Journal of Oral Biology and Craniofacial Research 16 (2026) 15–22

Contents lists available at ScienceDirect

Journal of Oral Biology and Craniofacial Research

journal homepage: www.elsevier.com/locate/jobcr

Diagnostic accuracy of artificial intelligence in the diagnosis of pemphigus and pemphigoid groups of disorders based on clinical images: A systematic review and meta-analysis

Pravallika Kakada, Monal Yuwanati *, Pratibha Ramani

Department of Oral and Maxillofacial Pathology, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, 600077, India

ARTICLE INFO

Keywords:
Pemphigus
Pemphigoid
Artificial intelligence
Diagnosis
Accuracy
Sensitivity
Specificity

ABSTRACT

Background: Pemphigus and pemphigoid disease pose diagnostic challenges to pathologists and clinician due to overlapping clinical and histopathological features. The systematic review evaluated the diagnostic accuracy of artificial intelligence (AI) in identifying pemphigus and pemphigoid disorders based on clinical images, compared to conventional diagnostic techniques.

Methods: A search was carried out in PubMed, Scopus, and Web of Science databases since inception using search strategy comprising of "Artificial Intelligence", "Pemphigus", "Pemphigoid", "Diagnosis", "sensitivity", "specificity", and "accuracy". Studies examined AI models' accuracy for diagnosing pemphigus and pemphigoid were included in the review. A meta-analysis was conducted using Meta-Disc software, with pooled sensitivity, specificity, and summary receiver operating characteristic (SROC) curve analysis.

Results: Out of 4094 search results, five studies met the eligibility criteria after screening and selection steps. The pooled diagnostic accuracy of AI-based systems was 0.83 (CI 0.73–0.90) with a of 0.86 % (CI: 0.76–0.91) pooled sensitivity and 0.84 % (CI: 0.79–0.87) pooled specificity, reflecting its moderate effectiveness in ruling out pemphigus and pemphigoid disorders. Further, the positive likelihood ratio (PLR) was 5.34 (95 % CI: 4.05–7.05), and the negative likelihood ratio (NLR) was 0.16 (95 % CI: 0.09–0.28), with a DOR of 32.53 (95 % CI: 14.70–71.94) indicates moderate to strong accuracy.

Conclusion: AI have moderate diagnostic efficacy in diagnosis of pemphigus and pemphigoid diseases. However, additional research is needed to develop standardized methodologies and ensure generalizability across different populations before integrating into clinical practice.

1. Introduction

The pemphigus and pemphigoid are characterized by blisters formation secondary to separation of epithelium and involves the skin, eyes, and oral mucosa.[1] Further, it has significant mortality rates which range between 6 % and 41 % in the first year following diagnosis.[2,3] Therefore, early interventions must be carried out in a timely and precise manner during diagnosis.[4] The identification of these disorders is very challenging due to considerable overlap with other vesiculobullous lesions.[5,6] Furthermore, blister formation can be caused by different etiologies, such as viral infections (herpes simplex), genetic defects (epidermolysis bullosa), and allergic reactions (contact dermatitis or drug reactions).[7–10] Not only that, the diagnostic features are also confounded by overlapping histopathological presentation with other vesiculobullous disorders, which may result in misdiagnosis and delayed treatment.[11,12] With the potentially life-threatening nature of delayed diagnosis, early detection is critical.[13] Conventional methods of diagnosis such as clinical examination, histopathology, and direct immunofluorescence microscopy, although effective, are usually time-consuming, costly, and reliant on technical skills.[14–16–]

Artificial intelligence (AI) has progressively advanced from its early role in diagnostic tools for conditions such as cardiovascular disease and

This article is part of a special issue entitled: AI matters published in Journal of Oral Biology and Craniofacial Research.

* Corresponding author. Department of Oral and Maxillofacial Pathology, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, 600077, India.

E-mail addresses: drpravallika2210@gmail.com (P. Kakada), monal9817@gmail.com (M. Yuwanati), pratibaramani@saveetha.com (P. Ramani).

https://doi.org/10.1016/j.jobcr.2025.10.023
Received 6 August 2025; Received in revised form 29 September 2025; Accepted 27 October 2025
Available online 8 November 2025
2212-4268/© 2025 The Authors. Published by Elsevier B.V. on behalf of Craniofacial Research Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

pneumonia to more sophisticated applications in dermatology and autoimmune disorders.[17,18] The development of deep learning algorithms, particularly convolutional neural networks (CNNs) and transformer-based architectures, has significantly improved the accuracy of clinical image analysis.[19–21] Within dermatopathology, these advances have enabled AI-based systems to begin addressing complex autoimmune blistering diseases such as pemphigus and pemphigoid, where timely and accurate diagnosis remains a considerable challenge.[22,23] To date, much of the work in this area has centered on laboratory-based applications, including direct immunofluorescence (DIF) and indirect immunofluorescence (IIF) image analysis.[24–26] While these techniques provide valuable diagnostic information, they demand substantial technical expertise, specialized infrastructure, and are often costly and time-consuming.[27] By contrast, clinical image-based AI applications offer the advantage of rapid, non-invasive screening and have particular relevance in resource-limited settings where access to advanced laboratory facilities may be restricted. Expanding research efforts from laboratory-based approaches to clinical image analysis is therefore essential to maximize the real-world diagnostic utility of AI in pemphigus and pemphigoid disorders.

