Clinical Nutrition 41 (2022) 202–210

Contents lists available at ScienceDirect

Clinical Nutrition

journal homepage: http://www.elsevier.com/locate/clnu

Original article

Explainable machine learning model for predicting the occurrence of postoperative malnutrition in children with congenital heart disease

Hui Shi [a, d], Dong Yang [b], Kaichen Tang [b], Chunmei Hu [c], Lijuan Li [c], Linfang Zhang [c], Ting Gong [c], Yanqin Cui [c, e, *]

[a] Guangzhou Women and Children's Medical Center, Institute of Pediatrics, Guangzhou Medical University, No.9 Jinsui Road, Zhujiang Newtown, Tianhe District, Guangzhou, 510623, China
[b] Guangzhou AID Cloud Technology, No. 68 Huacheng Avenue, Tianhe District, Guangzhou, China
[c] Cardiac Intensive Care Unit, The Heart Center, Guangzhou Women and Children Medical Center, Guangzhou Medical University, No.9 Jinsui Road, Zhujiang Newtown, Tianhe District, Guangzhou 510623, China
[d] Department of Biostatistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China
[e] Department of Pediatric Surgery, Guangdong Provincial Key Laboratory of Research in Structural Birth Defect Disease, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, 510623, Guangdong, China

ARTICLE INFO

Article history:
Received 4 June 2021
Accepted 5 November 2021

Keywords:
Malnutrition
Congenital heart disease
Machine learning
Children
Interpretability

SUMMARY

Background & aims: Malnutrition is persistent in 50%–75% of children with congenital heart disease (CHD) after surgery, and early prediction is crucial for nutritional intervention. The aim of this study was to develop and validate machine learning (ML) models to predict the malnutrition status of children with CHD. We used explainable ML methods to provide insight into the model's predictions and outcomes.
Methods: This prospective cohort study included consecutive children with CHD admitted to the hospital from December 2017 to May 2020. The cohort data were divided into the training and test data sets based on the follow-up time. The outcome of the study was CHD child malnutrition 1 year after surgery, the primary outcome was an underweight status, and the secondary outcomes were stunted and wasting status. We used five ML algorithms with multiple features to construct prediction models, and the performance of these ML models was measured by an area under the receiver operating characteristic curve (AUC) analysis. We also used the permutation importance and SHapley Additive exPlanations (SHAP) to determine the importance of the selected features and interpret the ML models.
Results: We enrolled 536 children with CHD who underwent complete repair. The proportions of children with an underweight, stunted, or wasting status 1 year after surgery were 18.1% (97/536), 12.1% (65/536), and 17.5% (94/536), respectively. All patients contributed to the generation of 115 useable features, which allowed us to build models to predict malnutrition. Five prediction algorithms were used, and the XGBoost model achieved the greatest AUC in all outcomes. The results obtained from the permutation importance and SHAP analyses showed that the 1-month postoperative WAZ-score, discharge WAZ score and preoperative WAZ score were the top 3 important features in predicting an underweight status in the XGBoost algorithm. Regarding the stunted status, the top 3 important features were the 1-month postoperative HAZ score, discharge HAZ score, and aortic clamping time. Regarding the wasting status, the top 3 important features were the hospital length of stay, formula intake, and discharge WHZ-score. We also used a narrative case report as an example to describe the clinical manifestations and predicted the primary outcomes of two children.
Conclusions: We developed an ML model (XGBoost) that provides accurate early predictions of malnutrition 1-year postoperatively in children with CHD. Because the ML model is explainable, it may better enable clinicians to better understand the reasoning underlying the outcome. Our study could aid in determining individual treatment and nutritional follow-up strategies for children with CHD.

* Corresponding author. Cardiac Intensive Care Unit, The Heart Center, Guangzhou Women and Children Medical Center, Guangzhou Medical University, Guangzhou, 510623, China.
E-mail address: cuiyanqin@gwcmc.org (Y. Cui).

https://doi.org/10.1016/j.clnu.2021.11.006

1. Introduction

Children with congenital heart disease (CHD) have been reported to present continued significant malnutrition postoperatively (50%–75%) [1–3]. Malnutrition plays a significant role in short- and long-term postoperative outcomes, including reduced muscle function, poor wound healing, impaired immunity, reduced cognition, brain dysfunction, and neurodevelopmental impairment [4–6]. Given the relatively high prevalence and pervasive impact of postoperative malnutrition in children with CHD, there is a strong emphasis on the importance of early, accurate projections of children's postoperative growth and malnutrition for informing customized medical nutrition management. The American Heart Association and the American Academy of Pediatrics also recommend systematic surveillance, evaluation, and management of developmental outcomes in children with CHD throughout childhood to promote the early detection of delays and optimize long-term outcomes [7].

Multidimensional risk factors, including genetics, residual lesions, cardiac dysfunction, energy intake, and other prenatal factors, are associated with postoperative malnutrition in children with CHD [8]. Leading to variations in the early postoperative growth trajectory of children with CHD, given such variability, identifying children at risk of developing malnutrition is challenging [4]. To date, no single diagnostic or treatment characteristic has been found to reliably predict developmental outcomes. Despite intense studies, patient- and treatment-related factors, such as diagnosis, birth history, and perioperative events, typically account for <40% of the variance in developmental outcomes in children with CHD [9]. Selecting the most informative predictors at different time points could help better predict the growth of children with CHD after surgery.

In recent years, machine learning (ML) models have demonstrated improved performance in predicting various diseases or clinical conditions [10–12]. Compared with traditional statistical methods, ML models can capture complex, nonlinear relationships and identify previously unknown correlations in big data, thereby enabling deeper insights to be gained from clinical data [13]; consequently, ML models show promising potential in clinical scenarios in which large amounts of data are collected and limited prior knowledge is available for determing the connections between clinical features and outcomes. Moreover, although ML approaches have good predictive accuracy, their application in actual clinical settings is limited because their predictions are difficult to interpret and hence are not actionable [14]. Interpretable methods explain why a certain prediction was made for a patient, i.e., the specific patient characteristics that led to the prediction.

In this study, we used multidimensional data collected at multiple time points in a prospective cohort study to establish an ML model for the early prediction of malnutrition in children with CHD 1 year after surgery and to explain and evaluate the interpretability of this ML model in assisting clinicians in making decisions regarding nutritional interventions.

