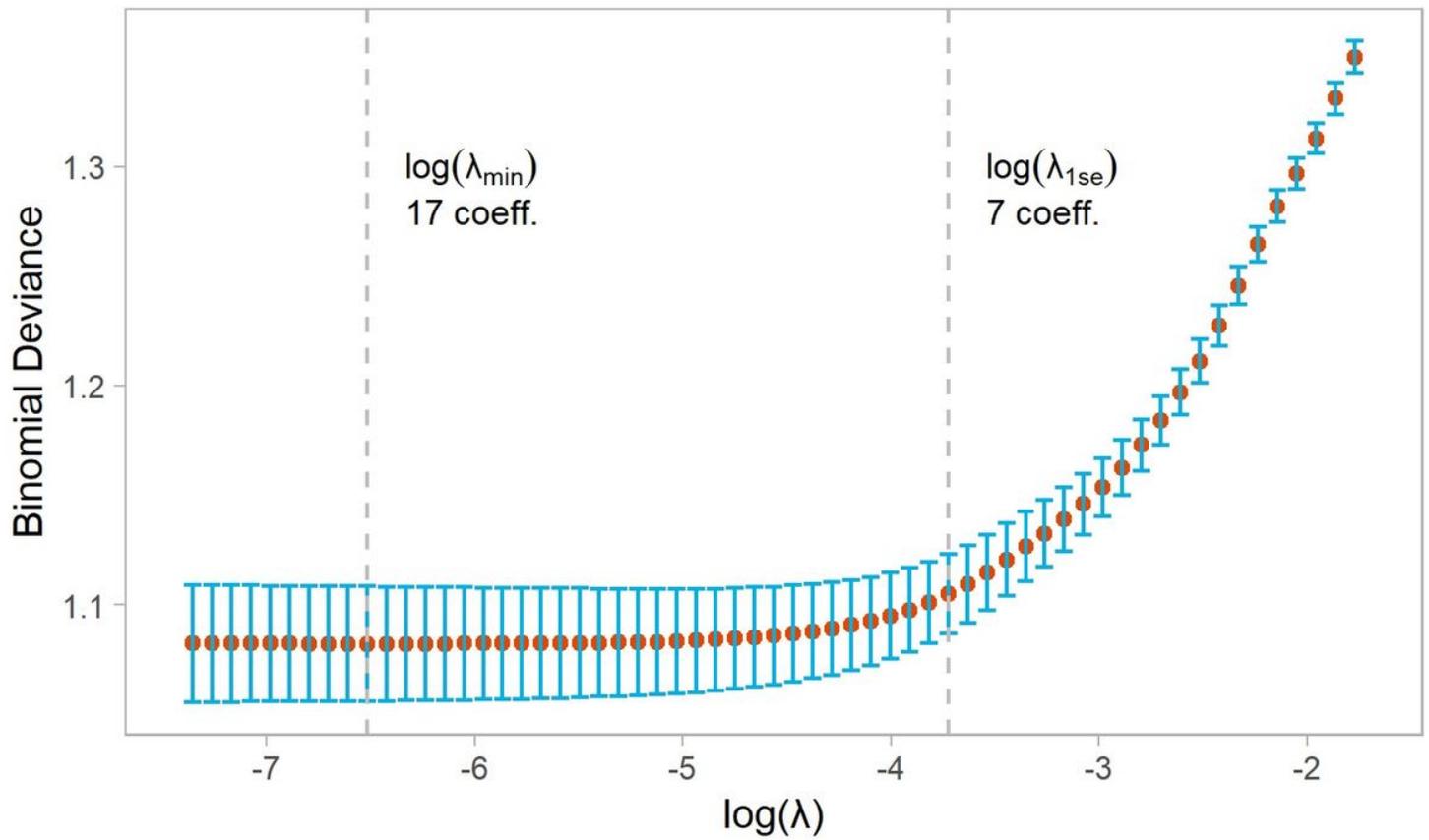


Figure 5

Feature selection in the LASSO model used tenfold cross-validation via minimum criteria.

