Prediction of coronary artery lesions in children with Kawasaki syndrome based on machine learning

Yaqi Tang[1†], Yuhai Liu[2,3†], Zhanhui Du[1†], Zheqi Wang[4] and Silin Pan[1*]

Abstract

Objective Kawasaki syndrome (KS) is an acute vasculitis that affects children < 5 years of age and leads to coronary artery lesions (CAL) in about 20-25% of untreated cases. Machine learning (ML) is a branch of artificial intelligence (AI) that integrates complex data sets on a large scale and uses huge data to predict future events. The purpose of the present study was to use ML to present the model for early risk assessment of CAL in children with KS by different algorithms.

Methods A total of 158 children were enrolled from Women and Children's Hospital, Qingdao University, and divided into 70–30% as the training sets and the test sets for modeling and validation studies. There are several classifiers are constructed for models including the random forest (RF), the logistic regression (LR), and the eXtreme Gradient Boosting (XGBoost). Data preprocessing is analyzed before applying the classifiers to modeling. To avoid the problem of overfitting, the 5-fold cross validation method was used throughout all the data.

Results The area under the curve (AUC) of the RF model was 0.925 according to the validation of the test set. The average accuracy was 0.930 (95% CI, 0.905 to 0.956). The AUC of the LG model was 0.888 and the average accuracy was 0.893 (95% CI, 0,837 to 0.950). The AUC of the XGBoost model was 0.879 and the average accuracy was 0.935 (95% CI, 0.891 to 0.980).

Conclusion The RF algorithm was used in the present study to construct a prediction model for CAL effectively, with an accuracy of 0.930 and AUC of 0.925. The novel model established by ML may help guide clinicians in the initial decision to make a more aggressive initial anti-inflammatory therapy. Due to the limitations of external validation and regional population characteristics, additional research is required to initiate a further application in the clinic.

Keywords Kawasaki syndrome, Coronary artery lesions, Machine learning, Random forest

†Yaqi Tang, Yuhai Liu and Zhanhui Du contributed equally to this work.

*Correspondence:
Silin Pan
silinpan@126.com

[1]Heart Center, Qingdao Women and Children's Hospital, Qingdao University, Qingdao, China
[2]Dawning International Information Industry Co., Ltd., No. 78 Zhuzhou Road, Laoshan District, Qingdao, China
[3]Sugon Nanjing Institute, Co., Ltd., No. 519 Chengxin Avenue, Fangyuan Road, Jiangning District, Nanjing, China
[4]School of Mathematics, Jilin University, Changchun, China

Tang *et al. BMC Pediatrics*   (2024) 24:158

Page 2 of 9

Introduction

Kawasaki syndrome (KS) is a mucocutaneous lymph node syndrome associated with vascular endothelial dysfunction and immune activation that mainly affects children under the age of 5 [1]. The disease is described in all continents but presents the highest annual incidence in Asian countries [2–3]. The incidence of KS in children under 5 years was 68.8 to 107.3 per 100,000 children from 2013 to 2017 in Shanghai [4]. The incidence of KS in children aged 0–4 years was 309 to 330.2 per 100,000 children from 2015 to 2016 in Japan [5]. Coronary artery lesion (CAL) is the most serious complication in the acute phase of KS, including coronary artery dilatation (CAD), coronary artery aneurysm (CAA), long-term coronary artery stenosis (CAS), and even myocardial infarction (MI) [6]. The latest national survey conducted by the Japan Circulation Society (JCS) shows that 7% of children with KS present vascular complications in the acute phase, and 2.3% develop cardiovascular sequelae after discharge [7]. About 20-25% of untreated cases will develop into severe CAL, which makes KS the most common cause of pediatric acquired heart disease in developed countries [8]. Therefore, it is necessary to carry out CAL risk prediction stratification for early diagnosis and intensive care.

Machine learning (ML) is a subset of artificial intelligence (AI), which identifies the features of data sets by constructing corresponding algorithms [9]. The programmer works to find out the characteristics of the features related to the outcome of each event and establishes the training set. The machine determines what characteristics each outcome is related to by learning the training set. In brief, it identifies which end category it belongs to when there is an unfamiliar string of data. With the size of the training sets increasing, the accuracy of ML improved gradually. ML has been extensively used in the field of medicine and health care. Takeuchi al [10]. used a random forest (RF) classifier to establish a prediction model for intravenous immunoglobulin (IVIG) resistant KS with a sensitivity of 79.7% and a specificity of 89.3%, which was significantly better than other traditional models. Xue et al. [11] established a Cox model to predict the risk associated with blood lipid profiles (Lp) in children with ST-segment elevation myocardial infarction (STEMI). It is pointed out that children with STEMI presented higher Lp (a) and lower HDL-C, as well as apoA1, which are more likely to have a higher risk of adverse cardiovascular disease. Sun et al. [12] designed a prediction model based on RF to estimate the probability of arrhythmia after transcatheter closure of atrial septal defect (ASD). Li et al. [13] used manual feature engineering to identify the degree of bone marrow invasion of acute myeloid leukemia (AML), and the sensitivity and specificity of the model were 87.6% and 89.5%, respectively.