Although studies have proven efficiency, there are gaps which require evidences. There is no one specific AL model or tool that can work without needing much technical jargon. Further, there is ambiguity as well as variability in diagnostic efficacy of AI based technique in pemphigus and pemphigoid disease. This systematic review assessed efficacy of AI-based diagnostic methods in pemphigus and pemphigoid diseases.

2. Methods

2.1. Protocol and registration

This systematic review was reported according to the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).[28] The systematic review is registered in PROSPERO (CRD42024580232).

2.2. Research question

The research question for the systematic review is:

What is the diagnostic accuracy (O) of AI (I) for pemphigus and pemphigoid disorders (P) compared to conventional diagnostic methods (C)?

2.3. Eligibility criteria

We included studies that involved patients with pemphigus and pemphigoid disorders, used AI technologies for diagnosing these disorders, compared AI methods to traditional histopathological methods, reported how well AI worked in diagnosing these disorders (including sensitivity, specificity, and overall accuracy), were either randomised controlled trials or observational studies, and were published. We left out studies that didn't focus on patients with pemphigus and pemphigoid disorders or included other unrelated conditions; didn't use AI technologies or only looked at traditional diagnostic methods; didn't provide important results about how accurate AI is for diagnosis; or were peer-reviewed articles, case reports, grey literature, reviews, and opinion pieces.

2.4. Search strategy and sources

A search strategy was developed using the following keywords and terms:

(((ALL=(Artificial intelligence)) OR ALL=(Ai)) OR ALL=(Deep learning)) OR ALL=(Machine learning)) OR ALL=(Convolutional Neural Network)) OR ALL= (CNN) AND (C((ALL=(Diagnosis)) OR ALL=(Diagnostic aid)) OR ALL=(Sensitivity)) OR ALL=(Specificity)) OR ALL=(Diagnostic accuracy)) OR ALL=(Accuracy)) OR ALL=(ROC curve)) OR ALL=(AUC curve) AND (C(ALL=(Pemphigus)) OR ALL=(Vesiculobullous lesions)) OR ALL=(Skin diseases))) OR ALL=(Autoimmune blistering disorders))) OR ALL=(Skin lesions)) OR ALL=(Blistering lesions)) OR ALL=(Dermatological diseases)) OR ALL=(Autoimmune bullous disease)

These keywords were used with Boolean operators (AND, OR). The search was carried out in PubMed, Scopus and Web of science without any restrictions or filters (Table 1). The search results were exported for screening and selection process.

2.5. Screening and selection

The search results were imported into Rayyan software[29] and duplicates were identified and removed. After removing the duplicates remaining results were screened for eligibility based on title, abstract and keywords by two reviewers. The eligible articles were then subjected to full text retrieval and screening. Finally, only those articles met eligibility criteria were included in the final analysis of the review.

2.6. Data extraction

Two reviewers independently extracted the data using the predefined criteria into excel sheet. Information extracted included author name, publication year, name of journal, country, study design, population characteristics if available, type of sample, number of cases, number of images, type of AI technology used, comparison with conventional diagnostic methods, and outcomes (sensitivity, specificity, diagnostic accuracy).

2.7. Risk of bias assessment

The risk of bias was performed with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[30] Each study was

Table 1
Search strategies.

