




A simple scoring model based on machine learning predicts intravenous immunoglobulin resistance in Kawasaki disease

Yuto Sunaga¹ · Atsushi Watanabe¹ · Nobuyuki Katsumata^{1,2} · Takako Toda^{1,3}  · Masashi Yoshizawa¹ · Yosuke Kono¹ · Yohei Hasebe¹ · Keiichi Koizumi^{1,4} · Minako Hoshiai^{1,5} · Eiryō Kawakami^{6,7,8} · Takeshi Inukai¹

Received: 22 September 2022 / Revised: 25 December 2022 / Accepted: 30 December 2022 / Published online: 11 January 2023
© The Author(s) 2023, corrected publication 2023

Abstract

Introduction In Kawasaki disease (KD), accurate prediction of intravenous immunoglobulin (IVIG) resistance is crucial to reduce a risk for developing coronary artery lesions.

Objective To establish a simple scoring model predicting IVIG resistance in KD patients based on the machine learning model.

Methods A retrospective cohort study of 1002 KD patients diagnosed at 12 facilities for 10 years, in which 22.7% were resistant to initial IVIG treatment. We performed machine learning with diverse models using 30 clinical variables at diagnosis in 801 and 201 cases for training and test datasets, respectively. SHAP was applied to identify the variables that influenced the prediction model. A scoring model was designed using the influential clinical variables based on the Shapley additive explanation results.

Results Light gradient boosting machine model accurately predicted IVIG resistance (area under the receiver operating characteristic curve (AUC), 0.78; sensitivity, 0.50; specificity, 0.88). Next, using top three influential features (days of illness at initial therapy, serum levels of C-reactive protein, and total cholesterol), we designed a simple scoring system. In spite of its simplicity, it predicted IVIG resistance (AUC, 0.72; sensitivity, 0.49; specificity, 0.82) as accurately as machine learning models. Moreover, accuracy of our scoring system with three clinical features was almost identical to that of Gunma score with seven clinical features (AUC, 0.73; sensitivity, 0.53; specificity, 0.83), a well-known logistic regression scoring model.

Conclusion A simple scoring system based on the findings in machine learning seems to be a useful tool to accurately predict IVIG resistance in KD patients.

Key Points

- In Kawasaki disease (KD), accurate prediction of intravenous immunoglobulin (IVIG) resistance is crucial to reduce a risk for developing coronary artery lesions.
- Machine learning model predicted IVIG resistance in KD patients, and Shapley additive explanation (SHAP) was a useful approach for explaining the outcome of the machine learning model.
- A simple scoring system using three clinical features (days of illness at initial therapy, serum levels of CRP, and total cholesterol at diagnosis) based on SHAP efficiently predicted IVIG resistance.

Keywords Intravenous immunoglobulin resistance · Kawasaki disease · Machine learning · Shapley additive explanation

✉ Takako Toda
takakotojp@yahoo.co.jp

¹ Faculty of Medicine, Department of Pediatrics, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi 409-3898, Japan

² Department of Neonatology, Yamanashi Prefectural Central Hospital, Kofu, Yamanashi, Japan

³ Department of Pediatric Cardiology, National Cerebral and Cardiovascular Center, Osaka, Japan

⁴ Department of Pediatrics, Fujiyoshida Municipal Hospital, Fujiyoshida, Yamanashi, Japan

⁵ Department of Pediatrics, Yamanashi Prefectural Central Hospital, Kofu, Yamanashi, Japan

⁶ Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan

⁷ Institute for Advanced Academic Research (IAAR), Chiba University, Chiba, Japan

⁸ Advanced Data Science Project, RIKEN Information R&D and Strategy Headquarters, Kanagawa, Japan

Abbreviations

AUC	Area under the ROC curve
CAL	Coronary artery lesions
CRP	C-reactive protein
IVIG	Intravenous immunoglobulin
KD	Kawasaki disease
Light GBM	Light gradient boosting machine
ROC	Receiver operating characteristic
SHAP	Shapley additive explanation
SMOTE	Synthetic minority over-sampling technique

Introduction

Kawasaki disease (KD) is an acute febrile illness in infants and children. Of clinical importance, it is characterized by systemic vasculitis and affects medium-sized arteries, especially the coronary arteries [1, 2]. To avoid the development of coronary artery lesions (CAL), high-dose (2 g/kg) intravenous immunoglobulin (IVIG) therapy has been established as a standard initial treatment for KD patients in the acute phase [2, 3]. However, approximately 20% of KD patients are resistant to the initial IVIG treatment [3], and IVIG resistance is a typical risk factor for developing CAL [1, 4–7]. Under these circumstances, several recent studies showed a possible clinical benefit of intensive initial IVIG therapy combined with other anti-inflammatory agents for the high-risk KD patients [6, 8–10]. For effective pre-treatment risk stratification, it is crucial to establish a scoring system to accurately predict IVIG resistance at the timing of clinical diagnosis of KD. Currently, there are several widely used Japanese scoring models for predicting IVIG resistance: Gunma score proposed by Kobayashi et al. [11], Kurume score proposed by Egami et al. [12], and Osaka score proposed by Sano et al. [13]. These scoring systems were developed by the logistic regression analysis of clinical profiles and laboratory findings before initial treatment, which were selected based on statistical assumptions.

To establish a more reliable and simple scoring system for the prediction of IVIG resistance in KD patients, an alternative approach using large data repositories is required. Recently developed machine learning approach has shown great potential for assisting the clinical diagnosis and predicting outcomes [14–19]. Machine learning utilizes sophisticated algorithms operating on large-scale, heterogeneous datasets to uncover informative patterns that would be difficult or impossible for even well-trained individuals to identify [20]. Indeed, several recent studies applied machine learning to predict IVIG resistance in KD patients and confirmed its usefulness [14–16]. However, in clinical practice, even if machine learning has a high degree of accuracy, a simple scoring system is more convenient for risk-stratified treatment.

In the present study, we applied machine learning to predict IVIG resistance in 1002 KD cases treated with single IVIG protocol in multiple institutes. Subsequently, using the three most important features associated with IVIG resistance in the machine learning, we developed a new simple scoring system and confirmed its utility by comparison with the three representative scoring systems.

Materials and methods

Study participants

The study is a retrospective review of multicenter registration database of 1141 consecutively diagnosed KD patients who were diagnosed between June 2010 and December 2020 in 12 inpatient facilities for the care of pediatric patients, as listed in Supplemental Table 1. Diagnosis of KD was retrospectively confirmed based on criteria defined in the fifth edition of the Japanese Kawasaki Disease Diagnostic Guidelines [21]. In brief, a diagnosis was made when the patients had at least five of the six major symptoms (fever, conjunctival congestion, oral mucosa alteration, cervical lymphadenopathy, swelling of extremities, and polymorphous rash), or when the patients had four major symptoms with the development of CAL. The first day of illness was defined as the day when at least one of the major symptoms appeared. Development of CAL was defined by quantifying the internal coronary artery dimension as per the Japanese Ministry of Health Criteria (a maximum absolute internal diameter > 3 mm in children < 5 years of age, or > 4 mm in children 5 years and older, or segment 1.5 times greater than an adjacent segment, or the presence of luminal irregularity) and whenever body surface area-adjusted Z score of any coronary artery was $\geq +2.5$ (including left main, left anterior descending, left circumflex arteries, and right coronary arteries) [2]. The facilities included all of the 11 pediatric inpatient facilities in Yamanashi Prefecture and 1 facility in Nagano Prefecture in Japan. The registration database was constructed with anonymized clinical records of all the diagnosed KD cases in each hospital that were collected at the end of every year. During the COVID-19 pandemic (from March 2020 to December 2020), 44 cases were diagnosed with Kawasaki disease. In this 10-month period, COVID-19 was uncommon in Yamanashi Prefecture (only 22 children were diagnosed with COVID-19). None of these 44 cases were considered to be multisystem inflammatory syndrome in children (MIS-C) since all cases had no history of direct contact with people with COVID-19 cases within 4 weeks prior to diagnosis, and 37 cases were directly confirmed to be negative for SARS-CoV2 at admission (PCR analysis, 35 cases; antigen test, 2 cases). The study was performed under the approval by the Research Ethics Committee of University of Yamanashi Hospital (Approval Number 1698).

Treatment of Kawasaki disease

Initial laboratory tests listed in Supplemental Table 2 had been standardized, and all of the patients were treated identically with a first-line regimen of 2 g/kg/dose of IVIG in combination with 30 mg/kg of oral aspirin immediately after the diagnosis of KD was made based on the above criteria. Standardized treatment workflow was confirmed in the meeting by each facility every year. IVIG therapy was completed within 24 h after diagnosis of KD in all of the patients. No patients were treated with glucocorticoids. The response to the initial treatment was evaluated 48 h after initiation of IVIG administration and was considered as “IVIG resistance” when the body temperature was over 37.5 °C, and the serum level of C-reactive protein (CRP) was higher than half of the peak value as previously reported [12, 22]. Body temperature was measured in the axilla using an electronic thermometer. IVIG-resistant patients were treated with second-line therapy comprising an additional 2 g/kg/dose of IVIG or 5 mg/kg of intravenous infliximab [23]. In addition, when the patients were considered to be resistant to the second-line therapy, plasma exchange was carried out after the patient was transferred to University of Yamanashi Hospital [24].

Machine learning

The predictors for IVIG resistance were chosen from routinely available data including 6 demographic variables, 22 laboratory data, and 2 echocardiographic parameters at diagnosis as listed in Table 1. For any missing laboratory data and echocardiography parameter values, the median value was complementary used in the machine learning. We used the random forest model [25], eXtreme Gradient Boosting (XGBoost) [26], and light gradient boosting machine (Light GBM) [27], which are tree-based nonparametric methods requiring no assumption about data distribution. We also performed logistic regression analysis and support vector machine analysis (SVM) [28]. We operated each machine

learning in the training set (approximately 80% of the random sample) using scikit-learn in Python software (version 3.8.3), and the optimal parameters (number of trees and the maximum depth of the tree) were determined according to the best area under the receiver operating characteristic (ROC) curve (AUC) in the validation set (approximately 20% of the random sample) as in the previous studies [14, 15] by using *k*-fold crossvalidation (*k* = 10) (Supplemental Fig. 1) [29]. Considering an imbalanced dataset of the IVIG response, we used synthetic minority over-sampling technique (SMOTE), which is a technique of over-sampling the minority class [30, 31].

Development of the scoring system

For development of the simple scoring system to predict IVIG resistance, we selected the features that influenced the prediction model in the Light GBM algorithm, in which the highest AUC was observed. We used the Shapley additive explanation (SHAP), which is a unified approach for explaining the outcome of machine learning model [32–34]. SHAP values evaluate the importance of the output, and a higher SHAP value indicates that a feature has a larger impact and is more important on the model [15, 35]. To determine the cutoff level of each variable, we used the SHAP dependence plot, which evaluates significance of each feature in the output of the Light GBM model [19]. Based on the SHAP value, we constructed a new predictive scoring model (Yamanashi score). To validate the accuracy of the new score system, we applied the score system in the above Yamanashi study cohort and compared it with three previously established score systems.

Statistical analysis

Statistical analyses were performed using EZR software (version 1.41) [36] and Python software (version 3.8.3). Spearman’s correlation coefficient was used to analyze the correlation of each score. Creation and comparison of the ROC curves were performed by using the EZR software.

Table 1 Variables in the model

Variables	
Clinical features	Sex, age in months, height, weight, days of illness at initial therapy (start day), patients with the absence or presence of five or more major symptoms
Laboratory data	White blood cell count (WBC); percentage of neutrophils (Neut); hemoglobin level (Hb); platelet count (Plt); serum levels of C-reactive protein (CRP), total protein (TP), albumin, globulin (IgG), sodium (Na), potassium (K), chloride (Cl), alanine aminotransferase (ALT), aspartate aminotransferase (AST), lactic acid dehydrogenase (LDH), total bilirubin (T. bilirubin), blood urea nitrogen (BUN), creatinine (Cre), creatine kinase (CK), total cholesterol (T. cholesterol), high-density lipoprotein cholesterol (HDL chol), and triglyceride(TG); D-dimer value
Echocardiographic parameters	Maximum coronary artery diameter before initial treatment (CA diameter), maximum coronary artery Z score before initial treatment (CA Z score)

Results

Prediction of IVIG resistance by machine learning

From June 2010 to December 2020, 1141 consecutive KD cases were enrolled in the Yamanashi study cohort. In the present study, 139 cases were excluded for further analyses due to a diagnosis of incomplete KD ($n=129$), severe lack of laboratory data ($n=2$), or delayed IVIG treatment after 10 days of onset ($n=8$) (Fig. 1). One hundred and ninety-three cases (19%) were diagnosed before day 5. In the remaining 1002 cases, 227 cases (22.7%) were resistant to the first course of IVIG treatment (demographics were indicated in Supplemental Tables 3 and 4). In the demographics of 12 facilities (Supplemental Table 5), variations in day of illness at initial therapy (median, 5.2 days; range, 4.9–5.8) and IVIG resistance (22.5%, 14–34%) were largely acceptable. We operated each machine learning using 6 demographic variables, 22 laboratory data, and 2 echocardiographic parameters at diagnosis listed in Table 1. The data of 1002 cases were divided at random into 801 cases of the training dataset (approximately 80%) and 201 cases of the test dataset (approximately 20%). Considering a relatively low frequency of IVIG resistance as an imbalanced dataset of machine learning, we applied SMOTE [30, 31]. Prediction values and ROC curves for IVIG resistance in each model are summarized in Table 2 and Fig. 2a. The highest AUC was observed in the Light GBM model (0.78) (Fig. 2a). In the Light GBM model, global accuracy, sensitivity, specificity, positive prediction value, negative prediction value, positive likelihood ratio, and negative likelihood ratio scores were 0.78 (95% confidence interval (CI), 0.72–0.84), 0.50 (0.36–0.64), 0.88 (0.82–0.93), 0.59 (0.43–0.74), 0.83 (0.77–0.89), 4.14 (2.48–6.90), and 0.57 (0.43–0.75), respectively. These observations demonstrated that machine learning models achieved good discriminating abilities to predict IVIG sensitivity in KD patients although the ability to predict IVIG resistance was relatively limited in this cohort.

Development of scoring system to predict IVIG resistance

We next evaluated the top 20 features among 30 items tested in the Light GBM model using SHAP (Fig. 2b). In SHAP summary plot, the higher the SHAP value of a feature, the higher the probability of IVIG resistance. In each SHAP value of a feature, each dot represents the feature attribution value of each patient, and red and blue dots represent higher and lower feature value, respectively. The highest SHAP value feature was days of illness at initial therapy (start day) (SHAP value [average of absolute value], 0.80). Additionally, serum levels of CRP (0.68) and total cholesterol

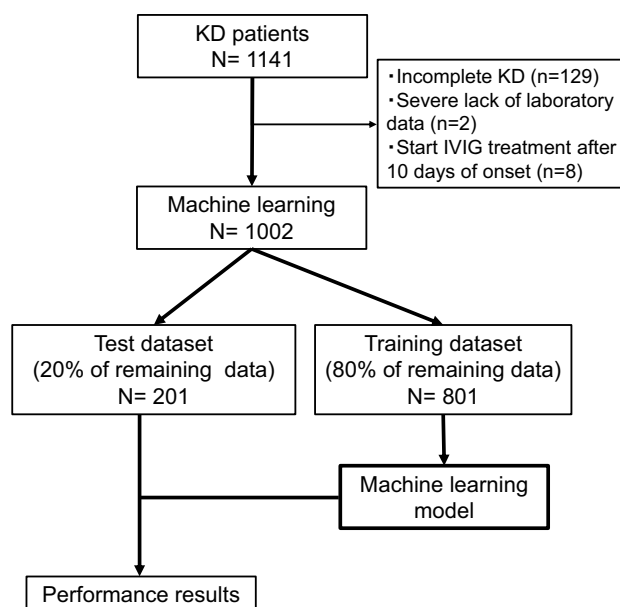


Fig. 1 Flowchart of machine learning. Among 1141 consecutive Kawasaki disease cases enrolled in the cohort of the Yamanashi study, 139 cases were excluded for machine learning due to indicated reasons. The remaining 1002 cases were separated randomly into training dataset (80%, 801 cases) and test dataset (20%, 201 cases). Synthetic minority over-sampling technique (SMOTE) was applied for an imbalanced dataset of intravenous immunoglobulin (IVIG) resistance, and Shapley additive explanation (SHAP) was applied to identify the variables that influenced the prediction model

(0.41) were the other top two features. Base on the SHAP dependence plot of these three features (Fig. 3), we tried to create a new score system to predict IVIG resistance. The cutoff levels for each variable were determined based on the intersecting line of “zero” SHAP value (Fig. 3) as follows: start day \leq day 4, CRP \geq 10 and 7 mg/dL, and total cholesterol \leq 131 mg/dL. Finally, we constructed a simple scoring model (Yamanashi score) using the three variables as follows: two points were scored for start day \leq day 4 and CRP \geq 10 mg/dL, while one point was given for CRP \geq 7 mg/dL (and $<$ 10 mg/dL) and total cholesterol \leq 131 mg/dL. The maximum total score was five points.

Validation of scoring systems to predict IVIG resistance

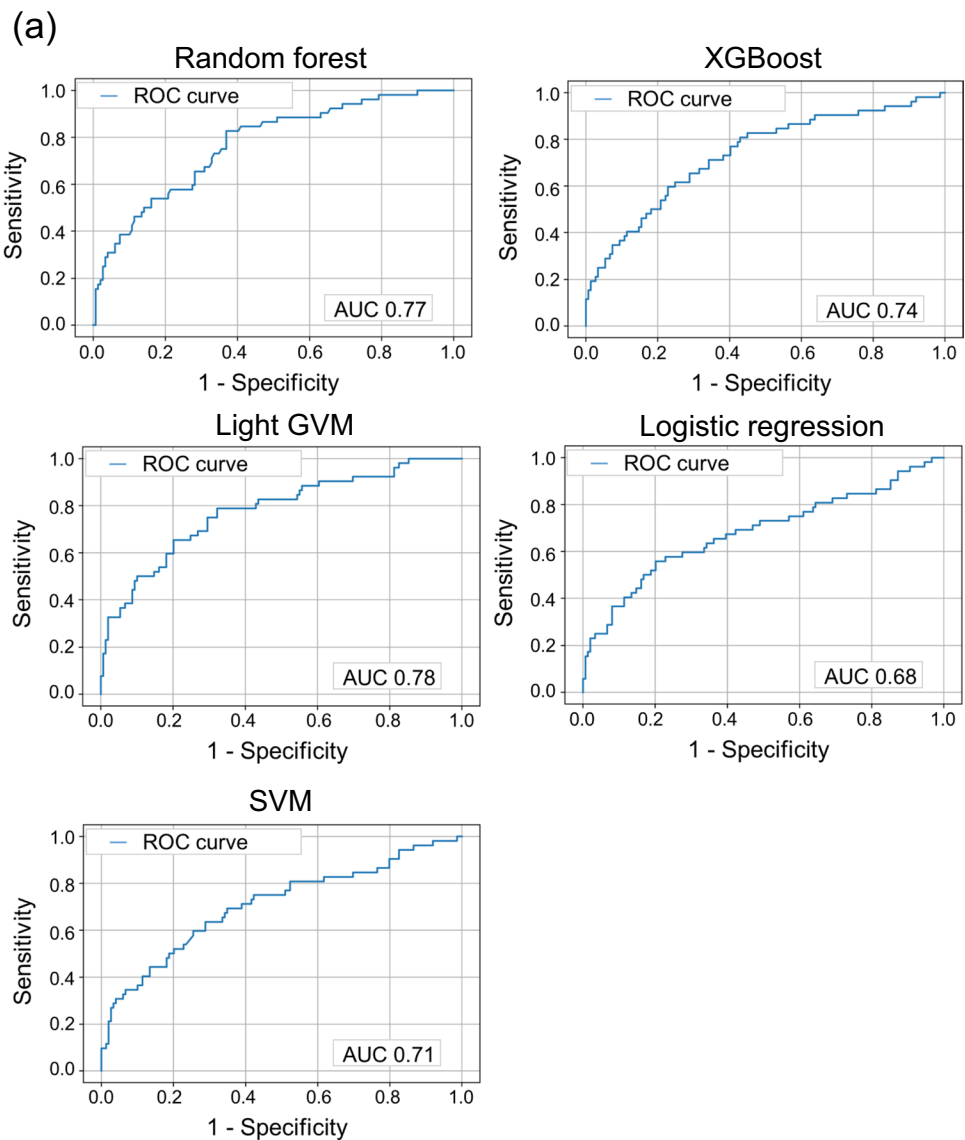
We validated the accuracy of the Yamanashi score in the prediction of IVIG resistance by comparing it with three representative scoring systems in the cohort of Yamanashi study. Among the 1002 KD cases, 545 cases were excluded for the validation due to lack of even one of variables for the 4 scoring systems, and thus the remaining 457 cases were available for further analyses. Among the 457 cases, 108 cases (23.6%) were resistant to initial IVIG treatment. Of note,

Table 2 Prediction values in each machine learning model

	Random forest	XGBoost	Light GBM	Logistic regression	SVM
Global accuracy	0.77 (0.71–0.83)	0.76 (0.69–0.81)	0.78 (0.72–0.84)	0.72 (0.65–0.78)	0.70 (0.63–0.76)
Sensitivity	0.42 (0.29–0.57)	0.50 (0.36–0.64)	0.50 (0.36–0.64)	0.58 (0.43–0.71)	0.60 (0.45–0.73)
Specificity	0.89 (0.83–0.94)	0.85 (0.78–0.90)	0.88 (0.82–0.93)	0.77 (0.69–0.83)	0.73 (0.65–0.80)
Positive predictive value	0.58 (0.41–0.74)	0.53 (0.38–0.68)	0.59 (0.43–0.74)	0.46 (0.34–0.59)	0.44 (0.32–0.56)
Negative predictive value	0.82 (0.75–0.87)	0.83 (0.76–0.89)	0.83 (0.77–0.89)	0.84 (0.77–0.90)	0.84 (0.76–0.90)
Positive likelihood ratio	3.94 (2.25–6.90)	3.23 (2.04–5.15)	4.14 (2.48–6.90)	2.46 (1.69–3.56)	2.22 (1.57–3.14)
Negative likelihood ratio	0.65 (0.51–0.82)	0.59 (0.45–0.78)	0.57 (0.43–0.75)	0.55 (0.40–0.77)	0.55 (0.39–0.78)

95% confidence interval is indicated in parenthesis

Fig. 2 Prediction ability of machine learning for IVIG resistance. **a** ROC curves for IVIG resistance in five models. The horizontal axis indicates false positive rate (1 - specificity), and the vertical axis indicates true positive rate (sensitivity). AUC is indicated at the bottom. **b** The 20 most important features for predicting IVIG resistance in the Light GBM model. Left: SHAP summary plot of the top 20 features. The higher the SHAP value of a feature, the higher the probability of IVIG resistance. Each dot represents the feature attribution value of each patient. A red dot represents higher feature value, and a blue dot represents lower feature value. Right: Importance matrix plot of the top 20 features. SHAP value of each feature is indicated in descending order



when three points for the total score was applied as a cutoff (Supplemental Fig. 2), the accuracy (AUC, 0.72 (95%CI, 0.67–0.77); sensitivity, 0.49 (0.39–0.59); specificity, 0.82 (0.78–0.86)) of the Yamanashi score was almost identical

to that (AUC, 0.78; sensitivity, 0.50; specificity, 0.88) of the Light GBM model using 30 clinical variables. Next, we compared the prediction accuracy of the Yamanashi score with three previous scoring systems (Fig. 4, Supplemental

Fig. 2 (continued)

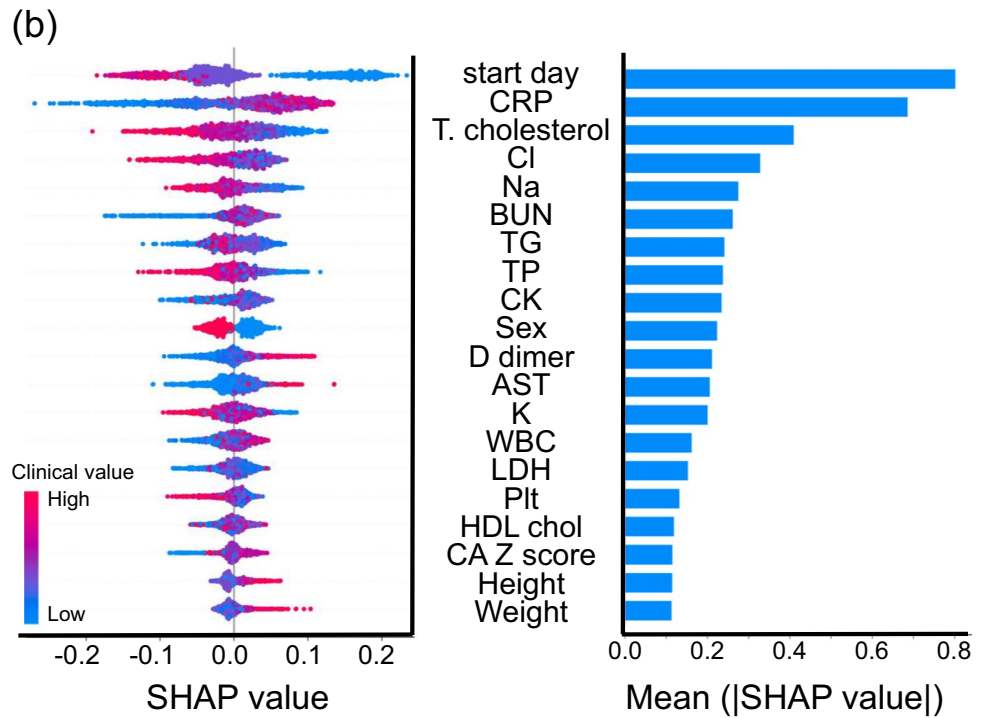


Fig. 3 SHAP dependence plot of top three features. The SHAP dependence plots of **a** days of illness at initial therapy (start day), serum levels of **b** C-reactive protein (CRP), and **c** total cholesterol (T. cholesterol). The horizontal axis indicates the actual value of the feature, and the vertical axis indicates the SHAP value of the feature. The horizontal line indicates the zero level of the SHAP value, and the vertical line indicates the cutoff level for each variable. Cutoff value is indicated on the top in each panel

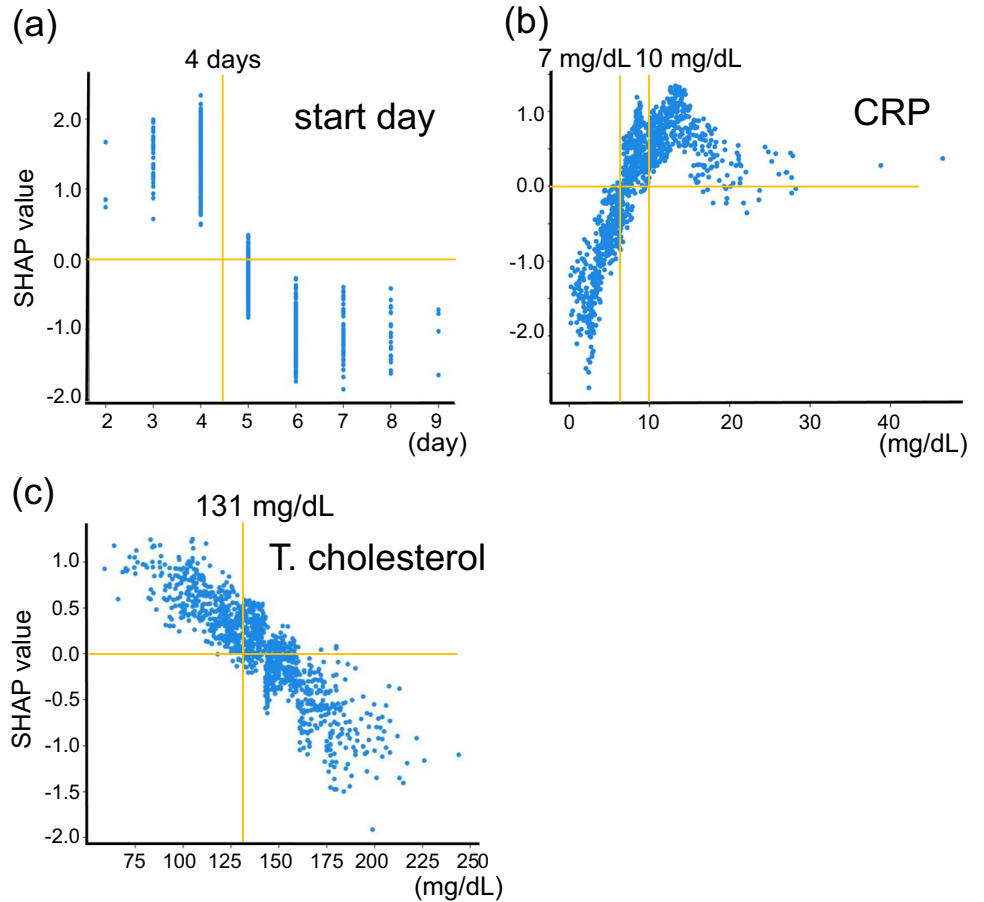


Table 6). Among three variables of the Yamanashi score, total cholesterol was not included in all of three previous scores, while start day and CRP were included in three and two previous scores, respectively (Supplemental Fig. 3). Interestingly, although only three variables were included in the Yamanashi score, ROC curve and AUC of the Yamanashi score were almost identical to those of the Gunma score (AUC, 0.73 (95% CI, 0.67–0.79)) (Fig. 4a), in which seven variables were included [11]. When the Gunma score was applied with a cutoff of five points for total score, sensitivity and specificity were 0.53 (95% CI, 0.43–0.63) and 0.83 (0.79–0.87), respectively. In the 457 cases of the Yamanashi cohort study, the Yamanashi score was significantly correlated with the Gunma score ($R^2=0.43$). In Kurume (Fig. 4b) [12] and Osaka (Fig. 4c) [13] scores, although the correlation coefficients (R^2) with the Yamanashi score were 0.36 and 0.39, respectively, AUC values (0.67 (95% CI, 0.61–0.73) and 0.68 (95% CI, 0.62–0.73)) were inferior ($p=0.04$ and $p=0.07$) to those of the Yamanashi score, respectively. These observations revealed that the simple scoring system using top three features in the machine learning model predicted IVIG resistance almost as accurately as the machine learning model itself as well as the widely used Gunma score, at least in the Yamanashi cohort study.

Discussion

Recently established machine learning has been widely applied in the field of clinical medicine such as outcome prediction, diagnosis, and image interpretation [14–19]. In the present study, we applied the machine learning models to predict IVIG resistance of the initial KD treatment in the Yamanashi cohort study in which clinical data of the 1002 cases were available. Compared with the conventional model such as the widely used logistic regression model in which variables are selected by a knowledge-oriented approach, machine learning is an unbiased approach using a large number of variables. Taking advantage of machine learning, we applied all the initial laboratory data without any assumptions. Considering an imbalanced dataset of IVIG resistance, we applied SMOTE [30, 31] and confirmed a good discriminating ability to predict IVIG resistance. To apply the accurate prediction ability of machine learning model to clinical practice, we established a new scoring system (Yamanashi score) based on the findings in the SHAP plot [32–34] of the Light GBM model, which showed the best prediction accuracy among five models we tested. We selected the following three features with high SHAP values: days of illness at initial therapy, serum levels of CRP, and total cholesterol at diagnosis. Surprisingly, this simple scoring system using the only three features predicted IVIG resistance almost as accurately as the Light GBM model itself.

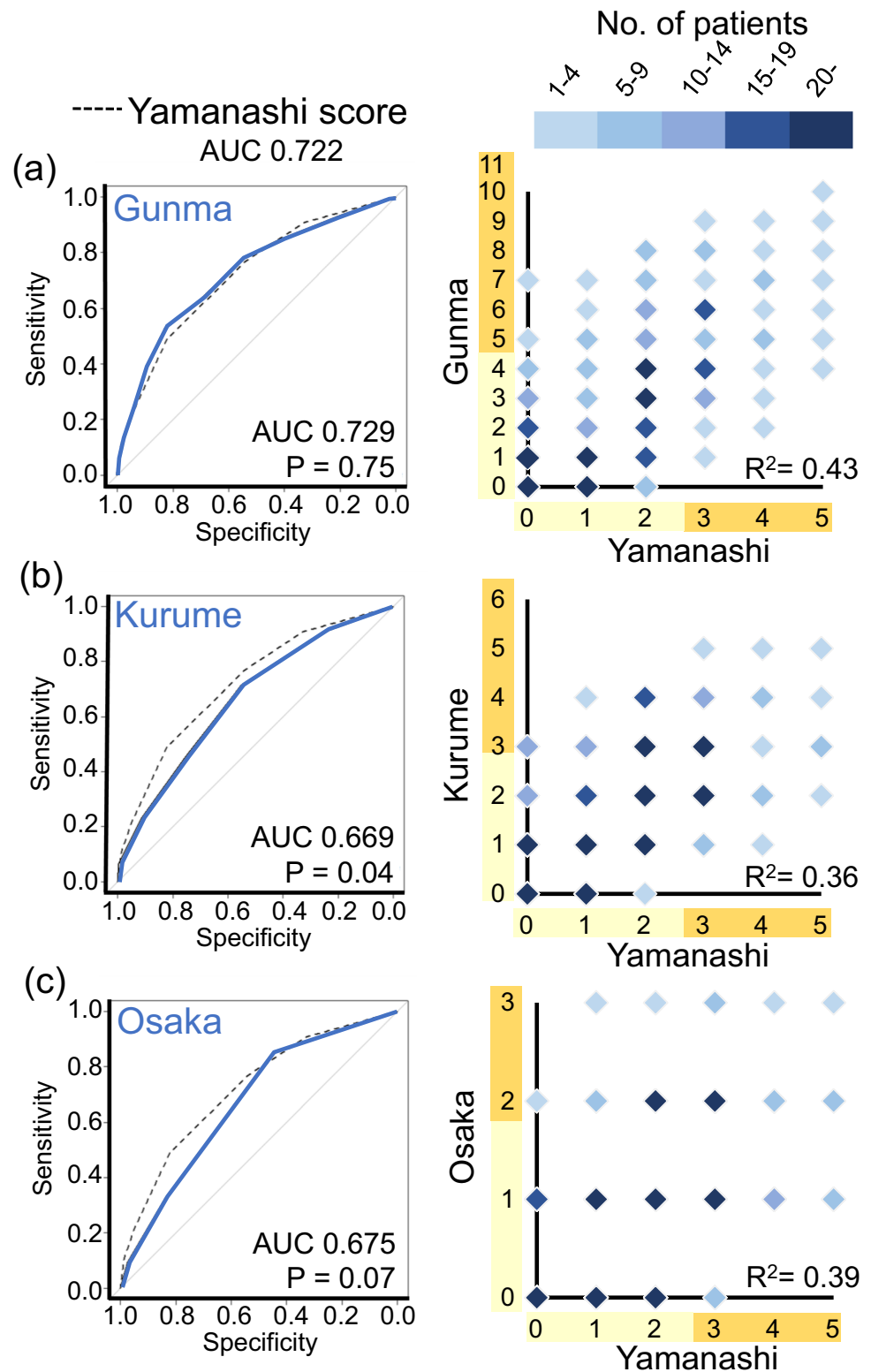
Moreover, Yamanashi score was as reliable as three previously established major scoring systems [11–13]. Among the three features of Yamanashi score, two features were included in three scoring systems [11–13] as follows: serum CRP level was included in all three scoring systems (Gunma [11], Kurume [12], and Osaka [13]), and days of illness at initial therapy was included in two scoring systems (Gunma [11] and Kurume [12]). In contrast, serum total cholesterol level was not included in the three previously established scoring systems. These observations suggest that serum total cholesterol level may make a significant and unique contribution for an accurate prediction of the Yamanashi score.