The purpose of the present study was to predict who among patients with KS will develop into severe CAL by constructing corresponding algorithms in ML based on clinical manifestation and clinical auxiliary examination.

Methods

Study population

Abstracted data from the eligible patients in Women and Children's Hospital, Qingdao University from May 2021 to June 2022 including patients who received a diagnosis of KS or patients with incomplete KS (IKS) or IVIG-resistant KS under 5 years old at enrollment. The 6th edition diagnostic guidelines revision was used as a guide in the present study [14]. To eliminate irrelevant clinical manifestations and make sure all patients were treated with the same therapy, excluded patients with infections or immunodeficient disease, or who did not receive IVIG within 10 days from the onset. Patients treated with glucocorticoids were also eliminated since the influence of glucocorticoids is not clear yet (15–16). Besides, the medical records with a missing value ratio > 70% were excluded. Echocardiography (Echo) was performed to evaluate the coronary artery before admission and discharge. Compared with the defined internal diameter measurement previously [17], coronary artery abnormality was defined as a Z-score of the coronary artery internal diameter of 2.5 or more according to the 6th edition diagnostic guidelines [14]. Patients at enrollment all accepted a continuous intravenous infusion of IVIG 2 g/kg/24 hours combined with oral aspirin 30 mg/kg/d from the onset, and 3 mg/kg/d after the normalization of C-reactive protein (CRP) and body temperature.

Feature vectors selected

Feature selections combined both clinical manifestations of KS and the high-risk factors of CAL confirmed by previous studies [18–21], excluding the unconventional clinical auxiliary examination. All features were collected before patients were given IVIG treatment. Medical records with missing counting features such as height and weight were analyzed by linear regression analysis and filled according to the test regression equation. The qualitative features of clinical manifestations were transformed into counting features. The K-nearest neighbor (KNN) filling method was used to complete the missing qualitative features or correct the outliers through the correlation of the data in each dimension.

Dimensionality reduction

The numerical features were standardized in one dimension and carried on the principal component analysis (PCA) of the standardized data, which aims to reduce dimensionality and extract features [22]. The high correlation between features had been eliminated. PCA

simplified the complexity of analysis but kept the original information, and integrated multiple vectors into a minority of comprehensive vectors. The total variance of P random features was divided into the sum of the variance of P unrelated random features and kept the first principal component maximum. The variance contribution rate was defined as the ratio of the variance of the principal component to the total variance, and the higher the value, the more original information was reflected in the component. The weight of the vectors was equal to the variance contribution rate of the principal component, and to make the sum of weights of coefficient of the vectors in the linear combination of principal components was 1 by keeping uniformization of the weighted average.

Artificial data synthesis

The SMOTE was proposed by Chawla [23] in 2002, which synthesized better classifier performance by using the interpolation method combining the over-sampling of the minority class and the under-sampling of the majority class. N was the number of the minority classified in the training set, the process was as follows:

1. Sampled x_i from the minority class. Its eigenvector was $x_i, i \in \{1, ..., N\}$. First of all, the K nearest neighbors of the sample x_i were found from all the N samples of the minority class. Marked it as $x_{i(near)}, near \in \{1, ..., K\}$.
2. Then sampled randomly from the K neighbors and regenerated into a random number μ from 0 to 1, thus a new sample $x_{i(near)}$ was synthesized: $x_{new} = x_i + \mu(x_{i(k)} - x_i)$
3. Synthesized T new samples by repeating steps 1 and 2 for T times.

Model building and verification

The software VScode was employed for programming and Python version 3.6 (Python Software Foundation) was performed for statistical analyses. Packages as sklearn 1.2.2, pandas 2.1.3, and numpy 1.24.3 were used to add 95% confidence intervals (CI) and complete cross-validation.

To avoid overfitting, the validation set approach was performed. The entire data sets were divided into 70% and 30% as the training sets and the test sets. The classifiers RF, logistic regression (LR), and eXtreme Gradient Boosting (XGBoost) were constructed using demographic variables processed by the input features of the training sets severally. The data from test sets were substituted into the models for validation. According to the prediction results of the models, the data was calculated one by one as positive examples in this order and

figured out the values of two important quantities each time. The receiver operating characteristic (ROC) curve was obtained by arranging the two quantities as horizontal and vertical coordinates respectively. The discriminatory capacity of the model was assessed using the area under the ROC curve (AUC). The vertical axis of the ROC curve was sensitivity or true positive rate (TPR), while the horizontal axis was false positive rate (FPR). Besides, the five-fold cross validation method was used through all the data to avoid overfitting. The entire data set was divided into five parts and each set divided contains the full category of labels. The cross_val_score function is used to assess the accuracy of the model. The 2x standard deviation method was used to calculate the boundary of the confidence interval. The average value of accuracy minus 2x standard deviation to calculate lower bounds and with the average value plus 2x the standard deviation to calculate the upper bound. Results are expressed as an odds ratio with a 95% confidence interval (CI).