2. Materials and methods

2.1. Data source and population

The data were obtained from a prospective cohort study carried out at Guangzhou Women and Children's Medical Center in China (registration number: NCT03626480). In total, 948 consecutive infants (<12 months of age) who were admitted with CHD and underwent complete surgical repair from December 2017 to May 2020 were eligible for enrollment. The patients were followed up for 1 year. Patients with any of the following characteristics were excluded: (1) any diagnosed chromosomal diseases (e.g., Down syndrome); (2) any diagnosed metabolic diseases (e.g., galactosemia or hypothyroidism); (3) a history of gastrointestinal surgery; (4) conditions that restrict oral feeding (e.g., cleft palate); and (5) illiteracy. A summary of the study procedures is presented in Fig. 1.

These cohort data contain extensive information and were collected from the following two sources: perioperative and follow-up electronic medical records (EMRs) and self-administered questionnaires. Specifically, the data included sociodemographic information, CHD-related treatments, ICU information, biochemical tests, children anthropometry, mother's reproductive history, gestational vitamin and mineral supplementation, health behaviors (tobacco and alcohol use), infant feeding information before and after surgery and during follow-up, and health status at different time points (health status was assessed by self-reports, from self-administered questionnaires). The cohort data were utilized to develop and validate the algorithm. We selected patients from our database who were treated from December 2017 to August 2019 as the development set, which was used for the model training and internal validation. The model parameters were tuned using 5-fold cross validation along with the grid-search method. The patients in our database who were treated from September 2019 to May 2020 were used as the test set.

This study was approved by the Ethics Committee and Institutional Review Board of Guangzhou Women's and Children's Medical Center, Guangzhou, China (No. GO-2017-017) and conducted in accordance with the ethical guidelines of the World Medical Association's Declaration of Helsinki. All data were deidentified before they were provided to the investigators.

2.2. Outcome variables

Anthropometric measurements (height and weight) were used to investigate the nutritional status of children with CHD. After CHD surgery, all participants were followed for 1 year, during which they invited to return to the medical center for detailed anthropometric measurements (at 1, 3, 6 and 12 months after surgery). The children's weight and height were measured by trained study staff using a standardized protocol based on the World Health Organization (WHO) guidelines.

The weight for age (WAZ), height for age (HAZ), and weight for height (WHZ) are widely used to evaluate the nutritional status of children, based on the World Health Organization child growth standards [15]. Our outcome was malnutrition, which was defined as a z-score below −2 [16]. The z-scores were further categorized as underweight, stunted, and wasting as defined by a WAZ, HAZ and WHZ below −2 z-scores [17], respectively. Here, the primary outcome of interest was an underweighted status 1 year after surgery, and the secondary outcomes were stunted and wasting status 1 year after surgery.

2.3. Predictors and feature extraction

The cohort database contains a total of 1532 available features at all time points. Records in the cohort database from three time points (preoperative, intraoperative/discharge and 1 month postoperative) were available for use as early predictors (803 features). Using these predictors, we aimed to develop an algorithm for the early prediction of malnutrition in the subsequent year. The specific process is shown in Fig. 1.

First, we removed potentially difficult-to-obtain variables with missing data rates greater than 20% [18]. Then, a systematic review of the literature was conducted, and a consensus meeting among the authors, including two senior physicians (CY and HC), was held to identify the possible factors that should be considered in the

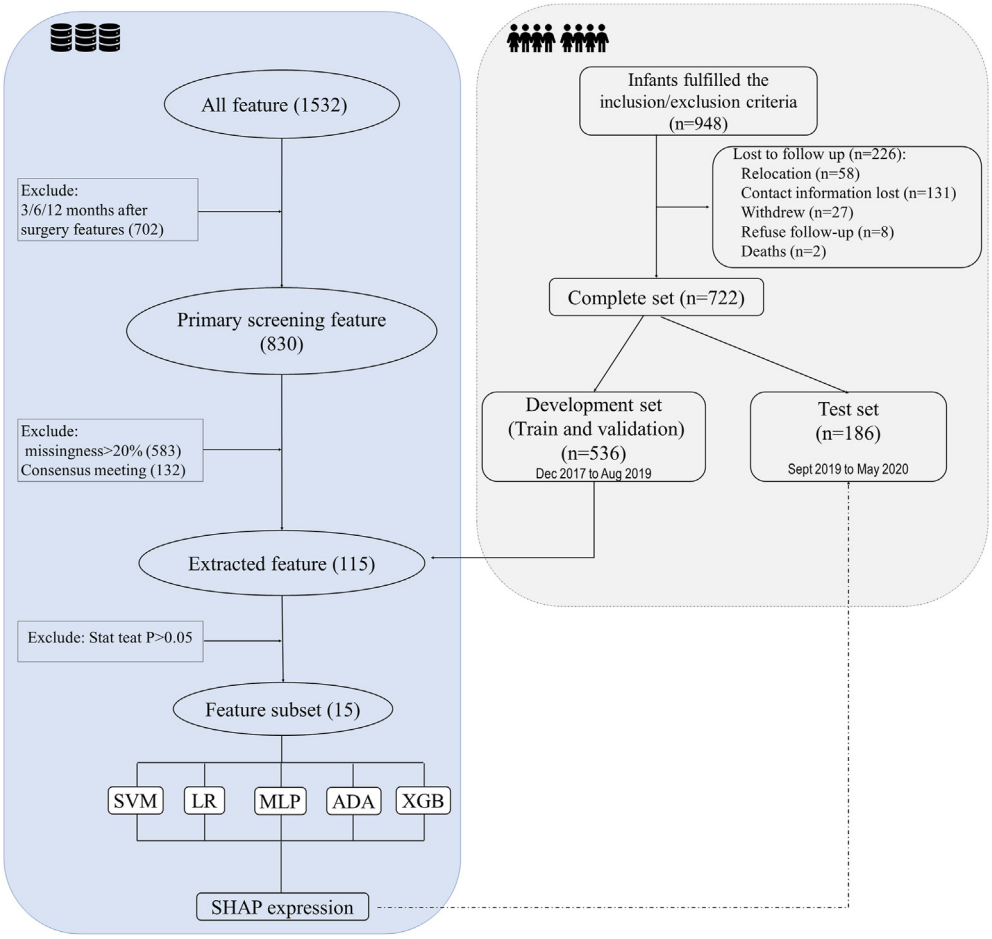

Fig. 1. Flow chart of the study design and analytic strategy used to train, test, and validate of machine learning algorithms to predict malnutrition in children with congenital heart disease. ADA = adaptive boosting; LR = logistic regression; MLP = multilayer perceptron; SHAP = SHapley Additive exPlanations; SVM = support vector machine.

