2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340801

A Deep Learning Framework for Image-Based Screening of Kawasaki Disease

Jonathan Y. Lam, John T. Kanegaye, Ellen Xu, Michael A. Gardiner, Jane C. Burns, Shamim Nemati, and Adriana H. Tremoulet

***Abstract*— Kawasaki disease (KD) is a leading cause of acquired heart disease in children and is characterized by the presence of a combination of five clinical signs assessed during the physical examination. Timely treatment of intravenous immunoglobin is needed to prevent coronary artery aneurysm formation, but KD is usually diagnosed when pediatric patients are evaluated by a clinician in the emergency department days after onset. One or more of the five clinical signs usually manifests in pediatric patients prior to ED admission, presenting an opportunity for earlier intervention if families receive guidance to seek medical care as soon as clinical signs are observed along with a fever for at least five days. We present a deep learning framework for a novel screening tool to calculate the relative risk of KD by analyzing images of the five clinical signs. The framework consists of convolutional neural networks to separately calculate the risk for each clinical sign, and a new algorithm to determine what clinical sign is in an image. We achieved a mean accuracy of 90% during 10-fold cross-validation and 88% during external validation for the new algorithm. These results demonstrate the algorithms in the proposed screening tool can be utilized by families to determine if their child should be evaluated by a clinician based on the number of clinical signs consistent with KD.**

***Clinical Relevance*— This screening framework has the potential for earlier clinical evaluation and detection of KD to reduce the risk of coronary artery complications.**

I. Introduction

Kawasaki disease (KD) is an idiopathic febrile disease primarily affecting children younger than 5 years of age that leads to coronary artery aneurysms (CAAs) in about 25% of untreated cases [1]. It is characterized by five clinical signs: rash, bilateral conjunctival erythema, cervical lymphadenopathy, changes in the lips and oral cavity, and changes in the extremities. KD is the most common cause of acquired heart disease in children in developed countries and is typically diagnosed in the emergency department (ED) after pediatric patients are evaluated following several days of fever. The longer the delay before the administration of intravenous immunoglobin (IVIG), the standard treatment for KD, the greater the risk for development of CAAs. Since early recognition of KD is vital for timely treatment with IVIG, a potential solution for decreasing the risk of CAAs is to reduce the delay in having a clinician evaluate a child for suspicion of KD. We propose to accomplish this by creating deep learning algorithms to screen images of potential KD clinical signs before a child is examined by a clinician or hospitalized. Parents will upload images of their child, and a recommendation will be made to seek medical advice if image analysis of the relevant clinical signs determines the child is at risk for KD.

In an earlier work, we presented a deep learning algorithm consisting of convolutional neural networks (CNNs) to separately assess the presence of KD clinical signs on a dataset of crowdsourced and publicly available images [2]. We demonstrated that transfer learning using a pre-trained VGG-16 model with ImageNet weights could accurately discriminate KD from similar febrile illnesses with a median accuracy of 82% across all signs. However, use of this algorithm requires classification of the clinical sign in an image, which would not be appropriate for families who lack clinical expertise. Here, we expand on the previous work by developing and externally validating a separate algorithm to automatically detect the KD clinical sign in an image before feeding the image into the convolution neural network for the classified clinical sign. We evaluate several CNN architectures with transfer learning as well as Vision Transformers (ViT) to assess their performance on this image classification task [3]. ViT is based on the Transformer architecture that has demonstrated comparable performance to CNNs with the advantage of fewer computational resources required for training [4]–[6].

The goal of this study was to develop a model for users without clinical expertise for a novel KD screening tool that will be made publicly available on the website of the non-profit, parent-based Kawasaki Disease Foundation. The intended use of this tool is for families of children who have KD-like illnesses to evaluate if their child's presentation is consistent with KD and take appropriate action if necessary.

II. Dataset

Two datasets were used: a dataset consisting of crowdsourced images and images available from the Internet as previously described [2], and a dataset of images acquired from patients with KD and patients with a similar phenotype admitted to Rady Children's Hospital San Diego (RCHSD

*Research supported by the Gordon and Marilyn Macklin Foundation, the National Institutes of Health (R01HL140898 to J. C. Burns and A. H. Tremoulet, R01LM013998 to S. Nemati, and T15LM011271 to J. Y. Lam), and a grant from the Patient-Centered Outcomes Research Institute (CER-1602-3447 to J. C. Burns).