Results

Demographic characteristics

A total of 158 children were enrolled in the present study, including 104 males (65.8%). The mean age was 779.91±654.63 days. 9.49% (*n*=15) of children were diagnosed as CAL according to the Z-score assessment by the Echo. 1.26% (*n*=2) of the research population present IKS and 1.26% (*n*=2) occurred IVIG-resistant KS. The detailed baseline characteristics are shown in Table 1.

Predictive model for CAL

The cohort for the model included demographic characteristics, signs, symptoms of KS, laboratory results, and diagnosis. The comparison between Non-CAL and CAL in the input qualitative variables is shown in Fig. 1. After excluding the variables that were not common in the clinical routine examination and deleting data that did not meet the requirements, 29 feature vectors were determined (Table S1). After PCA dimension reduction, 24 principal components are retained (not shown in the Figures). For solving the imbalance between the data, the SMOTE was used for equalization and the distribution of the training set after the data balance was shown in Table 2. The SMOTE was used for the construction of classifiers from imbalanced data consisting of 143 labels as non-CAL and 15 labels as CAL. There were 110 records in the training set, including 100 children with non-CAL and 10 with CAL. There were 48 records in the test set, including 43 children with non-CAL and 5 with CAL. The schematic diagram of the prediction model conduction is shown in Fig. 2. The AUC of the RF model was 0.925 according to the validation of the tests set and the confusion matrix was shown in Fig. 3. The accuracy

Table 1 Demographics of research population. *N* = 158

Characteristic		Characteristic		n (%)
Height (cm)	87.32 ± 16.63	Male		104 (65.82)
Weight (kg)	12.88 ± 7.10	Rash		104 (65.82)
Age (d)	779.91 ± 654.6	macular papule		18 (11.39)
Hospitalization time (d)	7.35 ± 4.57	millet		144 (91.14)
Fever Time (d)	6.09 ± 2.97	Bulbar conjunctival injection		68 (43.04)
WBC (×10⁹)	14.95 ± 5.56	Chapped lips		142 (89.87)
Hb (G/L)	111.52 ± 13.93	Reddening of lips		125 (79.11)
CRP (mg/l)	67.60 ± 50.13	Strawberry tongue		121 (76.58)
PLT(×10⁹)	331.74 ± 143.82	Cervical lymphadenopathy		135 (85.44)
ESR (mm/h)	62.04 ± 25.21	Changes in peripheral extremities		36 (22.78)
Na⁺ (mmol/l)	136.57 ± 3.25	Changes of Perineal		15 (9.49)
ALT (U/L)	56.60 ± 91.64	Z-sroce > 2.5		73 (46.20)
AST (U/L)	47.98 ± 83.73	Arrhythmia		2 (1.26)
LDH (U/L)	294.85 ± 72.02	IKS		2 (1.26)
CK (U/L)	54.96 ± 38.85	IVIG-resistant KS		
CK-MB (U/L)	17.99 ± 21.24			
cTNI (ng/ml)	0.15 ± 0.89			
PCT (n/ml)	1.14 ± 1.66			