analysis of the postoperative growth of children with CHD [1](shown in Table S1 Appendix). In total, 115 variables were chosen for the initial analysis based on empirical observations (shown in Table S2-S5 Appendix), and we defined the data rules of each variable(what needs to be explained is listed in Table S6 Appendix). For variables with missing proportions less than 20%, we imputed the categorical variables with the mode and the continuous variables with multiple imputation methods. To minimize potential overfitting resulting from the high dimensionality of the features, we first selected features that were statistically significant in the univariate test. Then, a recursive feature elimination (RFE) approach based on a random forest was applied to select low-dimensional features with which to build our models. After the removal of duplicates, in total, 15 variables were selected, of these variables, 11 were structured, and 4 were unstructured.

2.4. Model construction and interpretation

We used the following five representative supervised ML algorithms for the model construction: logistic regression (LR), support vector machine (SVM), adaptive boosting (ADA), multilayer perceptron (MLP) and extreme gradient boosting (XGBoost) [19]. The bootstrap method was implemented with 1000 replications to derive the confidence interval of the AUC, accuracy, sensitivity and specificity. The grid search method with fivefold cross validation was used to choose the best parameters for each model.

We provided explanations for the ML models using SHapley Additive exPlanations (SHAP) [20]. This method is based on the coalition game theory concept; the individual feature value of the data acts as a player in a coalition game (prediction task), and the SHAP value determines how to fairly distribute the gain (prediction performance) among the features [21]. The SHAP value has high potential for rationalizing the predictions made by complex ML models [22]. In the present study, we used the SHAP method to observe the influence of each feature on the prediction results during the prediction process applied to each sample. Additionally, we conducted an analysis to evaluate whether our results were impacted by the questionnaires.

2.5. Statistical analysis

The data cleaning was conducted using Python (Anaconda Distribution, Version 3.7) with Pandas (v1.0.1) and NumPy (1.18.1). The scikit-learn (v0.23.2, https://github.com/scikit-learn/scikit-learn) package was used to build the base models, including LR, SVM, ADA, MLP and XGBoost. The XGboost (v1.1.1, https://github.com/dmlc/xgboost) package was used to build the XGBoost model. The means and standard deviations or medians and interquartile ranges were used to analyze and express the continuous variables, which were tested using an independent-sample t test or Mann–Whitney U test. The categorical variables are expressed as quantities and percentages and were compared by a chi-square test.

3. Result

3.1. Basic characteristics

In total, 536 children with CHD (average age 4.56 months; 59.7% males) were included in this study. The proportions of patients with underweight, stunted, and wasting status 1 year after surgery were 18.1% (97/536), 12.1% (65/536), and 17.5% (94/536), respectively. The changing trend of the Z-score and the incidence of malnutrition in the children with CHD at each time point are shown in the supporting files (Table S7, Figure S1 Appendix). In total, 279 (52.1%) children had a primary or major diagnosis of septal defects. Table 1 shows the baseline characteristics of all subjects for by the primary outcome (underweight). The underweight children had lower birth heights than the non-underweight children ($p < 0.001$). The length of hospital stay and length of ICU stay in the underweight group were longer than those in the non-underweight group ($p < 0.001$; $p = 0.003$). A statistically significant difference in the WAZ scores, including the preoperative WAZ-score ($p < 0.001$), discharge WAZ-score ($p < 0.001$), and 1-month postoperative WAZ score ($p < 0.001$) was found between the two groups. A lower paternal education level (high school or lower) and lower household incomes (less than 5000 yuan (RMB)) were associated with a higher prevalence of an underweight status ($p = 0.034$; $p = 0.013$). The baseline characteristics of the study population by the secondary outcome (a stunted and wasting status) are shown in Table S2-S5 Appendix.

3.2. Model evaluation

Regarding the primary outcome, i.e., an underweight status, we established models to predict the occurrence of an underweight status in children with CHD 1 year after surgery based on cohort data assessed preoperatively, at discharge and 1 month postoperation. Five prediction algorithms, i.e., LR, SVM, ADA, MLP and XGBoost, were used, and the parameters with the best AUC were

Table 1
Baseline demographics and clinical characteristics by primary outcome (an underweight status).