Data base	SEARCH STRATEGY
SCOPUS	TITLE-ABS-KEY(Artificial intelligence) OR TITLE-ABS-KEY(Ai) OR TITLE-ABS- KEY(Machine learning) OR TITLE-ABS-KEY(deep learning) OR TITLE-ABS-KEY(convolutional neural network) OR TITLE-ABS-KEY(CNN) AND TITLE-ABS-KEY (autoimmune bullous lesions) OR TITLE-ABS-KEY(Autoimmune blistering disorders) OR TITLE-ABS-KEY(dermatological disease) OR TITLE-ABS-KEY (pemphigus) OR TITLE-ABS-KEY(Pemphigoid) AND TITLE-ABS-KEY(Sensitivity) OR TITLE-ABS-KEY(Diagnosis) OR TITLE-ABS-KEY(Diagnostic Aid) OR TITLE-ABS-KEY(Accuracy) OR TITLE-ABS-KEY(Specificity)
PUBMED	((((Artificial intelligence[Title/Abstract]) OR(Ai[Title/Abstract])) OR (machine learning [Title/Abstract])) OR (deep learning [Title/Abstract])) OR (convolutional neural network[Title/Abstract])) OR (CNN[Title/Abstract])) AND((((Pemphigus[Title/Abstract]) OR (Vesiculobullous lesions [Title/Abstract])) OR (skin disease [Title/Abstract])) OR (autoimmune blistering disease [Title/Abstract])) OR (autoimmune bullous disease [Title/Abstract]))) AND((((Diagnosis[Title/Abstract]) OR (Diagnostic aid [Title/Abstract])) OR (accuracy[Title/Abstract])) OR (sensitivity[Title/Abstract])) OR (specificity[Title/Abstract]))
WEB OF SCIENCE	(((ALL=(Artificial intelligence)) OR ALL=(Ai)) OR ALL=(Deep learning)) OR ALL=(Machine learning)) OR ALL=(Convolutional Neural Network)) OR ALL= (CNN) AND (C((ALL=(Diagnosis)) OR ALL=(Diagnostic aid)) OR ALL=(Sensitivity)) OR ALL=(Specificity)) OR ALL=(Diagnostic accuracy)) OR ALL=(Accuracy)) OR ALL=(ROC curve)) OR ALL=(AUC curve) AND (C (ALL=(Pemphigus)) OR ALL=(Vesiculobullous lesions)) OR ALL=(Skin diseases))) OR ALL=(Autoimmune blistering disorders))) OR ALL=(Skin lesions)) OR ALL=(Blistering lesions)) OR ALL=(Dermatological diseases)) OR ALL=(Autoimmune bullous disease)

independently reviewed by two assessors, and discrepancies were resolved through discussion. The tool assesses patient recruitment, index test, reference standard, and flow and timing of the study in each study. Studies were considered low, unclear, or high risk of bias based on the evaluation of domains.

2.8. Data synthesis

To assess the diagnostic accuracy of AI for diagnosis of pemphigus and pemphigoid, meta-analysis of diagnostic tests was carried out using sensitivity and specificity extracted from the included studies. Pooled sensitivity, and specificity was estimated alongside the generation of summary receiver-operating characteristic (SROC) curves with help of using the Meta-Disc 2.0.[31] A bivariate random-effects regression model was employed for the synthesis of pooled sensitivity and specificity, with SROC curves generated using this bivariate framework. An area under the curve (AUC) above 0.9 was considered as good diagnostic accuracy. Heterogeneity within studies was evaluated from Cochran's Q test ($\chi 2$) and I^2 test. An inconsistency value (I^2) value $>$ 75 % was considered as high heterogeneity.

3. Results

3.1. Search results

A total of 4094 records were identified through a comprehensive search of databases (PubMed, Web of Science, and Scopus) by April 2025. After removing 104 duplicates, 3990 studies were screened based on title, abstract and keywords which lead to exclusion of 3968 studies leaving the 22 studies for full-text screening. One study full text could not be retrieved. The author of the paper was contacted however one full text paper couldn't be retrieved. Remaining 21 articles were screened for eligibility. Out of 21, 14 were excluded since they didn't not include Pemphigus and Pemphigoid in skin disorders, 2 studies were excluded as they did not include the required data. Several studies, although indexed under dermatopathology or autoimmune blistering diseases, did not include cases of pemphigus or pemphigoid or lacked relevant diagnostic data. Finally, 5 were included in the summary synthesis. PRISMA (Fig. 1).[32]

3.2. Characteristics of included studies

Table 2T provides a detailed overview of several studies published between 2020 and 2024, analyzing a wide range of dermatological