WBC: White blood cell count; Hb: Hemoglobin; PLT: Platelet count; CRP: C-reactive protein; ESR: Erythrocyte Sedimentation Rate; Na⁺: Serum sodium concentration; ALT: Alanine aminotransaminase; AST: aspartate aminotransferase; LDH: lactate dehydrogenase; CK: creatine kinase; CK-MB creatine kinase-MB; cTNI: cardiac troponin I; PCT: procalcitonin
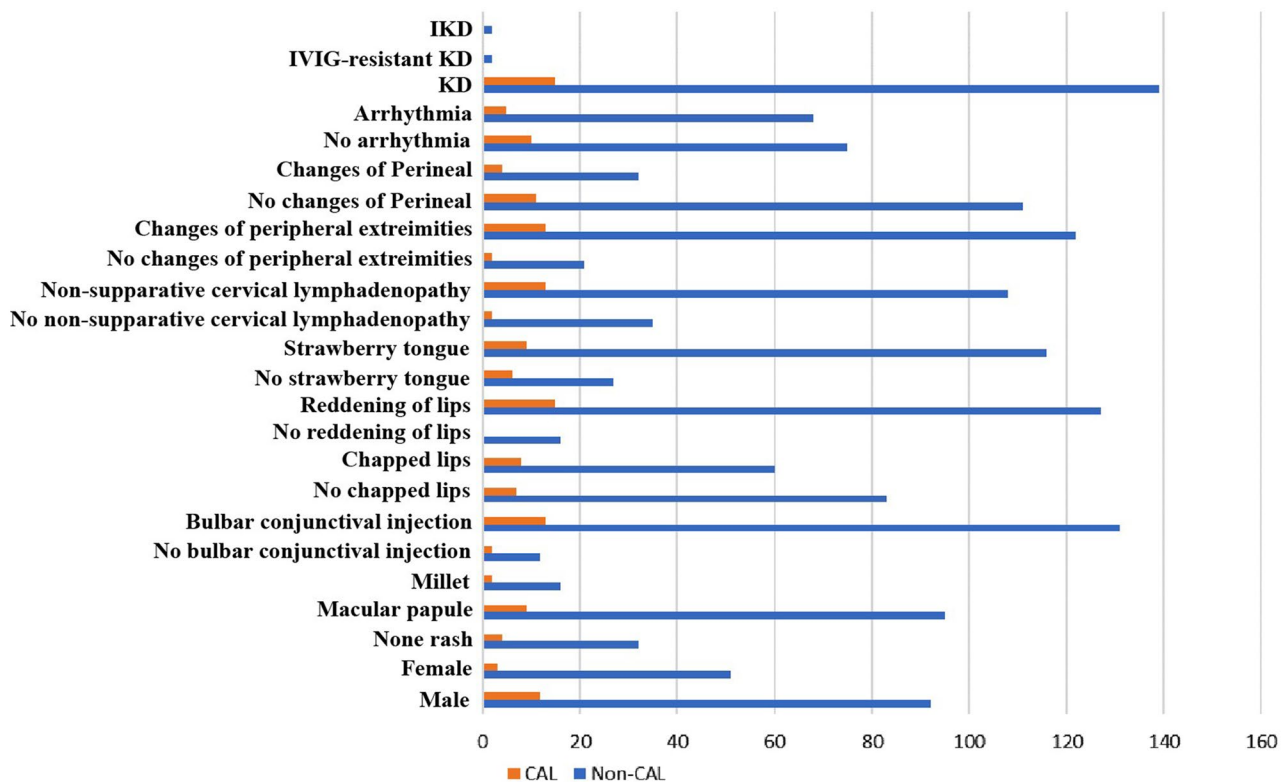
Fig. 1 Comparison between Non-CAL patients and CAL patients in the input qualitative variables

Table 2 Distribution before and after SMOTE

Category	Non-CAL	CAL
Proportion before SMOTE	0.91	0.09
Proportion after SMOTE	0.5	0.5

obtained from 5-fold cross validation was 0.947, 0.868, 0.972, 0.945, 0.919 and 0.930. The average accuracy was 0.930 (95% CI, 0.905 to 0.956). The AUC of the LG model was 0.888 and the average accuracy was 0.893 (95% CI, 0,837 to 0.950). The AUC of the XGBoost model was 0.879 and the average accuracy was 0.935 (95% CI, 0.891 to 0.980). To show the several results distinctly, we established data collection in Table 3.

Discussion

Prevention of CAL is a significant step in the treatment of KS. The standard therapy for KS is effective in reducing the incidence of CAL significantly [24]. However, even with timely initiation of IVIG treatment, coronary artery dilatation might occur in 30% of children, and 5–10% of children might eventually develop permanent coronary artery disease [25]. Early identification and possible intervention might reduce the risk of late coronary artery dilatation. It has been confirmed that corticosteroid therapy combined with IVIG as the initial treatment may intervene in CAL effectively in children with severe KS or IVIG-resistant [26–28]. However, it remains

controversial whether corticosteroids should be used as initial therapy [25]. One of the implications of our predictive algorithm is to recognize patients with a high risk of CAL in the early stage of the disease, which can help guide clinicians in the initial decision to make a more aggressive initial anti-inflammatory therapy. For patients with a high risk for CAL sought to consider a careful follow-up such as more frequent echocardiography.

