A machine-learning algorithm for diagnosis of multisystem inflammatory syndrome in children and Kawasaki disease in the USA: a retrospective model development and validation study

Jonathan Y Lam, Chisato Shimizu, Adriana H Tremoulet, Emelia Bainto, Samantha C Roberts, Nipha Sivilay, Michael A Gardiner, John T Kanegaye, Alexander H Hogan, Juan C Salazar, Sindhu Mohandas, Jacqueline R Szmuszkovicz, Simran Mahanta, Audrey Dionne, Jane W Newburger, Emily Ansusinha, Roberta L DeBiasi, Shiying Hao, Xuefeng B Ling, Harvey J Cohen, Shamim Nemati*, Jane C Burns*, the Pediatric Emergency Medicine Kawasaki Disease Research Group†, the CHARMS Study Group†

Summary

Background Multisystem inflammatory syndrome in children (MIS-C) is a novel disease that was identified during the COVID-19 pandemic and is characterised by systemic inflammation following SARS-CoV-2 infection. Early detection of MIS-C is a challenge given its clinical similarities to Kawasaki disease and other acute febrile childhood illnesses. We aimed to develop and validate an artificial intelligence algorithm that can distinguish among MIS-C, Kawasaki disease, and other similar febrile illnesses and aid in the diagnosis of patients in the emergency department and acute care setting.

Methods In this retrospective model development and validation study, we developed a deep-learning algorithm called KIDMATCH (Kawasaki Disease vs Multisystem Inflammatory Syndrome in Children) using patient age, the five classic clinical Kawasaki disease signs, and 17 laboratory measurements. All features were prospectively collected at the time of initial evaluation from patients diagnosed with Kawasaki disease or other febrile illness between Jan 1, 2009, and Dec 31, 2019, at Rady Children's Hospital in San Diego (CA, USA). For patients with MIS-C, the same data were collected from patients between May 7, 2020, and July 20, 2021, at Rady Children's Hospital, Connecticut Children's Medical Center in Hartford (CT, USA), and Children's Hospital Los Angeles (CA, USA). We trained a two-stage model consisting of feedforward neural networks to distinguish between patients with MIS-C and those without and then those with Kawasaki disease and other febrile illnesses. After internally validating the algorithm using stratified tenfold cross-validation, we incorporated a conformal prediction framework to tag patients with erroneous data or distribution shifts. We finally externally validated KIDMATCH on patients with MIS-C enrolled between April 22, 2020, and July 21, 2021, from Boston Children's Hospital (MA, USA), Children's National Hospital (Washington, DC, USA), and the CHARMS Study Group consortium of 14 US hospitals.

Findings 1517 patients diagnosed at Rady Children's Hospital between Jan 1, 2009, and June 7, 2021, with MIS-C (n=69), Kawasaki disease (n=775), or other febrile illnesses (n=673) were identified for internal validation, with an additional 16 patients with MIS-C included from Connecticut Children's Medical Center and 50 from Children's Hospital Los Angeles between May 7, 2020, and July 20, 2021. KIDMATCH achieved a median area under the receiver operating characteristic curve during internal validation of 98·8% (IQR 98·0–99·3) in the first stage and 96·0% (95·6–97·2) in the second stage. We externally validated KIDMATCH on 175 patients with MIS-C from Boston Children's Hospital (n=50), Children's National Hospital (n=42), and the CHARMS Study Group consortium of 14 US hospitals (n=83). External validation of KIDMATCH on patients with MIS-C correctly classified 76 of 81 patients (94% accuracy, two rejected by conformal prediction) from 14 hospitals in the CHARMS Study Group consortium, 47 of 49 patients (96% accuracy, one rejected by conformal prediction) from Boston Children's Hospital, and 36 of 40 patients (90% accuracy, two rejected by conformal prediction) from Children's National Hospital.

Interpretation KIDMATCH has the potential to aid front-line clinicians to distinguish between MIS-C, Kawasaki disease, and other similar febrile illnesses to allow prompt treatment and prevent severe complications.

Funding US Eunice Kennedy Shriver National Institute of Child Health and Human Development, US National Heart, Lung, and Blood Institute, US Patient-Centered Outcomes Research Institute, US National Library of Medicine, the McCance Foundation, and the Gordon and Marilyn Macklin Foundation.

Correspondence to:
Jonathan Y Lam, Department of
Biomedical Informatics,
University of California San
Diego, La Jolla, CA 92093, USA
j7lam@ucsd.edu

See **Online** for appendix

Research in context

Evidence before this study

Multisystem inflammatory syndrome in children (MIS-C) is a novel inflammatory disease identified during the COVID-19 pandemic with potential to cause permanent organ damage or life-threatening illness. Early detection of MIS-C remains a challenge given its clinical similarities to Kawasaki disease and other acute febrile childhood illnesses. Although there are reported differences in patients with MIS-C, such as older age, lower platelet count, and elevated inflammatory markers, there is no specific diagnostic test. Artificial intelligence (AI) has the potential to aid in early detection of MIS-C by modelling the complex relationships between clinical variables. We searched PubMed and medRxiv for research articles and preprints published between Jan 1, 2020, and Dec 1, 2021, using the terms "machine learning" OR "model" OR "score" AND "Kawasaki disease" AND "MIS-C". Of the five studies found, only one study focused on using clinical variables to distinguish between Kawasaki disease and MIS-C. Another study used clinical variables to distinguish between MIS-C, COVID-19, Kawasaki disease, and toxic shock syndrome. Both studies used a combination of demographic, laboratory, and clinical features to create diagnostic scores. However, the diagnostic scores from these studies were not tested against patients with other febrile illnesses and were not externally validated. Furthermore, some input features used in these studies might not be available in many clinical settings at the time of initial evaluation, such as

D-dimer and echocardiographic findings, and the data were acquired retrospectively without a standardised timeframe.

