Received: 14 August 2023 | Accepted: 29 September 2023

DOI: 10.1111/jdv.19639

ORIGINAL ARTICLE

JEADV JOURNAL OF THE EUROPEAN ACADEMY OF DERMATOLOGY & VENEREOLOGY

Automated classification of hidradenitis suppurativa disease severity by convolutional neural network analyses using calibrated clinical images

A. Wiala[1] | R. Ranjan[2] | H. Schnidar[2] | K. Rappersberger[1,3] | C. Posch[3,4,5]

[1]Department of Dermatology, Clinic Landstrasse, Vienna, Austria

[2]SCARLETRED Holding GmbH, Vienna, Austria

[3]School of Medicine, Sigmund Freud University, Vienna, Austria

[4]Department of Dermatology, Clinic Hietzing, Vienna, Austria

[5]Department of Dermatology and Allergy, School of Medicine, German Cancer Consortium (DKTK), Technical University of Munich, Munich, Germany

Correspondence
Christian Posch, Department of Dermatology, Clinic Hietzing, Wolkersbergenstraße 1, 1130 Vienna, Austria.
Email: christian.posch@gesundheitsverbund.at

Abstract

Background: The assessment of hidradenitis suppurativa (HS) severity requires detailed, and error-prone lesion counts. This proof-of-concept study aimed to automatically classify HS disease severity using machine learning of clinical smartphone images.

Methods: 777 ambient-light and size-controlled images were used to build a class-balanced synthetic dataset ($n = 7675$). Convolutional neural networks (CNN) were used for automated severity classification (scale 0–3), and to assess disease-dynamics. International Hidradenitis Suppurativa Severity Score System (IHS4) served as reference. A U-NET algorithm was implemented for automated localization of diseased skin.

Results: CNNs were able to distinguish no/mild from moderate/severe disease with an overall prediction accuracy of 78% [receiver operating curve (AUC) 0.85]. Correct IHS4 classification was achieved with an overall accuracy of 72% (AUC 0.84–0.89). In addition, disease dynamics using IHS4 numerical values aligned with CNN outputs (NRMSE 0.262). The UNET algorithm localized lesions with a pixel accuracy of 88.1% and test loss of 0.42.

Limitations: Limitations in assessing tattooed and hairy skin. Limited number of patients with dark skin colour and Hurley I.

Conclusion: CNNs were able to distinguish no/mild from moderate/severe disease, classify disease severity over time, and automatically identify diseased skin areas and the skin phototype. This study breaks new grounds for fast, reliable, reproducible and easy-to-use HS severity assessments using clinical images.

INTRODUCTION

Hidradenitis suppurativa (HS) is a chronic debilitating skin disease.[1,2] Skin lesions include inflammatory nodules, abscesses, draining fistulas and scars mainly affecting intertriginous areas. Patients suffer from excruciating pain, strictures, and malodorous discharge significantly impairing their quality of life.[3,4] HS disease severity and dynamic determine available treatment options, which comprise of antibiotics, TNF-alpha and IL-17 antibody treatment and surgical interventions.

To date, up to 30 different assessment tools have been proposed to evaluate disease severity, disease dynamics and treatment response.[5] Hurley staging is the most widely used method. It offers a 4-scale, crude assessment from no disease to severe disease, which is still used for treatment decisions today. It is, however, unable to monitor disease dynamics or response to treatment. Therefore, the Hidradenitis suppurativa Clinical Response Score (HiSCR) was developed in 2015. Achieved HiSCR is defined by at least 50% reduction in the total abscess and nodule (AN) count with no increase in abscesses and draining fistulas relative to baseline at a set

Linked Article: T. Tzellos et al. J Eur Acad Dermatol Venereol 2024;38:456–457. https://doi.org/10.1111/jdv.19785

time-point.[6] As repeated and reliable measures of disease dynamics over time are required for optimized patient care, HiSCR might be of limited value for clinical routine. One established score to address disease severity and dynamics is the International Hidradenitis Suppurativa Severity Score System (IHS4), calculated by the number of nodules (multiplied by 1) plus the number of abscesses (multiplied by 2) and the number of draining tunnels (multiplied by 4).[7]

The majority of HS scoring systems characterize and count different HS skin lesions, defining mild, moderate and severe HS. At this point, there are no widely accepted definitions for staging HS.[8] This study shows that both, severity staging and the assessment of disease dynamics of HS can be automated using calibrated clinical images and convolutional neural network (CNN) analyses. While the underlying technology itself is not new, as CNNs have been used in medicine and dermatology for some time,[9] the assessment of HS using such an approach has not been reported to date. This proof-of-concept study highlights that an automated HS-classification approach has the potential to simplify clinical practice and improve the consistency and reproducibility of clinical scores.