There were several predictive models with discriminative results established by researchers in different countries or regions for CAL children and IVIG-resistant KS. Chang et al. [29] established a scoring system based on the CRP, neutrophil/lymphocyte ratio, male gender, and IVIG resistance with a sensitivity and a specificity of 60.8% and 70.6%. Hua et al. [30] developed a CAL risk prediction model in children under 6 months of age. The AUC of the model was 0.731, with a sensitivity and specificity of 64.7% and 80.9%. Lee et al. [31] used N-terminal-pro-brain natriuretic protein (NT-proBNP) and polymorphonuclear Neutrophil (PMN) to create the prediction model for CAL, which presents a sensitivity of 73.3% and a specificity of 67.9%. In China, Yang et al. [32] constructed a predictive tool for the efficacy of IVIG therapy in children with KS, and the sensitivity and specificity were 56% and 79% in the internal verification, respectively. However, these published risk assessment
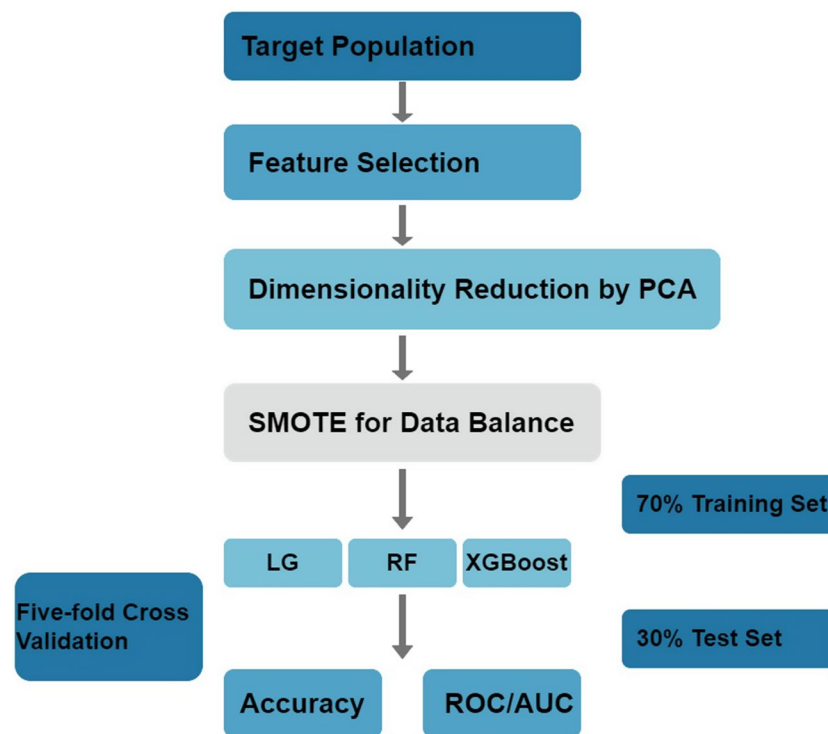
Tang *et al. BMC Pediatrics*   (2024) 24:158

Page 6 of 9

Fig. 2 Schematic diagram for model building

systems scoring of IVIG-resistance or CAL were short of the external data verification (33–34).

The traditional models are constructed with only a few features, while the model calculated by ML can integrate all aspects of clinic feature vectors. ML is suitable for many tasks and makes it easy for the model to retrain and update using the newest data [35]. Supervised learning in ML identifies the relationship between input data and output data in the training set, then summarizes the new data into a known label according to its features, which one we choose to employ (36–37). The input data used in the present study covered common clinical manifestations and auxiliary examinations based on the latest 6th edition diagnostic guidelines, which makes the clinical application of the model easier. In the present study, three classifiers were employed to establish the model. According to the result of validation studies, the accuracy of the RF model was more than 0.90 and the AUC was 0.925. The accuracy of the XGBoost model was 0.930 with a 0.879 AUC. The result of validation showed that our model has feasibility and application prospects in the clinical risk prediction of CAL. In addition, the validation set approach and five-fold cross validation method were performed throughout the process of model establishment and validation. These efforts allow us to avoid the problem of overfitting to the greatest extent and increase the credibility of the model.

This study has several limitations. A more stringent external validation from other institutions should be involved in further assessing the generalizability of the proposed scoring model before applying it. Besides, we have not subsumed hypoalbuminemia and hyperbilirubinemia which are newly added in the sixth edition of the guidelines in the features. The population studied in the present study appears strongly regional characteristics, which might represent the patient population in the east of China.