Added value of this study

To the best of our knowledge, our study is the first to use AI for screening of patients with MIS-C, Kawasaki disease, or similar febrile illnesses. We developed a deep-learning algorithm called KIDMATCH (Kawasaki Disease vs Multisystem Inflammatory Syndrome in Children) using clinical signs and laboratory data routinely collected during the initial evaluation of these patients. KIDMATCH showed consistent performance during external validation in patients with MIS-C from 16 hospitals across the USA. KIDMATCH is interpretable on a case-by-case basis by examining the most important features and whether they affect the MIS-C risk score positively or negatively. The conformal prediction framework identified outlier patients that might have been misclassified otherwise, thus identifying individuals for whom the model was not applicable.

Implications of all the available evidence

KIDMATCH showed the ability to distinguish between MIS-C, Kawasaki disease, and similar febrile illnesses using data available at the time of initial evaluation to the emergency department. These results highlight the potential of KIDMATCH as a clinical decision support system for diagnosing paediatric patients with MIS-C and Kawasaki disease in a timely manner to allow prompt treatment.

Introduction

As the COVID-19 pandemic spread, reports of children with a SARS-CoV-2-associated multisystem inflammatory condition emerged.[1–6] Clinical features of this new disorder, named multisystem inflammatory syndrome in children (MIS-C), include fever, gastrointestinal symptoms, conjunctival injection, rash, and elevated inflammatory markers.[7] Complications might include shock and multi-organ failure. According to the US Centers for Disease Control and Prevention (CDC), 8798 MIS-C cases and 71 MIS-C deaths had been reported nationwide as of Aug 1, 2022.[8] Despite its low prevalence, MIS-C is a serious condition with the potential to cause life-threatening illness, and the absence of a specific diagnostic test makes recognition of MIS-C a challenge.

Treatments for MIS-C include intravenous immunoglobulin, corticosteroids, and anti-inflammatory biological agents that rely on timely diagnosis of MIS-C to be most effective.[9,10] Kawasaki disease, an acute paediatric illness of unknown cause characterised by inflammation of the coronary arteries, is associated with fever and clinical criteria including rash, conjunctival injection, changes in lips or oropharyngeal mucosa, cervical lymphadenopathy, and changes in peripheral extremities.[11] Many of these clinical features overlap with MIS-C.[10,12] Although there are reported differences, such as older age, lower platelet count, and elevated inflammatory markers, in patients

with MIS-C compared with those with Kawasaki disease, none of these features alone or in combination is sufficient to diagnose MIS-C.[4] Artificial intelligence (AI) has the potential to aid in early detection of MIS-C by modelling the complex relationships between clinical variables, but, to the best of our knowledge, there is currently no machine-learning algorithm that differentiates Kawasaki disease from MIS-C.

In response to the difficulty clinicians have in diagnosis of and differentiation between MIS-C and Kawasaki disease, we aimed to develop and validate a clinical decision support system to distinguish among children with MIS-C, Kawasaki disease, and other febrile illnesses characterised by similar clinical and laboratory features in the emergency department.

Methods

Study design and participants

In this retrospective model development and validation study, we developed a two-stage AI model called KIDMATCH (Kawasaki Disease *vs* Multisystem Inflammatory Syndrome in Children) to classify patients as having MIS-C, Kawasaki disease, or other febrile illness using clinical signs and laboratory values that would be available at the time of a patient's initial evaluation. This study is reported in accordance with TRIPOD.

For internal validation, patients diagnosed with MIS-C between May 7, 2020, and July 20, 2021, were prospectively enrolled from Rady Children's Hospital (San Diego, CA, USA) and, to improve the generalisability of the results, Connecticut Children's Medical Center (Hartford, CT, USA) and Children's Hospital Los Angeles (Los Angeles, CA, USA; appendix p 3). To avoid the potential for misclassification, we prospectively enrolled patients with Kawasaki disease or other febrile illness who were diagnosed during an earlier time period, before the COVID-19 pandemic, between Jan 1, 2009, and Dec 31, 2019, at Rady Children's Hospital. All patients (ie, those with MIS-C, Kawasaki disease, or other illnesses) enrolled from Rady Children's Hospital were identified from a REDCap database at the Kawasaki Disease Research Center at the University of California San Diego (UCSD; San Diego, CA, USA). For external validation, patients with MIS-C were prospectively enrolled between April 22, 2020, and July 21, 2021, from Boston Children's Hospital (Boston, MA, USA), Children's National Hospital (Washington, DC, USA), and the CHARMS Study Group consortium (a 14-hospital database of patients with MIS-C funded by the Patient-Centered Outcomes Research Institute and housed at USCD).

Patients were diagnosed with MIS-C if they met the CDC case definition.[13] All patients with MIS-C had positive antibody testing for either the nucleocapsid or spike protein of SARS-CoV-2, and none had received a SARS-CoV-2 vaccine. Patients were diagnosed with Kawasaki disease if they met the case definition of the American Heart Association for either complete or incomplete Kawasaki disease.[14] All patients with Kawasaki disease were diagnosed and treated by one of two clinicians who are highly experienced in the treatment of patients with Kawasaki disease (JCB and AHT). Patients were diagnosed with other febrile illnesses if they met the following case definition: previously healthy child with fever for at least 3 days plus at least one of the clinical criteria for Kawasaki disease. More than 50% of patients with other febrile illnesses were referred for evaluation because of a clinical suspicion for Kawasaki disease. The final diagnoses for children with other febrile illnesses were adjudicated 2–3 months after enrolment by two experienced paediatric clinicians (JTK and JCB) who reviewed the clinical outcomes in the medical record and all available test results (appendix p 4). A viral syndrome was defined as a self-limited illness that resolved without treatment and without apparent sequelae. Written informed consent or assent as appropriate was obtained from parents and children, and the study was approved by the institutional review boards of the participating institutions. UCSD served as the central institutional review board of record for the CHARMS Study Group participants.