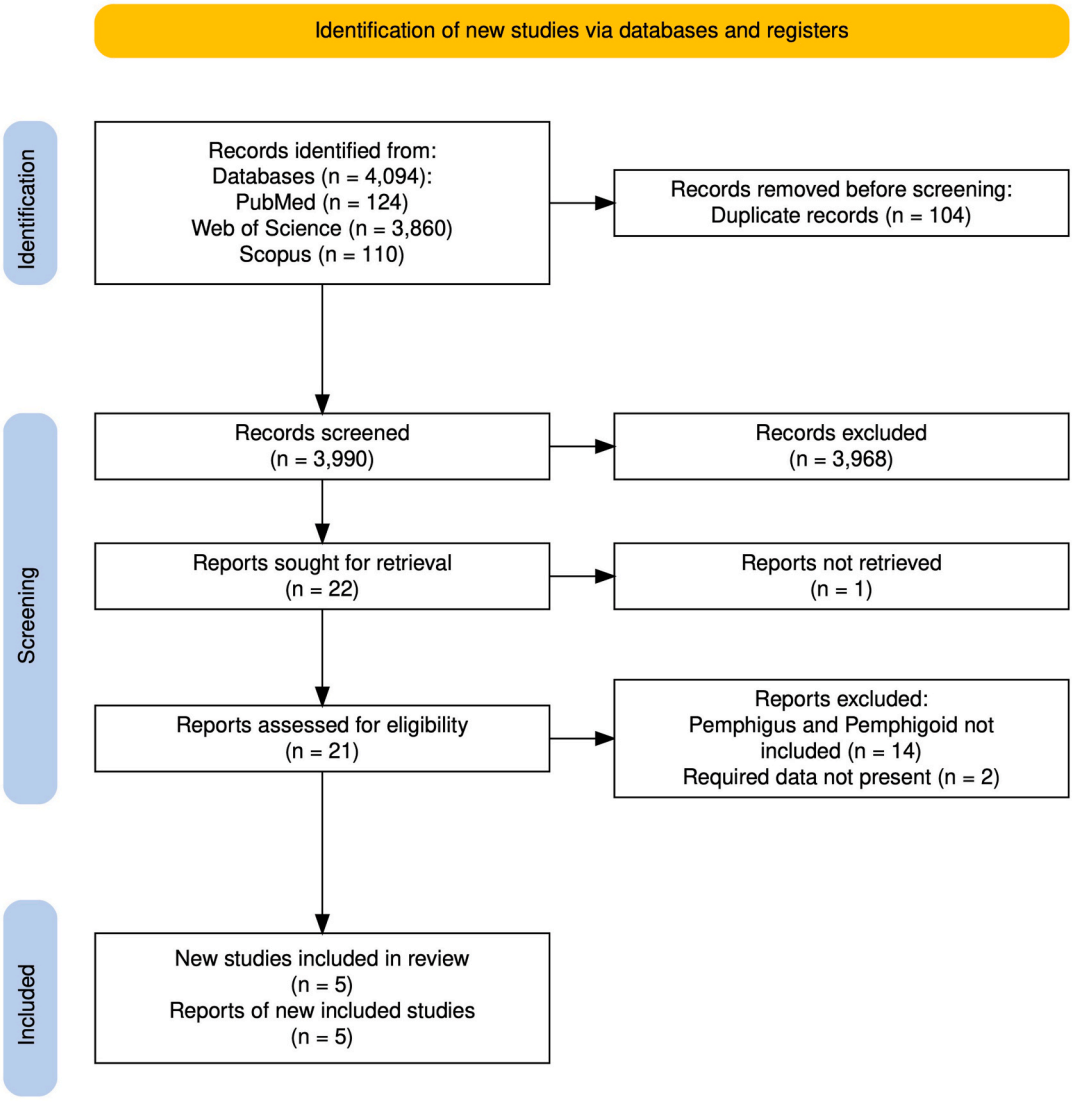

Fig. 1. PRISMA flow diagram illustrating the study selection process for the systematic review.

Table 2
Characteristics of included studies.

Sr No	Article	Country	Population	Disease	Type of CNN/AI	No of images	Competitor	Sensitivity	Specificity	Accuracy
1	Ahmed, M 2023	Bangladesh	web sources	Pemphigus Vulgaris	CNN	3500		0.545	0.702	0.93
2	Capurro N et al., 2024	Italy			six CNNs and the Swin Transformer	436	Pathology reports	0.953	0.975	0.946
3	Huang K et al., 2024	China	Xiangya-Derm database-20 years of data from 15 hospitals,	Pemphigus Pemphigoid	Xy-SkinNet. Xy-SkinNet	463 + 436 (899)	Histopathological examination	0.774	0.986	0.75
4	Noyan MA et al., 2020	Turkey	2260 Tzanck smear images		TzanckNet	2260	Two dermatologists	0.849	0.895	0.877
5	Schielein MC et al. 2022	Germany	university hospital	bullous pemphigoid	VGG-16,25 VGG-19,25 Xception,26 InceptionV3,27 ResNet5028 MobileNetV2.29	881	senior dermatologist, 1 junior dermatologist, 2 epidemiologists			
5	Schielein MC et al. 2022	Germany	university hospital	bullous pemphigoid	VGG- 16,25	881		0.694	0.922	0.803
5	Schielein MC et al. 2022	Germany	university hospital	bullous pemphigoid	VGG-19,25	881		0.694	0.892	0.789
5	Schielein MC et al. 2022	Germany	university hospital	bullous pemphigoid	Xception,26	881		0.856	0.794	0.826
5	Schielein MC et al. 2022	Germany	university hospital	bullous pemphigoid	InceptionV3,27	881		0.892	0.824	0.859
5	Schielein MC et al. 2022	Germany	university hospital	bullous pemphigoid	ResNet5028 and	881		0.811	0.853	0.831
5	Schielein MC et al. 2022	Germany	university hospital	bullous pemphigoid	MobileNetV2.29	881		0.766	0.892	0.826

conditions using AI and machine learning techniques. The studies involved data from various countries, including Bangladesh, Italy, China, Turkey and Germany with sample sizes ranging from 436 to 3500 clinical images.[17,24–27] Population sources included university hospitals, public databases, and pathology reports Schielein et al.[27] analyzed 4051 images, with 881 for bullous pemphigoid, while Huang et al.[25] worked with 899 images for pemphigus vulgaris and pemphigoid. Machine learning models and AI systems used in these studies included neural networks like CNNs, TzanckNet, Xy-SkinNet, and advanced architectures like VGG, Inception, and MobileNet (Table 3). Sensitivity, specificity, and accuracy metrics varied widely, reflecting differences in data sources, competitor benchmarks (e.g., dermatologists or pathology reports), and image datasets, which ranged from a few hundred to several thousand images (Table 2). The findings highlight the evolving potential of AI tools in enhancing diagnostic precision for dermatopathological diseases.

3.3. Risk of bias

The QUADAS-2 tool for risk of bias assessment for the five studies revealed variability in methodological quality across domains (Table 4). Ahmed et al. exhibited high-risk in-patient selection as it used datasets

Table 4
Risk of bias scoring using QUADAS-2.

Domain	Patient Selection	Index Test	Reference Standard	Flow and Timing	Overall
Ahmed et al.	3	2	3	3	3
Capurro et al.	1	1	1	1	1
Huang et al.	1	1	1	1	1
Noyan et al.	1	1	1	1	1
Schielein et al.	1	1	1	1	1

Table 3
True positive, false negative, false positive, true negative number, sensitivity, and specificity in each included study.

Name	Study	TP	FP	FN	TN	Country	AI System	Database
AhmedM et al., 2023	1	745	622	636	1497	Bangladesh	CNN	Web sources
Capurro N et al., 2024	2	47	3	1	42	Italy	CNN	Hospital
Huang K et al., 2024	3	696	203	13	887	China	Xy-SkinNet	Hospital
Noyan MA et al., 2020	4	118	21	23	197	Turkey	TzanckNet	Hospital
Schielein MC et al., 2022 (a)	5	278	122	38	443	Germany	VGG16	Hospital
Schielein MC et al., 2022 (b)	5	278	122	52	429	Germany	VGG19	Hospital
Schielein MC et al., 2022 (C)	5	342	58	99	382	Germany	Xception	Hospital
Schielein MC et al., 2022 (d)	5	357	43	85	396	Germany	InceptionV3	Hospital
Schielein MC et al., 2022 (e)	5	324	76	71	410	Germany	ResNet50	Hospital
Schielein MC et al., 2022 (f)	5	306	94	52	429	Germany	MobileNetV2	Hospital

from web sources, with possible exclusion of "hard-to-diagnose" cases, limiting generalizability.[24] The study also had "some concerns" regarding the index test, due to limited blinding and lack of a gold-standard reference, further impacting the reliability of their diagnostic accuracy findings. Ahmed et al. shows high risk overall due to patient selection issues, unclear index test methods and lack of a gold-standard reference.24 In contrast, Capurro et al., Huang et al., Noyan et al., and Schielein et al. were assessed as low risk across all domains.[17,25–27] These aspects ensured a closer alignment with the QUADAS-2 criteria, minimizing potential bias. Overall, while the findings of most studies can be considered reliable, the limitations identified in Ahmed et al. suggest that its results should be interpreted cautiously within the broader evidence base (Fig. 2).[24]

3.4. Meta-analysis of diagnostic accuracy of artificial intelligence in the diagnosis of pemphigus and pemphigoid

Five included studies in random-effect diagnostic meta-analysis comprising clinical 7976 images of pemphigus and pemphigoid disorders. The pooled diagnostic accuracy of AI-based systems was 0.83 with 95 % CI (0.73–0.90) (Fig. 3) with a pooled sensitivity of 0.86 % (CI: 0.76–0.91) (Fig. 4), pooled specificity of 0.84 % (CI: 0.79–0.87) (Fig. 5) , reflecting its moderate effectiveness in ruling out pemphigus and pemphigoid disorders. Further, the positive likelihood ratio (PLR) was 5.34 (95 % CI: 4.05–7.05), and the negative likelihood ratio (NLR) was 0.16 (95 % CI: 0.09–0.28), with a DOR of 32.53 (95 % CI: 14.70–71.94) (Fig. 6) indicates moderate to strong accuracy. However, forest plot analysis indicated high heterogeneity among studies (I^2 = 90.4 %), which could be due to variability of AI models across different diagnostic settings.

3.5. Heterogeneity

Significant heterogeneity was observed across the included studies (I^2 = 90.4 %). To explore potential sources, we compared dataset characteristics, AI model types, data sources, and validation methods. The analysis suggested that variations in sample size, image modality, and algorithm architecture contributed to the observed heterogeneity.[25–27] For example, studies using hospital-based datasets and clearly defined validation methods reported higher diagnostic reliability than those using web-sourced datasets .[17,24–27] Similarly, advanced CNN architectures (e.g., DenseNet, Swin Transformer, InceptionV3) demonstrated greater diagnostic accuracy compared with more basic CNN frameworks.[24,26,27] These methodological differences explain much of the observed variability and highlight the need for standardized dataset curation and model validation in future research.

4. Discussion

The diagnostic performance of AI in identifying pemphigus and pemphigoid diseases using clinical image datasets was examined in this meta-analysis and systematic review. The meta-analysis showed that AI has moderate diagnostic accuracy in identifying the true diagnosis of pemphigus and pemphigoid diseases while reducing false-positive and false-negative diagnoses. The pooled sensitivity and specificity for pemphigus and pemphigoid lesions indicate that AI can reliably detect characteristic clinical features, achieving diagnostic performance comparable to and in some cases exceeding, that of experienced clinicians. Although the initial search identified 4094 records, only five studies met the strict inclusion criteria.[17,24–27] Most studies included several other dermatopathology or vesiculobullous diseases, while others studies lacked sufficient diagnostic data specifically for pemphigus and pemphigoid. The limited number of studies highlights the growing but still moderate level of evidence emphasizing the need for more focused AI-based diagnostic assessment research in pemphigus and pemphigoid. One potentially relevant article could not be retrieved in full text despite contacting the corresponding author. Its exclusion may have introduced selection bias, although the final synthesis still captures the most representative and data-complete studies available.

The machine learning models can reduce reliance on tissue biopsy examination and supporting non-invasive diagnostics by utilizing the clinical photographs. Amani et al. utilized AI trained on clinical photographs to effectively differentiate pemphigus vulgaris from healthy mucosa.[33] Meanwhile, Ramesh et al. employed deep learning models, Xception, and MobileNetV2, to categories vesiculobullous lesions.[34] Unfortunately, they did not specifically focus on pemphigus or pemphigoid diseases. However, their approach demonstrated the potential of AI in diagnosing these lesions with higher diagnostic accuracy opened up new avenue in diagnostic utility of AI in pemphigus or pemphigoid diseases. Additionally, Achararit et al. used deep CNNs to successfully to identify oral lichen planus, an another type of vesiculobullous disease, highlighting the potential of AI to accurately differentiate between various mucosal diseases.[35] In the context of viral etiologies presenting with vesiculobullous lesions, Rahmani et al. documented the successful use of AI tools for distinguishing monkeypox lesions from other vesicular eruptions based on clinical photography.[36] Also, Hurvitz et al. demonstrated that AI models integrating heterogeneous clinical features significantly improved early recognition in rare bullous diseases.[37] Ehtesham et al. developed a case-based intelligent system for oral lesions, demonstrating high diagnostic agreement with expert clinicians for vesicular and bullous presentations.[30] Wells et al. emphasized the diagnostic utility of AI in dermatopathology for vesicular diseases.[31] Similarly, Dubuc et al. developed a machine learning model that combined clinical, immunological, and histopathological features to accurately diagnose bullous pemphigoid. Although these studies emphasize the use of AI in diagnosis, AI is unlikely to replace these confirmatory tests in the near future but rather serve as an adjunct, potentially guiding the need for more invasive diagnostics.

The diagnosis of pemphigus and pemphigoid lesions with help of AI can present a number of challenges, despite the fact that AI have shown promising findings. The heterogeneity in development of AI application approaches and models including variations in dataset sources and preprocessing techniques which makes direct comparisons across models impossible is one of the major challenges. The absence of standardized reporting frameworks further complicates the generalizability of these findings. Additionally, many authors did not adequately describe the training and validation of their image datasets in sufficient detail, resulting in evaluation bias.[24] Moreover, demographic factors such as gender and ethnic background were absent in the majority of studies, potentially affecting diagnostic accuracy and limiting the applicability of AI models across diverse patient populations.[38,39] To minimize variability and achieve more consistent results, researchers should use a homogeneous population and a standardized image annotation technique. In addition, large multicenter or multi-institutional diverse datasets are crucial for enhancing the applicability, reliability, and generalizability of diagnostic models across different populations

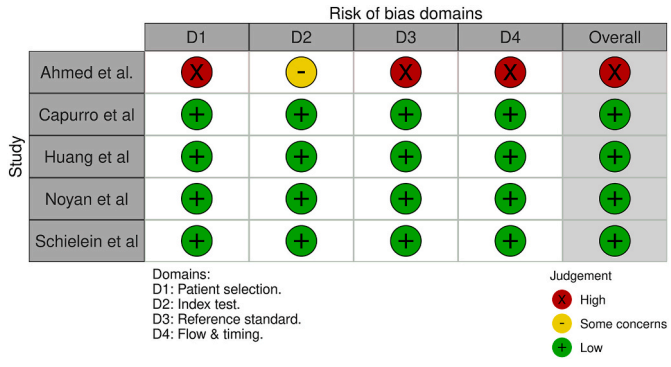

Fig. 2. Risk of bias assessment using the QUADAS-2 tool for included studies.

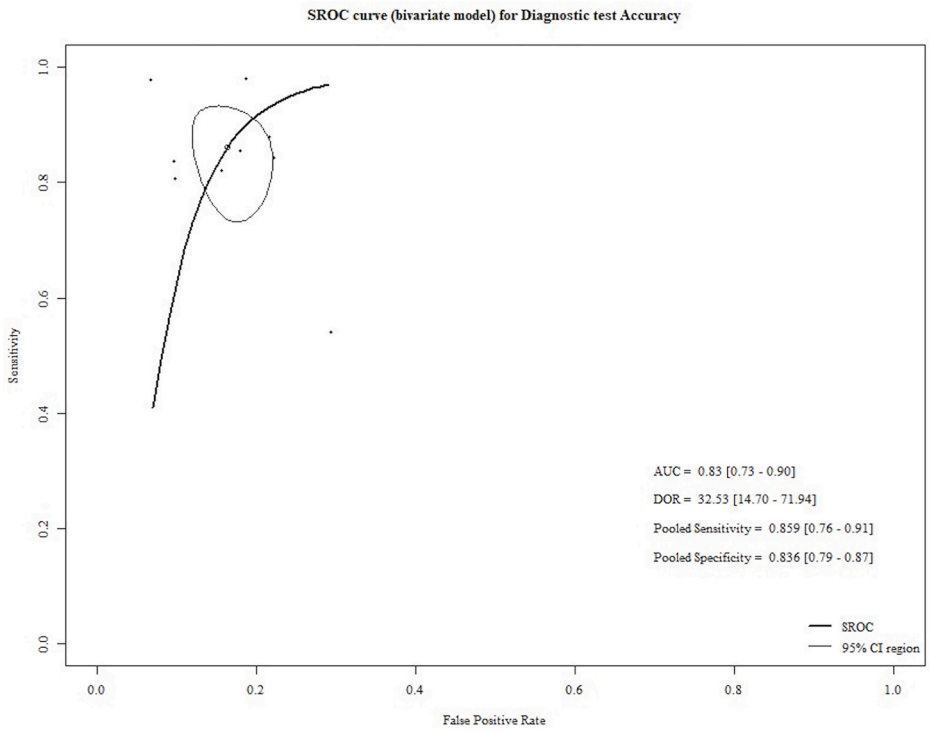

Fig. 3. Summary Receiver Operating Characteristic (SROC) curve showing the diagnostic accuracy of AI in detecting pemphigus and pemphigoid disorders.

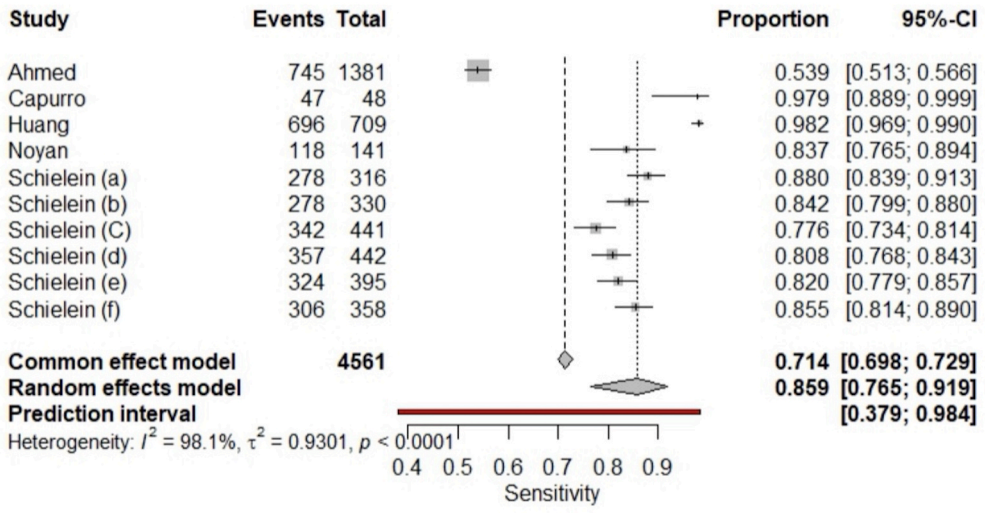

Fig. 4. Forest plot of pooled sensitivity estimates for AI-based diagnostic models.

and race.

A key limitation of this review is that most CNN architectures were represented by only a single study (e.g., TzanckNet, Xy-SkinNet, VGG variants, InceptionV3, ResNet50, MobileNetV2), with dataset sizes ranging from 436 to 2260 images.[17,24–27] This has hindered us to conduct sensitivity analyses or make robust comparisons between models. Further, the substantial heterogeneity observed (I^2 = 98.1 %) was attributable to differences in dataset size, imaging modality (cytology, histopathology, clinical images), and validation method rather than inherent performance disparities between these CNN models.[17,25,26] These constraints highlight the urgent need for standardized, multi-center datasets and harmonized validation frameworks to enable reliable benchmarking of AI models in dermatologic and mucosal disease diagnosis.

Bringing advanced technology into medical practice comes with its share of challenges. A key hurdle is the need for extensive, well-curated image datasets that can help build accurate diagnostic tools.[40] Simultaneously, we must handle ethical concerns with care, including patient privacy, data security, and informed consent. Further, there is a need for legal and regulatory bodies so that accountability remains clear in AI based medical diagnosis.[41] Looking into the future, clinicians and technology developers should focus on large-scale, multicenter studies to ensure diagnostic models work across different populations. Integrating different data sets such as clinical images, serological tests, and histological findings could make diagnoses more precise. Expanding remote diagnostic tools could significantly improve patient outcomes, particularly in underserved regions where early detection and timely treatment can make a significant difference.

5. Conclusion

The systematic review revealed that AI models are effective in

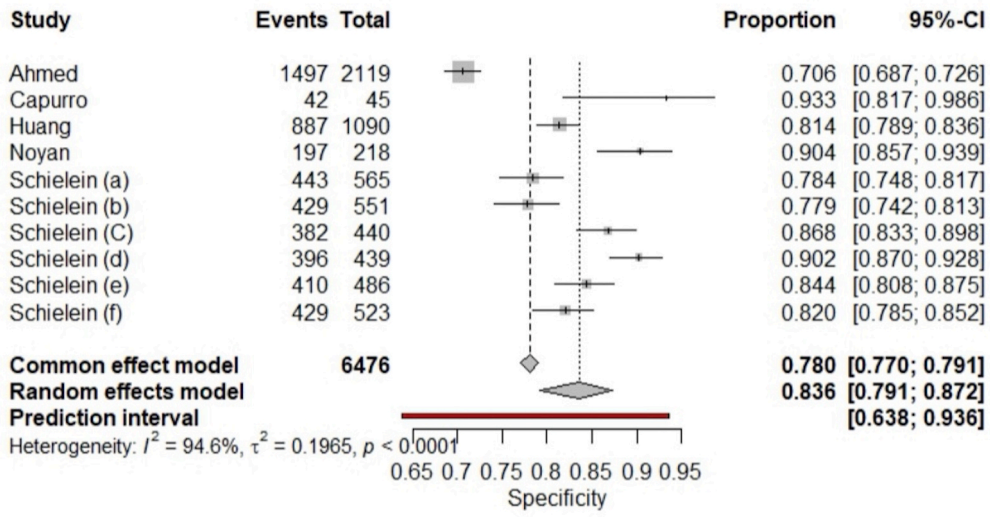

Fig. 5. Forest plot of pooled specificity estimates for AI-based diagnostic models.

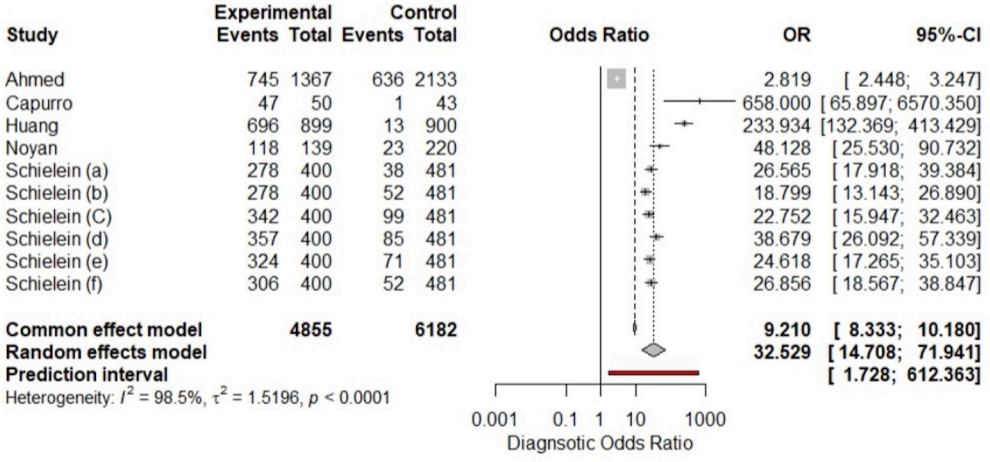

Fig. 6. Diagnostic Odds Ratio (DOR) analysis comparing AI-based diagnosis with conventional methods.

diagnosing Pemphigus and Pemphigoid diseases by improving speed, accuracy, and efficiency. Despite the promising results, additional research is needed to reduce the existing gaps. For AI to be a recognized diagnostic tool in the field of dermatopathology, there must be integration between AI developers, dermatologists, and pathologists to fully explore the capacity of AI in this field, for both clinical and resource-limited settings.

Informed patient consent

Not applicable as it is a Systematic Review.

Authors contributions

Pravallika Kakada, