Journal of Pediatric Surgery 56 (2021) 2165–2171

Contents lists available at ScienceDirect

Journal of Pediatric Surgery

journal homepage: www.elsevier.com/locate/jpedsurg.org

Understanding risk factors for postoperative mortality in neonates based on explainable machine learning technology

Yaoqin Hu[a], Xiaojue Gong[a], Liqi Shu[b], Xian Zeng[c], Huilong Duan[c], Qinyu Luo[a], Baihui Zhang[a], Yaru Ji[a], Xiaofeng Wang[a], Qiang Shu[a,*], Haomin Li[a,*]

[a] The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China
[b] Rhode Island Hospital, Brown University, United States
[c] The College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

ARTICLE INFO

Article history:
Received 16 December 2020
Revised 23 March 2021
Accepted 23 March 2021

Keywords:
Neonatal surgery
Postoperative mortality
Machine learning

ABSTRACT

Purpose: : We aimed to introduce an explainable machine learning technology to help clinicians understand the risk factors for neonatal postoperative mortality at different levels.
Methods: : A total of 1481 neonatal surgeries performed between May 2016 and December 2019 at a children's hospital were included in this study. Perioperative variables, including vital signs during surgery, were collected and used to predict postoperative mortality. Several widely used machine learning methods were trained and evaluated on split datasets. The model with the best performance was explained by SHAP (SHapley Additive exPlanations) at different levels.
Results: : The random forest model achieved the best performance with an area under the receiver operating characteristic curve of 0.72 in the validation set. TreeExplainer of SHAP was used to identify the risk factors for neonatal postoperative mortality. The explainable machine learning model not only explains the risk factors identified by traditional statistical analysis but also identifies additional risk factors. The visualization of feature contributions at different levels by SHAP makes the "black-box" machine learning model easily understood by clinicians and families. Based on this explanation, vital signs during surgery play an important role in eventual survival.
Conclusions: : The explainable machine learning model not only exhibited good performance in predicting neonatal surgical mortality but also helped clinicians understand each risk factor and each individual case.

1. Introduction

The first 28 days of life, which is also known as the neonatal period, represent the most vulnerable time for a child's survival. Overall, 2.7 million neonatal deaths were reported in the 2015 Global Report of Neonatal Mortality, accounting for 45% of under-five deaths [1]. The leading cause of mortality in infants in 2011 was congenital anomalies with a rate of 126.1/100,000 births [2]. Many neonates with critical congenital anomalies require surgery as soon as possible after birth. Advances in the sophistication of operative and anesthetic techniques and progress in pre- and post-operative care have contributed to the improved survival of infants undergoing surgery in the neonatal period. However, neonates exhibit increased risks of morbidity and mortality after surgery compared with any other age group [3]. Understanding the mortality risk factors for neonatal surgery and providing prediction tools for clinicians before deleterious outcomes ensue will help to save more neonates.

Several studies have developed logistic regression-based prediction models for postoperative mortality in neonates undergoing major noncardiac surgery [4,5]. A super learning model that selects the optimal regression algorithm among all weighted combinations of a set of candidate algorithms was also introduced in this task and provided improved or equivalent performance compared with individual regression and machine learning algorithms for predicting neonatal surgical mortality [6]. One of these studies used a hospital database (Kids' Inpatient Database), whereas the others used the American College of Surgeons' National Surgical Quality Improvement Program Pediatric (ACS-NSQIP-P) database [7], which contains standardized, validated preoperative clinical data and procedural data. Although these studies have demonstrated the potential of machine learning to predict postoperative neonatal mor-

Level of Evidence: Level III

* Corresponding authors.
E-mail address: hmli@zju.edu.cn (H. Li).

https://doi.org/10.1016/j.jpedsurg.2021.03.057
0022-3468/

tality, there are still some areas that have not been studied. First, none of these studies included vital signs during surgery in the prediction model. Second, although the traditional logistic regression models and super learning model developed in these studies achieved good discrimination, a good tool for clinicians to understand the specific risk of particular patients and quantitatively define the risk at different levels remains lacking.

With the development of clinical information systems, abundant high-dimensional dynamic data, such as vital signs during surgery, have been captured, and some of these data can be openly accessed [8]. There is untapped potential for data science to utilize more perioperative data to develop prediction models for postoperative mortality in neonates. In addition, machine-learning techniques have recently achieved great advances [9,10]. In this study, we included vital signs during surgery in the prediction model and introduced an explainable machine learning method that will help clinicians understand risk features not only at the feature level but also at the patient level.

2. Materials and Methods

2.1. Study Population and Data Description

This study was approved by the Institutional Review Board of the Children's Hospital of Zhejiang University School of Medicine with a waiver of informed consent. A total of 1481 neonatal surgeries performed between May 2016 and December 2019 at the children's hospital were enrolled in the present analyses based on the inclusion criteria of age at surgery $\leq$28 days.

All perioperative variables were directly extracted from related clinical information systems or manually reviewed by authors from the medical records. In addition to demographic, preoperative clinical, and procedural variables as well as postoperative ICU stay length and in-hospital mortality, similar to other related studies, we included vital signs, such as heart rate, blood pressure (SBP/DBP/MBP), temperature and pulse oxygen saturation (SPO_2) during the surgery, which were captured at 5-minute intervals by an anesthetic clinical information system. There were different lengths of such temporal series data for different surgeries. In this study, we used the summary features (mean and standard deviation) of these types of data as input to the prediction model. We also collected the features of birth of these neonates, such as the weight at birth, gestational age, cesarean section, amniotic fluid features and Apgar score at 1 min and 5 min.

2.2. Statistical analysis

The neonatal surgeries were categorized according to their in-hospital mortality status. Continuous variables (such as age and weight) of the patients who died or survived were reported as the mean$\pm$SD and were compared using the Mann-Whitney U test. Categorical variables (such as sex) were reported as counts (percentages) and compared using the chi-square test. A p-value < 0.05 was considered statistically significant. Risk ratios were calculated by unconditional maximum likelihood estimation, and their confidence intervals were calculated using a normal approximation.

2.3. Machine Learning and Explainable Method

All of the datasets were randomly divided into training and testing subsets using the python scikit-learn package [11]. The training dataset contains 80% of the cases, and the test dataset contains the remaining 20% cases. The missing values, as shown in Table 1, were imputed by modeling each feature with the missing values as a function of the other features in a round-robin fashion, which was implemented as IterativeImputer in the python scikit-learn package [12]. Several widely used machine learning methods, such as random forest, XGBoost, logistic regression, Gaussian naïve Bayes and k-neighbors, were used to train on the same training dataset and were validated on the test dataset. These five machine learning models were also from the scikit-learn python package.

To shift the prediction earlier, all characteristics shown in Table 1 except the "ICU stay length" were used as input features. After converting the category features using a one-hot encoding scheme, 261 features, including 211 distinct primary diagnoses (ICD-10 coded), were used in the prediction model. The performance of these models on the test dataset was evaluated by AUC (area under the ROC curve). The best performing model was enhanced with an interpretation method called SHAP, which is a game-theoretic approach to explain the output of any machine learning model by computing each feature to the prediction [13]. This explainable machine learning model will help clinicians understand the risk factors for a single prediction, a single variable and the entire dataset. These explanations have the potential to generate human actionable knowledge to improve clinical outcomes.

3. Result

A total of 1481 neonatal surgical cases were included in this study. The mean age at surgery was 8.51 $\pm$ 7.55 days. Greater than 58% of the neonates were male, and approximately 24% were born preterm. There were 211 distinct primary diagnosis ICD-10 codes among these neonates, and the top 5 diagnoses (ICD10, percentage) were congenital malformation of the anus (Q43.901, 8.17%), intestinal obstruction of newborn (P76.900, 6.75%), omphalos (Q79.201, 5.60%), necrotizing enterocolitis of newborn (P77.x00, 4.93%) and congenital heart disease (Q24.900, 4.79%). More diagnostic characteristics are shown in supplementary Table S1.

A total of 50 neonates died in the hospital after surgery with a mortality rate of 3.4% in the study cohort. The mean ICU stay after surgery was 8.79 $\pm$ 17.77 days. Most perioperative characteristics varied significantly between patients who died in the hospital and patients who survived (Table 1). Patients with a high risk of postoperative mortality were preterm and had a lower body weight, especially at birth; a lower blood pressure before surgery; bloody amniotic fluid; a lower Apgar score, especially at 5 min; a higher ASA class; a longer surgery duration; and worse vital signs or unstable vital signs during surgery. Moreover, there were no statistically significant differences in sex, age, delivery method, or meconium-stained amniotic fluid (MSAF) grades.

The mortality risk ratios of the different types of surgery are shown in Table 2. Cardiac surgery was the highest-risk surgery in neonates with a risk ratio of 2.938 [1.616, 5.341] compared with lower GIT surgery, which is the main type of surgery in the neonatal period in this children's hospital.

The random forest machine learning model achieved the best performance among the 5 machine learning models, as shown in Fig. 1. Random forest is one of the most popular tree-based supervised learning algorithms. It is also the most flexible and easiest to use. The random forest model was also reported as the best model in a previous super learning study [6]. Then, the random forest machine learning model was trained on all of the datasets, and SHAP TreeExplainer was used to explain these risk factors in the trained model.

The top 20 most important risk factors for neonatal postoperative mortality are shown in Fig. 2. The SHAP value (a value represents the impact of a feature on the model output) and colored corresponding feature value explained the feature contributions in the prediction. If clinicians want to explore the detailed relationship between these factors and the risk of postoperative mortal-

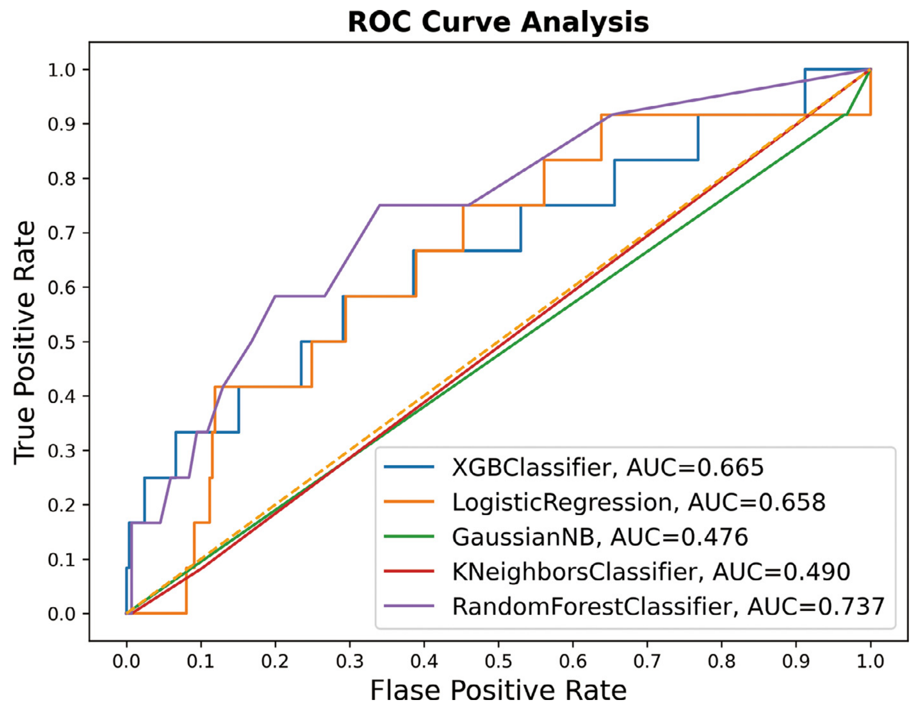

Fig. 1. The performance of different machine learning models on the test dataset. A larger AUC (area under the ROC curve) indicates better discrimination of the model.

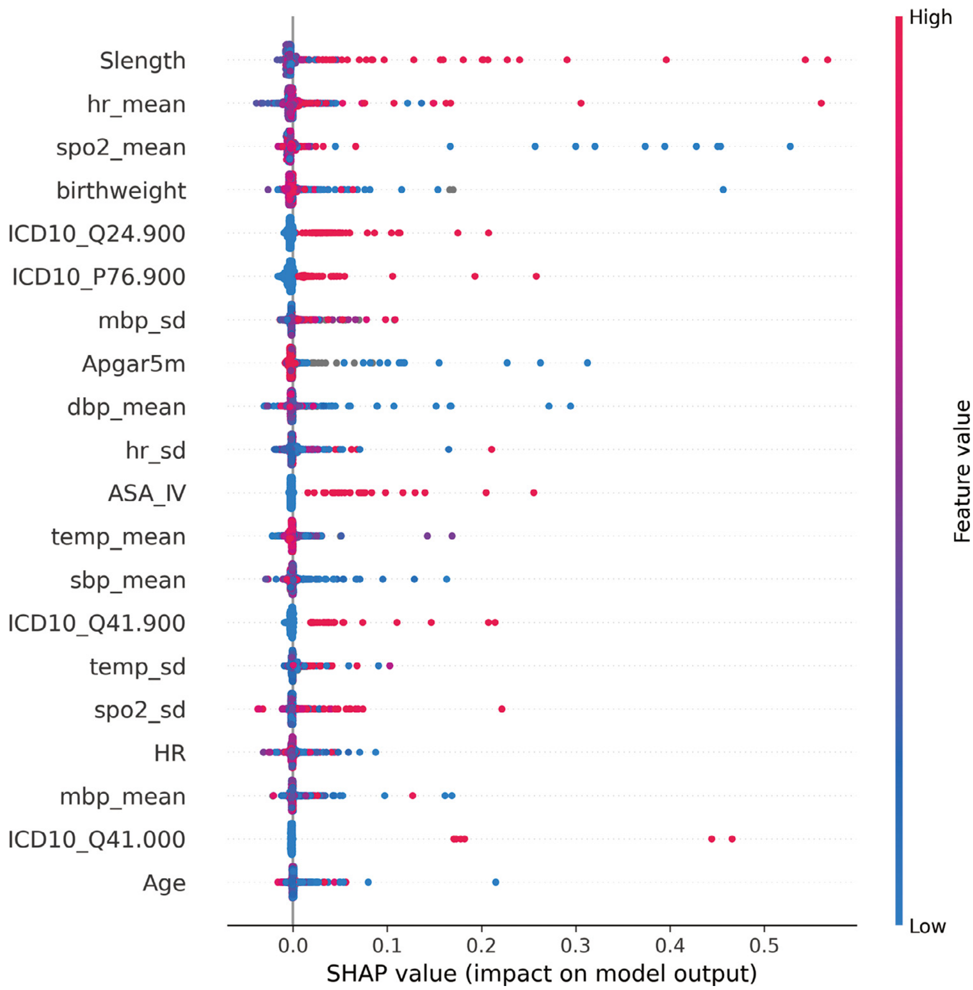

Fig. 2. Top 20 most important risk factors for neonatal postoperative mortality. The color of the points in the beeswarm plot indicate feature values.

Table 1
perioperative characteristics of the patient cohorts.

Characteristic	Died (n = 50)	Survived (n = 1431)	p-value	Missing rate
Gender			1	0.00%
Female	21 (42.0%)	601 (42.0%)		
Male	29 (58.0%)	830 (58.0%)		
Age in days	8.72 ± 7.01	8.50 ± 7.57	0.838	0.00%
Weight at surgery(kg)	2.74 ± 0.77	2.95 ± 0.70	0.038	0.07%
Weight at birth (g)	2737.96 ± 875.58	2975.92 ± 692.46	0.019	2.23%
SBP before surgery (mmHg)	64.66 ± 10.71	69.80 ± 12.71	0.005	0.47%
Heart Rate before surgery	150.70 ± 17.06	145.96 ± 17.17	0.055	0.27%
Gestational age class (%)			0.002	4.93%
Extremely Preterm	1(2.0%)	7 (0.5%)		
Very preterm	7(14.3%)	56 (4.1%)		
Late preterm	11(22.4%)	269 (19.8%)		
Term	30(61.2%)	1027 (75.6%)		
Cesarean Delivery	27 (54.0%)	645 (47.5%)	0.444	4.86%
Amniotic fluid (%)			0.004	44.56%
Grade 1 MSAF	0 (0.0%)	45 (5.5%)		
Grade 2 MSAF	0 (0.0%)	22 (2.7%)		
Grade 3 MSAF	0 (0.0%)	30 (3.7%)		
Bloody	3 (10.7%)	13 (1.6%)		
No special	25 (89.3%)	703 (86.5%)		
APGAR score at 1 min	8.43 (2.67)	9.20 (1.59)	0.013	55.03%
APGAR score at 5 min	8.79 (2.19)	9.61 (1.03)	<0.001	56.04%
ASA class			< 0.001	0.54%
ASA-I	0(0.0%)	95(6.7%)		
ASA-II	17(34.0%)	941(66.1%)		
ASA-III	25(50.0%)	366(25.7%)		
ASA-IV	6(12.0%)	21(1.5%)		
ASA-V	2(4.0%)	0(0.0%)		
Surgery Duration(mins)	119.30 ± 88.92	82.88 ± 54.17	<0.001	0.00%
Vital Signs in Surgery				
Mean of heart rate	146.56 ± 19.96	143.59 ± 13.77	0.148	0.81%
SD of heart rate	19.61 ± 13.58	16.07 ± 10.00	0.017	0.81%
Mean of temperature	33.28 ± 2.10	34.31 ± 1.81	0.002	48.28%
SD of temperature	1.84 ± 1.88	1.12 ± 1.42	0.009	48.35%
Mean of SBP	51.21 ± 7.40	57.00 ± 8.87	<0.001	30.11%
SD of SBP	10.47 ± 5.61	8.44 ± 4.63	0.008	30.18%
Mean of DBP	33.53 ± 8.45	36.67±9.13	0.037	30.11%
SD of DBP	7.41 ± 4.74	6.04 ± 3.69	0.027	30.11%
Mean of MBP	41.53 ± 7.63	45.95 ± 9.27	0.004	30.11%
SD of MBP	8.55 ± 4.86	7.09 ± 3.91	0.025	30.11%
Mean of SPO2	92.80 ± 6.86	95.15 ± 3.45	<0.001	0.88%
SD of SPO2	6.23 ± 4.46	4.17 ± 3.09	<0.001	1.01%
Type of surgery			<0.001	0.00%
Lower GIT	29 (58.0%)	789 (55.1%)		
Upper GIT	2 (4.0%)	171 (11.9%)		
Cardiac	15 (30.0%)	129 (9.0%)		
Thoracic	1 (2.0%)	120 (8.4%)		
Neurosurgery	2 (4.0%)	74 (5.2%)		
cranioplasty	0 (0.0%)	28 (2.0%)		
gynecological	0 (0.0%)	21 (1.5%)		
Plastic	1 (2.0%)	10 (0.7%)		
Other	0 (0.0%)	89 (6.2%)		
ICU stay(days)	12.51 ± 17.63	8.66 ± 17.76	0.136	1.49%

ity, the model can explain the risk factors at the feature level, as shown in Fig 3. As noted in Fig 3A, clinicians can easily observe that a length of surgery greater than 3 h (240 mins) will definitely increase the risk of postoperative mortality. Given that we occasionally need to consider the interaction effects among risk factors, the model will also automatically color the feature dependence plot with the strongest interaction. In Fig 3A, the surgery length interacted with the cardiac surgery, and most of the surgeries longer than 3 h were cardiac surgeries. The visualized explanations of other risk factors are shown in the supplemental material for reference. Based on this intuitive visualization, clinicians can understand not only which factors impact mortality but also how they quantitatively impact the outcome and define the threshold value (such as surgery length < 3 h) to guide clinical practice. Based on these feature dependence plots, the systolic blood pressure during the surgery should be maintained above 50 mmHg (supplementary Figure S3), the pulse oxygen saturation (SPO_2) during surgery should be maintained above 80% (supplementary Figure S6), and the heart rate should not exceed 150 bpm (supplementary Figure S9). If there are such events during surgery, clinicians should pay more attention to avoid them or prepare for potential mortality.

Furthermore, explainable machine learning identified more risk factors than traditional statistical analysis. For example, the age at surgery did not show a significant difference at the statistical level, as shown in Table 1. However, if we look closely at the correlation between age and SHAP value, as shown in Fig 3B, we notice that younger neonates and a few elderly neonates had relatively higher values. Surgery too early or too late for neonates may be risk factors for some conditions.

In addition to visualizing the features on all datasets, the model can visualize feature attributions as "forces" in each instance, as

Table 2
Type of surgery and their risk ratio of neonatal postoperative mortality.

Type of Surgery	Survived	Died	Risk Ratio (CI)	Chi.square *p*-value
Lower GIT	789(55.14%)	29(58.00%)	1 [Na, Na]	Na
Upper GIT	171(11.95%)	2(4.00%)	0.326 [0.079,1.354]	0.101
Cardiac	129(9.01%)	15 (30.00%)	2.938 [1.616,5.341]	<0.001
Thoracic	120(8.39%)	1(2.00%)	0.233 [0.032,1.696]	0.112
Neurosurgery	74(5.17%)	2(4.00%)	0.742 [0.181,3.051]	0.677
Cranioplasty	28(1.96%)	0(0.00%)	0.000 [0.000, Na]	0.311
Gynecological	21(1.47%)	0(0.00%)	0.000 [0.000, Na]	0.380
Plastic	10(0.70%)	1(2.00%)	2.564 [0.383,17.190]	0.328
other	89(6.22%)	0(0.00%)	0.000 [0.000, Na]	0.071

shown in Fig 4. Each feature value is a force that either increases or decreases the risk of mortality. The prediction starts from the baseline, which is the average risk of mortality in all study patients. In the plot, each feature value is an arrow that pushes to increase (positive value in red) or decrease (negative value in blue) the prediction. These forces balance each other out for the actual prediction of the instance. For example, as a low-risk neonate shown in Fig 4A, the main risk factor was age (age=0). However, diagnosis, surgery length, heart rate, and Apgar score decreased the risk of mortality after surgery. For the high-risk neonate shown in Fig 4B, the high fluctuation of SPO_2, diagnosis (omphalos), and low systolic blood pressure (SBP_mean=47.13) increased the risk of postoperative mortality. Based on these explanations, clinicians will understand more about the nature of the prediction, which will help them to control preventable risk factors in the future. All 50 patients who died are explained in the supplemental material. Each patient who died presented unique feature patterns that contributed to death. Among the top 1 impact features in these 50 cases, 18 (36%) had poor or unstable vital signs during surgery, 16 (32%) were diagnosed with certain conditions, 5 (10%) had high-level ASA classes, 5 (10%) had a long length of surgery, 2 (4%) had a low Apgar score at 5 min, 1 (2%) was too small for age and 1 (2%) had bloody amniotic fluid.

4. Discussion

As shown in Table 1 and Fig 2, vital signs, such as heart rate, blood pressure, temperature, and SPO_2, during surgery are important predictors of a high risk of death after surgery in neonates. These vital signs were severely affected in 36% of patients who died. In an additional test, the random forest models trained with and without vital sign data were compared. The area under the receiver operating characteristic curve (AUC) decreased from 0.701 to 0.596 when vital signs were not used (supplemental material Figure S21). The main reason for the lack of such factors in previous related studies is that such temporal series data are not collected in traditional databases for many reasons. First, the data size will increase to a level that cannot be easily handled by traditional data analysis methods. In this study, there were 282,360 vital sign observations among these patients. The temporal lengths also vary among different surgeries. These factors make it difficult for these data to be captured in databases and managed by traditional data table structures. Second, how to use these time-series data remains a challenge for traditional data analysis. In this study, the summary of these data is presented in a crude fashion and does not take complete advantage of all of the information available in them. Typically, recurrent neural networks have been applied in processing sequential data in many domains [14,15]. Despite its great success in these scenarios, significant barriers remain when applying it in this task since it requires regularly spaced data points, whereas the physiological data recorded are often sparse, noisy and incomplete. In this study, the missing rate of vital signs varied from 0.8% to 48%. More complicated models or feature extraction methods should be developed for these types of data in the future.

Machine learning methods, including logistic regression and super learning, have been reported to be good at predicting surgical mortality in neonates [4–6]. Different models trained on different datasets with different clinical features have reported different performances. Lillehei CW et al. reported an AUC of 0.9 in neonatal noncardiac surgical procedures [4]. The major limitation of that study is that it excluded cardiac surgery. As shown in Table 2, cardiac surgery, which is the riskiest surgery in neonates, contributed to approximately 30% of neonate deaths after surgery. Stey AM et al. reported an AUC of 0.77 when predicting postoperative adverse events in neonates [5]. Mortality is only a small component of adverse events, including transfusion, reintubation, and surgical site infection. The super learning model introduced by Cooper JN et al. reported an impressive AUC=0.87 in a relatively large dataset [6]. There are some reasons that contribute to such good performance. First, more mortality-associated clinical features were collected, such as ventilator dependence, oxygen support, nutritional support, inotropic support, open wounds, CPR or ECMO in the previous 7 days, transfusion in the previous 48 h, and the status of the patient, including the risk of cardiac (none, minor, major, severe), the grade of intraventricular hemorrhage (0–4 grade), and severe sepsis in the previous 48 h (SIRS, sepsis, septic shock). Some of these data require expert effort to evaluate in the patient and to be collected into a dedicated database. In this study, we only used the currently available EMR data with limited input features. Second, a relatively larger dataset (6499 cases) was used to train the model in that study. The size of the training data is critical for the success of any machine learning model. Third, the super learning approach, which selects the optimal regression algorithm among all weighted combinations of a set of candidate algorithms, can always achieve better results than a single algorithm. However, it also made the training of the model and explaining the prediction more complicated. In this study, we introduced the explainable machine learning technology SHAP to convert the “black-box” models into friendly visualized explanations. This approach explains feature contributions not only at feature levels but also at instance levels. The feature-level explanation will help clinicians understand the feature and generate practical knowledge about how to control the preventable risk at a specific level in practice. The odds ratio provided by traditional logistic regression does not have the capability to directly generate such knowledge. The instance-level explanation will help clinicians make decisions when faced with specific real instances that do not always equal the average results of the population. A “what-if” approach can be used before surgery based on mimic data to identify the potential risk(s) and prepare for them. Such a personalized prediction has the potential to manage risk in a more precise manner. We also noticed that such an approach can identify additional risk factors compared with traditional statistical analysis. As an explainable approach, clinicians can easily assess the reasonableness

