
Abstract

Explainability of Convolutional Neural Networks for Dermatological Diagnosis

Raluca Jalaboi¹, MSc; Mauricio Orbes Arteaga², MSc; Dan Richter Jørgensen², PhD; Ionela Manole², MD; Oana Ionescu Bozdog², MD; Andrei Chiriac², MD; Ole Winther¹, PhD; Alfiia Galimzianova², PhD

¹Technical University of Denmark, Kongens Lyngby, Denmark

²Omhu, Copenhagen, Denmark

Corresponding Author:

Raluca Jalaboi, MSc

Technical University of Denmark

Anker Engelunds Vej 1 Bygning 101A, 2800 Kgs

Kongens Lyngby

Denmark

Phone: 45 45 25 25 25

Email: rjal@dtu.dk

Abstract

Background: Convolutional neural networks (CNNs) are regarded as state-of-the-art artificial intelligence (AI) tools for dermatological diagnosis, and they have been shown to achieve expert-level performance when trained on a representative dataset. CNN explainability is a key factor to adopting such techniques in practice and can be achieved using attention maps of the network. However, evaluation of CNN explainability has been limited to visual assessment and remains qualitative, subjective, and time consuming.

Objective: This study aimed to provide a framework for an objective quantitative assessment of the explainability of CNNs for dermatological diagnosis benchmarks.

Methods: We sourced 566 images available under the Creative Commons license from two public datasets—DermNet NZ and SD-260, with reference diagnoses of acne, actinic keratosis, psoriasis, seborrheic dermatitis, viral warts, and vitiligo. Eight dermatologists with teledermatology expertise annotated each clinical image with a diagnosis, as well as diagnosis-supporting characteristics and their localization. A total of 16 supporting visual characteristics were selected, including basic terms such as *macule*, *nodule*, *papule*, *patch*, *plaque*, *pustule*, and *scale*, and additional terms such as *closed comedo*, *cyst*, *dermatoglyphic disruption*, *leukotrichia*, *open comedo*, *scar*, *sun damage*, *telangiectasia*, and *thrombosed capillary*. The resulting dataset consisted of 525 images with three rater annotations for each. Explainability of two fine-tuned CNN models, ResNet-50 and EfficientNet-B4, was analyzed with respect to the reference explanations provided by the dermatologists. Both models were pretrained on the ImageNet natural image recognition dataset and fine-tuned using 3214 images of the six target skin conditions obtained from an internal clinical dataset. CNN explanations were obtained as activation maps of the models through gradient-weighted class-activation maps. We computed the fuzzy sensitivity and specificity of each characteristic attention map with regard to both the fuzzy gold standard characteristic attention fusion masks and the fuzzy union of all characteristics.

Results: On average, explainability of EfficientNet-B4 was higher than that of ResNet-50 in terms of sensitivity for 13 of 16 supporting characteristics, with mean values of 0.24 (SD 0.07) and 0.16 (SD 0.05), respectively. However, explainability was lower in terms of specificity, with mean values of 0.82 (SD 0.03) and 0.90 (SD 0.00) for EfficientNet-B4 and ResNet-50, respectively. All measures were within the range of corresponding interrater metrics.

Conclusions: We objectively benchmarked the explainability power of dermatological diagnosis models through the use of expert-defined supporting characteristics for diagnosis.

Acknowledgments: This work was supported in part by the Danish Innovation Fund under Grant 0153-00154A.

Conflict of Interest: None declared.

(*iproc* 2021;7(1):e35437) doi: [10.2196/35437](https://doi.org/10.2196/35437)

KEYWORDS

dermatology; explainability; convolutional neural networks

Multimedia Appendix 1

Explainability of ResNet-50 and EfficientNet-B4 models in terms of sensitivity between dermatologists-provided segmented supporting characteristics and model activation maps. All activation maps were computed based on the gold standard diagnosis using gradient-weighted class-activation maps. Interrater sensitivity is computed as the pairwise average for dermatologist-provided supporting characteristic segmentations.

[[PNG File , 415 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

Examples of explanations for images where both models correctly predicted the gold standard diagnosis. From left to right: the original image, the union of all characteristics selected by all dermatologists annotating the image, an EfficientNet-B4 gradient-weighted class-activation map (Grad-CAM) visualization, and a ResNet-50 Grad-CAM visualization. In all cases, the EfficientNet-B4 visualization was closer to the dermatologist map than the ResNet-50 visualization. ResNet-50 appears to be more specific, focusing on smaller, more noticeable lesions.

[[PNG File , 1059 KB-Multimedia Appendix 2](#)]

Edited by T Derrick; this is a non-peer-reviewed article. Submitted 03.12.21; accepted 03.12.21; published 10.12.21.

Please cite as:

*Jalaboi R, Orbes Arteaga M, Richter Jørgensen D, Manole I, Bozdog OI, Chiriac A, Winther O, Galimzianova A
Explainability of Convolutional Neural Networks for Dermatological Diagnosis
iproc 2021;7(1):e35437*

URL: <https://www.iproc.org/2021/1/e35437>

doi: [10.2196/35437](https://doi.org/10.2196/35437)

PMID: [27739472](https://pubmed.ncbi.nlm.nih.gov/27739472/)

©Raluca Jalaboi, Mauricio Orbes Arteaga, Dan Richter Jørgensen, Ionela Manole, Oana Ionescu Bozdog, Andrei Chiriac, Ole Winther, Alfiia Galimzianova. Originally published in Iproceedings (<https://www.iproc.org>), 10.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in Iproceedings, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.iproc.org/>, as well as this copyright and license information must be included.