U.PORTO

MESTRADO INTEGRADO EM MEDICINA

Artificial Intelligence and Hidradenitis Suppurativa- A Systematic Review of Diagnosis and Severity Assessment

Mariana Fernandes
Soares de Almeida

M

2024/2025

FACULDADE DE **MEDICINA**

U.PORTO
FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

UC Dissertação - DECLARAÇÃO DE INTEGRIDADE

Eu, Mariana Fernandes Soares de Almeida, abaixo assinado, nº mecanográfico 201904603, estudante do 6º ano do Ciclo de Estudos Integrado em Medicina, na Faculdade de Medicina da Universidade do Porto, declaro ter actuado com absoluta integridade na elaboração do meu trabalho de Dissertação ou Monografia.

Neste sentido, confirmo que **NÃO** incorri em plágio (acto pelo qual um indivíduo, mesmo por omissão, assume a autoria de um determinado trabalho intelectual, ou partes dele). Mais declaro que todas as frases que retirei de trabalhos anteriores pertencentes a outros autores, foram referenciadas, ou redigidas com novas palavras, tendo colocado, neste caso, a citação da fonte bibliográfica.

Faculdade de Medicina da Universidade do Porto, 17/02/2025

Assinatura conforme cartão de identificação:

Mariana F. S. de Almeida

U.PORTO
FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

UC Dissertação – DECLARAÇÃO DE REPRODUÇÃO

NOME

Mariana Fernandes Soares de Almeida

NÚMERO DE ESTUDANTE	E-MAIL
201904603	Up201904603@up.pt

DESIGNAÇÃO DA ÁREA DO PROJECTO

Ciências médicas e da saúde- Biotecnologia médica

TÍTULO DISSERTAÇÃO/~~MONOGRAFIA~~ (riscar o que não interessa)

Artificial Intelligence and Hidradenitis Suppurativa- A Systematic Review of Diagnosis and Severity Assessment

ORIENTADOR

Carmen Maria Lisboa da Silva

COORIENTADOR (se aplicável)

José Alberto da Silva Freitas

ASSINALE APENAS UMA DAS OPÇÕES:

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTE TRABALHO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.	☒
É AUTORIZADA A REPRODUÇÃO PARCIAL DESTE TRABALHO (INDICAR, CASO TAL SEJA NECESSÁRIO, Nº MÁXIMO DE PÁGINAS, ILUSTRAÇÕES, GRÁFICOS, ETC.) APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.	☐
DE ACORDO COM A LEGISLAÇÃO EM VIGOR, (INDICAR, CASO TAL SEJA NECESSÁRIO, Nº MÁXIMO DE PÁGINAS, ILUSTRAÇÕES, GRÁFICOS, ETC.) NÃO É PERMITIDA A REPRODUÇÃO DE QUALQUER PARTE DESTE TRABALHO.	☐

Faculdade de Medicina da Universidade do Porto, 17/02/2025

Assinatura conforme cartão de identificação: Mariana F. S. de Almeida

U.PORTO
FMUP FACULDADE DE MEDICINA UNIVERSIDADE DO PORTO

UC Dissertação - DECLARAÇÃO DE TRANSPARÊNCIA RELATIVAMENTE À UTILIZAÇÃO DE FERRAMENTAS DE CHATBOT GENERATIVO BASEADAS EM LARGE LANGUAGE MODELS

Eu, Mariana Fernandes Soares de Almeida abaixo assinado, nº mecanográfico 201904603, estudante do 6º ano do Ciclo de Estudos Integrado em Medicina, na Faculdade de Medicina da Universidade do Porto, declaro que:

☒ Não procedi à utilização de ferramentas de *chatbox* generativo baseadas em *large language models* para nenhuma das tarefas no contexto do meu trabalho de Dissertação ou Monografia

☐ Procedi à utilização de ferramentas de *chatbox* generativo baseadas em *large language models* no contexto do meu trabalho de Dissertação ou Monografia, encontrando-se todas as interacções (*prompts* e respostas) transcritas em anexo bem como a indicação das aplicações utilizadas.

Neste sentido, confirmo que a eventual utilização de ferramentas de *chatbox* generativo baseadas em *large language models* no contexto do meu trabalho de Dissertação ou Monografia foi exclusivamente descrita na sequência de *prompts* e respostas transcritos em anexo e nas aplicações indicadas.

Faculdade de Medicina da Universidade do Porto, 17/02/2025

Assinatura conforme cartão de identificação:

Mariana F. S. de Almeida

ARTIFICIAL INTELLIGENCE AND HIDRADENITIS SUPPURATIVA- A SYSTEMATIC REVIEW OF DIAGNOSIS AND SEVERITY ASSESSMENT

Mariana Almeida[1], Patrícia Gomes[2], Alberto Freitas[3], Carmen Lisboa[2,4,*]

AFFILIATIONS:

[1] Faculty of Medicine, University of Porto

[2] Department of Dermatology and Venereology, São João Local Health Unit, Porto

[3] RISE-Health, Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto

[4] RISE-Health, Department of Pathology, Microbiology, Faculty of Medicine, University of Porto

***CORRESPONDING AUTHOR:**

Carmen Lisboa; email: carlis@med.up.pt

FUNDING INFORMATION

None.

CONFILTS OF INTEREST

None declared.

ETHICS STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the articles listed in the References.

AUTHOR CONTRIBUTIONS

Mariana Almeida (MA): Conceptualization, Methodology, Investigation, Writing - Original draft preparation, Writing - review and editing, Visualization; **Patrícia Gomes (PG):** Methodology, Investigation, Writing - review and editing**; Alberto Freitas (AF)**: Investigation, Writing - review and editing, Visualization; **Carmen Lisboa (CL):** Conceptualization, Methodology, Writing - review and editing, Visualization, Supervision.

Keywords: Hidradenitis suppurativa, Artificial intelligence, Diagnosis, Severity assessment, Machine learning algorithms.

ABSTRACT

Background: Hidradenitis Suppurativa (HS) requires accurate and timely diagnosis and severity assessment to improve clinical management. However, this remains a challenge due to the lack of standardized clinical tools. Artificial Intelligence (AI) allows for the creation of tools that assist the management of HS, with the potential to improve the diagnostic and severity assessment process.

Aim: This systematic review aimed to synthesize and evaluate the current applications of AI in the diagnosis and severity assessment process of HS.

Methods: A systematic search was conducted across PubMed, Web of Science, SCOPUS, IEEE Xplore, and ACM Digital Library databases from inception to 13 November 2024.

Results: Five articles met the inclusion criteria of the review. A study evaluated the diagnostic accuracy of an AI tool by image processing method. Two studies applied machine learning algorithms to differentiate lesions of HS from other entities. Two proof-of-concept studies explore the potential use of AI to automate the severity assessment of HS.

Conclusions: These articles offer different perspectives on the potential applications of AI. This review supports the potential of AI-based tools to standardize and assist in HS's diagnosis and severity assessment process. However, further improvement and clinical validation of these tools are needed.

Keywords: Hidradenitis suppurativa, Artificial intelligence, Diagnosis, Severity assessment, Machine learning algorithms.

INTRODUCTION

Hidradenitis suppurativa (HS) is a chronic, inflammatory skin disease that significantly impairs patients' quality of life [1]. It's characterized by recurrent skin lesions such as painful inflammatory nodules, deep abscesses, draining fistulas, and scars, primarily in the axillary, inguinal, and anogenital areas [2]. The pathogenesis of this condition is still not completely understood. However, it's believed that inflammation plays a leading role in its development. Additionally, immunologic, environmental, and genetic factors contribute to the pathogenesis of HS [3].

The clinical heterogeneity of HS can lead to misdiagnosis and delays in clinical evaluation, impacting the ability to standardize its diagnosis and severity assessment processes properly. On average, patients with HS experience on average 7 and up to 10-year diagnostic delay [4, 5]. Timely clinical evaluation at the early beginning of treatment is crucial for more effective management and controlling the disease's progression and several systemic comorbidities [6]. It's also important to mention that the decision on the course of treatment relies on the disease´s staging. The Hurley staging system classifies HS in three stages. However, a critical limitation of this staging system is that it isn't a dynamic evaluation [7]. Following this, the International Hidradenitis Suppurativa Severity Score System (IHS4) was created, providing a dynamic disease severity scoring system [8]. These manual score systems are inherently subjective and time-consuming because they rely on dermatologist's expertise and thorough evaluation.

In today's rapidly advancing technological world, the role of Artificial Intelligence (AI) and its medical applications, particularly in Dermatology, is becoming increasingly significant. The visual component inherent to the dermatology field, with a large clinical image database, creates an opportunity for the implementation of developing AI image processing techniques [9]. Machine Learning (ML) and its subset Deep Learning (DL), both components of the broader framework of AI, automate processes by inferring patterns from data. DL relies on Artificial Neural Networks, such as Convolutional Neural Networks (CNNs), to perform more complex tasks like the interpretation of medical images [10, 11]. This way, CNNs emerge as a promising tool in Dermatology, by efficiently identifying complex patterns and features in clinical images, they have the potential to revolutionize the diagnostic process of dermatological conditions [12].

In this systematic review, we aimed to provide a detailed and updated review of the current applications of AI in the diagnosis and severity assessment process of HS. As one of the first comprehensive syntheses incorporating the latest advances in AI, it will fill the knowledge gaps in this area.

MATERIAL AND METHODS

A systematic review of original studies was carried out in compliance with the current Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [13].

Search strategy and databases

We searched PubMed, Web of Science, and SCOPUS databases, from inception to November 13, 2024. The IEEE Xplore and the ACM Digital Library databases of the engineering, computer science, and information technology fields were also screened. We also hand-searched the bibliographies of relevant publications. The search terms selected to ensure the inclusion of all relevant articles were ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks") AND ("hidradenitis suppurativa" OR "acne inversa") (*Suppl. File*). The query results were uploaded and saved in *CoVidence*; an online platform designed for managing systematic reviews. To ensure the proper results of the research and to follow high standards of conducting systematic reviews, all duplicate bibliographic records were checked and removed.

Eligibility criteria

The search results, restricted to the title and abstract fields in this phase, were per the following inclusion criteria: (1) studies whose population is patients diagnosed with Hidradenitis Suppurativa or patients with genetic syndromes associated with HS, including the whole wide range of HS severity levels, without any restriction of age, gender; (2) studies that use artificial intelligence (AI), or related technologies in the context of the diagnosis or the severity assessment processes of Hidradenitis Suppurativa; (3) studies assessing the effectiveness, accuracy, or validity of AI in the diagnosis and/or in the clinical assessment, evaluation by improving diagnosis and/ or severity assessments in HS. The search findings might or might not include a comparator (traditional methods of severity assessment, that is, clinical assessment). There weren't any restrictions on the publication date, and studies published in any language could be included. The exclusion criteria consisted of the following: (1) studies whose population was non-human: animal or in vitro; (2) study design as editorials, commentaries, opinion pieces, or review articles without original data; (3) abstract-only text; (4) studies focusing on conditions other than HS and where applications of artificial intelligence to hidradenitis suppurativa clinical assessment of Hidradenitis Suppurativa, namely the diagnosis process and the standardization of the severity assessment process were not the primary focus.

Screening and data extraction

Firstly, two independent reviewers (PG, MA) screened the search results based on the title and abstract. The full text was obtained from the articles that met the inclusion criteria. Subsequently, the full-text screening of the articles was conducted further to assess compliance with the inclusion criteria and quality. Any discrepancies between the two reviewers were resolved through discussion with the third and fourth authors. There was no blinding process of the identity of the articles.

A data extraction sheet was created, and one reviewer (MA) performed the initial data extraction of the articles that met the inclusion criteria and quality evaluation. The study's aim, methods, assessment of complete reporting, reproducibility of results, data source and sample size, AI performance metrics, main results, and main limitations were extracted. This data extraction was reviewed by the other reviewers (AF, CL, PG).

Quality assessment

Two reviewers (CL, MA) independently assessed the quality of each study. Any discrepancies were resolved through discussion between them.

QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool was used to assess the risk of bias in studies that evaluate diagnostic accuracy. Four specific domains (patient selection, index test, reference standard, flow, and timing) were evaluated [14]. The PROBAST (Prediction model Risk Of Bias Assessment Tool) was chosen to assess the risk of bias in studies due to its pertinence in assessing studies focused on prediction models. The domains of participants, predictors, outcomes, and analysis were evaluated [15].