In the SHAP dependence plot of the present study, lower serum total cholesterol level (cutoff value, 131 mg/dL) was associated with higher risk of IVIG resistance. Our finding seems to be consistent with a previous finding showing that levels of serum total cholesterol decreased in the acute phase of KD patients due to abnormal lipid metabolism [37]. In particular, recent report by Shao et al. [38] revealed that serum total cholesterol level before the initial IVIG treatment was significantly lower in the cases of IVIG resistance in a single-center prospective cohort study. Although the underlying mechanism for association between dyslipidemia and the severity of systemic inflammation in KD remains unclear, a recent study by Zhang et al. [39] revealed that dyslipidemia during acute phase of KD was associated with aberrant levels of adipokines including adiponectin, omentin-1, and chemerin. In the above study by Shao et al. [38], alterations in the other lipid proteins were also associated with IVIG resistance: a higher level of triglyceride and lower levels of high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and apolipoprotein A. Thus, although the lipid profile was not fully evaluated in the present study, dyslipidemia due to systemic inflammation in the acute phase of KD patients may be a rational explanation for the usefulness of serum total cholesterol level as one of predictors for IVIG resistance in the Yamanashi score.

Three previous studies applied machine learning to predict the IVIG resistance in the patients with Kawasaki disease [14–16]. However, no scoring systems were proposed in these previous reports. Moreover, in two previous reports [14, 16], a list of the top features was unavailable. In the previous report by Wang et al. [15], the top three features were reported to be platelet count, serum calcium level, and the ratio of serum albumin level to globulin level. Of note, in their SHAP values that were partially consistent with those in our study, days of fever, serum cholesterol level, and serum CRP level were listed as the fourth, eighth, and nineteenth features in the 20 most important features among 82 variables, respectively.

This study has several limitations. First, since the majority of the subjects in the present study were of Japanese

Fig. 4 Comparison of Yamanashi score with three previously established scoring systems. Comparison of ROC curves (left panels) and correlations (right panels) between Yamanashi and Gunma (a), Kurume (b), and Osaka (c) scores in the 457 cases of the Yamanashi cohort study. Color scale indicates the number of cases



ethnicities, further validation is required before the present scoring system can be applied to other ethnicities and different populations. Second, although the patients were treated with a standardized protocol, the study was based

on retrospective data collection from a number of hospitals. Third, several known predictive factors such as neutrophil-to-lymphocyte and platelet-lymphocyte ratios [40] were not evaluated due to the lack of data collection. Recently,

utilities of coagulation profile [41], hepcidin [42], and genetic variants of the interleukin gene [43] have been also reported. Thus, machine learning using these factors as additional variables might improve the accuracy. Feature engineering of clinical variables is another possibility to further improve the accuracy [44]. Forth, insufficient reduction in the serum CRP level was additionally included in the definition of IVIG resistance in the present study as previously reported by others [12, 22], while only persistent fever was evaluated in many studies [45, 46]. Fifth, external validation for the practical use of our scoring system might be difficult since serum cholesterol level is not routinely evaluated in the other cohorts. In this context, however, when we applied this system to the recent 73 cases (IVIG resistance, 14 cases) who were diagnosed in the same 12 facilities from January 2021 to May 2022 and confirmed them to be negative for SARS-CoV2 at diagnosis by PCR or antigen test, prediction values in this system were similar to those in the previous three scoring systems (Supplemental Table 7) [11–13].

In conclusion, we implemented the machine learning algorithm to predict IVIG resistance in KD patients and confirmed its potential. Moreover, using only three features of the machine learning model, we designed a simple scoring system to predict IVIG resistance. Of note, in spite of its simplicity, the scoring system predicted IVIG resistance almost as accurately as the machine learning approach as well as three previously established major scoring systems.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10067-023-06502-1>.

Acknowledgements All authors express our sincere gratitude to all the members in the Yamanashi Kawasaki Disease Research Group, who supported acquisition of data. In addition to those listed as authors, the following investigators participated and cooperated in the acquisition of data for this study: Tomohiro Saito (Yamanashi Prefectural Central Hospital), Sho Hokibara (Kofu Municipal Hospital), Koji Kobayashi (Yamanashi Kosei Hospital), Tomoaki Sano and Toshie Nishijima (Yamanashi Red Cross Hospital), Hiroki Sato and Hiroaki Kanai (Suwa Central Hospital), Miwa Goto (National Hospital Organization Kofu National Hospital), Makoto Tsuruta (Kofu-Kyoritsu Hospital), Satoru Kojika and Makoto Nakamura (Fujiyoshida Municipal Hospital), Sonoko Mizorogi and Kinuko Saito (Nirasaki City Hospital), Masanori Ohta and Kazuya Takahashi (Tsuru Municipal General Hospital), and Kazumasa Sato and Mie Mochizuki (Kyonan Medical Center Fujikawa Hospital).

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Drs. Masashi Yoshizawa, Yosuke Kono, Yohei Hasebe, Keiichi Koizumi, and Minako Hoshiai. Drs. Takako Toda, Atsushi Watanabe, Nobuyuki Katsumata, and Prof. Eiryō Kawakami coordinated and supervised data collection. The first draft of the manuscript was written by Dr. Yuto Sunaga and Prof. Takeshi Inukai, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The datasets generated during and analyzed during the current study are not publicly available due to the risk of revealing the identity of the subjects but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate All authors confirm that approval was granted by the Research Ethics Committee of University of Yamanashi Hospital (Approval Number 1698) and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All authors confirm that all persons gave their informed consent prior to their inclusion in the study. Details that might disclose the identity of the subjects under study have been omitted.

Disclosures None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Kawasaki T, Kosaki F, Okawa S, Shigematsu I, Yanagawa H (1974) A new infantile acute febrile mucocutaneous lymph node syndrome (MLNS) prevailing in Japan. *Pediatrics* 54(3):271–276
2. McCrindle BW, Rowley AH, Newburger JW, Burns JC, Bolger AF, Gewitz M, Baker AL, Jackson MA, Takahashi M, Shah PB et al (2017) Diagnosis, treatment, and long-term management of Kawasaki Disease: a scientific statement for health professionals from the American Heart Association. *Circulation* 135(17):e927–999. <https://doi.org/10.1161/CIR.0000000000000484>
3. Newburger JW, Takahashi M, Beiser AS, Burns JC, Bastian J, Chung KJ, Colan SD, Duffy CE, Fulton DR, Glode MP, et al. (1991) A single intravenous infusion of gamma globulin as compared with four infusions in the treatment of acute Kawasaki syndrome. *New Engl J Med* 324:1633–1639. <https://doi.org/10.1056/NEJM199106063242305>
4. Tremoulet AH, Best BM, Song S, Wang S, Corinaldesi E, Eichenfield JR, Martin DD, Newburger JW, Burns JC (2008) Resistance to intravenous immunoglobulin in children with Kawasaki disease. *J Pediatr* 153:117–121. <https://doi.org/10.1016/j.jpeds.2007.12.021>
5. Muta H, Ishii M, Furui J, Nakamura Y, Matsuishi T (2006) Risk factors associated with the need for additional intravenous gamma-globulin therapy for Kawasaki disease. *Acta Paediatr* 95:189–193. <https://doi.org/10.1080/08035250500327328>
6. Kobayashi T, Saji T, Otani T, Nakamura T, Arakawa H, Kato T, Hara T, Hamaoka K, Ogawa S, Miura M et al (2012) 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* 379:1613–1620. [https://doi.org/10.1016/S0140-6736\(11\)61930-2](https://doi.org/10.1016/S0140-6736(11)61930-2)

7. Burns JC, Capparelli EV, Brown JA, Newburger JW, Glode MP (1998) Intravenous gamma-globulin treatment and retreatment in Kawasaki disease: US/Canadian Kawasaki Syndrome Study Group. *Pediatr Infect Dis J* 17:1144–1148. <https://doi.org/10.1097/00006454-199812000-00009>
8. Ogata S, Ogihara Y, Honda T, Kon S, Akiyama K, Ishii M (2012) Corticosteroid pulse combination therapy for refractory Kawasaki disease: a randomized trial. *Pediatrics* 129:e17–23. <https://doi.org/10.1542/peds.2011-0148>
9. Tremoulet AH, Jain S, Jaggi P, Jimenez-Fernandez S, Pancheri JM, Sun X, Kanegaye JT, Kovalchin JP, Printz BF, Ramilo O, Burns JC (2014) Infliximab for intensification of primary therapy for Kawasaki disease: a phase 3 randomised, double-blind, placebo-controlled trial. *Lancet* 383:1731–1738. [https://doi.org/10.1016/S0140-6736\(13\)62298-9](https://doi.org/10.1016/S0140-6736(13)62298-9)
10. Burns JC, Konecny P, Kuijpers T, Shimizu C, Tremoulet A, Arditi M (2017) Found in translation: international initiatives pursuing interleukin-1 blockade for treatment of acute Kawasaki disease. *Arthritis Rheumatol* 69:268–276. <https://doi.org/10.1002/art.39975>
11. Kobayashi T, Inoue Y, Takeuchi K, Okada Y, Tamura K, Tomomasa T, Kobayashi T, Morikawa A (2006) Prediction of intravenous immunoglobulin unresponsiveness in patients with Kawasaki disease. *Circulation* 113(22):2606–2612. <https://doi.org/10.1161/CIRCULATIONAHA.105.592865>
12. Egami K, Muta H, Ishii M, Suda K, Sugahara Y, Iemura M, Matsushita T (2006) Prediction of resistance to intravenous immunoglobulin treatment in patients with Kawasaki disease. *J Pediatr* 149(2):237–240. <https://doi.org/10.1016/j.jpeds.2006.03.050>
13. Sano T, Kurotobi S, Matsuzaki K, Yamamoto T, Maki I, Miki K, Kogaki S, Hara J (2007) Prediction of non-responsiveness to standard high-dose gamma-globulin therapy in patients with acute Kawasaki disease before starting initial treatment. *Eur J Pediatr* 166(2):131–137. <https://doi.org/10.1007/s00431-006-0223-z>
14. Kuniyoshi Y, Tokutake H, Takahashi N, Kamura A, Yasuda S, Tashiro M (2020) Comparison of machine learning models for prediction of initial intravenous immunoglobulin resistance in children with Kawasaki disease. *Front Pediatr* 8:570834. <https://doi.org/10.3389/fped.2020.570834>
15. Wang T, Liu G, Lin H (2020) A machine learning approach to predict intravenous immunoglobulin resistance in Kawasaki disease patients: a study based on a Southeast China population. *PLoS ONE* 15(8):e0237321. <https://doi.org/10.1371/journal.pone.0237321>
16. Liu J, Zhang J, Huang H, Wang Y, Zhang Z, Ma Y, He X (2021) A machine learning model to predict intravenous immunoglobulin-resistant Kawasaki disease patients: a retrospective study based on the Chongqing population. *Front Pediatr* 8(9):756095. <https://doi.org/10.3389/fped.2021.756095>
17. Takeuchi M, Inuzuka R, Hayashi T, Shindo T, Hirata Y, Shimizu N, Inatomi J, Yokoyama Y, Namai Y, Oda Y et al (2017) Novel risk assessment tool for immunoglobulin resistance in Kawasaki disease: application using a random forest classifier. *Pediatr Infect Dis J* 36(9):821–826. <https://doi.org/10.1097/INF.0000000000001621>
18. Kawakami E, Tabata J, Yanaiharu N, Ishikawa T, Koseki K, Iida Y, Saito M, Komazaki H, Shapiro JS, Goto C et al (2019) Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. *Clin Cancer Res* 25(10):3006–3015. <https://doi.org/10.1158/1078-0432.CCR-18-3378>
19. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, Chen KL, Yang CY, Lee OK (2020) Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care* 24(1):478. <https://doi.org/10.1186/s13054-020-03179-9>
20. Goecks J, Jalili V, Heiser LM, Gray JW (2020) How machine learning will transform biomedicine. *Cell* 181(1):92–101. <https://doi.org/10.1016/j.cell.2020.03.022>
21. Ayusawa M, Sonobe T, Uemura S, Ogawa S, Nakamura Y, Kiyosawa N, Ishii M, Harada K, et al. (2005) Revision of diagnostic guidelines for Kawasaki disease (the 5th revised edition). *Pediatr Int* 47:232–234. <https://doi.org/10.1111/j.1442-200x.2005.02033.x>
22. Moran AM, Newburger JW, Sanders SP, Parness IA, Spevak PJ, Burns JC, et al. (2000) Abnormal myocardial mechanics in Kawasaki disease: rapid response to gamma-globulin. *Am Heart J* 139:217–2. <https://doi.org/10.1067/mhj.2000.101221>
23. Koizumi K, Hoshiai M, Katsumata N, Toda T, Kise H, Hasebe Y, Kono Y, Sunaga Y, Yoshizawa M, Watanabe A et al (2018) Infliximab regulates monocytes and regulatory T cells in Kawasaki disease. *Pediatr Int* 60(9):796–802. <https://doi.org/10.1111/ped.13555>
24. Koizumi K, Hoshiai M, Moriguchi T, Katsumata N, Toda T, Kise H, Hasebe Y, Kono Y, Sunaga Y, Yoshizawa M et al (2019) Plasma exchange downregulates activated monocytes and restores regulatory T cells in Kawasaki disease. *Ther Apher Dial* 23(1):92–98. <https://doi.org/10.1111/1744-9987.12754>
25. Breiman L (2001) Random forests. *Mach Learn* 45:5–32
26. Chen T (2016) XGBoost: a scalable tree boosting system. 785–794. <https://doi.org/10.1145/2939672.2939785>
27. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Qi, Liu T-Y (2017) Light GBM: a highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* 30. (NIP 2017) 3149–3157
28. Noble WS (2006) What is a support vector machine? *Nat Biotechnol* 24(12):1565–1567. <https://doi.org/10.1038/nbt1206-1565>
29. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI* 14(2):1137–1145
30. Chawla NV, Bowyer KW, Hall LO (2002) SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357. <https://doi.org/10.1613/jair.953>
31. Fernández A, García S, Herrera F, Chawla NV (2018) SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *J Artif Intell Res* 61:863–905. <https://doi.org/10.1613/jair.1.11192>
32. Shapley LS (1953) A value for n-person games. In: Kuhn HW and Tucker AW (eds) *Contributions to the Theory of Games II*, Princeton University Press, Princeton 28:307–317
33. Rodríguez-Pérez R, Bajorath J (2020) Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J Comput Aided Mol Des* 34:1013–1026. <https://doi.org/10.1007/s10822-020-00314-0>
34. Lundberg SM, Lee S (2017) A unified approach to interpreting model predictions. <https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
35. Scott ML, Gabriel GF, Su-In L (2018) Consistent individualized feature attribution for tree ensembles. <https://arxiv.org/abs/1802.03888>
36. Kanda Y (2013) Investigation of the freely available easy-to-use software ‘EZ’ for medical statistics. *Bone Marrow Transplant* 48:452–458. <https://doi.org/10.1038/bmt.2012.244>
37. Salo E, Pesonen E, Viikari (1991) Serum cholesterol levels during and after Kawasaki disease. *J Pediatr* 119(4):557–561. [https://doi.org/10.1016/s0022-3476\(05\)82404-7](https://doi.org/10.1016/s0022-3476(05)82404-7)
38. Shao S, Zhou K, Liu X, Liu L, Wu M, Deng Y, Duan H, Li Y, Hua Y, Wang C (2021) Predictive value of serum lipid for intravenous immunoglobulin resistance and coronary artery lesion in Kawasaki disease. *J Clin Endocrinol Metab* 10: dgab230. <https://doi.org/10.1210/clinem/dgab230>

39. Zhang XY, Yang TT, Hu XF, Wen Y, Fang F, Lu HL (2018) Circulating adipokines are associated with Kawasaki disease. *Pediatr Rheumatol Online J* 16(1):33. <https://doi.org/10.1186/s12969-018-0243-z>
40. Kanai T, Takeshita S, Kawamura Y, Kinoshita K, Nakatani K, Iwashima S, Takizawa Y, Hirono K, Mori K, Yoshida Y et al (2020) The combination of the neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios as a novel predictor of intravenous immunoglobulin resistance in patients with Kawasaki disease: a multicenter study. *Heart Vessels* 35(10):1463–1472. <https://doi.org/10.1007/s00380-020-01622-z>
41. Shao S, Yang L, Liu X, Liu L, Wu M, Deng Y, Duan H, Li Y, Hua Y, Luo L et al (2021) Predictive value of coagulation profiles for both initial and repeated immunoglobulin resistance in Kawasaki disease: a prospective cohort study. *Pediatr Allergy Immunol* 32(6):1349–1359. <https://doi.org/10.1111/pai.13495>
42. Ishikawa T, Wada Y, Namba H (2021) Kawai T (2021) Hepcidin in Kawasaki disease: upregulation by acute inflammation in patients having resistance to intravenous immunoglobulin therapy. *Clin Rheumatol* 40:5019–5024. <https://doi.org/10.1007/s10067-021-05822-4>
43. Amano Y, Akazawa Y, Yasuda J, Yoshino K, Kojima H, Kobayashi N, Matsuzaki S, Nagasaki M, Kawai Y, Minegishi N, et al. (2019) A low-frequency IL4R locus variant in Japanese patients with intravenous immunoglobulin therapy-unresponsive Kawasaki disease. *Pediatr Rheumatol* 17(1). <https://doi.org/10.1186/s12969-019-0337-2>
44. Zheng A, Casari A (2018) Feature engineering for machine learning: principles and techniques for data scientists
45. Hamada H, Suzuki H, Onouchi Y, Ebata R, Terai M, Fuse S, Okajima Y, Kurotobi S, Hirai K, Soga T et al (2019) Efficacy of primary treatment with immunoglobulin plus ciclosporin for prevention of coronary artery abnormalities in patients 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* 393:1128–1137. [https://doi.org/10.1016/S0140-6736\(18\)32003-8](https://doi.org/10.1016/S0140-6736(18)32003-8)
46. Miyata K, Miura M, Kaneko T, Morikawa Y, Sakakibara H, Matsushima T, Misawa M, Takahashi T, Nakazawa M, Tsuchihashi T, et al (2021) Risk factors of coronary artery abnormalities and resistance to intravenous immunoglobulin plus corticosteroid therapy in severe Kawasaki Disease: an analysis of post RAISE. *Circ Cardiovasc Qual Outcomes* 14:e007191. <https://doi.org/10.1161/CIRCOUTCOMES.120.007191>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.