Conclusion

Three different algorithms were used in the present study to construct a prediction model for CAL effectively. The accuracy of the RF model was more than 0.90 and the AUC was 0.925, which shows that our model has feasibility and application prospects in the clinical risk prediction of CAL. The novel model established by ML may help guide clinicians in the initial decision to make a more aggressive initial anti-inflammatory therapy. Compared with models established by traditional methods in other regions of China, the verification results of this model showed higher accuracy. Due to the limitations of external validation and regional population characteristics, additional research is required to initiate a further application in the clinic.
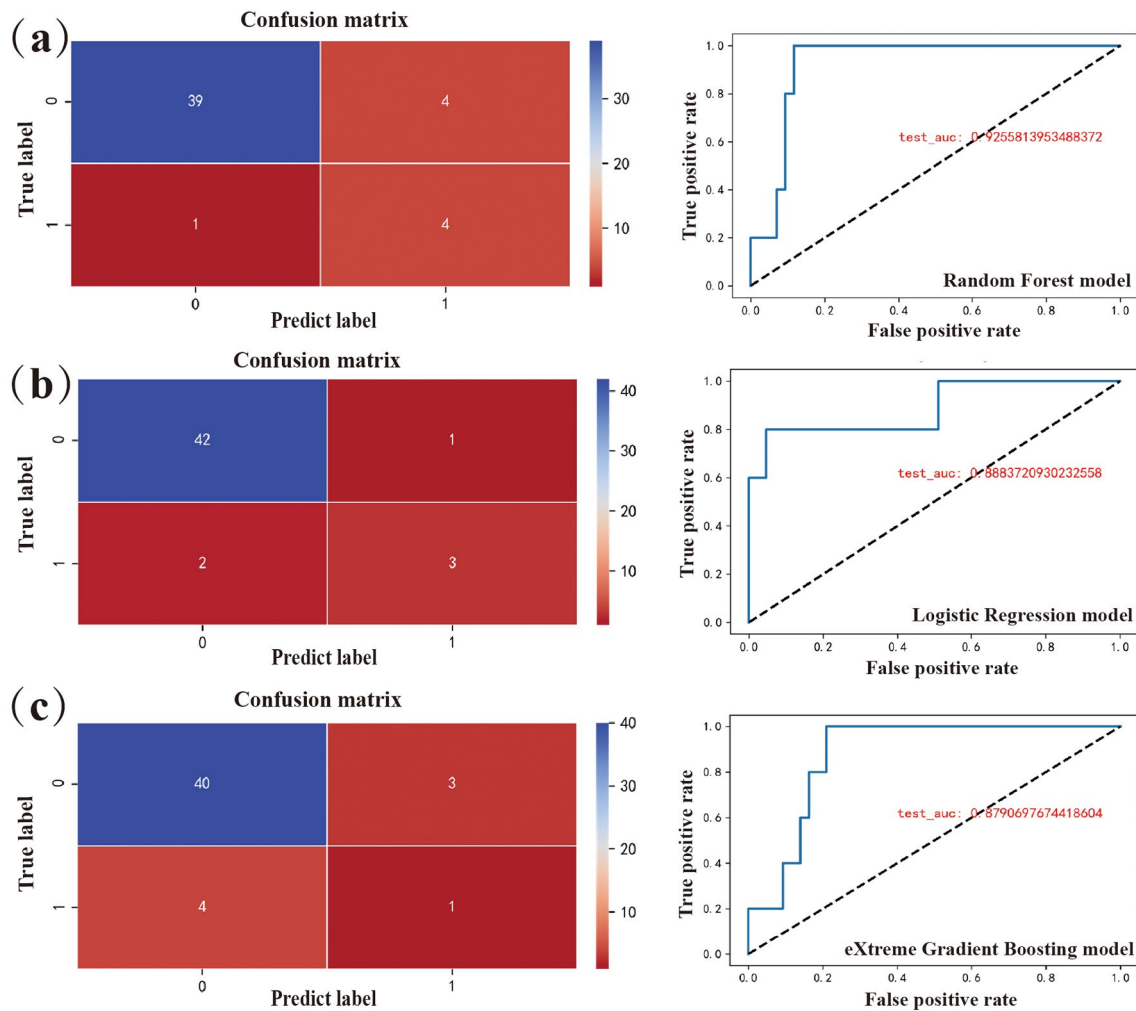
Tang *et al. BMC Pediatrics*   (2024) 24:158

Page 7 of 9

Fig. 3 Demonstration of the prediction effect of the 3 models. (**a**) Demonstration of the prediction effect of RF model. (**b**) Demonstration of the prediction effect of LG model. (**c**) Demonstration of the prediction effect of XGBoost model

Table 3 Demonstration of the prediction effect of the 3 models

Model	AUC	Average Accuracy	95% CI	Accuracy
RF	0.925	0.930	0.905–0.956	0.947/0.868/0.972/0.945/0.919
LG	0.888	0.893	0.837–0.950	0.921/0.842/0.891/0.919/0.919
XGBoost	0.879	0.935	0.891–0.980	0.921/0.973/0.946/0.918/0.919

Abbreviations

AI	artificial intelligence
ASD	atrial septal defect
AML	acute myeloid leukemia
AUC	area under the curve
ALB	albumin ratio
CAL	coronary artery lesion
CAD	coronary artery dilatation
CAA	coronary artery aneurysm
CAS	coronary artery stenosis
CRP	C-reactive protein
DT	Decision Tree
Echo	Echocardiography
FPR	false positive rate
IVIG	intravenous immunoglobulin
IKS	incomplete KS
JCS	Japan Circulation Society
JPS	Japan Pediatric Society
KS	Kawasaki syndrome
KNN	K-nearest neighbor
MI	myocardial infarction
LR	logistic regression
NPV	negative predictive value
NT-proBNP	N-terminal-pro-brain natriuretic protein
PCA	principal component analysis
PPV	positive predictive value
PMN	polymorphonuclear Neutrophil
RF	random forest
ROC	receiver operating characteristic
STEMI	ST-segment elevation myocardial infarction
TPR	true positive rate
XGBoost	eXtreme Gradient Boosting