Red-dotted vertical lines were drawn at the optimal values using the minimum criteria (minimize the mean-squared error). The value 7 represents that 19 features were reduced to 7 nonzero features.

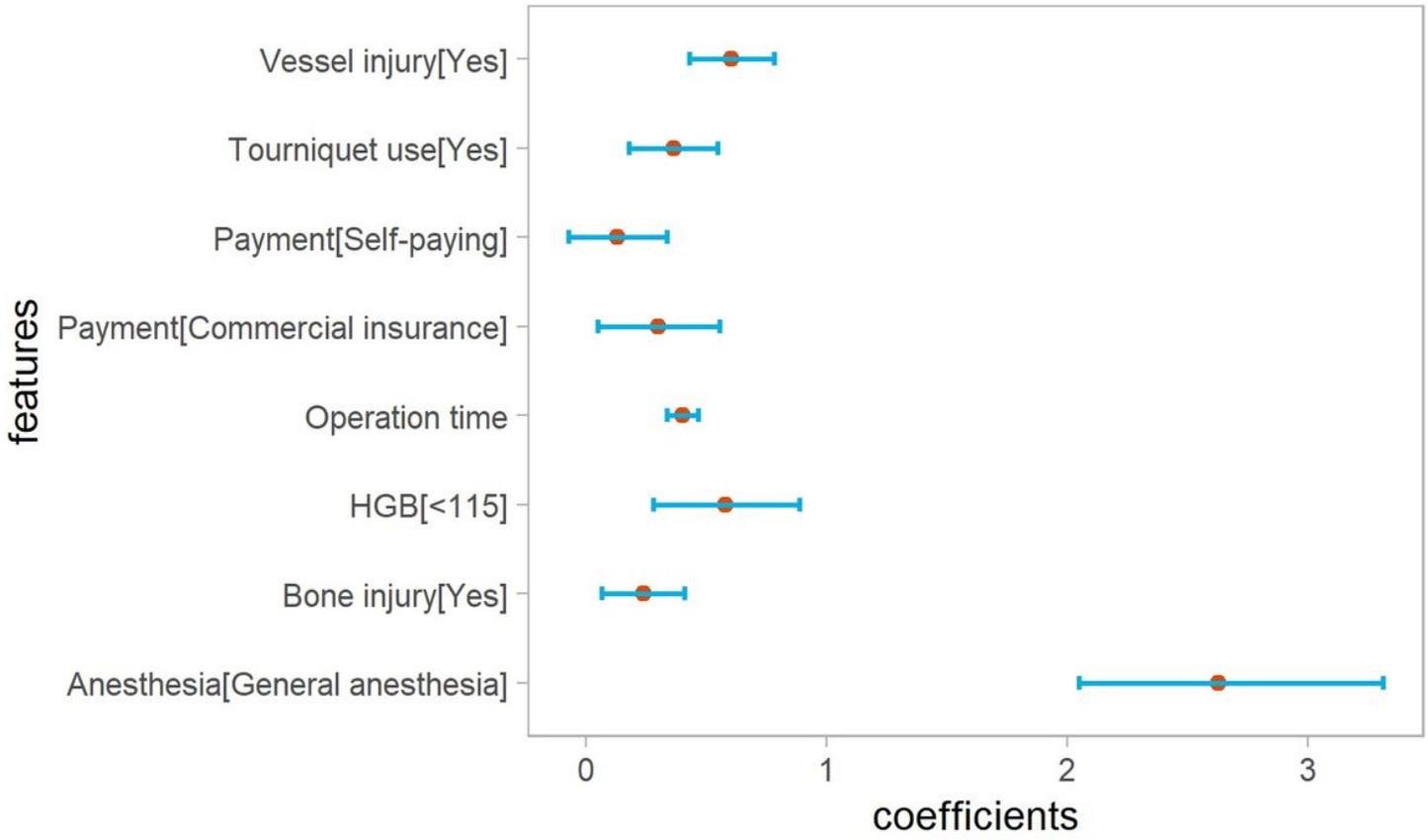


Figure 6

Coefficients of the multidimensional binary LR model.

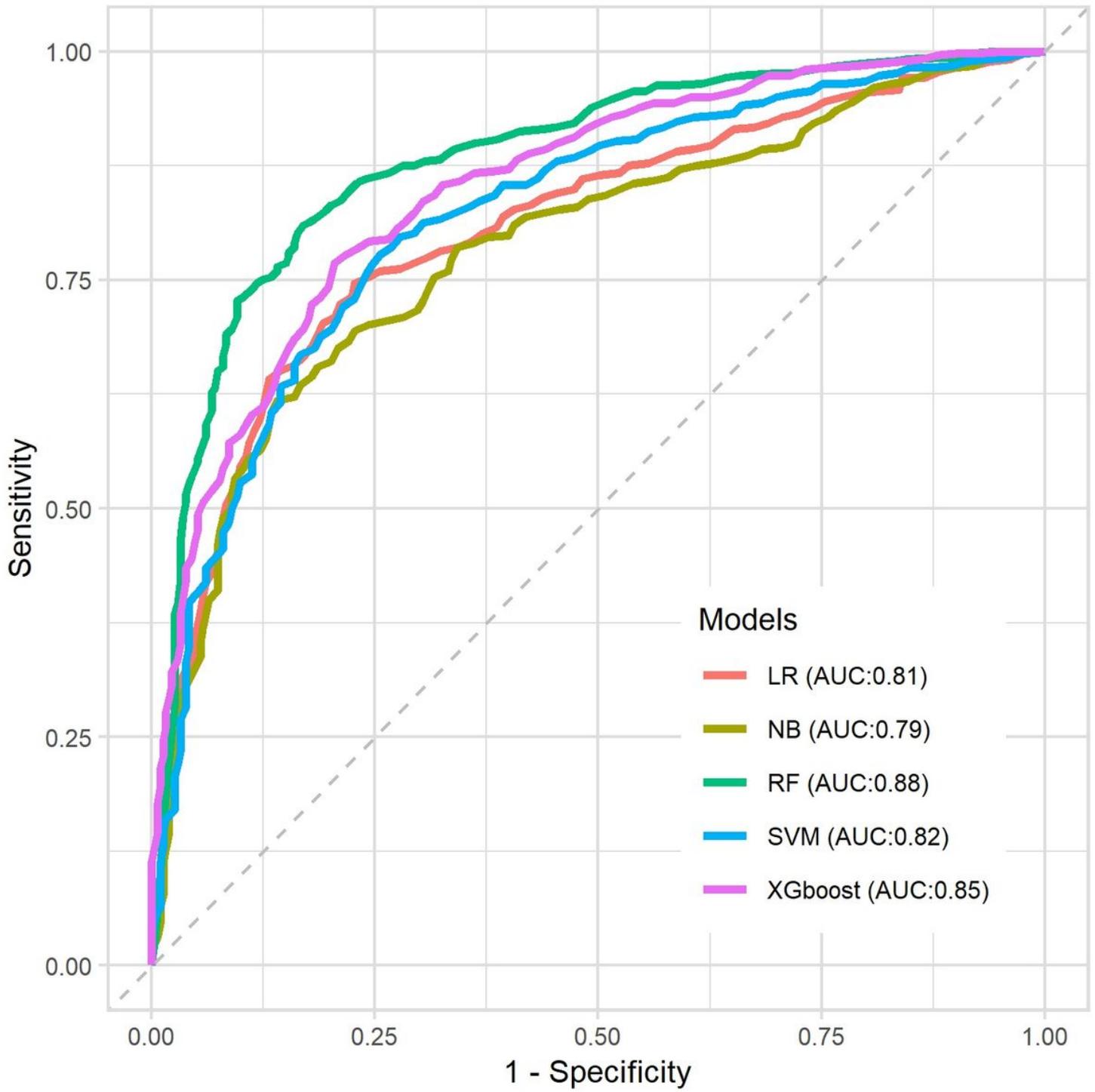


Figure 7

ROC plots for four machine learning methods and the LR method.

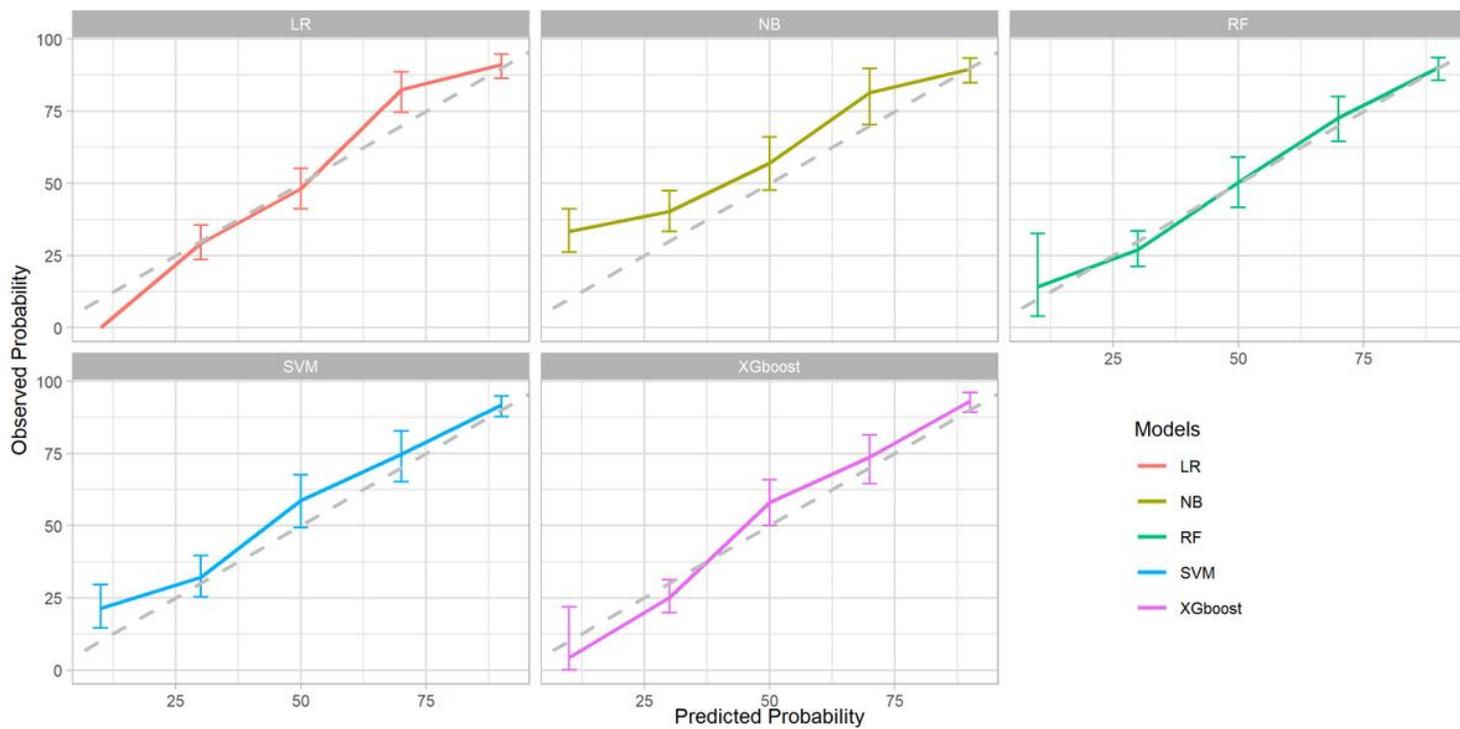


Figure 8

Calibration curve of the four machine learning and LR models shows the relationships between observed and predicted LOS. Data are presented as mean and 95% CIs.