J. Y. Lam and S. Nemati are with the Department of Biomedical Informatics, University of California San Diego, La Jolla, CA, USA (e-mail: j7lam@ucsd.edu).

J. T. Kanegaye, E. Xu, M. A. Gardiner, J. C. Burns, and A. H. Tremoulet are with the Department of Pediatrics, University of California San Diego and Rady Children's Hospital, San Diego, CA, USA.

dataset) from 2017-2022 who gave consent for photographs for a study approved by the University of California San Diego Institutional Review Board. No demographic information is available for the crowd-sourced dataset. There were 605 images acquired from 164 patients in the RCHSD dataset with a mean age of 3.64 years (standard deviation: 2.89). 58% of the patients were male and ethnicity was reported as follows: 37% Hispanic, 23% more than two races or other, 23% White, 14% Asian, and 3% African American. For the RCHSD dataset, images were usually acquired by a clinician at the time of initial clinical encounter in the ED and prior to formal hospital admission. Clinicians acquired images with smartphones and uploaded them to a patient's electronic health record using the Epic Haiku application. Only images of positive KD clinical signs were taken. Images were grouped into the five KD clinical signs and adjudicated by a pediatric KD specialist to ensure accuracy. The number of images for each clinical sign is described in Table I. Sample images from each dataset are presented in Fig. 1 and 2.

TABLE I.
NUMBER OF IMAGES PER CLINICAL CRITERIA IN EACH DATASET

Dataset	Clinical Sign					Total
	Rash	*Eyes*[a]	*Lymph*[b]	*Mouth*[c]	*Extremities*	
Crowd-Sourced	450	317	140	460	537	1904
RCHSD	72	202	47	147	137	605

a. Eyes: bilateral conjunctival erythema, b. Lymph: cervical lymphadenopathy, c. Mouth: changes in the lips and oral cavity

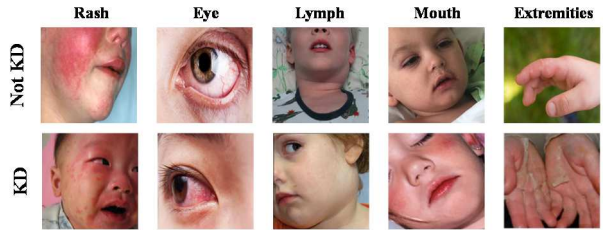

Figure 1. Sample images from the crowd-sourced dataset.

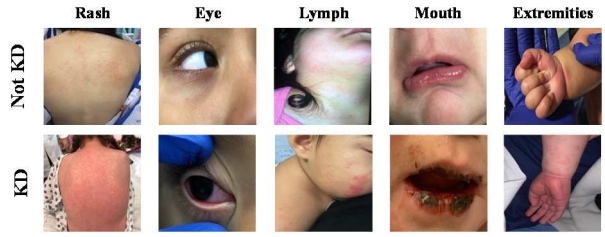

Figure 2. Sample images from the RCHSD dataset.

III. METHODS

A. Data Preprocessing

In the crowd-sourced dataset, images from the erythema of peripheral extremities and peeling of peripheral extremities were combined into a single extremities class. Images with any computer-generated text such as legends were cropped to remove text or excluded from the dataset if text could not be removed. In the RCHSD dataset, images obtained by clinicians from a parent's phone or electronic screen were excluded.

B. Models

We evaluated the VGG16 [7], Big Transfer (BiT) [8], and Inception V3 [9] architectures with transfer learning using pre-trained weights from ImageNet and the original ViT architecture [3] using pre-trained weights from the JFT-300M dataset. BiT and Inception V3 perform significantly better than ResNet50 [10] on several image classification tasks, so we chose these two as well as VGG16 based on its prior use for classification of the individual KD clinical criteria in the crowd-sourced dataset for the CNN architectures. ViT utilizes Transformers, a self-attention based architecture dominant in natural language processing models [11], and takes advantage of the computational efficiency and scalability of Transformers to train large models with substantially less computational resources compared to CNNs. Images in a ViT model are split into patches and the respective linear embeddings are combined with position embeddings and a classification token as input to the Transformer encoder.

