

DR KAVITA YANG SARIN (Orcid ID : 0000-0001-5363-3053)

Article type : Research Letter

Automated detection of skin reactions in epicutaneous patch testing using machine learning

W.H. Chan,^{1,2,*} R. Srivastava,^{1,*} N. Damaraju,³ H. Do,¹ G. Burnett,³ J. MacFarlane,³ S.M. Xie,⁴ J.K. Chen,^{1,†} G. Honari^{1,†} and K.Y. Sarin^{1,†}

¹Department of Dermatology, Stanford University School of Medicine, Stanford, California, USA

²Department of Medicine, Baylor College of Medicine, Houston, Texas, USA

³Department of Biomedical Informatics, Stanford University School of Medicine, Stanford, California, USA

⁴Department of Computer Science, Stanford University, Stanford, California, USA

*Co-first authors

†Co-senior authors

Correspondence: Kavita Y. Sarin

Email: ksarin@stanford.edu

ORCID: <https://orcid.org/0000-0001-5974-642X> (Warren H. Chan)
<https://orcid.org/0000-0002-0441-8389> (Rushil Srivastava)
<https://orcid.org/0000-0001-7846-4671> (Nandita Damaraju)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/BJD.20141](https://doi.org/10.1111/BJD.20141)

This article is protected by copyright. All rights reserved

<https://orcid.org/0000-0002-5060-8191> (Hanh Do)

<https://orcid.org/0000-0002-0820-2753> (Sang Michael Xie)

<https://orcid.org/0000-0003-2666-8407> (Jennifer K. Chen)

<https://orcid.org/0000-0002-4447-9151> (Golara Honari)

<https://orcid.org/0000-0001-5363-3053> (Kavita Y. Sarin)

Funding sources: WHC is a Howard Hughes Medical Institute (HHMI) Medical Research Fellow. SMX is supported by a National Defense Science and Engineering (NDSEG) Graduate Fellowship. KYS is the D. G. “Mitch” Mitchell Clinical Investigator supported by the Damon Runyon Cancer Research Foundation (CI-104-19).

Conflicts of interest: GH is a clinical investigator for AbbVie, Sanofi, Dermira, NiPera and is a consultant for AbbVie and Regeneron. JKC is a clinical investigator for AbbVie, Sanofi Regeneron, Dermira, and NiPera.

Dear Editor,

Patch testing is the diagnostic gold standard to identify causes of allergic contact dermatitis (ACD); however, the paucity of specialized patch testing clinics along with the need for multiple clinic visits can be burdensome for patients and healthcare systems. Automated classification of allergic reactions from photographs offers the opportunity to reduce patient, physician, and clinic time and improve access to care. Machine learning approaches have evolved over the past decade to classify skin cancer lesions, risk-stratify psoriasis patients, and identify promising new therapies. Here, we demonstrate a new proof-of-principle application of machine learning in dermatology for accurate diagnosis of allergic reactions from patch testing photographs.

77 patients [Fitzpatrick I (n=2), II (n=28), III (n=29), IV (n=14), V (n=4)] who underwent patch testing with the American Contact Dermatitis Society (ACDS) Core Screening Allergen Series at Stanford Hospital & Clinics between February 2019 and January 2021 were recruited for this study. Allergens were placed in eight 5x2 panels on the back and removed after 48 hours per routine clinical care (**Figure 1A**). De-identified images of each 5x2 panel were obtained four days after allergen application. We developed a cropping algorithm to perform automated cropping of each panel into images of individual allergens. Ground truth labels [1+ (n=83), 2+ (n=33), 3+ (n=2), IR (n=16)] for each allergen image were derived from Day 4 patch test scoring by two dermatologists (JKC, GH) according to International Contact Dermatitis Research Group (ICDRG) guidelines and further categorized into positive (1+, 2+, 3+) and negative (0, IR) reactions. Doubtful (?+) reactions and images of insufficient quality were excluded, resulting in a labeled dataset of 3695 images (118 positive, 3577 negative).

The dataset was separated into a training set (1118 total images: 51 positive, 1067 negative), validation set (373 total images: 17 positive, 356 negative), and test set (2204 total images: 50 positive, 2154 negative), with no overlap in patients between the test set and the training/validation data. Training was performed using Google's Xception convolutional neural network (CNN) architecture¹ pretrained to 94.5% top-five accuracy on the 1,000 object classes (1.28 million images) of the 2014 ImageNet Large Scale Visual Recognition Challenge.² Following removal of the final classification layer, the CNN was retrained using transfer learning, leveraging the natural-image features learned by the ImageNet pretrained network.³

On the test set of 50 positive and 2154 negative images, the CNN achieves an accuracy of 99.5% and F1 score of 0.89. A precision-recall curve was generated and compared to that derived from a colors-based computer vision approach (**Figure 1B**). 21 unique allergens (out of the 80 ACDS Core Screening Allergens) comprise the 50 positive images in the test set. 13 out of the 21 allergens were correctly classified with 100% accuracy by the CNN (that is, the CNN classification was consistent with clinician clinical interpretation) in all test patients. 7 allergens (neomycin, epoxy resin bisphenol A, nickel, thimerosal, cobalt chloride hexahydrate, budesonide, and benzoyl peroxide) were each correctly classified in 96.7% of patients, while 1 allergen (Balsam Peru) was classified correctly in 93.3% of cases. Of the misclassifications, 4 were false positives and 7 were false negatives. The CNN correctly classified all dyes (p-Phenylenediamine, Disperse Yellow 3, and Disperse Blue 106/124 Mix) in all patients. All 6 IR images in the test set were correctly classified as negative reactions.

Several benefits underlie the application of machine learning for automated detection in patch testing. First, a standardized machine learning algorithm addresses low inter-individual concordance^{4,5} between dermatologists trained in patch testing and allows for the standardization of diagnoses across institutions. Second, a robust machine learning application can facilitate the emerging possibilities of at-home diagnosis, a welcome alternative to multiple clinic visits at sparse specialized patch testing locations. Third, the properties of deep learning⁶ enable the creation of schemata that include skin-sensitizing potency⁷ and regional prevalence and clinical relevance⁸ (**Figure 1C**), which may motivate improvements in future classification accuracy.

Our study provides a proof-of-principle demonstration of a machine learning approach that can be used to classify patch test photographs with a high degree of accuracy in a limited dataset. To train a classifier that can distinguish beyond positive vs. negative reactions (i.e. between positive grades (1+,2+,3+), IR, and ?+ reactions), multi-institution efforts focused on expanding the training dataset of patch test photographs are needed. This letter calls for academic centers to aid in the accrual of patch testing images. Together with the Xception-based CNN and machine learning schema, these elements serve as the building blocks towards a future in which aspects of patch testing reading can be automated, standardized, and virtualized, with the potential to reduce cost for patients and physicians, improve detection accuracy, and enhance access to care.

References

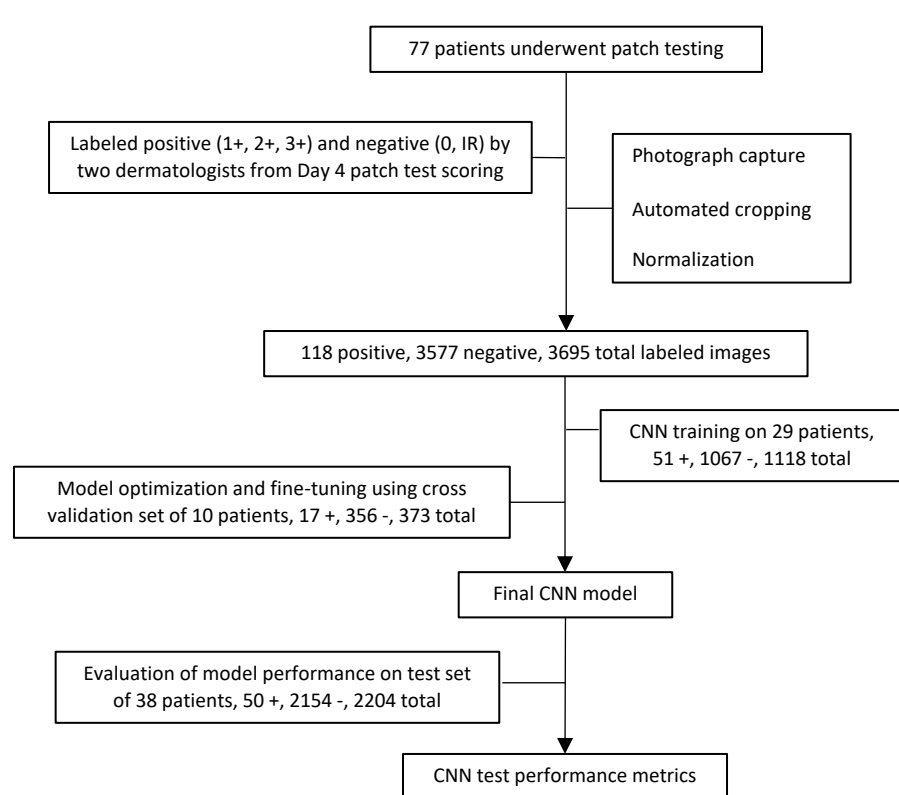
1. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. Vol 2017-January. Institute of Electrical and Electronics Engineers Inc.; 2017:1800-1807. doi:10.1109/CVPR.2017.195
2. Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis*. 2015;115(3):211-252. doi:10.1007/s11263-015-0816-y
3. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22(10):1345-1359. doi:10.1109/TKDE.2009.191
4. Svedman C, Isaksson M, Björk J, Mowitz M, Bruze M. 'Calibration' of our patch test reading technique is necessary. *Contact Dermatitis*. 2012;66(4):180-187. doi:10.1111/j.1600-0536.2011.02044.x
5. Bruze M, Isaksson M, Edman B, Björkner B, Fregert S, Möller H. A study on expert reading of patch test reactions: inter-individual accordance. *Contact Dermatitis*. 1995;32(6):331-337. doi:10.1111/j.1600-0536.1995.tb00621.x
6. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500-510. doi:10.1038/s41568-018-0016-5
7. Basketter DA, Alépée N, Ashikaga T, et al. Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis*. 2014;25(1):11-21. doi:10.1097/DER.0000000000000003
8. DeKoven JG, Warshaw EM, Zug KA, et al. North American Contact Dermatitis Group Patch Test Results: 2015-2016. *Dermatitis*. 2018;29(6):297-309. doi:10.1097/DER.00000000000000417

Figure legend

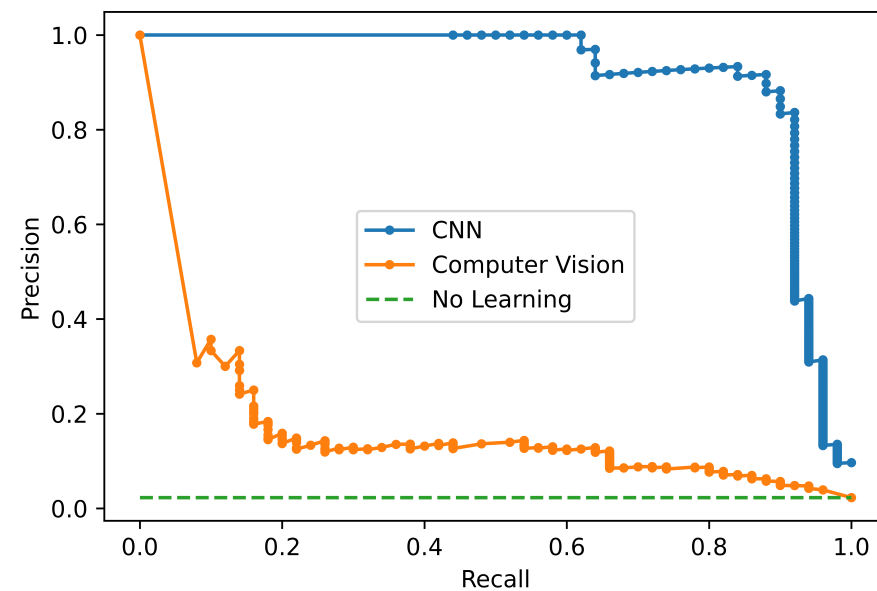
Fig 1. CNN workflow, performance, and schema. A, Data pre-processing, training, and evaluation workflow. Photographs were collected using a 12-megapixel camera, cropped into individual allergen images, and normalized by temperature, hue, and intensity. B, Precision-recall curves showing

convolutional neural network (CNN) performance (F1 score 0.887, AUC 0.915) on the test set of 2204 images, compared to a color-based computer vision (CV) representation (F1 score 0.027, AUC 0.166). C, Schema for future automated scoring of patch testing results. Iteration over each allergen (e.g. *Myroxylon pereirae*) and identification of the image corresponding to the allergen is followed by a CNN based-image feature extraction and integration with allergen skin-sensitizing potency and significance-prevalence index number (a composite measure of allergen prevalence and clinical relevance) to arrive at a probability distribution over ICDRG scores.

A



bjd_20141_f1.pdf



C