Variables	All Cases (n = 536)	Non-underweight (n = 439)	Underweight (n = 97)	*P-value*
Characteristics of the children				
Age at operation, mo	4.56 ± 3.12	4.58 ± 3.13	4.51 ± 3.12	0.831
Male (%)	320(59.70%)	260(59.23%)	60(61.86%)	0.716
Birth weight, kg	3.27 ± 1.05	3.31 ± 1.15	3.17 ± 0.89	0.089
Birth height, cm	49.19 ± 3.91	49.54 ± 3.76	47.58 ± 4.20	<0.001
Preterm (%)	72(13.43%)	56(12.76%)	16(16.49%)	0.416
Spontaneous delivery (%)	348(64.92%)	285(64.92%)	63(64.95%)	0.924
Exclusive breastfeeding[a] (%)	159(29.66%)	134(30.52%)	25(25.77%)	0.421
Hospital length of stay, d	15.39 ± 7.28	14.74 ± 6.53	18.34 ± 9.50	<0.001
ICU length of stay, d	5.49 ± 5.74	4.95 ± 4.64	7.96 ± 8.84	0.003
RACHS	2.33 ± 0.72	2.32 ± 0.70	2.34 ± 0.78	0.912
Preoperative WAZ score	−1.90 ± 1.42	−1.66 ± 1.33	−3.00 ± 1.32	<0.001
Discharged WAZ score	−2.27 ± 1.39	−2.03 ± 1.30	−3.47 ± 1.16	<0.001
1-month postoperative WAZ-score	−2.20 ± 1.29	−1.91 ± 1.18	−3.46 ± 0.93	<0.001
CHD diagnosis				
Septal defects (%)	279(52.05%)	239(54.44%)	40(41.24%)	0.025
Right heart lesions (%)	84(15.67%)	56(12.76%)	28(28.87%)	<0.001
Left heart lesions (%)	20(3.73%)	16(3.64%)	4(4.12%)	0.944
Double outlet right ventricle (%)	16(2.99%)	11(2.51%)	5(5.15%)	0.29
Transposition of the great arteries (%)	25(4.66%)	22(5.01%)	3(3.09%)	0.586
Single ventricle (%)	5(0.93%)	4(0.91%)	1(1.03%)	0.637
Thoracic arteries and veins (%)	81(15.11%)	68(15.49%)	13(13.40%)	0.717
Pulmonary venous malformation (%)	21(3.92%)	18(4.10%)	3(3.09%)	0.862
Laboratory parameters(preoperative)				
PA(g/L)	143.59 ± 41.45	143.4 ± 41.48	144.55 ± 41.59	0.985
ALB(u/L)	39.86 ± 4.45	39.95 ± 4.32	39.47 ± 4.96	0.691
NT-proBNP(ng/ml)	5121.59 ± 8770.74	4983.37 ± 8663.25	5737.05 ± 9256.78	0.263
Parental and family characteristics				
Mother's age, y	29.24 ± 7.24	29.52 ± 6.32	27.97 ± 11.83	0.685
Father's age, y	32.13 ± 7.90	31.92 ± 5.38	33.08 ± 14.64	0.741
Mother's education (%)				
High school or lower	373(69.59%)	301(68.56%)	72(74.23%)	0.33
College or higher	163(30.41%)	138(31.44%)	25(25.77%)	
Father's education (%)				
High school or lower	381(71.08%)	303(69.02%)	78(80.41%)	0.034
College or higher	155(28.92%)	136(30.98%)	19(19.59%)	
Household incomes[b] (month, RMB) (%)				
<5000	372(69.40%)	294(66.97%)	78(80.41%)	0.013
≥5000	164(30.60%)	145(33.03%)	19(19.59%)	
Preoperative comorbidities				
Diarrhea (%)	99(18.47%)	81(18.45%)	18(18.56%)	0.904
Bronchiolitis (%)	149(27.80%)	113(25.74%)	36(37.11%)	0.033
Pathological jaundice (%)	93(17.35%)	81(18.45%)	12(12.37%)	0.20
Pneumonia (%)	232(43.28%)	184(41.91%)	48(49.48%)	0.212
Thalassemia (%)	18(3.36%)	14(3.19%)	4(4.12%)	0.88
Iron deficiency anemia (%)	31(5.78%)	24(5.47%)	7(7.22%)	0.669

ICU = intensive care unit, WAZ = weight for age, RACHS = risk adjustment for congenital heart surgery, CHD = congenital heart disease, PA = prealbumin, ALB = serum albumin.
[a] Breastfeeding from birth to 6 months.
[b] Household incomes was divided by the median value.

selected for each model. In our study, XGBoost outperformed the other algorithms in predicting an underweight status (Table 2, Fig. 2A). In the test data set, the XGBoost model achieved the greatest AUC (0.87, CI 0.80 to 0.91), the highest accuracy (0.81, CI 0.74 to 0.85), the best sensitivity (0.85, CI 0.79 to 0.89) and the best specificity (0.88, CI 0.86 to 0.91). Therefore, we used the XGBoost model for all downstream analyses.

Then, we trained and tested our XGBoost model using settings with the following different end points: 3, 6 and 12 months after discharge. The AUC remained high (0.87–0.91) up to 1 year (Table S8, Figure S2 Appendix). Meanwhile, to investigate whether the questionnaire features had an effect on the performance of the prediction model, we withdrew six features from the questionnaire and trained and evaluated the XGBoost model with the remaining 9 variables. The results showed that the model without the questionnaire features had an AUC of 0.81, which was significantly lower than the of model with all selected features ($p = 0.041$) (Table S9, Figure S3 Appendix). Overall, the performance of the models that only utilized variables from one source was lower that than of the models using variables from a combination of sources, demonstrating the importance of defining complex predictors and combining their efficiencies in nonlinear models.

3.3. Feature importance of the XGBoost model and external validation

We computed the feature importance score of all variables to identify the important features used by the models that predict an underweight status in CHD children (Fig. 3). A feature's position on the y-axis indicates its importance. These variables included growth features (preoperative WAZ score, discharge WAZ score, and 1-month postoperative WAZ score), disease features (septal defects, right heart lesions, hospital length of stay, ICU length of stay, history of bronchiolitis, hypoglycemia, and 1-month postoperative NT-proBNP), family status features (household income and paternal education level), feeding status and self-rated health status. The permutation importance results demonstrated that the top five risk features were the 1-month postoperative WAZ score, discharge WAZ score, preoperative WAZ score, formula intake, and hospital length of stay.

To further understand and obtain an overview of the importance of the features, we applied the SHAP method. SHAP global explanations are based on calculations of the SHAP explanations for all individual patients and then averaging them by feature to obtain a cohort view. The larger the mean absolute SHAP value of a feature, the more important that feature is to the model prediction. In Fig. 3, the x-axis represents the SHAP value, and the color from blue to red represents the feature's value from low to high. For instance, patients with higher 1-month postoperative WAZ scores (red dots on the plot) are much less likely to develop malnutrition than those with lower 1-month postoperative WAZ scores (blue dots on the plot). Similarly, patients with right heart lesions are more likely to develop malnutrition than those with other types of CHDs (such as anomalous pulmonary venous drainage). Additionally, patients whose average family income is lower than 5000 Yuan/month are more likely to develop malnutrition.

The performance of the XGBoost algorithm in the test data set is presented in Fig. 2B. As shown in the figure, the model had an AUC of 0.84 in the external validation set, which is similar to the AUC value in the training set. This finding indicates that the model did not suffer from overfitting and that the model's performance could be generalized to a temporal external test set. The baseline clinical features of the training and testing sets are shown in Table S10.

Table 2
Evaluation of the model performance.