RESULTS

The initial sensitivity search identified 66 articles for screening from PubMed, SCOPUS, Web of Science, IEEE Xplore, and ACM Digital Library databases, and two from citation searching. After excluding 34 duplicates, 34 articles were screened independently by two authors based on the title and abstract using the *CoVidence* systematic review software. Of these, 14 articles were considered pertinent for a full-text review and assessed for eligibility. After thoroughly examining the relevant studies, five studies were found suitable for qualitative synthesis. We excluded four articles for wrong intervention, three for wrong outcome, one for wrong study design, and one article for abstract-only text. After following the review protocol, five articles were included in this systematic review for qualitative synthesis (*Figure 1*).

Figure 1

Characteristics of the included studies

Table 1

The main characteristics of the 5 studies included are summarized in *Table 1.* A study, Cirone et al. [16], assessed the diagnostic accuracy of an AI-based tool (*VisualDX*), assessing HS with high accuracy. Four studies [17-20] were prediction models. Kirby et al. [17] and Yamanaka-Takaichi et al. [18] explored the application of ML to assist in the differential diagnosis of HS. Kirby et al. [17] identified the age, gender, and risk factors of patients as the strongest predictive features of HS. Yamanaka-Takaichi et al. [18] identified 15 clinical features as the most important for distinguishing perianal HS from perianal IBD, with high sensitivity. Hernández Montilla et al. [20] with the creation of the Automatic International Hidradenitis Suppurativa Severity Score System (AIHS4), automated the calculation of IHS4 scores from clinical images by using a deep-learning model for lesion detection. AIHS4 displayed an overall performance comparable to the less experienced specialists. However, it was comparable to the senior dermatologist assessing moderate to severe HS cases. Wiala et al.[19] applied CNNs for clinical image analysis and severity assessment, effectively classifying HS, and distinguishing no/mild from moderate/severe disease with high accuracy. Complete reporting and reproducibility of results, as well as AI performance metrics, were described in *Suppl. Table.*

Quality assessment

The risk of bias in the study by Cirone et al.[16] was low in all the domains assessed, following the QUADAS-2 tool (*Table 2*). However, the study's applicability regarding patient selection and reference standard is unclear. These concerns stem mainly from using processed images to represent darker phenotypes, and it relied only on data from medical claims.

PROBAST was chosen to assess the quality of the remaining four studies [17-20] with a prediction model structure (*Table 3, Figure 2*). The study by Yamanaka-Takaichi et al.[18] had an

unclear overall risk of bias and applicability. The dataset was too small with an overrepresentation of patients in the later stages of the disease. Wiala et al. [19] had an overall low risk of bias. The underrepresentation of patients with low-severity HS stages and with darker skin phenotypes may not accurately represent the landscape of HS patients. Both Yamanaka-Takaichi et al.[18] and Kirby et al.[17] relied on claims data which may limit the generalization of results. The overall risk of bias of Hernández Montilla et al.[20] was high due to the bias introduced in the patient selection, namely the small dataset and inconsistent image quality, which affected the annotation and training process of the AI-based model for severity assessment.

Table 2

Table 3

Figure 2

DISCUSSION

To the best of our knowledge, this is one of the first systematic reviews of the applications of AI to the diagnosis and severity assessment of HS. We highlight the different perspectives on the use of AI-based tools for the clinical assessment of HS, namely the application of ML algorithms to assist in the differential diagnosis of HS, and AI-based tools for automating the severity assessment processes.

Challenges

Considering the novelty of this topic, the number of available studies was still limited. Another significant challenge we faced was the heterogeneity among studies, not only in the AI technologies used but also in their study designs and characteristics.

The main limitations of the studies included in this review also reveal important challenges that still limit the applicability of this technology in clinical practice. However promising, the application of AI to the clinical assessment of HS is still in the early stages.

One of the main challenges was ensuring the quality and quantity of the datasets used to train and validate the AI tools. The quality of the data used to train and validate AI tools directly impacts the quality of the tool developed. A considerable-sized dataset that is representative of the population is challenging to obtain but is essential to ensure the performance of these AI tools in real clinical settings. If the dataset is too small, bias may be introduced, limiting AI performance.

Considering that diagnosis is, on average, delayed by 7 years, the early stages of the disease are often underrepresented in the datasets. Therefore, insufficient AI training in this disease stage will perpetuate this problem. It is crucial to contradict this tendency and create datasets that allow for AI tools capable of early diagnosis to prevent disease progression and improve long-term outcomes.

The non-standardization of clinical images taken in different clinical settings also creates a barrier to these models' training and development process. ML algorithms, when trained with pictures with different light exposure, settings, perspective, and variability in other technical parameters, will struggle to discern the real differences and not those due to this technical variability. The International Skin Imaging Collaboration aimed to achieve consensus on standards for image acquisition, however, the application in the clinical practice is challenging particularly when images are patient-generated [10, 21].

Also, the dependency on a specific data structure when training AI models in datasets, such as the use of claim data, may limit the generalizability of the findings to different healthcare systems and populations.

The lack of integration of patient-reported outcomes, such as pain impact on quality of life, which are essential in the HS clinical assessment, is also an important issue. The

evaluation and tracking of disease progression is also lacking. Without their integration in training and validating AI models, we won't obtain tools capable of a complete clinical assessment.

The use of proprietary technology, non-disclosure of algorithms or technical parameters, and limited accessibility to some databases make it difficult to reproduce the results and independently validate them in the future. The included studies lacked external validation which raises a valid concern for future applicabilities. This means further clinical validation in real-world settings of AI-based tools in different healthcare settings and the promotion of accessible models are also required.

Current AI Applications and Future Potential

Hernández Montilla et al. [20] and Wiala et al. [19] findings support AI's potential to integrate clinical practice potentially reducing the time spent on manual scoring systems and lowering the inter-observer variability of the "traditional" severity assessment methods. The capability of automatically assessing the severity of HS just by analyzing clinical images, such as ones taken by a smartphone, can open the potential for clinical assessments even in remote conditions, improving accessibility and reducing the burden on healthcare systems.

The dynamic evaluation is time-consuming and requires regular clinical assessments. AI-based algorithms, such as the U-Net algorithm displayed in Wiala et al. [19], emerging as tools to assist in lesion detection, potentially allowing for better tracking of disease patterns, areas, and progression of HS.

To further develop applications of AI in the clinical management of HS, interdisciplinary collaboration between medical and engineering professionals is essential. This cooperation can ensure that the tools developed are both clinically relevant and effectively integrated into healthcare systems. The integration of AI-based tools into clinical workflow is an essential step. For example, in Hernández Montilla et al. [20] study, the integration of AISH4 into the CADx system is what makes this tool accessible for the user, by providing an automatic AIHS4 score following clinical image upload.

Regarding the differential diagnosis of HS, the studies of Kirby et al.[17] and Yamanaka-Takaichi et al. [18] contributed with innovative approaches. Using machine learning algorithms to develop diagnostic prediction models offers a complementary approach to differential diagnosis, potentially contributing to reducing misdiagnosis, which is a major problem in the clinical assessment of HS.

Additionally, Cirone et al. [16] highlight the need to develop more inclusive tools that perform equally well in diverse skin phenotypes, overcoming the challenges of effectively diagnosing darker skin and not perpetuating this discrepancy [22].

Besides the included articles, discussing some of the findings of the excluded articles of Ezanno et al. [23] and Walss et al. [24] is relevant. They contributed with an interesting approach to the development of chatbots to provide clinical information about HS, improving patients´ literacy on this condition. Although this is not aligned with the aim of this study, these tools can potentially complement clinical practice. They can empower patients with information and clinical guidance. However, both studies revealed important limitations of the chatbots, such as the superficial answers and lack of integration with guidelines.

CONCLUSION

Despite the limited number and heterogeneity of the studies included in this systematic review, this work supports the potential of AI-based tools to improve diagnostic accuracy and standardize the severity assessment process. These advancements could help overcome challenges in HS management, such as misdiagnosis and the lack of standardized tools for clinical assessment. However, further studies with the development of comprehensive datasets and robust clinical validation of these AI tools in real-world settings are needed.

BIBLIOGRAPHY

1. McKenzie SA, Harview CL, Truong AK, Grogan TR, Shi VY, Bennett RG, et al. Physical symptoms and psychosocial problems associated with hidradenitis suppurativa: correlation with Hurley stage. Dermatol Online J. 2020;26(9).
2. Zouboulis CC, Del Marmol V, Mrowietz U, Prens EP, Tzellos T, Jemec GB. Hidradenitis suppurativa/acne inversa: criteria for diagnosis, severity assessment, classification and disease evaluation. Dermatology. 2015;231(2):184-90.
3. Frew JW. Hidradenitis suppurativa is an autoinflammatory keratinization disease: A review of the clinical, histologic, and molecular evidence. JAAD International. 2020;1(1):62-72.
4. Jemec GBE. Hidradenitis Suppurativa. New England Journal of Medicine. 2012;366(2):158-64.
5. Saunte DM, Boer J, Stratigos A, Szepietowski JC, Hamzavi I, Kim KH, et al. Diagnostic delay in hidradenitis suppurativa is a global problem. Br J Dermatol. 2015;173(6):1546-9.
6. Snyder CL, Chen SX, Porter ML. Obstacles to Early Diagnosis and Treatment of Hidradenitis Suppurativa: Current Perspectives on Improving Clinical Management. Clin Cosmet Investig Dermatol. 2023;16:1833-41.
7. Dagenet CB, De DR, Shih T, Brooks B, Fixsen D, Park SE, et al. Hurley Staging Training for Hidradenitis Suppurativa Patients. Skin Appendage Disorders. 2024;10(6):524-7.
8. Zouboulis CC, Tzellos T, Kyrgidis A, Jemec GBE, 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(5):1401-9.
9. De A, Sarda A, Gupta S, Das S. Use of Artificial Intelligence in Dermatology. Indian J Dermatol. 2020;65(5):352-7.
10. Yu K, Syed MN, Bernardis E, Gelfand JM. Machine Learning Applications in the Evaluation and Management of Psoriasis: A Systematic Review. J Psoriasis Psoriatic Arthritis. 2020;5(4):147-59.
11. Murphree DH, Puri P, Shamim H, Bezalel SA, Drage LA, Wang M, et al. Deep learning for dermatologists: Part I. Fundamental concepts. J Am Acad Dermatol. 2022;87(6):1343-51.
12. Eapen BR. Artificial Intelligence in Dermatology: A Practical Introduction to a Paradigm Shift. Indian Dermatol Online J. 2020;11(6):881-9.
13. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Bmj-British Medical Journal. 2009;339.

14. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
15. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170(1):51-8.
16. Cirone K, Akrout M, Simpson R, Lovegrove F. Investigating the Performance of VisualDx on Common Dermatologic Conditions in Skin of Color. SKIN The Journal of Cutaneous Medicine. 2024;8(5):1788-96.
17. Kirby J, Kim K, Zivkovic M, Wang S, Garg V, Danavar A, et al. Uncovering the burden of hidradenitis suppurativa misdiagnosis and underdiagnosis: a machine learning approach. Frontiers in Medical Technology. 2024;6.
18. Yamanaka-Takaichi M, Nadalian S, Loftus EV, Ehman Jr EC, Todd A, Grimaldo AB, et al. Differentiating clinical characteristics of perianal inflammatory bowel disease from perianal hidradenitis suppurativa. International Journal of Dermatology.n/a(n/a).
19. 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. Journal of the European Academy of Dermatology and Venereology. 2024;38(3):576-82.
20. Hernández Montilla I, Medela A, Mac Carthy T, Aguilar A, Gómez Tejerina P, Vilas Sueiro A, et al. Automatic International Hidradenitis Suppurativa Severity Score System (AIHS4): A novel tool to assess the severity of hidradenitis suppurativa using artificial intelligence. Skin Research and Technology. 2023;29(6):e13357.
21. Katragadda C, Finnane A, Soyer HP, Marghoob AA, Halpern A, Malvehy J, et al. Technique Standards for Skin Lesion Imaging: A Delphi Consensus Statement. JAMA Dermatology. 2017;153(2):207-13.
22. Adamson AS, Smith A. Machine Learning and Health Care Disparities in Dermatology. JAMA Dermatology. 2018;154(11):1247-8.
23. Ezanno A-C, Fougerousse A-C, Pruvost-Balland C, Maccari F, Fite C. AI in Hidradenitis Suppurativa: Expert Evaluation of Patient-Facing Information. Clinical, Cosmetic and Investigational Dermatology. 2024;Volume 17:2459-64.
24. Walss M, Anzengruber F, Arafa A, Djamei V, Navarini Alexander A. Implementing Medical Chatbots: An Application on Hidradenitis Suppurativa. Dermatology. 2021;237(5):712-8.

FIGURES

Figure 1

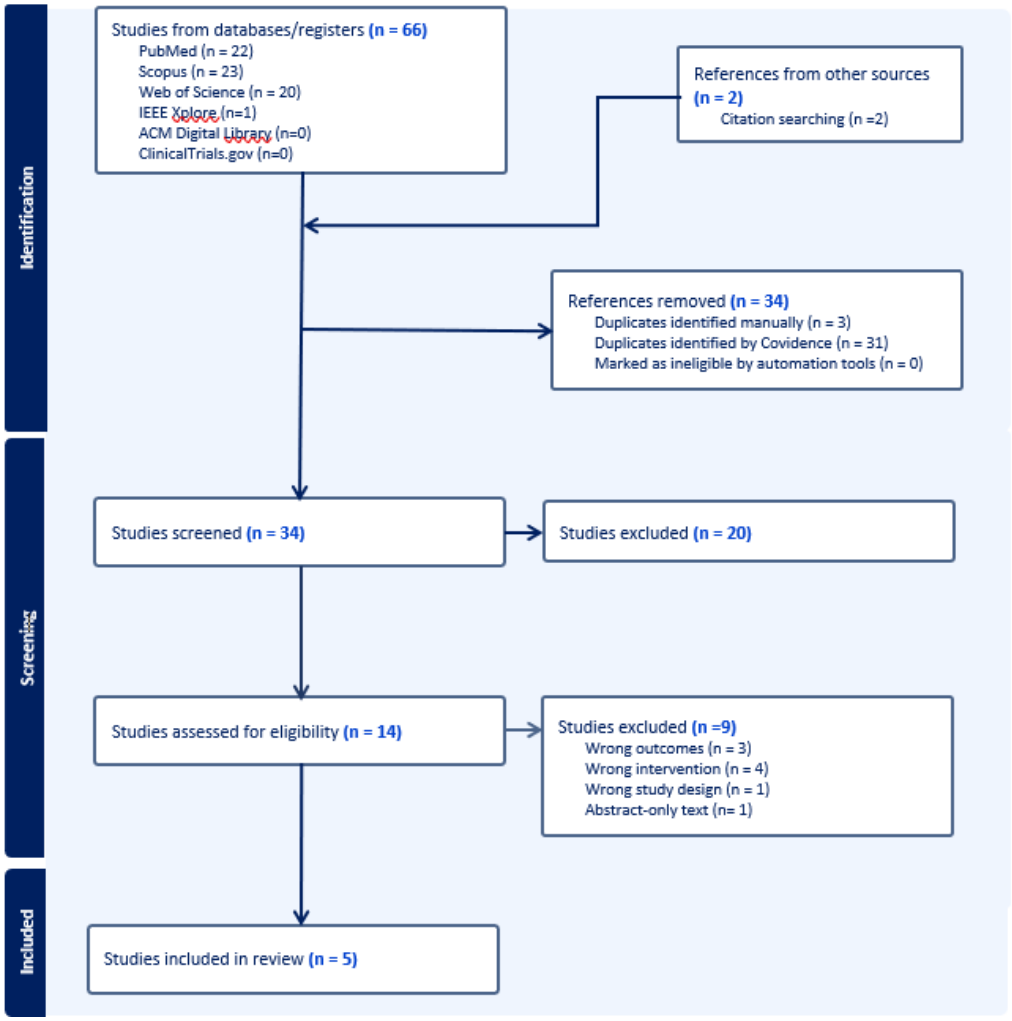
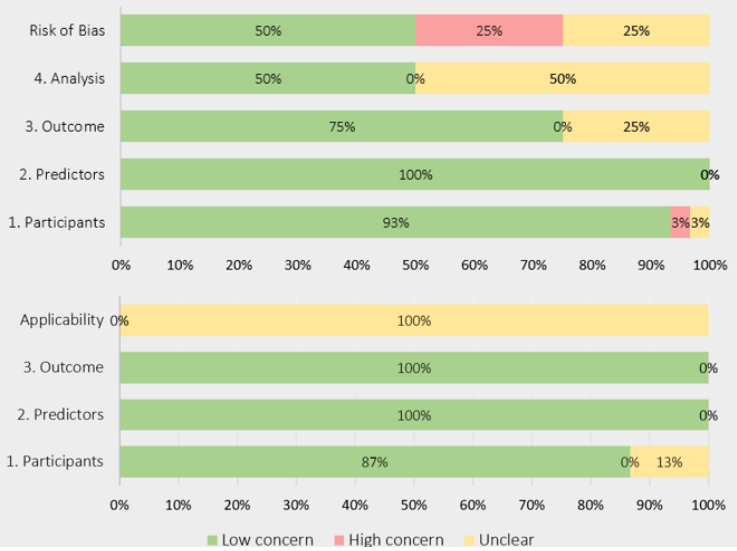


Figure 2



Legend for figures:

Figure 1- PRISMA flowchart for the selection of studies [13]

Figure 2- Summary of risk of bias and applicability assessment of the remaining included studies [17-20].

TABLES

Table 1-Main characteristics of the included studies

Study ID	Study Aim	Methods	Dataset	Data Source	Main Results
Cirone et al. (2024)	**Diagnosis:** Evaluation of diagnostic accuracy of VisualDx for dermatological conditions in diverse skin phenotypes.	AI-based diagnostic clinical decision-support system. Deep learning techniques for image processing.	480 images (16 conditions, 10 images each of subgroups: Fitzpatrick I-III, IV-VI, artificially darkened).	Partially open-source dermatological images. Image selection by a board-certified dermatologist. .	• HS showed the highest sensitivity (97%), across all subgroups. • Sensitivity differences between Fitzpatrick I-III and the other subgroups were statistically significant ($p < 0.001$).
Yamanaka-Takaichi et al. (2024)	**Diagnosis**: Assist in distinguishing perianal draining tunnels in HS from perianal fistulizing IBD, using ML.	ML algorithms for feature selection: random forest (trial-and-error setting); decision tree, XGB, recursive feature elimination. Validation by leave-one-out. Data visualization using t-SNE.	263 patients (98 with HS, 100 with IBD, and 65 with both HS and IBD).	Mayo Clinic's electronic health records from January 1st, 1998, to July 31st, 2021.	• Identification of 15 clinical features as the most important for distinguishing perianal HS from perianal IBD, with high sensitivity (93.4%). • Patients with HS had more lesions such as tunnels, axilla, and groin involvement.
Kirby et al. (2024)	**Diagnosis:** AI-based model to support clinical decision-making by identifying potential HS cases among patients with abscesses or cellulitis.	ML models: Penalized logistic regression using LASSO, RF, MLP, Boosting Algorithms (AdaBoost, XGBoost, LightGBM) and Ensemble Techniques (MaxVoting, Weighted Average). Models were validated using data from 2018 to 2019.	55989 cases; for the controls, 278,483 patients with abscesses and 1,431,524 patients with cellulitis.	IBM MarketScan Research Databases from 2000 to 2018 to train and test; from 2018 to 2019 to validate models.	• ML models differentiating HS and cellulitis (73%) performed better on all metrics than those differentiating HS and abscess patients (65%). • The strongest predictive features that identified HS were the age, gender, and risk factors of patients.
Wiala et al. (2023)	**Severity assessment-** Automatization of HS severity assessment by using CNNs to analyze	CNNs for severity classification (scale 0–3) and disease dynamics assessment. U-Net for automated lesion localization.	149 patients; 777 images (with assigned IHS4 scores). Dataset expanded to 7,675	Outpatient clinic in Vienna (May 2017 and January 2020).	• CNNs could effectively classify HS, distinguishing no/mild from moderate/severe disease with high accuracy (78%).

	clinical smartphone images		images using data augmentation.	Images taken by three dermatologists with smartphones. Expert dermatologist assigned IHS4 scores.	• CNN-based system identified disease dynamics and localized affected skin areas with a pixel accuracy of 88.1%.
Hernández Montilla et al. (2023)	**Severity assessment-** AI-based tool (AIHS4) to automate the severity assessment of HS by automatically calculating IHS4 scores from clinical images.	Deep learning model for lesion detection (*legit. Health-ISH4net)*, based on the *YOLOv5* architecture. Unification algorithm for inter-observer variability in HS lesion assessment.	219 clinical images, each with six label sets per image.	Legit.Health-HS-IHS4 dataset from DermQuest and DermNetNZ subsets. Six dermatologists annotated each image	• AIHS4 performed comparably to the less experienced specialists overall but demonstrated expertise comparable to that of senior dermatologist in assessing the severity of moderate and severe cases.

Abbreviations: AI, Artificial Intelligence; AIHS4, Automatic International Hidradenitis Suppurativa Severity Score System; AUC, Area Under the Curve; CNN, Convolutional Neural Network; CV, Coefficient of Variation; HS, Hidradenitis Suppurativa; IBD, inflammatory bowel disease; IHS4, International Hidradenitis Suppurativa Severity Score System; LASSO, least absolute shrinkage and selection operator; MAE, Mean Absolute Error; ML, Machine Learning; MLP, Multilayer Perceptron Neural Network; RF, Random Forest; t-SNE, t-distributed Stochastic Neighbor Embedding; XGB, extreme gradient boost.

Table 2- Risk of bias and applicability assessment of Cirone et al. (16) using QUADAS-2 tool

Author, Year	Risk of Bias				Applicability concerns		
	1. Patient selection	2. Idex test	3. Reference standard	4. Flow and timing	1. Patient selection	2. Index test	3. Reference standard
Cirone et al. (2024)	+	+	+	+	?	+	?

Low RoB High RoB Unclear

Table 3- Risk of bias and applicability assessment of the remaining included studies

Author, Year	Risk of Bias				Applicability			Overall	
	1. Participants	2. Predictors	3. Outcome	4. Analysis	1. Participants	2. Predictors	3. Outcome	Risk of Bias	Applicability
(2) Yamanaka-Takaichi et al., 2024	?	+	?	?	?	+	+	?	?
(3) Kirby et al. , 2024	+	+	+	+	?	+	+	+	?
(4) Wiala et al., 2023	+	+	+	+	?	+	+	+	?
(5) Hernández Montilla et al., 2023	-	+	+	?	?	+	+	-	?

Low RoB High RoB Unclear

Supplementary Files for Review

Search Strategy:

From inception to November 13, 2024, we obtained 22 articles in PubMed by using the following search terms ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks" OR "artificial intelligence"[MeSH Terms] OR "machine learning"[MeSH Terms] OR "deep learning"[MeSH Terms] OR "neural networks, computer"[MeSH Terms]) AND ("hidradenitis suppurativa" OR "acne inversa" OR "hidradenitis suppurativa"[MeSH Terms]).

The remaining databases were searched, from inception to November 13, 2024, using the following search terms: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks") AND ("hidradenitis suppurativa" OR "acne inversa"). In SCOPUS, we obtained 23 articles. In Web of Science, 20 articles; in IEEE Xplore 1 article and in ACM Digital Library and Clinicaltrials.gov, none.

Supplementary Table S1- Assessment of AI in the included studies.

Study ID	Complete reporting and reproducibility of results	AI performance metrics
Cirone et al. (2024)	Supported by: - Well-described methodology with specification of image processing techniques, including algorithms and parameter values. - Dataset sources are publicly available. Limited by: - Proprietary nature of VisualDx software. - Non-disclosure of the chosen images from the dataset.	Sensitivity: -Top-1: 100% for Fitzpatrick I-III and IV-VI, 90% in the processed images. -Top-3 and Top-5: 97%-100% for all subgroups. Specificity reduced in the processed images
Yamanaka-Takaichi et al. (2024)	Supported by: - Well-described methodology: model performance´s description with performance metrics used for its evaluation (accuracy and feature importance rankings); validation methods used. - Detailed description of the dataset selection and characteristics. - Use of free tools such as Python and RStudio. Limited by: - Non-availability of additional data from the electronic health records. - Missing hyperparameters of the machine learning algorithms used.	Random forest classifier: -Sensitivity: 93.4%; -Specificity: 90%; -High accuracy.
Kirby et al. (2024)	Supported by: - Well-described methodology including feature selection, training, validation, and general scheme processes. - Software used SAS (Software as Service) data preparation and Python for modeling—is open-sourced. Limited by: - Use of proprietary claims data (IBM MarketScan) limits accessibility. - Specific codes, features, or model weights were not shared, limiting reproducibility.	-Diagnostic accuracy: 65% and 73% of models trained on abscess and cellulitis controls, respectively. In top-performing models: -Precision: 60-80%. -Sensitivity: 55-76%. -AUC (Area Under the Curve): 81%–82%.
Wiala et al. (2023)	Supported by: - Well-described methodology, describing the specific performance metrics used, the dataset selection and data augmentation process, and the results obtained, including test accuracy results and the AI architectures used in each process. - The dataset is available at the author's reasonable request. Limited by:	-Binary classification: Accuracy 78%; AUC 0.85. -Multiclass classification: Accuracy 72%; AUC 0.84–0.89. -Lesion localization: Pixel accuracy 88.1%; test loss 0.42.

	- Non-open-source software Scarletred Vision. - Information regarding the AI model architecture, allows for only partially replicating the training process.	-Disease dynamics evaluation: NRMSE (Normalized Root Mean Square Error) 0.26.
Hernández Montilla et al. (2023)	Supported by: - Well-described methodology: Detailed description of the dataset; the methods used for data augmentation and algorithm evaluation metrics, e.g., MAE (Mean Absolute Error) or CV (Coefficient of Variation). - Open-source algorithm (YOLOv5). - Both datasets (DermQuest and DermnetNZ) used to create Legit.Health-HS-IHS4 (Legit.Health-HS-IHS4) are open source. Limited by: - Lack of full disclosure of criteria used for selecting clinical images within the dataset. - Lack of details regarding certain thresholds, such as the IoU (Intersection over Union) threshold.	-Evaluation of Legit.Health-IHS4: MAE (vs. the ground truth) and CV. -Precision: 0.44±0.09 to 0.46±0.10. -Recall: 0.39±0.07 to 0.41±0.11.

REPORTING GUIDELINES

The PRISMA 2020 checklist of items that should be included in reports of a systematic review report.

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 1: “Artificial Intelligence and Hidradenitis Suppurativa- A Systematic Review of Diagnosis and Severity Assessment”
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 3: “In today’s rapidly advancing technological world, the role of Artificial Intelligence (AI) and its medical applications, particularly in Dermatology, is becoming increasingly significant.” The visual component inherent to the dermatology field, with a large clinical image database, creates an opportunity for the implementation of developing AI image processing techniques”; “This way, CNNs emerge as a promising tool in Dermatology, by efficiently identifying complex patterns and features in clinical images, they have the potential to revolutionize the diagnostic process of dermatological conditions.”
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 3: “In this systematic review, we aimed to provide a detailed and updated review of the current applications of AI in the diagnosis and severity assessment process of HS. As one of the first comprehensive syntheses incorporating the latest advances in AI, it will fill the knowledge gaps in this area.”
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 4: “The search results, restricted to the title and abstract fields in this phase, were per the fol lowing inclusion criteria: (1) studies whose population is patients diagnosed with Hidradenitis Suppurativa or patients with genetic syndromes associated with HS, including the whole wide range of HS severity levels, without any restriction of age, gender; (2) studies that use artificial intelligence (AI), or related technologies in the context of the diagnosis or the severity assessment processes of Hidradenitis Suppurativa; (3) studies assessing the effectiveness, accuracy, or validity of AI in the diagnosis and/or in the clinical assessment, evaluation by improving diagnosis and/ or severity assessments in HS. The search findings might or might not include a comparator (traditional methods of severity assessment, that is, clinical assessment). There weren’t any restrictions on the publication date, and studies published in any language could be included. The exclusion criteria consisted of the following: (1) studies whose population was non-human: animal or in vitro; (2) study design as editorials, commentaries, opinion pieces, or review articles without original data; (3) abstract-only text; (4) studies focusing on conditions other than HS and where applications of artificial intelligence to hidradenitis suppurativa clinical assessment of Hidradenitis Suppurativa, namely the diagnosis process and the standardization of the severity assessment

Section and Topic	Item #	Checklist item	Location where item is reported
			process were not the primary focus."
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 4: "We searched PubMed, Web of Science, and SCOPUS databases, from inception to November 13, 2024. The IEEE Xplore and the ACM Digital Library databases of the engineering, computer science, and information technology fields were also screened. We also hand-searched the bibliographies of relevant publications."
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 4: "The search terms selected to ensure the inclusion of all relevant articles were: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks") AND ("hidradenitis suppurativa" OR "acne inversa")." Page 19: Supplemental File
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 5: "Firstly, two independent reviewers (PG, MA) screened the search results based on the title and abstract. The full text was obtained from the articles that met the inclusion criteria. Subsequently, the full-text screening of the articles was conducted further to assess compliance with the inclusion criteria and quality. Any discrepancies between the two reviewers were resolved through discussion with the third and fourth authors. There was no blinding process of the identity of the articles."
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 5: "A data extraction sheet was created, and one reviewer (MA) performed the initial data extraction of the articles that met the inclusion criteria and quality evaluation. (…)This data extraction was reviewed by the other reviewers (AF, CL, PG)."
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 5: "The study's aim, methods, assessment of complete reporting, and reproducibility of results, data source and sample size, AI performance metrics, main results, and main limitations were extracted."
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Not applicable
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable,	Page 5: "Two reviewers (CL, MA) independently assessed the quality of each study. Any discrepancies were resolved through discussion between them. QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool was used to assess the risk of bias in studies that evaluate diagnostic accuracy. (…) The PROBAST (Prediction model Risk Of Bias Assessment Tool) was chosen to assess the risk

Section and Topic	Item #	Checklist item	Location where item is reported
		details of automation tools used in the process.	of bias in studies due to its pertinence in assessing studies focused on prediction models."
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Not applicable
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Not applicable
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Not applicable
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Not applicable
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Not applicable
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not applicable
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not applicable
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not applicable
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not applicable
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 6: "The initial sensitivity search identified 66 articles for screening from PubMed, SCOPUS, Web of Science, IEEE Xplore, and ACM Digital Library databases, and two from citation searching. After excluding 34 duplicates, 34 articles were screened independently by two authors (…) Of these, 14 articles were considered pertinent for a full-text review and assessed for eligibility. After thoroughly examining the relevant studies, five studies were found suitable for qualitative synthesis." Page 14: Figure 1