Data preprocessing
We used age, the five classic clinical Kawasaki disease signs, and 17 laboratory measurements as features for KIDMATCH based on guidance from the clinician collaborators and availability of laboratory test results for the majority of the training cohort (appendix pp 4–5). Features were collected prospectively at the time of initial evaluation in the emergency department at all sites. The five clinical signs were rash, conjunctival injection, changes in lips or oropharyngeal mucosa, cervical lymphadenopathy, and changes in peripheral extremities. Laboratory data were white blood cell count, age-adjusted haemoglobin, platelets, neutrophils, bands, lymphocytes, atypical lymphocytes, monocytes, eosinophils, absolute neutrophil count, absolute band count, erythrocyte sedimentation rate, C-reactive protein, alanine aminotransferase, γ-glutamyl transferase, albumin, and sodium. Because of the absence of bands and atypical lymphocytes in patients with automated differentials for the complete blood count, an indicator variable was added for the type of differential (0 was manual, and 1 was automated). For samples with automated differentials, we imputed the values for bands, atypical lymphocytes, and absolute band counts using the mean of the respective feature. Outlier values defined as less than the 0·5th percentile or greater than the 99·5th percentile were set to the values of 0·5th percentile if lower or 99·5th percentile if higher. All other missing laboratory values were imputed using K-nearest neighbours as the mean of the respective feature for the ten most similar samples from the training data. Data were normalised for each laboratory feature, except haemoglobin, after these transformations by subtracting the mean and dividing by the SD. Haemoglobin was normalised for age.

Model design
In KIDMATCH, we separated the classification of MIS-C, Kawasaki disease, and other febrile illnesses into two stages (figure 1). The model in stage 1 was trained to differentiate between MIS-C and other paediatric febrile conditions because of the clinical need to prioritise the identification of patients with MIS-C in the emergency department during admission and drop in performance when first differentiating between patients with febrile illnesses and those without in stage 1 (appendix p 6). The model in stage 2 was trained to further classify patients falling into the "other" category as other febrile illness or Kawasaki disease. Because Kawasaki disease and MIS-C data distributions could vary across different sites, we incorporated a conformal prediction framework within the model.[15] Conformal prediction reduces false alarms by identifying unfamiliar samples in new patient populations when compared with the training cohort and assigns indeterminate labels rather than making spurious predictions. If a test sample was rejected by the conformal prediction framework, no prediction was calculated. In stage 1, the model calculated an MIS-C risk score between 0 and 1 for test samples, with 1 being the highest MIS-C risk. In stage 2, the model calculated
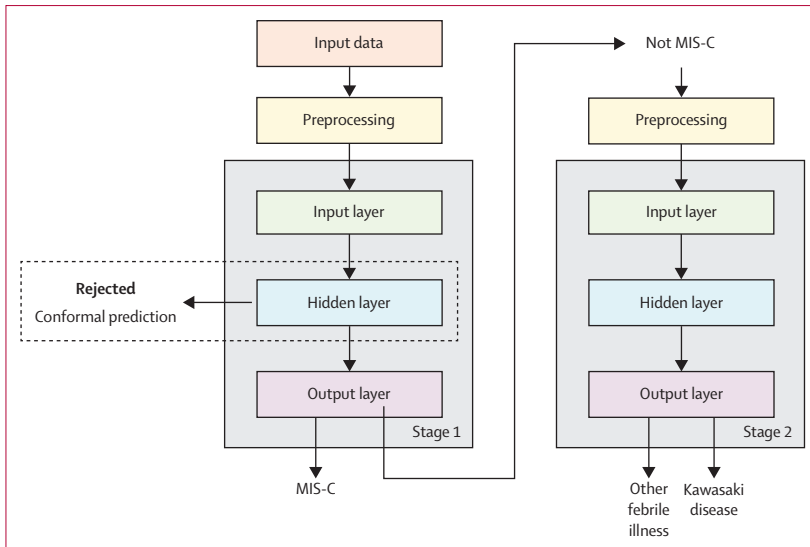
Figure 1: Model architecture
A patient could be classified as having MIS-C, other febrile illnesses, or Kawasaki disease if the input data were not rejected by the conformal prediction framework. MIS-C=multisystem inflammatory syndrome in children.

a Kawasaki disease risk score between 0 and 1, with 1 as the highest risk for Kawasaki disease.

We trained a feedforward neural network on each of the stages using Tensorflow (version 2.3.1) with a logistic regression using scikit-learn (version 0.24.2) as the baseline model. For both stages, the neural network was trained with the Adam optimiser at a learning rate of 0·01 and equally weighted batches of 100 from each class for stage 1 and 200 samples from each class for stage 2. Each neural network consisted of an input layer (23 units), a single hidden layer (ReLu activation function, L2 regularisation, 20% dropout rate), and a softmax output layer (2 units, binary cross-entropy loss function). The optimal number of units in the hidden layer was 12 for stage 1 and 16 for stage 2.

Model training and evaluation

We split the patients into training and internal validation cohorts using an 80:20 split and used stratified tenfold cross-validation to assess performance. Patients with any missing values were not considered for the internal test set based on the original design of the system user interface to generate a risk score from a complete set of features. Additionally, we observed a statistically non-significant decline in performance for both stages when including such patients in the internal validation test set (appendix p 7). All patients were included from the training cohort in stage 1, whereas patients with MIS-C were omitted from the training cohort in stage 2. Performance of the models was evaluated using accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) calculated at a minimum 95% sensitivity for the MIS-C classifications

for stage 1 and Kawasaki disease classifications for stage 2. We chose 95% sensitivity based on clinician feedback to avoid missing true positive cases. For external validation, patient data were first passed through the conformal prediction framework to determine whether a prediction should be made. Patient data that were not rejected were fed into the two-stage model to generate a final classification (MIS-C, other febrile illness, or Kawasaki disease) based on thresholds established during internal validation.

Conformal prediction

The trust sets used in the conformal prediction framework were constructed by filtering patients in the training cohort who had more than one missing value and MIS-C risk scores greater than the 95th percentile (appendix p 8). The weighted F1 score was calculated as the mean of the F1 score of all classes weighted by the support. We then constructed a trust set for each of the three classifications (other febrile illness, Kawasaki disease, and MIS-C). 200 randomly sampled patients from the internal training set were used for the other febrile illnesses and Kawasaki disease trust sets, and all patients with MIS-C were included in the MIS-C trust set. We generated feature representations for the conformal prediction framework by passing the preprocessed features of the training samples through the first hidden layer of the neural network. Feature representations from the test samples were compared with each of the conformal trust sets using cosine similarity, and if a test sample was rejected by all three trust sets based on a p value cutoff value of 0·05, then the model did not calculate risk scores for the test sample.

Shapley values

To explain the model predictions, we calculated the Shapley values for the test set using the Shapley Additive Explanations (SHAP) Python library.[16] Preprocessed data from the training set were used as the background to compare preprocessed test set data for stage 1 and stage 2. 100 random background samples from the internal training set were used to calculate the Shapley values for each feature in the internal test set.

Statistical analysis

p values were calculated with the Mann-Whitney U test for continuous variables between two groups, the Kruskal-Wallis test for three groups, the χ^2 test for categorical variables within the cohort used for internal validation, and the DeLong test[17] for AUC. Non-parametric statistical tests were chosen based on the mixture of normally and non-normally distributed features as determined by the Shapiro-Wilk test.

Devising a diagnostic test for two diseases for which there is no gold standard test presents special challenges. As a sensitivity analysis, we tested patients with Kawasaki disease who had coronary artery aneurysms

	MIS-C (n=135)	Kawasaki disease (n=775)	Other febrile illnesses (n=673)	p value*
Age†, years	8·4 (4·3–11·3)	2·9 (1·5–4·8)	3·4 (1·5–5·9)	<0·0001
Sex	..	..	..	NS
Male	77 (57%)	463 (60%)	394 (59%)	..
Female	58 (43%)	312 (40%)	279 (41%)	..
Ethnicity	..	..	..	<0·0001
Asian	6 (4%)	124 (16%)	47 (7%)	..
African American	14 (10%)	23 (3%)	10 (1%)	..
White	9 (7%)	167 (22%)	167 (25%)	..
Hispanic	94 (70%)	275 (35%)	272 (40%)	..
More than two races or other	12 (9%)	186 (24%)	114 (17%)	..
No information	0	0	63 (9%)	..
Maximum Z score‡	1·6 (0·9–2·4)	1·7 (1·2–2·4)	NA	NS
Lowest left ventricle ejection fraction	57% (48–61)	66% (63–70)	NA	<0·0001
Illness day of sample collection§	4 (3–6)	5 (4–7)	5 (4–7)	NS
Automated differential†	32 (24%)	33 (4%)	91 (14%)	<0·0001
Clinical signs†				
Rash	76 (56%)	715 (92%)	442 (66%)	<0·0001
Conjunctival injection	81 (60%)	725 (94%)	343 (51%)	<0·0001
Changes in lips or oropharyngeal mucosa	50 (37%)	723 (93%)	304 (45%)	<0·0001
Cervical lymphadenopathy	22 (16%)	274 (35%)	151 (22%)	<0·0001
Peripheral extremity changes	18 (13%)	625 (81%)	131 (19%)	<0·0001
Laboratory data				
White blood cell count†, 10³ cells per µL	9·8 (6·9–13·0)	13·2 (10·4–17·0)	9·5 (6·4–13·3)	<0·0001
Neutrophils†	70% (58–80)	58% (47–69)	47% (31–62)	<0·0001
Bands†	12% (2–24)	7% (2–15)	5% (2–11)	<0·0001
Lymphocytes†	11% (6–20)	20% (12–31)	32% (18–47)	<0·0001
Atypical lymphocytes†	0% (0–1)	0% (0–1)	1% (0–3)	<0·0001
Monocytes†	3% (1–5)	6% (3–8)	8% (5–11)	<0·0001
Eosinophils†	1% (0–3)	2% (1–4)	0% (0–1)	<0·0001
Absolute neutrophil count†, cells per µL	7980 (5476–10 416)	8892 (6305–11 768)	4800 (2599–8208)	<0·0001
Absolute band count†, cells per µL	1160 (236–2167)	891 (299–2095)	438 (148–1205)	<0·0001
Absolute lymphocyte count, cells per µL	962 (585–1797)	2565 (1412–3963)	2730 (1633–4356)	<0·0001
Haemoglobin concentration normalised for age†	−1·6 (−2·9 to −0·7)	−1·3 (−2·3 to −0·5)	−0·3 (−1·3 to 0·5)	<0·0001
Platelet count†, 10³ per L	160 (111–222)	339 (267–426)	247 (184–322)	<0·0001
Erythrocyte sedimentation rate†, mm/h	49 (31–75)	60 (39–75)	29 (15–45)	<0·0001
C-reactive protein†, mg/dL	18·7 (8·2–25·7)	7·0 (4·3–16·9)	2·9 (1·2–6·0)	<0·0001
Alanine aminotransferase†, IU/L	39 (21–63)	46 (26–117)	27 (19–38)	<0·0001
γ-glutamyl transferase†, IU/L	36 (24–85)	46 (18–128)	15 (12–20)	<0·0001
Albumin†, g/dL	3·6 (3·1–4·0)	3·8 (3·5–4·2)	4·1 (3·8–4·4)	<0·0001
Sodium†, mmol/L	133 (130–135)	137 (134–139)	138 (136–139)	<0·0001

Data are n (%) or median (IQR). Percentages might not sum to 100 as a result of rounding. MIS-C=multisystem inflammatory syndrome in children. NA=not applicable. NS=not significant. *p values were calculated with the Mann-Whitney U test for continuous variables between two groups, the Kruskal-Wallis test for three groups, and the x^2 test for categorical variables. †Model feature. ‡Maximum Z score (internal diameter normalised for body surface area) for the right and left anterior descending coronary arteries. §Illness day 1 was the first day of fever.

Table 1: Demographic and clinical characteristics of patients used in internal validation from Rady Children's Hospital San Diego (CA, USA; n=1517), Connecticut Children's Medical Center (Hartford, CT, USA; n=16), and Children's Hospital Los Angeles (CA, USA; n=50)

and patients with MIS-C who had a reduced left ventricular ejection fraction as characteristic patient subsets that were unlikely to be misclassified by clinicians. We also tested patients with other febrile illnesses diagnosed with bacterial infection as an additional sensitivity analysis based on the low prevalence of this population. To assess how the algorithm performed compared with human judgement, we extracted the raw feature values from 101 random samples of the internal test set for the final model and had two experienced paediatric infectious disease clinicians with Kawasaki disease expertise at Rady Children's Hospital (AHT and JCB) assign diagnoses based on the features alone.

Role of the funding source

The funders of the study were not involved in the study design, the collection, analysis, and interpretation of data, the writing of the report, or the decision to submit the paper for publication.

Results

1517 patients diagnosed at Rady Children's Hospital between Jan 1, 2009, and June 7, 2021, with MIS-C (n=69), Kawasaki disease (n=775), or other febrile illnesses (n=673) were identified for internal validation. Laboratory tests and clinical signs were obtained at the time of initial evaluation and before treatment for all patients. We added patients with MIS-C from Connecticut Children's Medical Center (n=16) and Children's Hospital Los Angeles (n=50), who were enrolled between May 7, 2020, and July 20, 2021, when training the model for a total of 135 patients with MIS-C during internal validation (table 1). External validation was performed using MIS-C clinical data (n=175) from patients enrolled between April 22, 2020, and July 21, 2021, at Boston Children's Hospital (n=50), Children's National Hospital (n=42), and the CHARMS Study Group consortium (n=83). In a comparison of data among the groups, patients with MIS-C had higher band counts, lower sodium concentration, lower platelet counts, and higher C-reactive protein (p<0·0001), and were older, than those in the other febrile illness and Kawasaki disease cohorts, consistent with previous reports (table 1).[1,4]

The stratified tenfold cross-validation results for stages 1 and 2 are shown in table 2. During internal validation, KIDMATCH achieved a median AUC of 98·8% (IQR 98·0–99·3) in the first stage and 96·0% (95·6–97·2) in the second stage. The neural network in stage 1 had similar accuracy, sensitivity, specificity, PPV,

and NPV compared with the logistic regression baseline in the validation set when classifying samples as MIS-C or not MIS-C. In stage 2, the neural network compared favourably with the logistic regression baseline in terms of accuracy, PPV, and specificity when classifying samples as other febrile illness or Kawasaki disease on the basis of thresholds set at 95% sensitivity for Kawasaki disease samples.

We selected the models in the cross-validation with the highest stage 1 accuracy to use in the final model. To ensure model generalisability and similar performance across external sites, we constructed a conformal prediction framework using the training samples from the final model.[15] Briefly, we selected the parsimonious combination of missingness and risk score with the highest weighted F1 score to construct the conformal trust sets (appendix p 8). This approach rejected three (2%) of the 149 other febrile illness samples, six (4%) of the 165 Kawasaki disease samples, and none of the MIS-C samples in the internal validation test set. The receiver operating characteristic curves for the final model indicated that the neural networks had high sensitivity and specificity with an AUC of 0·982 in stage 1 and 0·950 in stage 2 (figure 2).

The neural networks trained for each stage in the final model showed robust performance when setting thresholds (stage 1 –0·36, stage 2 –0·60) at a 95% sensitivity level. Although there was no statistical difference in the AUC for stage 1 (p=0·17) and stage 2 (p=0·59), the neural networks were chosen for the final model as the conformal prediction framework relied on feature representations that could not be calculated with logistic regression. In addition, the majority of patients at UCSD had a complete blood count with a manual differential, but the external sites had a significant proportion of automated differentials. The neural

	Accuracy	AUC	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Stage 1						
Training						
Logistic regression	95·4% (95·0–95·5)	98·6% (98·5–98·7)	92·4% (91·6–93·3)	95·6% (95·4–96·0)	70·9% (69·4–72·2)	99·1% (99·0–99·2)
Neural network	96·4% (96·1–97·2)	99·5% (99·4–99·6)	100·0% (99·2–100·0)	96·3% (96·0–96·9)	74·7% (73·9–78·8)	100·0% (99·9–100·0)
Validation						
Logistic regression	96·4% (96·1–96·8)	98·5% (98·0–98·8)	93·8% (87·5–100·0)	97·0% (95·8–97·3)	65·0% (59·3–66·7)	99·6% (99·2–100·0)
Neural network	96·4% (96·1–97·8)	98·8% (98·0–99·3)	93·8% (93·8–100·0)	97·0% (95·8–98·1)	63·6% (59·3–75·0)	99·6% (99·6–100·0)
Stage 2						
Training						
Logistic regression	91·4% (91·2–91·5)	97·2% (97·0–97·3)	95·1% (95·0–95·1)	86·7% (86·1–86·9)	89·8% (89·6–90·0)	93·4% (93·3–93·5)
Neural network	91·7% (91·4–91·8)	97·4% (97·3–97·4)	95·0% (95·0–95·1)	87·6% (86·9–88·0)	90·4% (90·0–90·7)	93·4% (93·4–93·5)
Validation						
Logistic regression	88·6% (88·2–90·1)	96·1% (95·5–96·7)	94·6% (94·6–95·3)	80·3% (78·9–84·3)	86·1% (85·2–88·1)	92·4% (92·0–92·9)
Neural network	90·1% (89·4–90·9)	96·0% (95·6–97·2)	94·6% (94·6–94·6)	84·3% (82·5–86·1)	88·6% (87·6–89·7)	92·4% (92·2–92·5)

Data are median (IQR). AUC=area under the receiver operating characteristic curve.

Table 2: Tenfold stratified cross-validation performance metrics for the training and validation cohorts in stage 1 and stage 2

networks were able to adjust for the difference between manual and automated differential complete blood counts effectively by incorporating an indicator variable as input to the model.

In a sensitivity analysis, we tested patients with Kawasaki disease who had coronary artery aneurysms and patients with MIS-C who had a reduced left ventricular ejection fraction as characteristic patient subsets that were unlikely to be misclassified by clinicians. The final model correctly assigned 124 (87%) of 142 patients with Kawasaki disease who had coronary artery aneurysms (31 [84%] of 37 in the test set) and 19 (95%) of 20 patients with MIS-C who had a reduced left ventricular ejection fraction (three [75%] of four in the test set; appendix p 9). We also tested patients with other febrile illnesses diagnosed with bacterial infection as an additional sensitivity analysis. Of the children with other febrile illnesses with documented bacterial infection, 69 (83%) of 83 (11 [79%] of 14 in the test set) were classified correctly. We extracted the raw feature values from 101 random samples of the internal test set for the final model and had two experienced paediatric infectious disease clinicians with Kawasaki disease expertise at Rady Children's Hospital (AHT and JCB) assign diagnoses based on the features alone. The algorithm outperformed both clinicians with an accuracy of 86% (87 of 101 cases correctly assigned) compared with 79% (80 of 101) and 80% (81 of 101; appendix p 10).

To determine how the features contributed to the model predictions, we used Shapley values—specifically the SHAP method.[16] In stage 1, the most important features that distinguished patients with MIS-C from those without were serum sodium, platelet count, neutrophils, and C-reactive protein (figure 3). The patterns observed were consistent with published reports of the laboratory testing characteristics of patients with MIS-C.[14] In stage 2, changes in peripheral extremities, conjunctival injection, erythrocyte sedimentation rate, and changes in the lips or oropharyngeal mucosa were the most important features for differentiating between other febrile illnesses and Kawasaki disease. Three of the four features mentioned previously are clinical signs used by clinicians to diagnose Kawasaki disease, so it is not surprising that the presence of one or more of these clinical signs contributed to a higher stage 2 risk score and higher probability of Kawasaki disease. The next most important features were age, with younger patients more likely to have Kawasaki disease, and γ-glutamyl transferase, higher levels of which indicate hepatobiliary inflammation, which is often observed in Kawasaki disease.[18]

We externally validated KIDMATCH using patients with MIS-C from the CHARMS Study Group consortium, Boston Children's Hospital, and Children's National Hospital. Our conformal prediction framework rejected two (2%) of 83 samples from the CHARMS Study Group consortium, one (2%) of 50 from Boston's Children's
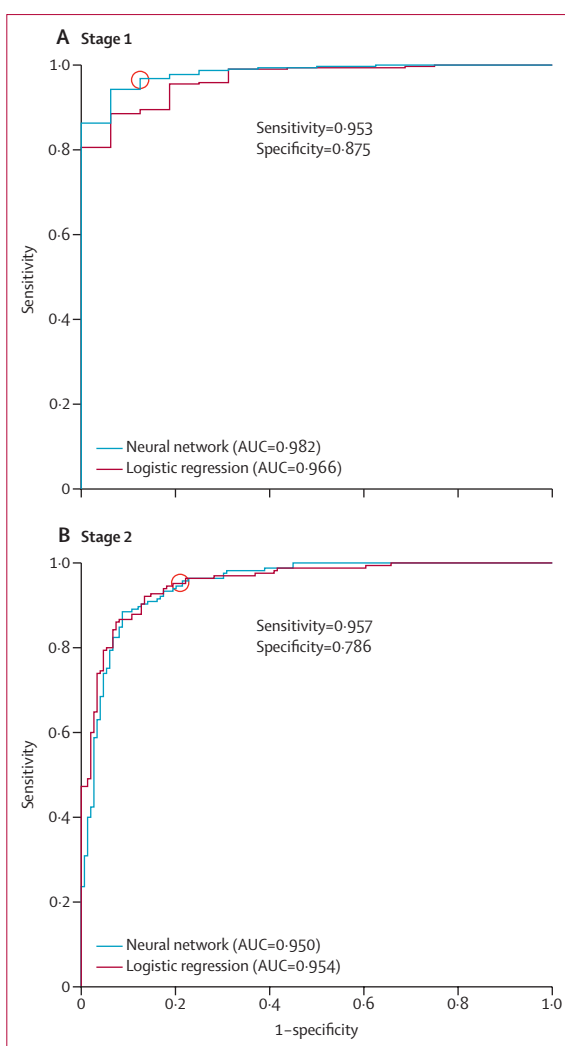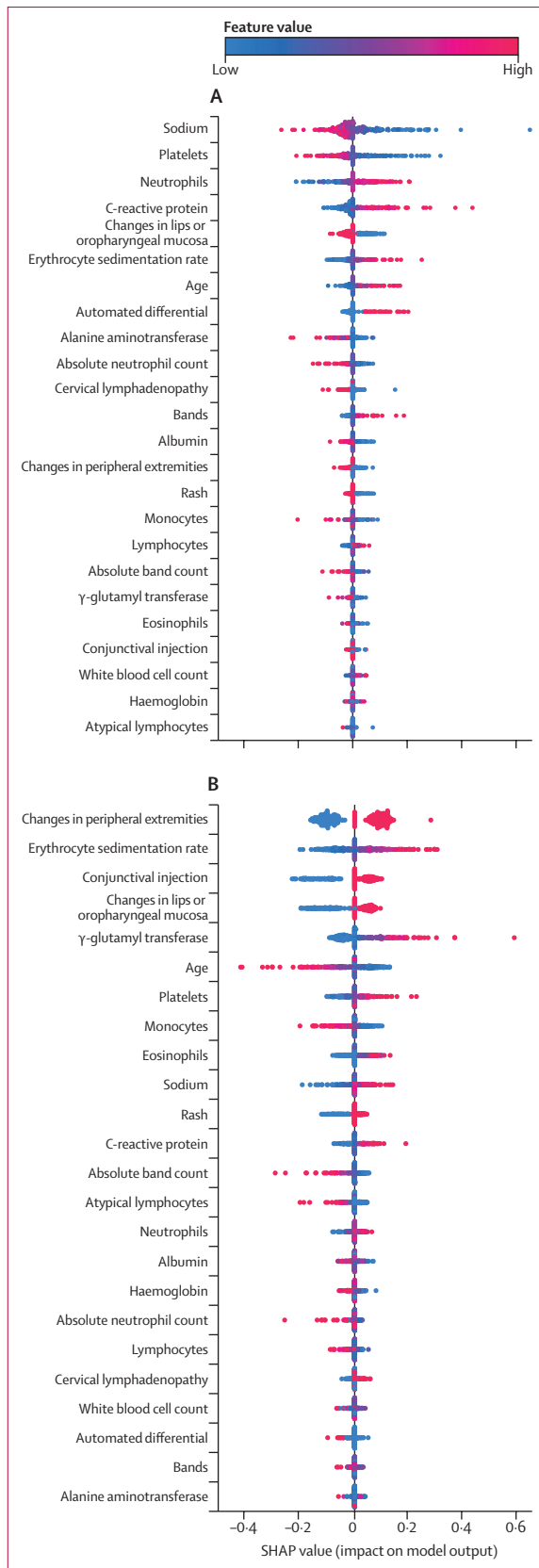

Figure 2: **ROCs for stage 1 (A) and stage 2 (B) in the final model**
Thresholds for each stage were set based on the red circle on the ROC for the neural network. AUC=area under the receiver operating characteristic curve. MIS-C=multisystem inflammatory syndrome in children. ROC=receiver operating characteristic curve.