Data was augmented using a combination of random horizontal or vertical flips, 90-degree rotations, and zoom by a factor of up to 20%. Each model was trained using the Adam optimizer and categorical cross entropy loss. For the CNN architectures, the output from the initially frozen pre-trained layers was pooled using the global average followed by a dropout layer and two feedforward layers with rectified linear unit and softmax activation respectively. The ViT model was similarly fine-tuned by removing the pre-trained classification head and adding feedforward layers with Gaussian Error Linear Unit activation and dropout layers. Class weights were added as a parameter during model training to address class imbalance. Hyperparameters optimized for each model included learning rate, units in the feedforward layers, number of epochs, dropout rate, batch size, and weight decay. Further hyperparameters for the ViT model included patch size, number of layers, number of heads, and projection dimensions. All models were developed in Tensorflow [12].

The performance of the models was assessed using accuracy with predictions based on the class with the maximum output probability. Models were evaluated using 10-fold cross-validation [13] by dividing the crowd-sourced dataset into train and test sets using a 90:10 ratio. Once optimized parameters were identified, models were trained on the entire crowd-sourced dataset and externally validated on the RCHSD dataset.

IV. RESULTS

The 10-fold cross-validation performance of the models on the crowd-sourced dataset is summarized in Table II.

TABLE II.
MODEL ACCURACY DURING 10-FOLD CROSS VALIDATION. VALUES REPORTED AS MEAN±STANDARD DEVIATION.

Model	Accuracy
VGG16	88.13±1.01

Model	Accuracy
BiT	90.12±0.92
Inception V3	89.87±0.86
ViT	89.56±0.87

Models had comparable performance except for VGG16 which underperformed. The performance of the models on the 605 images on the RCHSD dataset is summarized in Table III. The models accurately classified changes in extremities, changes in the lips/oral cavity, and bilateral conjunctival injection with mean accuracy across models greater than 88% but had difficulty with rash and cervical lymphadenopathy with mean accuracy below 77%. The ViT model underperformed the CNNs in terms of combined accuracy by at most 1.48%.

TABLE III.
MODEL ACCURACY DURING EXTERNAL VALIDATION

Clinical Sign	Model				*Mean*
	VGG16	*BiT*	*Inception V3*	*ViT*	
Rash	73.61	79.17	80.56	72.22	76.39
Eyes[a]	92.08	91.58	87.62	88.61	89.97
Lymph[b]	70.21	76.60	78.72	68.09	73.41
Mouth[c]	88.44	87.76	87.07	91.16	88.62
Extremities	89.78	91.24	94.89	91.97	91.97
Combined[d]	86.78	87.93	87.60	86.45	87.19

a. Eyes: bilateral conjunctival erythema, b. Lymph: cervical lymphadenopathy, c. Mouth: changes in the lips and oral cavity, d. Overall accuracy across all clinical signs

V. Discussion

Transfer learning with traditional CNNs performs well in discriminating between KD clinical criteria with ViT displaying slightly lower performance, consistent with prior comparisons of CNNs and ViT [4]–[6]. We did not benchmark more recent ViT advancements that outperform the base model such as a scaled ViT model [14] and token labeling [15] that could potentially lead to performance exceeding the CNNs. The worst performing clinical signs during external validation were rash and cervical lymphadenopathy. Analysis of images in the RCHSD dataset revealed that almost all incorrectly classified rash images were classified as changes in extremities and that incorrectly classified images of cervical lymphadenopathy were classified as rash. Performance across all models is affected by overlap between criteria. For example, rash develops across the trunk and extremities [16] but can also occur on the face (Fig. 1). Similarly, cervical lymphadenopathy and changes in the lips both occur in the lower part of the face. There is no provided guidance for how images should be taken in either dataset, so variability exists in photographic technique and the anatomic locations of the clinical criterion depicted.

Clinicians are more accurate than the reported model performance in determining which clinical criteria are present in patients [17], [18]. However, identification of findings traditionally requires an in-person or telemedicine encounter with a clinician. The advantage of the proposed screening tool is that families can evaluate whether they should seek clinical advice by simply uploading an image of their child at no cost and with no delay. An algorithm-driven recommendation to seek medical care could lead to earlier diagnosis and treatment of KD, thus reducing the risk of coronary artery complications [19].