Supplementary Information

Supplementary Material 1: Numerical range of input features

Data availability
All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval and consent to participate
The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki and has been approved and informed by the institutional review board of the ethics committee of the Qingdao Women and Children's Hospital. Informed consent was obtained from the children's parents through telephone follow-up. Confidentiality of the information was secured throughout the study process. Furthermore, the collected data were anonymous.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Kainth R, Shah P. Kawasaki disease: origins and evolution. Arch Dis Child. 2021;106(4):413–4. https://doi.org/10.1136/archdischild-2019-317070.
2. Skochko SM, Jain S, Sun X, et al. Kawasaki Disease Outcomes and response to Therapy in a Multiethnic Community: a 10-Year experience. J Pediatr. 2018;203:408–415e3. https://doi.org/10.1016/j.jpeds.2018.07.090.
3. Piram M. Epidemiology of Kawasaki Disease in Europe. Front Pediatr. 2021;9:673554. https://doi.org/10.3389/fped.2021.673554.
4. Xie LP, Yan WL, Huang M, et al. Epidemiologic Features of Kawasaki Disease in Shanghai from 2013 through 2017. J Epidemiol. 2020;30(10):429–35. https://doi.org/10.2188/jea.JE20190065.
5. Makino N, Nakamura Y, Yashiro M, et al. Nationwide epidemiologic survey of Kawasaki disease in Japan, 2015–2016. Pediatr Int. 2019;61(4):397–403. https://doi.org/10.1111/ped.13809.
6. Tsuda E, Tsujii N, Hayama Y. Stenotic Lesions and the Maximum Diameter of Coronary Artery aneurysms in Kawasaki Disease. J Pediatr. 2018;194:165–170e2. https://doi.org/10.1016/j.jpeds.2017.09.077.
7. Fukazawa R, Kobayashi J, Ayusawa M, et al. JCS/JSCS 2020 Guideline on diagnosis and management of Cardiovascular Sequelae in Kawasaki Disease. Circ J off J Jpn Circ Soc. 2020;84(8):1348–407. https://doi.org/10.1253/circj.CJ-19-1094.
8. Kuo HC. Diagnosis, Progress, and treatment update of Kawasaki Disease. Int J Mol Sci. 2023;24(18):13948. https://doi.org/10.3390/ijms241813948.
9. Bini SA, Artificial Intelligence M, Learning D, Learning, Computing C. What do these terms Mean and how will they Impact Health Care? J Arthroplasty. 2018;33(8):2358–61. https://doi.org/10.1016/j.arth.2018.02.067.
10. Takeuchi M, Inuzuka R, Hayashi T, et al. Novel risk Assessment Tool for Immunoglobulin Resistance in Kawasaki Disease: application using a Random Forest Classifier. Pediatr Infect Dis J. 2017;36(9):821–6. https://doi.org/10.1097/INF.0000000000001621.
11. Xue Y, Shen J, Hong W, et al. Risk stratification of ST-segment elevation myocardial infarction (STEMI) children using machine learning based on lipid profiles. Lipids Health Dis. 2021;20(1):48. https://doi.org/10.1186/s12944-021-01475-z.
12. Sun H, Liu Y, Song B, Cui X, Luo G, Pan S. Prediction of arrhythmia after intervention in children with atrial septal defect based on random forest. BMC Pediatr. 2021;21(1):280. https://doi.org/10.1186/s12887-021-02744-7.
13. Li H, Xu C, Xin B, et al. 18F-FDG PET/CT Radiomic Analysis with Machine Learning for identifying bone marrow involvement in the children with suspected relapsed Acute Leukemia. Theranostics. 2019;9(16):4730–9. https://doi.org/10.7150/thno.33841.
14. Kobayashi T, Ayusawa M, Suzuki H, et al. Revision of diagnostic guidelines for Kawasaki disease (6th revised edition). Pediatr Int off J Jpn Pediatr Soc. 2020;62(10):1135–8. https://doi.org/10.1111/ped.14326.
15. Miura M. Role of glucocorticoids in Kawasaki disease. Int J Rheum Dis. 2018;21(1):70–5. https://doi.org/10.1111/1756-185X.13209.
16. Okubo Y, Michihata N, Morisaki N, et al. Association between dose of glucocorticoids and coronary artery lesions in Kawasaki Disease. Arthritis Care Res (Hoboken). 2018;70(7):1052–7. https://doi.org/10.1002/acr.23456.
17. Ayusawa M, Sonobe T, Uemura S, et al. Revision of diagnostic guidelines for Kawasaki disease (the 5th revised edition). Pediatr Int. 2005;47(2):232–4. https://doi.org/10.1111/j.1442-200x.2005.02033.x.
18. Tsai CM, Yu HR, Tang KS, Huang YH, Kuo HC. C-Reactive protein to albumin ratio for Predicting Coronary Artery lesions and Intravenous Immunoglobulin Resistance in Kawasaki Disease. Front Pediatr. 2020;8:607631. https://doi.org/10.3389/fped.2020.607631.
19. Lee HY, Song MS. Predictive factors of resistance to intravenous immunoglobulin and coronary artery lesions in Kawasaki disease. Korean J Pediatr. 2016;59(12):477–82. https://doi.org/10.3345/kjp.2016.59.12.477.
20. Türkuçar S, Yıldız K, Acarı C, Dundar HA, Kır M, Ünsal E. Risk factors of intravenous immunoglobulin resistance and coronary arterial lesions in Turkish children with Kawasaki disease. Turk J Pediatr. 2020;62(1):1–9. https://doi.org/10.24953/turkjped.2020.01.001.
21. Liu HH, Chen WX, Niu MM, et al. A new scoring system for coronary artery abnormalities in Kawasaki disease. Pediatr Res. 2022;92(1):275–83. https://doi.org/10.1038/s41390-021-01752-8.
22. Mi JX, Zhang YN, Lai Z, Li W, Zhou L, Zhong F. Principal component analysis based on nuclear norm minimization. Neural Netw. 2019;118:1–16. https://doi.org/10.1016/j.neunet.2019.05.020.
23. Nguyen T, Mengersen K, Sous D, Liquet B. SMOTE-CD: SMOTE for compositional data. PLoS ONE. 2023;18(6):e0287705. https://doi.org/10.1371/journal.pone.0287705.
24. Pan Y, Fan Q, Hu L. Treatment of immunoglobulin-resistant kawasaki disease: a bayesian network meta-analysis of different regimens. Front Pediatr. 2023;11:1149519. https://doi.org/10.3389/fped.2023.1149519.
25. Chan H, Chi H, You H, et al. Indirect-comparison meta-analysis of treatment options for patients with refractory Kawasaki disease. BMC Pediatr. 2019;19(1):158. https://doi.org/10.1186/s12887-019-1504-9.
26. Hamada H, Suzuki H, Onouchi Y, et al. Efficacy of primary treatment with immunoglobulin plus ciclosporin for prevention of coronary artery abnormalities in children with Kawasaki disease predicted to be at increased risk of non-response to intravenous immunoglobulin (KAICA): a randomised controlled, open-label, blinded-endpoints, phase 3 trial. Lancet Lond Engl. 2019;393(10176):1128–37. https://doi.org/10.1016/S0140-6736(18)32003-8.
27. Kobayashi T, Saji T, Otani T, et al. Efficacy of immunoglobulin plus prednisolone for prevention of coronary artery abnormalities in severe Kawasaki disease (RAISE study): a randomised, open-label, blinded-endpoints trial. Lancet Lond Engl. 2012;379(9826):1613–20. https://doi.org/10.1016/S0140-6736(11)61930-2.
28. Zheng X, Li J, Yue P, et al. Is there an association between intravenous immunoglobulin resistance and coronary artery lesion in Kawasaki disease?—Current evidence based on a meta-analysis. PLoS ONE. 2021;16(3):e0248812. https://doi.org/10.1371/journal.pone.0248812.
29. Chang LS, Lin YJ, Yan JH, Guo MM, Lo MH, Kuo HC. Neutrophil-to-lymphocyte ratio and scoring system for predicting coronary artery lesions of Kawasaki disease. BMC Pediatr. 2020;20(1):398. https://doi.org/10.1186/s12887-020-02285-5.

30. Hua W, Ma F, Wang Y, et al. A new scoring system to predict Kawasaki disease with coronary artery lesions. Clin Rheumatol. 2019;38(4):1099–107. https://doi.org/10.1007/s10067-018-4393-7.
31. Lee HY, Song MS. Predictive factors of resistance to intravenous immunoglobulin and coronary artery lesions in Kawasaki disease. Korean J Pediatr. 2016;59(12):477. https://doi.org/10.3345/kjp.2016.59.12.477.
32. Yang S, Song R, Zhang J, Li X, Li C. Predictive tool for intravenous immunoglobulin resistance of Kawasaki disease in Beijing. Arch Dis Child. 2019;104(3):262–7. https://doi.org/10.1136/archdischild-2017-314512.
33. Arane K, Mendelsohn K, Mimouni M, et al. Japanese scoring systems to predict resistance to intravenous immunoglobulin in Kawasaki disease were unreliable for caucasian Israeli children. Acta Paediatr. 2018;107(12):2179–84. https://doi.org/10.1111/apa.14418.
34. Fabi M, Andreozzi L, Corinaldesi E, et al. Inability of Asian risk scoring systems to predict intravenous immunoglobulin resistance and coronary lesions in Kawasaki disease in an Italian cohort. Eur J Pediatr. 2019;178(3):315–22. https://doi.org/10.1007/s00431-018-3297-5.
35. Bayliss L, Jones LD. The role of artificial intelligence and machine learning in predicting orthopaedic outcomes. Bone Jt J. 2019;101–B(12):1476–8. https://doi.org/10.1302/0301-620X.101B12.BJJ-2019-0850.R1.
36. Rauschert S, Raubenheimer K, Melton PE, Huang RC. Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. Clin Epigenetics. 2020;12(1):51. https://doi.org/10.1186/s13148-020-00842-4.
37. Deo RC. Machine learning in Medicine. Circulation. 2015;132(20):1920–30. https://doi.org/10.1161/CIRCULATIONAHA.115.001593.

Publisher's Note