Hospital, and two (5%) of 42 from the Children's National Hospital (table 3). KIDMATCH accurately predicted MIS-C in 76 (94%) of 81 samples from the CHARMS Study Group consortium (appendix p 11), 47 (96%) of 49 samples from Boston Children's Hospital, and 36 (90%) of 40 samples from the Children's National Hospital after conformal prediction (table 3). Examination of the stage 1 risk scores for each site revealed that most patients with MIS-C were confidently classified as MIS-C (risk score of >0·8) with conformal prediction successfully identifying false alarms from the rejected patients (appendix p 12).

KIDMATCH generalised well to external MIS-C cohorts, with 90% or greater accuracy at all three external sources despite missingness of one to four features, most often (88%) γ-glutamyl transferase. The model had the lowest prediction accuracy for Children's National

Figure 3: **SHAP summary plot for stage 1 (A) and stage 2 (B) with raw feature values**
A feature for a patient with a SHAP value below 0 decreases the risk score. In stage 1, a higher risk score indicates a higher probability of MIS-C. In stage 2, a higher risk score indicates a higher probability of Kawasaki disease. Features are ranked in order of importance from top to bottom. MIS-C=multisystem inflammatory syndrome in children. SHAP=Shapley Additive Explanations.

Hospital at 90% (table 3). Further investigation revealed that their laboratory values for albumin were significantly lower than the MIS-C training distribution (median 2·9 g/dL [IQR 2·5–3·2] vs 3·6 g/dL [3·2–4·0], p<0·0001) due to differences in the test platform used by that clinical laboratory (appendix p 13). In addition, all misclassified patients with MIS-C from Children's National Hospital had a normal serum sodium of 138 mmol/L or higher, and the distribution of serum sodium values from this laboratory was significantly higher than from the other MIS-C clinical sites (median 136 [IQR 134–139] vs 133 [130–135], p<0·0001). Although these values deviated from those observed in other sites, outlier serum albumin and sodium values were observed in patients with MIS-C in the training cohort, and the model showed consistent performance when handling samples with outlier values. The reliance of the stage 1 algorithm on characteristic MIS-C laboratory test values such as low serum sodium and low platelet count increases the probability of misclassification when presented with normal values from a patient with MIS-C (appendix p 13). However, the model enables clinicians to explore how the relevant features are contributing to the risk score and adjust their clinical judgement accordingly.

Discussion

We present a machine-learning model for screening of patients with MIS-C, Kawasaki disease, or similar febrile illnesses using clinical signs and laboratory data routinely collected during the initial evaluation of these patients. To the best of our knowledge, this is the first application of AI to aid in the diagnosis of MIS-C and differentiate it from Kawasaki disease and other febrile illnesses. KIDMATCH has the ability to reject test samples that are outside the distribution in the training set, which provides a measure of confidence by statistically identifying outlier inputs. It is interpretable on a case-by-case basis by examining the most important features and whether they affect the MIS-C risk score positively or negatively. KIDMATCH showed consistent performance across different hospitals, and the conformal prediction framework identified outlier patients that would have been misclassified otherwise. A web calculator for KIDMATCH was developed using Streamlit, an open-source framework for building applications in Python, to assist clinicians with calculation of the proposed risk scores and assessment of the top factors contributing to risk (appendix pp 14–18). It was internally deployed at

	CHARMS Study Group consortium* (n=83)	Boston Children's Hospital (MA, USA; n=50)	Children's National Hospital (Washington, DC, USA; n=42)
Rejected	2 (2%)†	1 (2%)‡	2 (5%)†
Other febrile illnesses	3/81 (4%)	2/49 (4%)	3/40 (8%)
Kawasaki disease	2/81 (2%)	0	1/40 (3%)
MIS-C	76/81 (94%)	47/49 (96%)	36/40 (90%)

Data are n (%) or n/N (%). Percentages might not sum to 100 as a result of rounding. Percentages are based on the total number of patients from each site who were not rejected by conformal prediction. MIS-C=multisystem inflammatory syndrome in children. *Consisted of patients from 14 US hospitals. †Classifications were other febrile illnesses and Kawasaki disease. ‡Classification was other febrile illness.

Table 3: Predicted classifications of patients with MIS-C from external sites

Rady Children's Hospital and is currently being updated with clinician feedback in an ongoing single-site prospective implementation study. The laboratory tests incorporated into KIDMATCH (complete blood count, comprehensive metabolic panel, erythrocyte sedimentation rate, C-reactive protein, and γ-glutamyl transferase) are commonly obtained for paediatric patients in many outpatient and inpatient medical settings, and the clinical features are easy to assess by front-line clinicians. The use of such readily available data enables KIDMATCH to be potentially deployable immediately across the USA without the need for specialised laboratory tests.

A strength of our work is the universal availability of the required features in the majority of health-care settings and the validation using external cohorts. Two recent studies[19,20] have created diagnostic scores to distinguish between Kawasaki disease and MIS-C. Both studies have the same approach as our model in using a combination of demographic information, laboratory tests, and clinical features for risk assessment. However, models in both studies were not tested against control paediatric patients with other febrile illnesses and were not externally validated. Laboratory values for both studies were collected without a standardised timeframe, using the highest or lowest values instead of extracting values at a set timepoint, whereas our study used the first result from the initial evaluation. In addition, the Kostik score[19] included D-dimer, which might not be available in many clinical settings, and the Godfred-Cato scores[20] include pericardial effusion and other echocardiographic findings that would not be readily available in an emergency room. KIDMATCH uses routinely ordered laboratory studies and assessable clinical features, making it an effective screening tool at the point of initial evaluation before more costly testing is ordered.