	AUC	Accuracy	Sensitivity	Specificity
LR	0.83(0.75, 0.86)	0.77(0.7, 0.82)	0.73(0.65, 0.79)	0.80(0.78, 0.83)
SVM	0.83(0.72, 0.86)	0.77(0.7, 0.82)	0.72(0.66, 0.74)	0.79(0.75, 0.83)
ADA	0.84(0.78, 0.88)	0.78(0.74, 0.83)	0.75(0.5, 0.64)	0.87(0.8, 0.93)
MLP	0.86(0.71, 0.86)	0.79(0.72, 0.84)	0.76(0.71, 0.79)	0.88(0.77, 0.92)
XGBoost	0.87(0.80, 0.91)	0.81(0.74, 0.85)	0.85(0.79, 0.89)	0.88(0.86, 0.91)

LR = logistic regression, SVM = support vector machine, ADA = adaptive boosting, MLP, multilayer perceptron, AUC = area under the receiver operating characteristic curve.

3.4. SHAP values of individual prediction for interpretation

The SHAP method can be used to not only obtain explanations for a cohort but also provide an explanation for individual patients. In our study, to understand why the XGBoost model drew certain conclusions, we used the SHAP method to explain the individual predictions of two children in our cohort. Patient No. 1, who belonged to the “true positive” group, was correctly predicted to have a high probability of malnutrition (Fig. 4), and patient No. 2,

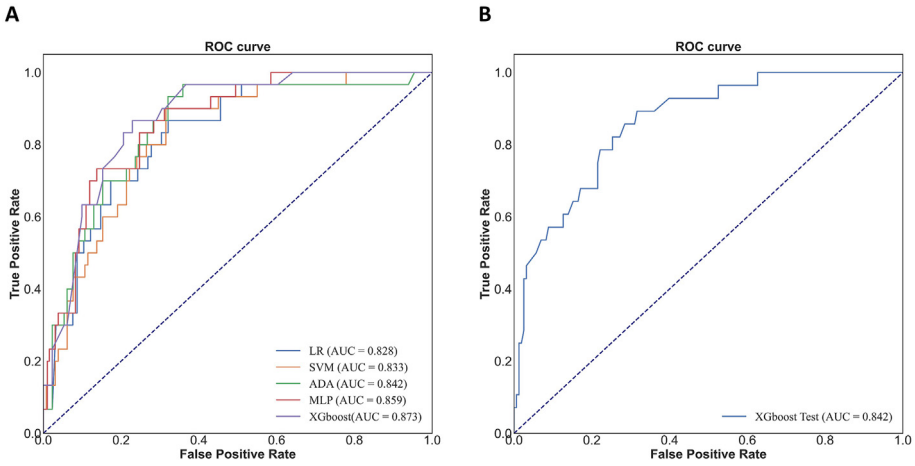

Fig. 2. Receiver operating characteristic curves showing underweight status predictive performance of the machine learning algorithms based on the selected significant features in the training (**A**) and test (**B**) datasets. ADA = adaptive boosting; AUC = area under the receiver operating characteristic curve; LR = logistic regression; MLP = multilayer perceptron; SVM = support vector machine.

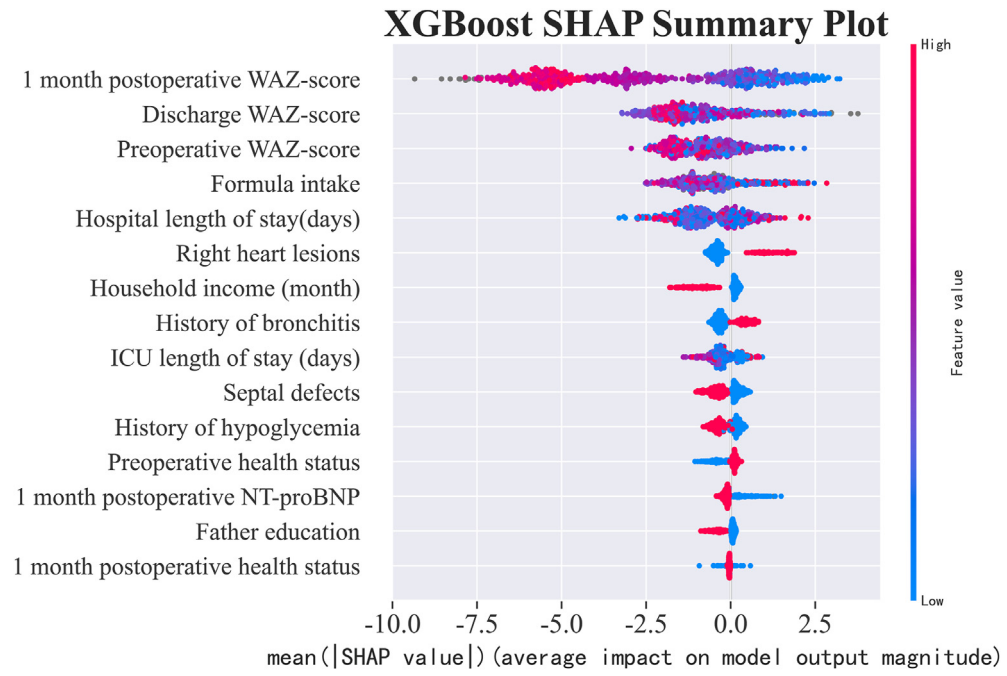

Fig. 3. Feature importance ranking based on SHapley Additive exPlanations (SHAP) values in XGBoost model. The features are ranked according to the sum of the SHAP values in all children, and the SHAP values are used to show the distribution of the effect of each feature on the XGBoost model outputs. Red indicates that the value of a feature is high, and blue indicates that the value of a feature is low. The x-axis indicates the effect of the SHAP values on the model output. The larger the value of the x-axis, the greater the probability of delayed remission. WAZ = weight for age. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

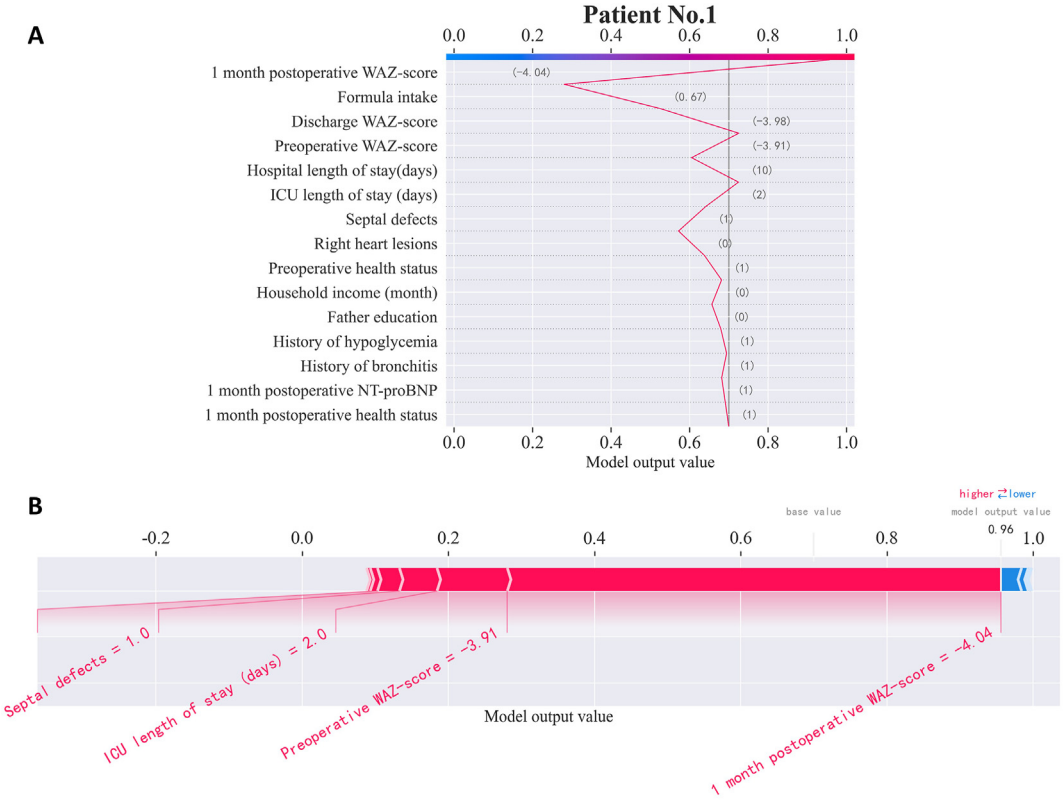

Fig. 4. SHAP explanation decision plots (**A**) and force plot (**B**) of Patient No. 1 (true positive) of the ML model. The decision plots (**A**) show how the individual features contribute to the classification into each of the two classes (a prediction path); The force plot (**B**) depicts the contribution of each feature to the process of moving the value of the decision score from the base value to the value predicted by the classifier.

who belonged to the "true negative" group, was correctly predicted to have a low probability of malnutrition (Fig. 5).

Patient No. 1 was a boy, who was diagnosed with ventricular septal defect (VSD). He was admitted to the hospital at the age of 6 months and 4 days and weighed 3.1 kg. He had a nonpremature delivery, no heart failure, and a history of bronchitis. The preoperative Aristotle score was 6.5, the RACHS-1 score was 2, ejection fraction (EF) was 66%, hemoglobin was 109 g/L, and NT-proBNP was 589.13 pg/ml. Surgical treatment for ventricular septal defects with pulmonary atresia was performed 7 days after admission. The patient received intraoperative extracorporeal circulation for 58 min and had a postoperative OR stay of 37 min, an ICU stay of 2 days, complete oral intake after surgery, a total hospital stays of 10 days, and good heart recovery after surgery (the postoperative ejection fraction at 3 months was 65%). The questionnaire responses indicated an average household income of less than 2000 RMB/month and a paternal education level of junior high school. This patient was predominantly bottle fed with formula for 1 month after discharge, and the formula intake was approximately 250 ml/24 h.

Our prediction model showed that this child's risk of malnutrition at 1 year after surgery was as high as 96%. The child's WAZ values before the operation and 1 month after the operation were −3.91 and −4.04, respectively. These indicators considerably contributied to the prediction results. The actual result was that the child's WAZ 1 year after surgery was −2.2, and he was clinically diagnosed with malnutrition (underweight status).

Patient No. 2 was a girl, who was diagnosed with VSD. She was admitted to the hospital at the age of 7 months and 19 days and weighed 3.2 kg. She was not born prematurely. She had mild heart failure and a history of pneumonia. The preoperative Aristotle score was 6, the RACHS-1 score was 2, EF was 65%, hemoglobin was 98 g/L, and NT-pro BNP = 559.4 pg/ml. VSD patch repair was performed 4 days after admission. The patient received intraoperative extracorporeal circulation for 58 min, and an ICU stay of 2 days. The questionnaire responses indicated that the average household income was 2000–5000/month and that the father had a junior college education level. The patient was predominantly breastfed for 1 month after discharge, and the formula intake was approximately 90 ml/24 h.

Our prediction model showed that the child's risk of malnutrition 1 year after the operation was only 13%. As shown in the figure, poor growth 1 month after surgery (WAZ = −3.46) increased the risk of the outcome. However, the child's preoperative growth and development (WAZ = −1.95) and her relatively short length of hospital stay (12 days) were protective factors, with larger contribution. The actual result was that the child's WAZ 1 year after surgery was −1.6, and the clinical diagnosis was normal.

3.5. ML prediction model for secondary outcomes

We also identified the predictive factors of the secondary outcomes (stunted and wasting status) using an ML prediction model. First, five algorithms (LR, SVM, ADA, MLP and XGBoost) were used to predict the occurrence of children with CHD with stunted or wasting status 1 year after surgery respectively. Fig. S4 shows the AUC of stunted (Figure S4A Appendix) and wasting (Figure S4B Appendix) status obtained with each algorithm, and the XGBoost algorithms performed well and had good predictions of the stunted (AUC = 0.72) and wasting (AUC = 0.80) status. Subsequently, by the feature selection approach, the top 10 variables were ranked based on their predictive importance in XGBoost prediction model. Regarding the stunted status (Figure S5 Appendix), the results demonstrated that the top five important features were the 1-month postoperative HAZ score, discharge HAZ score, aortic clamping time, preoperative HAZ score, and hospital length of stay. Regarding the wasting status (Figure S6 Appendix), the results