MATERIALS AND METHODS

Study design

This prospective, single-centre proof-of-concept study investigated whether HS severity can be assessed automatically using artificial intelligence (AI)-supported image analyses. The study design was reviewed and approved by the ethics committee of the city of Vienna (EK18-100-0618). Guidelines and recommendations for the use of AI-based technologies were taken into account.[10] Patients were recruited at the HS outpatient Clinic Landstrasse Vienna between May 2017 and January 2020. All patients gave informed consent to photo documentation and web-based assessments. Photos of involved skin areas were taken with commercial smartphones (iPhone 7, X and 11 Pro Max; Apple) in daily clinical routine following the recommendations for standardized photographic documentation of HS.[11] Images were taken by two HS-expert attendings and one resident in dermatology with prior experience in HS pathophysiology and management. HS patients of all disease stages were included in the study. Areas affected by other inflammatory skin conditions as well as areas with tattooed skin were excluded from the study.

Image analysis and signal intensity mapping

To reflect a clinical setting, the study configuration anticipated variable conditions in terms of illumination, viewpoint and background. To reflect a more realistic clinical scenario, the study design foresaw variable acquisition conditions in terms of illumination, viewpoint and background. To address this, images were taken within the environment of the CE-class 1 m certified medical device software Scarletred® Vision (V3.4). Before taking an image, a standardized skin patch (M-size) was applied to adjacent healthy skin. This patch was automatically recognized by the software to enable standardization of exposure, colours, imaging distance and angle. This approach overcomes the pitfalls of previous described image- and spectra-based methods by introducing a standardized erythema value (SEV*) derived from the algorithm ($L^*\text{max} - L^*$) $x + a^*$ in combination with a colour normalization sticker.[12–14] Next, the image was uploaded to the Scarletred® web platform. Diseased skin areas were manually annotated online by an HS-expert, and a reference area was selected to indicate healthy skin and individual skin tone. For each pixel inside a drawn area, the L^*, $+a^*(posA)$, $+b^*(posB)$ coordinates of the CIELAB-colour space and the standardized erythema value (SEV^*) were calculated. The resulting signal intensity map was extracted from the original images. Pseudo-grayscale images have been transformed into pseudo-colour images to optimize image analysis. Each signal strength value has been mapped to a colour designed and matched to the underlying signal. Each map defines a colour gradient in the minimum and maximum signal range.

Convolutional and mixed-input deep neural network

A CNN architecture was built to distinguish no/mild and moderate/severe disease along with estimating severity grade (Figure S1). The input comprised of pseudo-colour images coming from L^*, $+a^*(posA)$, $+b^*(posB)$ and SEV^* signal mapped colour-space.[15] In this study, the training set consisted of 80% of the dataset, of which 15% was considered for validation purposes to optimize the model. The remaining 20% of the data formed an independent test set, exclusively reserved for evaluating the model's performance. For data augmentation, input images were rescaled (1./255), rotated (5°–45°), shuffled, batched, resized (128 × 128), zoomed (by a factor of 0.05–0.5), horizontally/vertically flipped, width/height shifted and sheared. This set of augmented images were used to balance the image distribution in each class to prevent overfitting. Regularization, early stopping, batch normalization, dropout, reduced architecture complexity were used to prevent overfitting and improve the generalizability of the deep neural network. All deep neural networks were developed, implemented and run on in-house servers. Physician IHS4 assessment was used as a reference and to train the CNN model. IHS4-assessments were performed by one expert physician. All models consisted of convolutional, pooling, dropout and dense layers with a sigmoid and SoftMax activation layer to bring the output within the probabilistic range of 0 and 1 (supplemental information on request). A

Kruskal–Wallis test was conducted on a random subset of the data to see if the differences are significant.

RESULTS

Cohort characteristics and visual data exploration

In total, 149 patients were included in this study (male 55%, female 45%). Participants had a mean age of 65.9 years ±12.6 (SD). Images were assigned to three different body areas: 218 (28%) axillary region, 264 (34%) groins and 295 (38%) 'others' including the buttock or genital area. Twelve percent,[18] 48% (72) and 40% (59), had Hurley grade I, II and III. In this study, 77% were Fitzpatrick skin type I–II. Expert-assessed Fitzpatrick skin types were validated using the Individual Typology Angle (Figure S2), showing no significant differences ($p = 0.09$).