Section and Topic	Item #	Checklist item	Location where item is reported
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Page 6: ". We excluded four articles for wrong intervention, three for wrong outcome, one for wrong study design, and one article for abstract-only text. After following the review protocol, five articles were included in this systematic review for qualitative synthesis."
Study characteristics	17	Cite each included study and present its characteristics.	Page 16-17: Table 1 Page 6: "A study, Cirone et al. (16), assessed the diagnostic accuracy of an AI-based tool (VisualDX). Four studies (17-20) were prediction models. Kirby et al. (17) and Yamanaka-Takaichi et al. (18) explored the application of ML to assist in the differential diagnosis of HS. Hernández Montilla et al. (20) and Wiala et al. (19) focused on creating AI-based tools for automated HS severity assessment. Hernández Montilla et al. (20) with the creation of AIHS4, automated the calculation IHS4 scores from clinical images by using a deep-learning model for lesion detection. Wiala et al. (19) applied CNNs for clinical image analysis and severity assessment." Page 19-20: Suplementary Table S1
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Page 6-7:" The risk of bias in the study by Cirone et al. (16) was low in all the domains assessed, following the QUADAS-2 tool (Table 2)." "The study by Yamanaka-Takaichi et al. (18) had an unclear overall risk of bias and applicability. (…). Wiala et al. (19) had an overall low risk of bias and an unclear applicability assessment. (…) The overall risk of bias of Hernández Montilla et al. (20) was high (…)." Page 17-18: Table 2 and 3
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Not applicable
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Not applicable
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Not applicable
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Not applicable
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not applicable

Section and Topic	Item #	Checklist item	Location where item is reported
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not applicable
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not applicable
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 8-9: “Hernández Montilla et al. (20) and Wiala et al. (19) findings support AI’s potential to integrate clinical practice potentially reducing the time spent on manual scoring systems and lowering the inter-observer variability of the “traditional” severity assessment methods.” “The dynamic evaluation is time-consuming and requires regular clinical assessments. AI-based algorithms, such as the U-Net algorithm displayed in Wiala et al. (19), emerging as tools to assist in lesion detection, potentially allowing for better tracking of disease patterns, areas, and progression of HS.” “For example, in Hernández Montilla et al. (20) study, the integration of AISH4 into the CADx system is what makes this tool accessible for the user, by providing an automatic AIHS4 score following clinical image upload.” “Regarding the differential diagnosis of HS, the studies of Kirby et al.(17) and Yamanaka-Takaichi et al. (18) contributed with innovative approaches.” “Additionally, Cirone et al. (16) highlight the need to develop more inclusive tools that perform equally well in diverse skin phenotypes, overcoming the challenges of effectively diagnosing darker skin and not perpetuating this discrepancy (22).” “Besides the included articles, discussing some of the findings of the excluded articles of Ezanno et al. (23) and Walss et al. (24) is relevant. They contributed with an interesting approach to the development of chatbots to provide clinical information about HS, improving patients´ literacy on this condition.”
	23b	Discuss any limitations of the evidence included in the review.	Page 8-9: “The main limitations of the studies included in this review also reveal important challenges that still limit the applicability of this technology in clinical practice.” “One of the main challenges was ensuring the quality and quantity of the datasets used to train and validate the AI tools.” “One of the main challenges was ensuring the quality and quantity of the datasets used to train and validate the AI tools.” “The non-standardization of clinical images taken in different clinical settings also creates a barrier to these models' training and development process. ML algorithms, when trained with pictures with different light exposure, settings, perspective, and variability in other technical parameters, will struggle to discern the real differences and not those due to this technical variability.” “Also, the dependency on a specific data structure when training AI models in datasets, such as the use of claim data, may limit the generalizability of the findings to different healthcare systems and

Section and Topic	Item #	Checklist item	Location where item is reported
			populations." "The lack of integration of patient-reported outcomes, such as pain impact on quality of life, which are essential in the HS clinical assessment, is also an important issue. The evaluation and tracking of disease progression is also lacking. Without their integration in training and validating AI models, we won't obtain tools capable of a complete clinical assessment." "The use of proprietary technology, non-disclosure of algorithms or technical parameters, and limited accessibility to some databases make it difficult to reproduce the results and independently validate them in the future. The included studies lacked external validation which raises a valid concern for future applicabilities."
	23c	Discuss any limitations of the review processes used.	Page 8: "Considering the novelty of this topic, the number of available studies was still limited. Another significant challenge we faced was the heterogeneity among studies, not only in the AI technologies used but also in their study designs and characteristics."
	23d	Discuss implications of the results for practice, policy, and future research.	Page 8-10: "However promising, the application of AI to the clinical assessment of HS is still in the early stages." "The capability of automatically assessing the severity of HS just by analyzing clinical images, such as ones taken by a smartphone, can open the potential for clinical assessments even in remote conditions, improving accessibility and reducing the burden on healthcare systems." "To further develop applications of AI in the clinical management of HS, interdisciplinary collaboration between medical and engineering professionals is essential. This cooperation can ensure that the tools developed are both clinically relevant and effectively integrated into healthcare systems." "Using machine learning algorithms to develop diagnostic prediction models offers a complementary approach to differential diagnosis, potentially contributing to reducing misdiagnosis, which is a major problem in the clinical assessment of HS." "Besides the included articles, discussing some of the findings of the excluded articles of Ezanno et al. (23) and Walss et al. (24) is relevant. (…) these tools can potentially complement clinical practice. They can empower patients with information and clinical guidance. " "These advancements could help overcome challenges in HS management, such as misdiagnosis and the lack of standardized tools for clinical assessment. However, further studies with the development of comprehensive datasets and robust clinical validation of these AI tools in real-world settings are needed."
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Not applicable
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not applicable

Section and Topic	Item #	Checklist item	Location where item is reported
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not applicable
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Not applicable
Competing interests	26	Declare any competing interests of review authors.	Not applicable
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Not applicable

FORMATTING RULES OF CHOSEN JOURNAL

The *Journal of the European Academy of Dermatology and Venereology (JEADV)* requirements for the systematic reviews were followed.

Systematic review and meta-analysis, see **example**	Substantial body of clinical or translational work that provides novel findings	n/a	Follow **EQUATOR** Reporting Guidelines **Title Page** Structured **Abstract** of ≤ 300 words **Key points** **Graphical Abstract** ≤ 3000 words in the **Main text**[1] No limits for **tables and figures** (with the relevant **legends**) **PROSPERO** or a similar database registration before data extraction commences. **PRISMA** (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram using the file designation 'Figure' during submission **PRISMA checklist** using the file designation 'Supplementary files for review' during submission No Limits for **References**