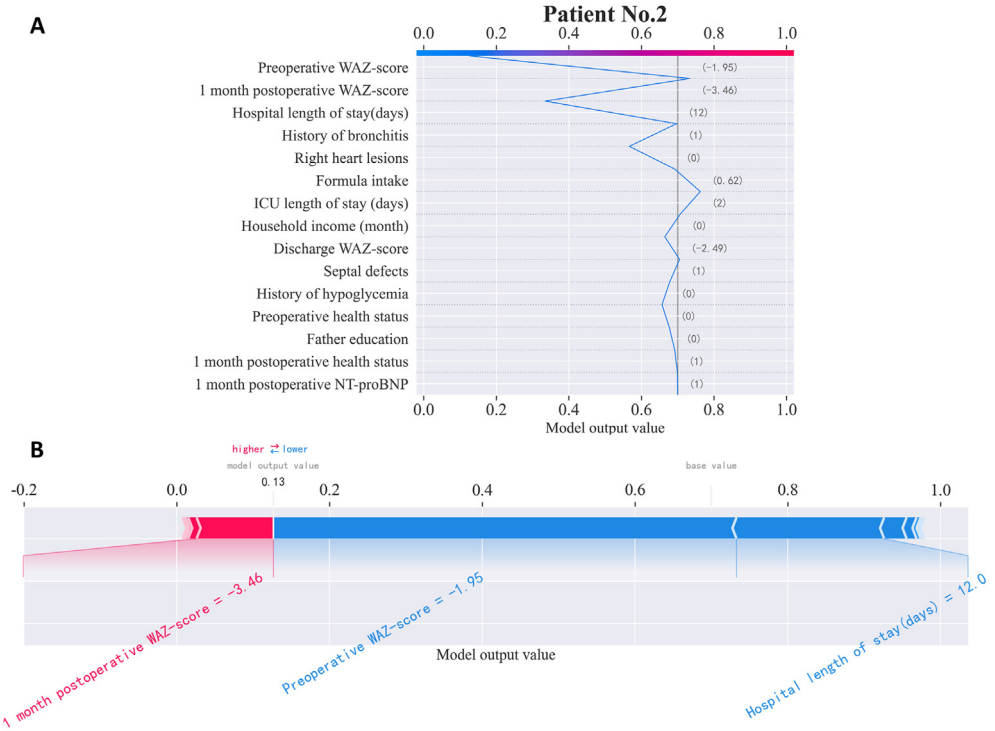

Fig. 5. SHAP explanation decision plots (**A**) and force plot (**B**) of Patient No. 2 (true negative) of the ML model. The decision plots (**A**) show how the individual features contribute to the classification into each of the two classes (a prediction path); The force plot (**B**) depicts the contribution of each feature to the process of moving the value of the decision score from the base value to the value predicted by the classifier.

demonstrated that the top five important features were hospital length of stay, formula intake, discharge WHZ score, preoperative WHZ score, and length of operation.

4. Discussion

In this study, we propose a prediction model developed based on the XGBoost algorithm to accurately identify malnutrition in children with CHD 1 year after surgery. Our research shows that, compared with different algorithms, the XGBoost algorithm has a more accurate prediction performance in the prediction of malnutrition in children with CHD. Then, based on three outcomes (underweight, stunted, and wasting) we constructed three prediction models with the XGBoost algorithm. Furthermore, our study provides meaningful explanations based on the samples provided using feature importance and the SHAP method.

As previously described, children with CHD still show significant malnutrition postoperatively [1]. Because surgery corrects hemodynamic derangements, the maximum catch-up growth occurs in the first year after the correction of CHD, and then the growth curves start to plateau [2,23]. Therefore, predicting the nutritional status, growth and development after 1 year is critical. Such predictions also reveal the best time for nutrition interventions in hospitals, communities or families. It is generally believed that the prognosis should not be determined by only one risk factor and that a combined analysis of multiple variables is more valuable [24]; consequently, the ML method can provide a more comprehensive prediction and can identify subtle and nonlinear relationships in the data. To date, however, studies investigating the prediction of child malnutrition using ML methods primarily included general pediatric populations, and no predictive models for children with CHD have been developed. Therefore, in the present study, we prospectively included 536 children with CHD and established ML prediction models based on 15 early features. A prediction is meaningful only when it is accurate and early enough to provide an added clinical benefit [25].

Our study suggests that compared with other features, perioperative growth (the WAZ score) play a more important role in the ML prediction an underweight status in children with CHD. Specifically, the top three features in the prediction models are the 1 month postoperative WAZ score, the discharge WAZ score, and the preoperative WAZ score. Regarding the stunted and wasting statuses, the prediction models show that the corresponding z-value features are also ranked higher in the models. This result is consistent with clinical knowledge and the results of previous studies [4] and clinical practice. Children with CHD with perioperative malnutrition have reduced postoperative resilience, likely due to a diminished immune response and functional reserve, leading to impaired metabolic control [26,27].

Several factors have consistently emerged as predictors of poorer developmental outcomes, including a longer duration of hospital stay, problems with feeding, and socioeconomic risks [9]. Our study also demonstrates that formula intake, length of stay in the hospital/ICU, type of disease, household income, and paternal education level were significant predictors in the ML model. We controlled the feeding styles and normalized the formula intake values to [0, 1] in the model, although we could not quantify the breast milk intake. We also noted that only 159 (29.6%) patients consumed breast milk, and most children consumed formula; therefore, this feature has some predictive value in nutrition interventions.

The clinical interpretability of ML models is of utmost importance in clinical practice because such models are often considered a black box; thus, their learning process and outputs do not offer a transparent interpretation, and the relationship between the clinical features and the response is invisible to doctors [28,29]. SHAP connects game theory with local explanations, combines several previously available methods, and presents possible, consistent and locally accurate results of the additive feature attribution method based on expected outcomes [20]. Therefore, in this study, we provided explanations for our XGBoost model using SHAP [20] and used SHAP to determine whether the influence of a feature is positive or negative, clarify the results obtained for two patients whose outcomes were correctly predicted, and explain the reasons for the model's correct prediction. Such efforts could greatly increase clinicians trust in the model's behavior and performance. The individual explanations were consistent with clinical knowledge and the results of previous studies and clinical practice [5], further verifying the reliability of the XGBoost model.