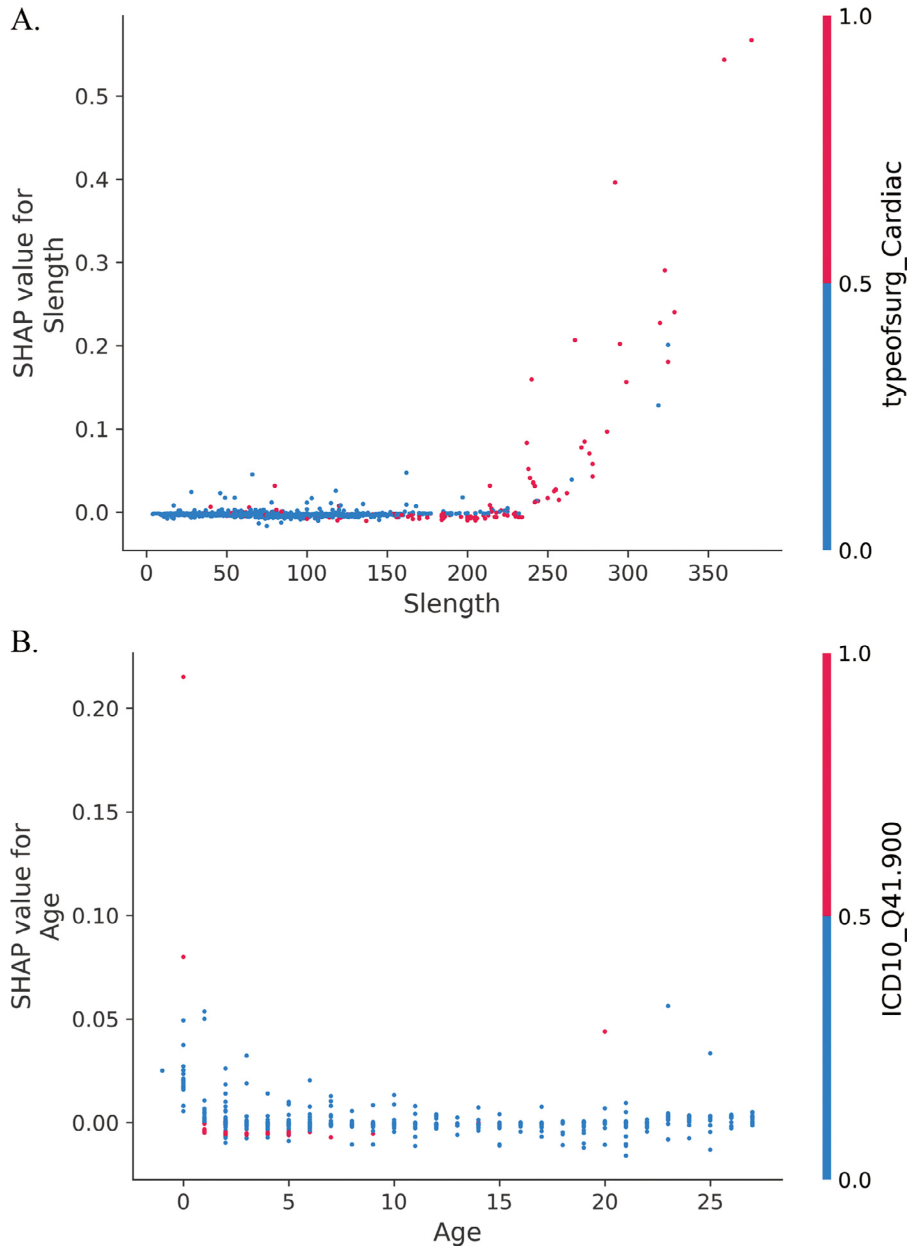

Fig. 3. Understanding of risk factors based on the SHAP dependence plot. A. How the model depends on the length of surgery. B How the model depends on the age of neonates at surgery.

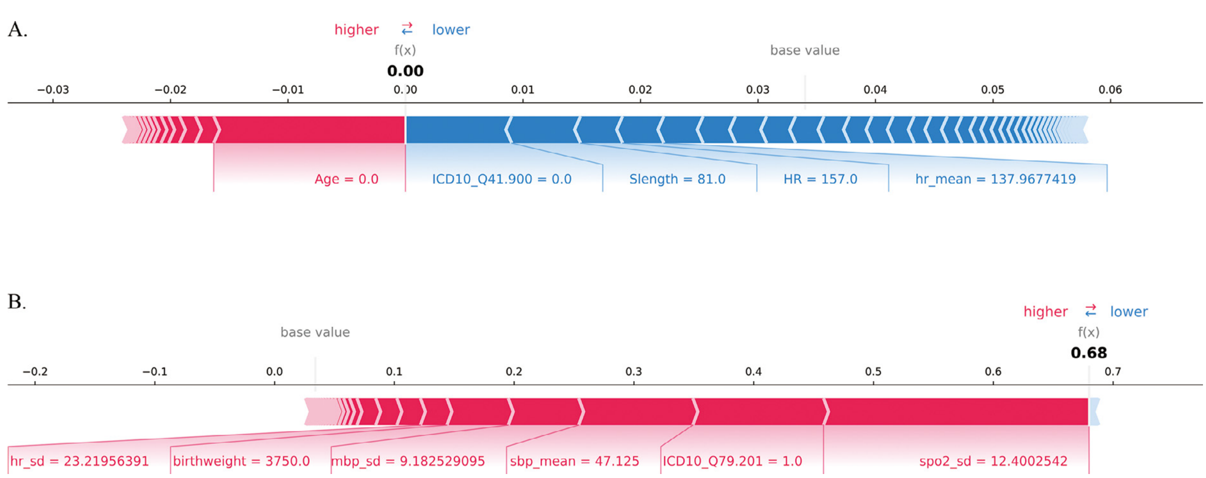

Fig. 4. Explaining instance risk factors. A. A neonate with a low risk of postoperative mortality. B. A neonate with a high risk of postoperative mortality.

of these newly identified factors based on their experience and knowledge.

Several studies in both adult and pediatric surgery have shown that hospital and surgeon characteristics are associated with patient outcomes after surgery [16,17]. Therefore, the use of machine learning technology should consider that the training dataset should be collected from similar hospitals or the same hospital as the practice hospital. The period of data generation may also affect the model performance. Different hospitals may obtain different models and different explanations of the model. A machine learning model trained on datasets from multiple centers should also consider hospitals and surgeons as features. When focused on improving local healthcare quality, we recommend training the machine learning model based on real-world data from a single center, not a registry database.

This study had several limitations. First, although the features used in this study included vital signs during surgery, there was a lack of many important features included in ACS-NSQIP-P, such as transfusion, ECMO and the sepsis status of the neonates. In this study, we also noticed that multiple surgeries conducted on neonates also increased the risk of death (the mortality rate increased to 7.69%). In addition, many neonates had multiple birth defects, but only the primary diagnosis was used in the prediction. In summary, we did not capture all of these important features for different reasons, and including these features in the prediction model will achieve better performance in the future. Second, only relatively small datasets ($n = 1481$) were available to train and validate the method. These limitations also partially explain the performance of the machine learning model in this study, which is not as good as some previous studies. After all, our focus in this study was on using explainable machine learning techniques to provide clinicians a better understanding of risk factors at different levels.

Conclusions

Vital signs during surgery contain important information about the status of neonates and should be included in the neonatal postoperative mortality prediction model. Explainable machine learning not only provides good performance in predicting neonatal surgical mortality but also helps clinicians understand each risk factor and each individual case. These explanations can be used to inform preoperative preparation of the patient or caregiver team and could shed light on potential resource needs for postoperative care. Given that the explanation is so straightforward, it can also help families understand the realistic risk of the surgery.

Funding

This work was supported by the National Natural Science Foundation of China [81871456, 81701944] and the Natural Science Foundation of Zhejiang Province [LY20H100005].

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jpedsurg.2021.03.057.

References

[1] UNICEF Committing to Child Survival: A Promise Renewed. 2015; 2015.
[2] MacDorman MF, Hoyert DL, Mathews TJ. Recent Declines in Infant Mortality in the United States, 2005-2011. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2013.
[3] Bucher BT, Duggan EM, Grubb PH, France DJ, Lally KP, Blakely ML. Does the American College of Surgeons National Surgical Quality Improvement Program pediatric provide actionable quality improvement data for surgical neonates? J Pediatr Surg 2016;51:1440–4. doi:10.1016/j.jpedsurg.2016.02.084.
[4] Lillehei CW, Gauvreau K, Jenkins KJ. Risk adjustment for neonatal surgery: a method for comparison of in-hospital mortality. Pediatrics 2012;130:e568–74. doi:10.1542/peds.2011-3647.
[5] Stey AM, Kenney BD, Moss RL, Hall BL, Berman L, Cohen ME, et al. A risk calculator predicting postoperative adverse events in neonates undergoing major abdominal or thoracic surgery. J Pediatr Surg 2015;50:987–91. doi:10.1016/j.jpedsurg.2015.03.023.
[6] Cooper JN, Minneci PC, Deans KJ. Postoperative neonatal mortality prediction using superlearning. J Surg Res 2018;221:311–19. doi:10.1016/j.jss.2017.09.002.
[7] Raval MV, Dillon PW, Bruny JL, Ko CY, Hall BL, Moss RL, et al. American College of Surgeons National Surgical quality improvement program pediatric: a phase 1 report. J Am Coll Surg 2011;212:1–11. doi:10.1016/j.jamcollsurg.2010.08.013.
[8] Zeng X, Yu G, Lu Y, Tan L, Wu X, Shi S, et al. PIC, a paediatric-specific intensive care database. Sci Data 2020;7:14. doi:10.1038/s41597-020-0355-4.
[9] Zeng X, An J, Lin R, Dong C, Zheng A, Li J, et al. Prediction of complications after paediatric cardiac surgery. Eur J Cardio-Thoracic Surg 2020;57:350–8. doi:10.1093/ejcts/ezz198.
[10] Shamout FE, Zhu T, Sharma P, Watkinson PJ, Clifton DA. Deep Interpretable Early Warning System for the Detection of Clinical Deterioration. IEEE J Biomed Heal Informatics 2020;24:437–46. doi:10.1109/JBHI.2019.2937803.
[11] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine Learning in Python. J Mach Learn Res 2012;12:2825–30.
[12] van Buuren S, mice Groothuis-Oudshoorn K. Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45:1–67. doi:10.18637/jss.v045.i03.
[13] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2:56–67. doi:10.1038/s42256-019-0138-9.
[14] Li K, Daniels J, Liu C, Herrero P, Georgiou P. Convolutional recurrent neural networks for glucose prediction. IEEE J Biomed Heal Informatics 2020;24:603–13. doi:10.1109/JBHI.2019.2908488.
[15] Ma F, Chitta R, Zhou J, You Q, Sun T, Dipole Gao J. Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In: Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. ACM; 2017. p. 1903–11. doi:10.1145/3097983.3098088.
[16] McAteer JP, LaRiviere CA, Drugas GT, Abdullah F, Oldham KT, Goldin AB. Influence of surgeon experience, hospital volume, and specialty designation on outcomes in pediatric surgery. JAMA Pediatr 2013;167:468. doi:10.1001/jamapediatrics.2013.25.
[17] Reames BN, Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and operative mortality in the modern era. Ann Surg 2014;260:244–51. doi:10.1097/SLA.0000000000000375.