We outline how the proposed tool will work in Fig. 3. First, one or more of the KD clinical criteria are observed in a child along with a minimum of five days of fever. Next, a device with capacity for digital image capture and upload is used to take a photograph of the child with sample images displayed in the web screening tool for guidance. If the user is uncertain which criteria is present in a child, the user will upload the image, and the model will determine the criterion before feeding the image into the respective CNN developed previously. Users also have the option of directly uploading their image to the appropriate CNN which will then calculate the KD risk for a given image. If risk exceeds a threshold, the tool will note that the corresponding clinical sign is present.

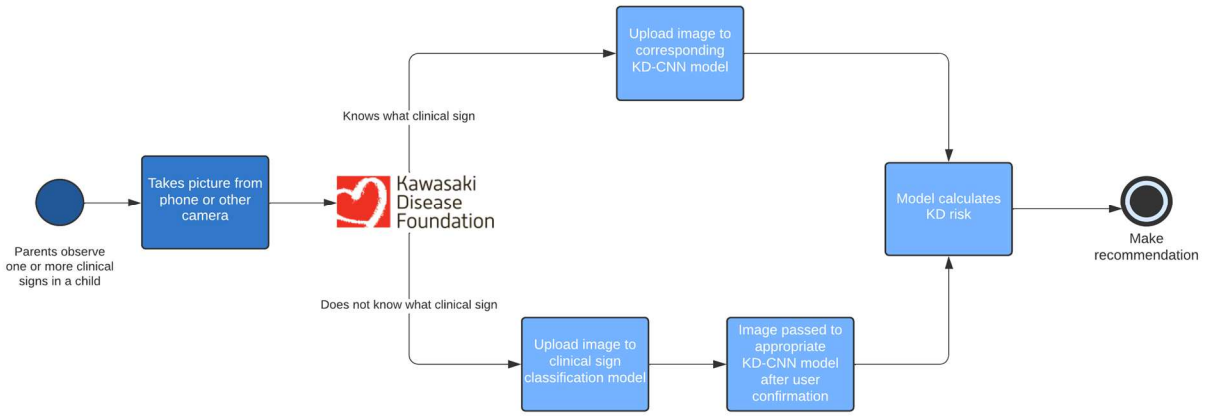

Fig. 3. Schematic overview of the Kawasaki disease screening tool

If two or more clinical signs are present, the tool will provide a recommendation to the user to seek medical attention because of potential risk for KD based on diagnostic guidelines [1], [16], [20].

The following example demonstrates one potential use case. A parent observes cutaneous changes in the distal extremities of their febrile child who has had fever for six days and conducts an Internet search to evaluate potential diagnoses. They find the KD screening tool on the Kawasaki Disease Foundation website and acquire images of the cutaneous changes with their smartphone following provided instructions. The parent does not know which criteria are present, so the parent uploads a photograph of the forearm to the model which predicts that the image is most likely rash. The image is then passed to the rash CNN, and the presence of rash is confirmed. Since the cutaneous changes extend to the hand, a second image is uploaded. The model predicts the image of the hand is consistent with changes in extremities and passes the image to a CNN which confirms erythema of the palms. The KD screening tool then notifies the parent that two clinical criteria consistent with KD have been observed and to seek medical attention for their child given suspected risk for KD.

There are several outstanding issues regarding the tool that need to be addressed. Variation in image quality and acquisition could impact model performance, so one solution is to provide a standardized set of KD clinical sign images in the uploading instructions to ensure consistency. Not all signs can be identified accurately, and the external validation dataset was limited in size, so limitations of the tool and non-intended uses should also be provided. In addition, it remains unclear how families can be made aware of this online screening tool. Despite these issues, the promising performance of the algorithms highlights their potential to empower parents to take steps towards earlier KD diagnosis.

VI. Conclusion

In this study, we developed and validated a deep learning model to accurately identify the specific KD clinical criteria in an image. This model will be utilized as part of a proposed KD screening tool for parents without clinical expertise to assist in determining which downstream CNN to send an uploaded image from a child with KD-like symptoms. With the reported model performance, there is potential for this screening tool to reduce the risk of CAAs in patients with KD by enabling earlier clinical consultation and treatment with IVIG instead of delayed KD diagnosis. Further work is ongoing to implement the model within a tool on the website of the non-profit Kawasaki Disease Foundation (https://kdfoundation.org/).

Acknowledgment

The authors would like to thank patients and their families for contributing photos and clinicians from the Pediatric Emergency Medicine Kawasaki Disease Research Group at Rady Children's Hospital who acquired the images.