Three different dermatologists (two attendings, one resident) took pictures of affected HS-areas, clinically assessed disease severity and managed HS-treatment. 777 images were included in the analysis. 276 images were excluded due to visible tattoos or presence of non-HS skin diseases and postoperative wounds. All images were assigned IHS4-scores by one expert dermatologist. The validity of the image-based IHS4 scores was confirmed by cross-validation of image-based scores with scores from medical records. According to IHS4, most images (49.6%) showed mild HS (Table 1).

Image augmentation and parameter assessment

A data augmentation process was used to build a class-balanced synthetic dataset consisting out of 7675 images (final dataset). CNN assessment was then compared to reference IHS4 categories and values. Heat maps comparing physicians' assessment using the IHS4 and CNN outputs based on the parameters SEV*, L*, +a* and +b* values revealed that, SEV*_meandt-refdt and +a*_meandt-refdt were most reliable ($p < 0.001$) for discriminating the IHS4 categories (Figure S3).[12]

TABLE 1 Cohort characteristics.

Patients; $n = 149$	
Age (years)	Median 65.9 ± 12.6
Sex	Female: 45%, Male 55%
Fitzpatrick skin type	I–II: 77%, III: 17%, IV: 5%, V: 1%
Hurley grades	I: 12%, II: 48%, III: 40%
Images; $n = 777$	
IHS4 classes	Clear (score 0): 15.1%; mild (score 1–3): 49.6%; moderate (score 4–10): 27.3%; severe (score ≥11): 8.0%

Binary (2-class) classification

First, the CNN was trained to distinguish between two categories: clear/mild and moderate/severe HS. This level of separation is critical for clinical treatment decisions as most systemic treatments are approved for moderate/severe HS. The overall test accuracy was calculated with 78%. 79% of images categorized as '0' (mild/clear, IHS4 ≤3) and 77% categorized as '1' (moderate/severe, IHS4 >3) were correctly classified by the CNN. The ROC curves showed an AUC of 0.85 (Figure 1).

Multiclass (4-class) classification

Similar to the Hurley classification, the performance of the CNN was tested on a scale from 0 to 3. Again, IHS4 classes served as reference: clear (0), mild,[1] moderate[2] and severe.[3] The highest accuracy for correctly identifying disease severity was observed for mild disease (82%). Moderate and severe disease were correctly identified in 56% and 59%. The overall test accuracy was calculated with 72%. The ROC revealed an AUC of 0.89 for clear, 0.84 for mild, 0.85 for moderate and 0.88 for severe disease (Figure 2). The CNN performed best (≥0.9) when identifying healthy skin and severe disease.

Disease dynamics assessment

A mixed input neural network consisting of multi-layer perceptrons including categorical and numerical data was trained to provide a more granular representation of disease activity (IHS4 score as reference). In five patients (3.4%) follow-up images of the affected areas were available. The mixed input CNN aligned with patient specific disease dynamics, highlighting that the automated results generated by the CNN matched physician-assessed disease dynamics (Figure S4). Predicted scores reflected changes in IHS4 values with a normalized root mean squared error (NRMSE) value of 0.2618. Considering that a NRMSE of 1 stands for a performance equivalent to a random model, the NRMSE of 0.26 indicated that the CNN is capable of registering slight changes of the inflammatory activity over time. Analyses show that it is possible to capture disease dynamics over time using automated disease severity assessments, which has the potential to simplify clinical practice.

Diseased skin area detection

The above assessments required physician labelling of diseased skin within clinical images. For the automated detection of lesional skin, a UNET algorithm was implemented. Comparing skin areas labelled by physicians with disease areas identified by the algorithm revealed the correct identification of HS-lesions with a pixel accuracy

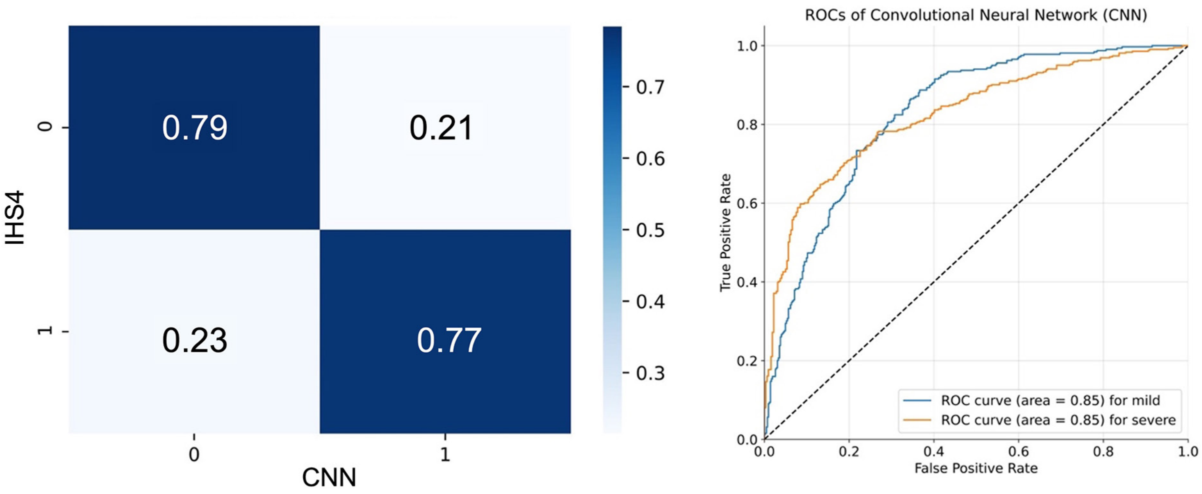

FIGURE 1 Distinction between clear/mild (0) and moderate/severe[1] disease: Confusion matrix revealed a robust performance and ROC curves indicated high sensitivity and specify at once.

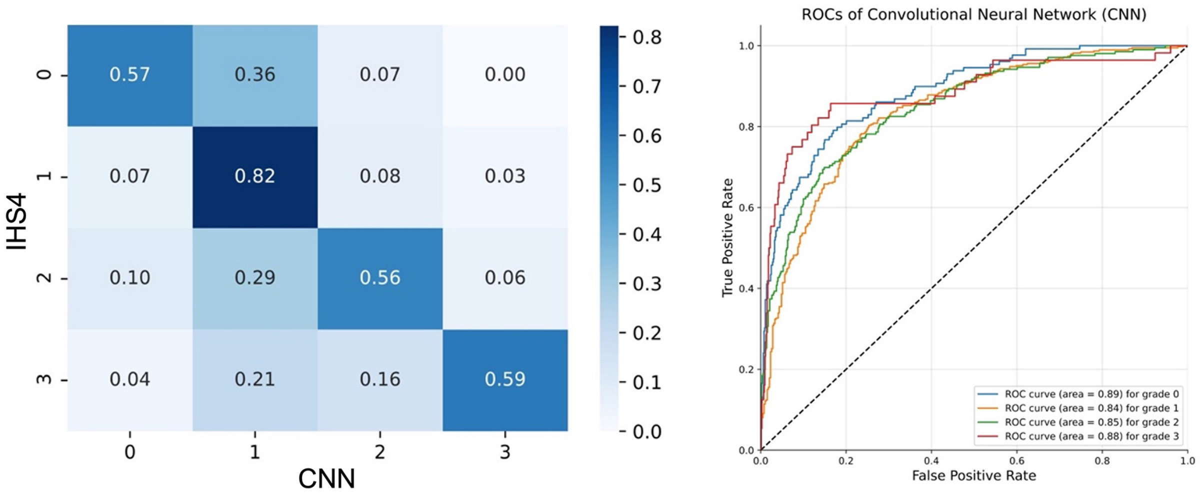

FIGURE 2 Distinction between 4 HS severity classes (0 = clear, 1 = mild, 2 = moderate, 4 = severe). The highest congruence was observed for mild disease (82%, dark blue square). The ROC curves revealed an excellent performance with AUC ranging from 0.84 to 0.89.

of 88.1% and a test loss of 0.42 (Figure 3). This approach may reduce manual labelling of different skin areas in the future.

DISCUSSION

Today, HS disease severity assessment is ill defined. Clinicians need to choose from many different visual assessment tools, mainly focusing on lesion counts. These tools have in common that they are subjective, require training and are thus challenging to use in daily clinical routine. This proof-of-concept study demonstrates that AI-based scoring of HS is feasible, with the promise to offer fast, reliable and reproducible HS assessments.

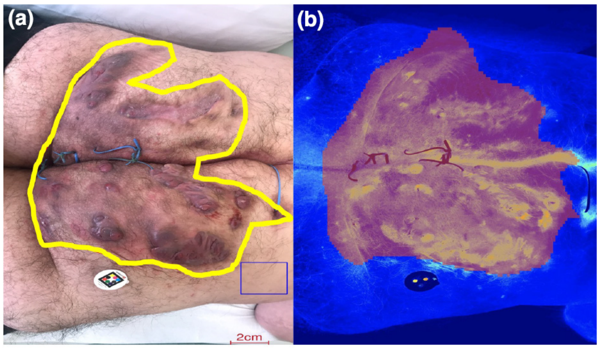

FIGURE 3 (a) Detection of lesioned skin: labelling of diseased skin performed by a physician (yellow line); (b) HS lesions before localization of affected skin by an UNET algorithm (red area, 88.1% accuracy).

Lesion counts have already been shown to be prone to inter- and intra-rater variability. An observational study found coefficients of 0.68–0.78 (inter-rater) and 0.70–0.78 (intra-rater) for abscess and fistula counts. Only inflammatory nodules were reliably distinguished in most cases.[16] Another study demonstrated only fair inter-rater reliability for the IHS4 when assessed by even experienced HS experts.[17] Results illustrate, that physicians struggle to agree on the type and number of lesions, which can be further complicated by the presence of complex communicating fistulas and abscesses, bridged scars and double-ended pseudo-comedos conflating into large inflammatory plaques. Inclusion of other metrics, for example, capturing affected surface areas and standardized signs of inflammation might be more helpful for evaluating the extent of disease and treatment response, but are rarely assessed in clinical routine.

Automated, AI-supported disease classification might offer a solution to the above problem. This study highlights that CNNs can reliably score HS lesions from clinical images taken with a commercial smartphone. Integrated into an analysis pipeline offering image-to-score automation, the CNN trained in this study has the potential to offer instantaneous and thus time-saving HS scoring. Despite not being within the scope of this study, automated scoring may also reduce inter- and intra-rater variability. However, CNNs represent just one of many AI-driven tools for computer-based image analysis previously applied in different disciplines of medicine and research but have not been used for the assessment of HS severity to date.

The results of this proof-of-concept study demonstrate that CNNs can be harnessed for the assessment of HS. By training CNNs with a synthetic dataset of 7675, generated from 777 individual clinical images, this model was able to reliably distinguish between no/mild and moderate/severe HS using IHS4 as a reference. The AUC, typically used to express the level of accuracy when comparing CNNs and physician assessments, revealed a score of 0.85. This distinction is critical, as moderate to severe HS implies approved, systemic treatment. Physicians less familiar with HS severity assessments may face uncertainty concerning the referral of patients or the use of effective systemic treatments including biologicals. Such therapeutic delays have been shown to be a significant risk factor for failing to achieve HiSCR.[18] Automatized support tools may aid to address this medical need.

The CNNs used in this study were able to distinguish multiple classes of HS disease severity: The system correctly identified all four categories (no/mild/moderate/severe) with an AUC of 0.84–0.89, indicating high sensitivity and specificity. Data are in line with results from previously published investigations evaluating the reliability of lesion-based HS scoring methods and highlight the robustness of CNN-based HS disease classification.[16,17,19] Additionally, HS is a chronic, recurrent inflammatory skin disease for which continuous monitoring of disease activity is crucial. Therefore, dynamic disease assessment is needed in clinical practice as well as in clinical trials. The performance of our mixed-input network using a reference IHS4-based scale returned a NRMSE of 0.26. Low NRMSE values and particularly values <1 indicate less residual variance and thus indicate a very good performance of this model. Even though the number of patients with follow-up images was small in this study, the CNN output aligned well with clinically assessed inflammatory disease dynamics. Results highlight the potential to simplify clinical practice using automated HS severity assessments.

This study was designed to test if HS disease severity and dynamics in a given anatomical area can automatically be calculated using clinical images. To reduce the need for manual annotation of diseased and healthy skin within an image, full automation can only be achieve if an algorithm is capable of detecting diseased skin areas. Using a UNET algorithm, we found that affected skin areas were recognized by the machine and aligned with physician annotations, with a pixel accuracy of 88.1%. This means that this approach achieved high congruence between physician and machine annotation of diseased skin.

Results offer a fundamentally new way of assessing HS severity in the future. Already at this point, data from this study demonstrate the potential of CNN-based HS scoring: First, the output of CNNs are likely unaffected by the users' individual levels of training and experience, and may offer an objective evaluation of HS severity at any given time-point. Second, no specialized hardware equipment is needed. The high availability of smartphones allows for automated assessments also in remote areas or areas with limited resources and access to specialized care. Third, taking clinical images are already a standard in dermatology and do not pose an additional working step in daily clinical routine. Fourth, automated image-based scoring methods could reduce the time spent on characterizing and counting HS lesions. Lastly, CNN-based assessments could offer reliable disease evaluations not only for physicians but also for patients. This might further help to improve disease management for patients who have readily been diagnosed with HS.

This study has limitations. The current CNN had troubles when assessing very hairy skin. Additionally, the presented model is to large degrees based on measuring different shades of red. This may limit the applicability to assess HS in patients with very dark skin tones (Fitzpatrick type V/VI). The dataset consisted of 777 unique images with a relative imbalance for low severity HS-classes. This is due to the referral-only patient selection in this specialized outpatient service. Yet, the performance of our model in more severe disease indicates, that even small differences in patients with a complex HS situs can be assessed reliably, making it likely that also lower grade HS can be evaluated using this tool. As with any other AI-based algorithm, this CNN will improve with a more diverse and larger data set. In this study, all outcome values from the CNN refer to an individual image. It is possible that disease severity for a patient might be underestimated by assessing only one image. Finally, digital outcome assessment from clinical images do not take other outcome measures including pain, discharge and quality of life into account. Particularly patient-centred outcome has not been part of this study but might affect HS disease severity. It needs to be noted, that such measures are also not included in any of the currently used assessment tools. Nonetheless, they are relevant for therapeutic decision-making and could also be added to AI-based models for a more holistic disease assessment. Finally, the robustness and validity of the CNNs used in this proof-of-concept study must be tested in future studies and should be compared to alternative, AI-based approaches assessing HS severity.

Data from this study highlight that CNNs can be used to distinguish site-specific no/mild and moderate/severe HS. This can be used to support therapeutic decision making, for example, confirming the indication for biologics. Additionally, automated severity assessment has the potential to capture disease dynamics over time, which is important for future pharmaceutical trials and outcome assessments in clinical routine. It is possible that also patient management can be improved with similar, patient-centred applications in the future. This study breaks new grounds and provides an outlook for a future with fast, reliable, reproducible and easy-to-use disease severity assessments in HS patients.

ACKNOWLEDGEMENT

This study was supported by the *Verein zur Förderung der Dermatologie und Venerologie*, Vienna, Austria. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

Hardware and software were provided by the SCARLETRED Holding GmbH.

CONFLICT OF INTEREST STATEMENT

HS is the founder and RR is an employee of SCARLETRED Holding GmbH. CP received honoraria and travel support from BMS, MSD, Novartis, Pierre Fabre, SunPharma, Almirall, Pelpharma, ScarletRed, Amgen, Sanofi Genzyme, AbbVie, Celgene and Pfizer all unrelated to this project. AW received honoraria and travel support from AbbVie, Biogen, Janssen, Novartis and Sanofi unrelated to this project. KR has no conflicts to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

Reviewed and ethically approved by the ethics committee of the city of Vienna; EK18-100-0618. Patient consent for the publication of recognizable patient photographs or other identifiable material was obtained by the authors and included at the time of article submission to the journal stating that all patients gave consent with the understanding that this information may be publicly available.