This study has several limitations. First, this single-center prospective study involved 536 children, and more patients from multiple sources are needed to verify the robustness and repeatability of our model. Second, the follow-up period (1 year postoperation) was relatively short. Third, despite the broad range of data used, we were unable to include information regarding factors, such as racial origin. This study is only the first step to building a postoperative malnutrition predictive model. In future studies, clinical ML models should be combined with data from different domains (e.g., genetics epigenetics, EEG, ERP, and imaging) to identify the individual growth trajectories of children with CHD and build a more comprehensive and accurate predictive model.

5. Conclusion

In summary, our study used an ML model based on a CHD cohort to predict postoperative malnutrition and identified the factors that helped the model make certain predictions, thereby making the model more reliable and transparent. The results could aid in determining individual nutritional treatment and follow-up strategies for children with CHD.

Funding statement

This research was funded by the Guangzhou high level clinical key specialty.

Author contribution

Study concept and design: Yanqin Cui, Hui Shi, and Dong Yang.
Investigation, methodology, and project administration: Yanqin Cui, Hui Shi, Chunmei Hu, Lijuan Li, Linfang Zhang, and Ting Gong.
Data analysis and interpretation: Kaichen Tang and Dong Yang.
Drafting of the manuscript: Hui Shi and Kaichen Tang.
Critical revision of the manuscript: Yanqin Cui, Hui Shi, Dong Yang, and Kaichen Tang.

Conflict of interest

The authors read and understood the journal's policy regarding the declaration of interests and declare that there are no conflicts of interest.

Acknowledgments

We thank the participating children and their parents. We also thank all nursing and medical staff who participated in our research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clnu.2021.11.006.

References

[1] Li L, Li K, An C, Fan J, Guo C, Liang S, et al. Identification of risk factors affecting catch-up growth after infant congenital heart disease surgery: rationale and design of a multicentre prospective cohort study in China. BMJ Open 2019;9:e030084.

[2] Vaidyanathan B, Radhakrishnan R, Sarala DA, Sundaram KR, Kumar RK. What determines nutritional recovery in malnourished children after correction of congenital heart defects? Pediatrics 2009;124:e294–9.

[3] Medoff-Cooper B, Ravishankar C. Nutrition and growth in congenital heart disease: a challenge in children. Curr Opin Cardiol 2013;28:122–9.

[4] Shi H, Hu C, Zhang L, Tong M, Li L, Cui Y. Early growth trajectory of infants with simple congenital heart disease and complex congenital heart disease undergoing cardiac repair: a prospective cohort study in China. JPEN - J Parenter Enter Nutr 2021;45(6):1181–91.

[5] Hassan BA, Albanna EA, Morsy SM, Siam AG, Al Shafie MM, Elsaadany HF, et al. Nutritional status in children with un-operated congenital heart disease: an Egyptian center experience. Front Pediatr 2015;3:53.

[6] Schoenbuchner SM, Dolan C, Mwangome M, Hall A, Richard SA, Wells JC, et al. The relationship between wasting and stunting: a retrospective cohort analysis of longitudinal data in Gambian children from 1976 to 2016. Am J Clin Nutr 2019;110:498–507.

[7] Marino BS, Lipkin PH, Newburger JW, Peacock G, Gerdes M, Gaynor JW, et al. Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association. Circulation 2012;126:1143–72.

[8] Mussatto KA, Hoffmann RG, Hoffman GM, Tweddell JS, Bear L, Cao Y, et al. Risk and prevalence of developmental delay in young children with congenital heart disease. Pediatrics 2014;133:e570–7.

[9] Mussatto KA, Hoffmann R, Hoffman G, Tweddell JS, Bear L, Cao Y, et al. Risk factors for abnormal developmental trajectories in young children with congenital heart disease. Circulation 2015;132:755–61.

[10] Macesic N, Bear Don't Walk OI, Pe'er I, Tatonetti NP, Peleg AY, Uhlemann AC. Predicting phenotypic polymyxin resistance in Klebsiella pneumoniae through machine learning analysis of genomic data. mSystems 2020;5.

[11] Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, et al. Predicting clinical outcomes of large vessel occlusion before mechanical thrombectomy using machine learning. Stroke 2019;50:2379–88.

[12] Quesada JA, Lopez-Pineda A, Gil-Guillen VF, Durazo-Arvizu R, Orozco-Beltran D, Lopez-Domenech A, et al. Machine learning to predict cardiovascular risk. Int J Clin Pract 2019;73:e13389.

[13] Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. Nat Med 2020;26:71–6.

[14] Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018;2:749–60.

[15] Talukder A, Ahammed B. Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. Nutrition 2020;78:110861.

[16] de Onis M, Blossner M. The World health organization global database on child growth and malnutrition: methodology and applications. Int J Epidemiol 2003;32:518–26.

[17] Group Whomgrs. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006;450:76–85.

[18] Miettinen T, Mantyselka P, Hagelberg N, Mustola S, Kalso E, Lotsch J. Machine learning suggests sleep as a core factor in chronic pain. Pain 2021;162:109–23.

[19] Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18:e323.

[20] Lundberg S, Lee SI. A unified approach to interpreting model predictions. 2017.

[21] Strumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst 2014;41:647–65.

[22] Rodriguez-Perez R, Bajorath J. Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values. J Med Chem 2020;63:8761–77.

[23] Carmona F, Hatanaka LS, Barbieri MA, Bettiol H, Toffano RB, Monteiro JP, et al. Catch-up growth in children after repair of Tetralogy of Fallot. Cardiol Young 2012;22:507–13.

[24] Fan Y, Liu Z, Hou B, Li L, Liu X, Liu Z, et al. Development and validation of an MRI-based radiomic signature for the preoperative prediction of treatment response in patients with invasive functional pituitary adenoma. Eur J Radiol 2019;121:108647.

[25] Lin H, Long E, Ding X, Diao H, Chen Z, Liu R, et al. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: a retrospective, multicentre machine learning study. PLoS Med 2018;15:e1002674.

[26] Golden MH. Proposed recommended nutrient densities for moderately malnourished children. Food Nutr Bull 2009;30:S267–342.

[27] Mehta NM, Duggan CP. Nutritional deficiencies during critical illness. Pediatr Clin 2009;56:1143–60.

[28] Wang X, Wang D, Yao Z, Xin B, Wang B, Lan C, et al. Machine learning models for multiparametric glioma grading with quantitative result interpretations. Front Neurosci 2018;12:1046.

[29] Song X, Yu ASL, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. Nat Commun 2020;11:5668.