We recognise limitations of our work due to the absence of a gold standard for Kawasaki disease or MIS-C diagnosis and the limited availability of febrile illness and Kawasaki disease data for external validation. We cannot exclude some degree of misdiagnosis in either the training or test set. However, the internal validation performance showed consistency with known discriminating features that were used as input for KIDMATCH. It is unknown whether a simpler model or addition of other routine laboratory tests would have a similar or better performance. It is also unknown how the model would perform on patients with other febrile illnesses or Kawasaki disease from other hospitals because we trained our model on pre-pandemic febrile illness and Kawasaki disease data from a single site. The thresholds established during internal validation might not be generalisable to different sites, and shifting the threshold might be required to adjust for different prevalence rates. However, the high model AUC means that the model can be used to effectively prioritise patients with a febrile illness for further evaluation of MIS-C or Kawasaki disease. A key step for deployment will be to establish standardised conditions for use so that the algorithm is applied to the appropriate patients. The current algorithm is only optimised for laboratory test values collected at the time of initial evaluation, and it is unknown how it would perform with data collected at a later timepoint. It is also unknown how end users should deal with patients flagged as indeterminate, but a possible solution could be to order more specialised tests such as ferritin, troponin, B-type natriuretic peptide or N-terminal pro B-type natriuretic peptide, and D-dimer, as well as IgG antibody against SARS-CoV-2, as is routine practice for patients with suspected MIS-C.[21]

On the basis of these results, the proposed algorithm is a generalisable and accurate tool for the diagnosis of MIS-C and Kawasaki disease during the initial evaluation of patients with suspected disease. Future work will include retrospective validation in external patients with other febrile illnesses or Kawasaki disease and prospective validation in patients with MIS-C, as well as refining the implementation of KIDMATCH within the clinical workflow. As the first, to the best of our knowledge, externally validated machine-learning solution for the diagnosis of MIS-C, KIDMATCH has the potential to aid front-line clinicians and improve patient outcomes through timely diagnosis.

References

1   Chiotos K, Bassiri H, Behrens EM, et al. Multisystem inflammatory syndrome in children during the coronavirus 2019 pandemic: a case series. *J Pediatric Infect Dis Soc* 2020; **9:** 393–98.

2   Verdoni L, Mazza A, Gervasoni A, et al. An outbreak of severe Kawasaki-like disease at the Italian epicentre of the SARS-CoV-2 epidemic: an observational cohort study. *Lancet* 2020; **395:** 1771–78.

3   Belhadjer Z, Méot M, Bajolle F, et al. Acute heart failure in multisystem inflammatory syndrome in children in the context of global SARS-CoV-2 pandemic. *Circulation* 2020; **142:** 429–36.

4   Whittaker E, Bamford A, Kenny J, et al. Clinical characteristics of 58 children with a pediatric inflammatory multisystem syndrome temporally associated with SARS-CoV-2. *JAMA* 2020; **324:** 259–69.

5   Feldstein LR, Rose EB, Horwitz SM, et al. Multisystem inflammatory syndrome in U.S. children and adolescents. *N Engl J Med* 2020; **383:** 334–46.

6   Dufort EM, Koumans EH, Chow EJ, et al. Multisystem inflammatory syndrome in children in New York State. *N Engl J Med* 2020; **383:** 347–58.

7   Jiang L, Tang K, Levin M, et al. COVID-19 and multisystem inflammatory syndrome in children and adolescents. *Lancet Infect Dis* 2020; **20:** e276–88.

8   US Centers for Disease Control and Prevention. Health Department-reported cases of multisystem inflammatory syndrome in children (MIS-C) in the United States. https://covid.cdc.gov/covid-data-tracker/#mis-national-surveillance (accessed Aug 25, 2022).

9   McArdle AJ, Vito O, Patel H, et al. Treatment of multisystem inflammatory syndrome in children. *N Engl J Med* 2021; **385:** 11–22.

10 Sharma C, Ganigara M, Galeotti C, et al. Multisystem inflammatory syndrome in children and Kawasaki disease: a critical comparison. *Nat Rev Rheumatol* 2021; **17:** 731–48.

11 Godfred-Cato S, Bryant B, Leung J, et al. COVID-19-associated multisystem inflammatory syndrome in children—United States, March–July 2020. *MMWR Morb Mortal Wkly Rep* 2020; **69:** 1074–80.

12 Elias MD, McCrindle BW, Larios G, et al. Management of multisystem inflammatory syndrome in children associated with COVID-19: a survey from the International Kawasaki Disease Registry. *CJC Open* 2020; **2:** 632–40.

13 US Centers for Disease Control and Prevention. Multisystem inflammatory syndrome in children (MIS-C) associated with coronavirus disease 2019 (COVID-19). https://emergency.cdc.gov/han/2020/han00432.asp (accessed Nov 10, 2021).

14 McCrindle BW, Rowley AH, Newburger JW, et al. Diagnosis, treatment, and long-term management of Kawasaki disease: a scientific statement for health professionals from the American Heart Association. *Circulation* 2017; **135:** e927–99.

15 Shashikumar SP, Wardi G, Malhotra A, Nemati S. Artificial intelligence sepsis prediction algorithm learns to say "I don't know". *NPJ Digit Med* 2021; **4:** 134.

16 Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017; **30:** 4768–77.

17 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44:** 837–45.

18 Eladawy M, Dominguez SR, Anderson MS, Glodé MP. Abnormal liver panel in acute Kawasaki disease. *Pediatr Infect Dis J* 2011; **30:** 141–44.

19 Kostik MM, Bregel LV, Avrusin IS, et al. Distinguishing between multisystem inflammatory syndrome, associated with COVID-19 in children and the Kawasaki disease: development of preliminary criteria based on the data of the retrospective multicenter cohort study. *Front Pediatr* 2021; **9:** 787353.

20 Godfred-Cato S, Abrams JY, Balachandran N, et al. Distinguishing multisystem inflammatory syndrome in children from COVID-19, Kawasaki disease and toxic shock syndrome. *Pediatr Infect Dis J* 2022; **41:** 315–23.

21 Abrams JY, Godfred-Cato SE, Oster ME, et al. Multisystem inflammatory syndrome in children associated with severe acute respiratory syndrome coronavirus 2: a systematic review. *J Pediatr* 2020; **226:** 45–54.