ORCID

A. Wiala https://orcid.org/0000-0003-1029-2302
C. Posch https://orcid.org/0000-0003-0296-3567

TWITTER

C. Posch PoschChristian

REFERENCES

1. Jemec GB, Kimball AB. Hidradenitis suppurativa: epidemiology and scope of the problem. J Am Acad Dermatol. 2015;73:S4–7. https://doi.org/10.1016/j.jaad.2015.07.052
2. Goldburg SR, Strober BE, Payette MJ. Hidradenitis suppurativa: epidemiology, clinical presentation, and pathogenesis. J Am Acad Dermatol. 2020;82:1045–58. https://doi.org/10.1016/j.jaad.2019.08.090
3. von der Werth JM, Jemec GB. Morbidity in patients with hidradenitis suppurativa. Br J Dermatol. 2001;144:809–13. https://doi.org/10.1046/j.1365-2133.2001.04137.x
4. Gooderham M, Papp K. The psychosocial impact of hidradenitis suppurativa. J Am Acad Dermatol. 2015;73:S19–22. https://doi.org/10.1016/j.jaad.2015.07.054
5. Ingram JR, Hadjieconomou S, Piguet V. Development of core outcome sets in hidradenitis suppurativa: systematic review of outcome measure instruments to inform the process. Br J Dermatol. 2016;175:263–72. https://doi.org/10.1111/bjd.14475
6. Kimball AB, Sobell JM, Zouboulis CC, Gu Y, Williams DA, Sundaram M, et al. HiSCR (Hidradenitis Suppurativa Clinical Response): a novel clinical endpoint to evaluate therapeutic outcomes in patients with hidradenitis suppurativa from the placebo-controlled portion of a phase 2 adalimumab study. J Eur Acad Dermatol Venereol. 2016;30:989–94. https://doi.org/10.1111/jdv.13216
7. Zouboulis CC, Tzellos T, Kyrgidis A, Jemec G, Bechara FG, Giamarellos-Bourboulis EJ, et al. Development and validation of the International Hidradenitis Suppurativa Severity Score System (IHS4), a novel dynamic scoring system to assess HS severity. Br J Dermatol. 2017;177:1401–9. https://doi.org/10.1111/bjd.15748
8. Kokolakis G, Sabat R. Distinguishing mild, moderate, and severe hidradenitis suppurativa. JAMA Dermatol. 2018;154:971–2. https://doi.org/10.1001/jamadermatol.2018.1599
9. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56. https://doi.org/10.1038/s41591-018-0300-7
10. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28:924–33. https://doi.org/10.1038/s41591-022-01772-9
11. Zouboulis CC, Nogueira da Costa A. Standardized photographic documentation of hidradenitis suppurativa/Acne Inversa. Dermatology. 2019;235:51–4. https://doi.org/10.1159/000493342
12. Partl R, Jonko B, Schnidar S, Schöllhammer M, Bauer M, Singh S, et al. 128 SHADES OF RED: objective remote assessment of radiation dermatitis by augmented digital skin imaging. Stud Health Technol Inform. 2017;236:363–74.
13. Ranjan R, Partl R, Erhart R, Kurup N, Schnidar H. The mathematics of erythema: development of machine learning models for artificial intelligence assisted measurement and severity scoring of radiation induced dermatitis. Comput Biol Med. 2021;139:104952. https://doi.org/10.1016/j.compbiomed.2021.104952
14. Schnidar Harald AN. Methods for assessing erythema. Patent EP2976013A1 2016.
15. Schaller M, Riel S, Bashur R, Kurup N, Schnidar H, Fehrenbacher B. Ivermectin treatment in rosacea: how novel smartphone technology can support monitoring rosacea-associated signs and symptoms. Dermatol Ther. 2022;35:e15869. https://doi.org/10.1111/dth.15869
16. Kimball AB, Ganguli A, Fleischer A. Reliability of the hidradenitis suppurativa clinical response in the assessment of patients with hidradenitis suppurativa. J Eur Acad Dermatol Venereol. 2018;32:2254–6. https://doi.org/10.1111/jdv.15163
17. Thorlacius L, Garg A, Riis PT, Nielsen SM, Bettoli V, Ingram JR, et al. Inter-rater agreement and reliability of outcome measurement instruments and staging systems used in hidradenitis suppurativa. Br J Dermatol. 2019;181:483–91. https://doi.org/10.1111/bjd.17716

18. Marzano AV, Genovese G, Casazza G, Moltrasio C, Dapavo P, Micali G, et al. Evidence for a 'window of opportunity' in hidradenitis suppurativa treated with adalimumab: a retrospective, real-life multicentre cohort study. Br J Dermatol. 2021;184:133–40. https://doi.org/10.1111/bjd.18983
19. Wlodarek K, Stefaniak A, Matusiak L, Szepietowski JC. Could residents adequately assess the severity of hidradenitis suppurativa? Interrater and intrarater reliability assessment of major scoring systems. Dermatology. 2020;236:8–14. https://doi.org/10.1159/000501771

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wiala A., Ranjan R., Schnidar H., Rappersberger K., Posch C.. Automated classification of hidradenitis suppurativa disease severity by convolutional neural network analyses using calibrated clinical images. J Eur Acad Dermatol Venereol. 2024;38:576–582. https://doi.org/10.1111/jdv.19639