References

[1] B. W. McCrindle *et al.*, "Diagnosis, Treatment, and Long-Term Management of Kawasaki Disease: A Scientific Statement for Health Professionals From the American Heart Association," *Circulation*, vol. 135, no. 17, pp. e927–e999, Apr. 2017. doi: 10.1161/CIR.0000000000000484.

[2] E. Xu, S. Nemati, and A. H. Tremoulet, "A deep convolutional neural network for Kawasaki disease diagnosis," *Sci Rep*, vol. 12, no. 1, p. 11438, Jul. 2022, doi: 10.1038/s41598-022-15495-x.

[3] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 2020, doi: 10.48550/ARXIV.2010.11929.

[4] S. Paul and P.-Y. Chen, "Vision Transformers are Robust Learners," 2021, doi: 10.48550/ARXIV.2105.07581.

[5] M. Springenberg, A. Frommholz, M. Wenzel, E. Weicken, J. Ma, and N. Strodthoff, "From CNNs to Vision Transformers -- A Comprehensive Evaluation of Deep Learning Models for Histopathology," 2022, doi: 10.48550/ARXIV.2204.05044.

[6] M. Filipiuk and V. Singh, "Comparing Vision Transformers and Convolutional Nets for Safety Critical Systems," in *Proceedings of the Workshop on Artificial Intelligence Safety 2022 (SafeAI 2022) co-located with the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI2022), Virtual, February, 2022*, 2022, vol. 3087. [Online]. Available: http://ceur-ws.org/Vol-3087/paper_31.pdf

[7] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, doi: 10.48550/ARXIV.1409.1556.

[8] A. Kolesnikov *et al.*, "Big Transfer (BiT): General Visual Representation Learning," 2019, doi: 10.48550/ARXIV.1912.11370.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2015, doi: 10.48550/ARXIV.1512.00567.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015, doi: 10.48550/ARXIV.1512.03385.

[11] A. Vaswani *et al.*, "Attention Is All You Need," 2017, doi: 10.48550/ARXIV.1706.03762.

[12] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," 2016, doi: 10.48550/ARXIV.1605.08695.

[13] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *IJCAI*, 14, pp. 1137–1143.

[14] X. Zhai, A. Kolesnikov, N. Houlsby, and L. Beyer, "Scaling Vision Transformers," 2021, doi: 10.48550/ARXIV.2106.04560.

[15] Z. Jiang *et al.*, "All Tokens Matter: Token Labeling for Training Better Vision Transformers," 2021, doi: 10.48550/ARXIV.2104.10858.

[16] J. W. Newburger *et al.*, "Diagnosis, Treatment, and Long-Term Management of Kawasaki Disease: A Statement for Health Professionals From the Committee on Rheumatic Fever, Endocarditis and Kawasaki Disease, Council on Cardiovascular Disease in the Young, American Heart Association," *Circulation*, vol. 110, no. 17, pp. 2747–2771, Oct. 2004, doi: 10.1161/01.CIR.0000145143.19711.78.

[17] İ. Devrim et al., "Reliability and accuracy of smartphones for paediatric infectious disease consultations for children with rash in the paediatric emergency department," BMC Pediatr, vol. 19, no. 1, p. 40, Dec. 2019, doi: 10.1186/s12887-019-1416-8.

[18] J. Sink et al., "A novel telemedicine technique for evaluation of ocular exam findings via smartphone images," J Telemed Telecare, vol. 28, no. 3, pp. 197–202, Apr. 2022, doi: 10.1177/1357633X20926819.

[19] M. S. Wilder, L. A. Palinkas, A. S. Kao, J. F. Bastian, C. L. Turner, and J. C. Burns, "Delayed Diagnosis by Physicians Contributes to the Development of Coronary Artery Aneurysms in Children With Kawasaki Syndrome," Pediatric Infectious Disease Journal, vol. 26, no. 3, pp. 256–260, Mar. 2007, doi: 10.1097/01.inf.0000256783.57041.66.

[20] Council on Cardiovascular Disease in the Young; Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease; American Heart Association., "Diagnostic Guidelines for Kawasaki Disease," Circulation, vol. 103, no. 2, pp. 335–336, Jan. 2001, doi: 10.1161/01.CIR.103.2.335.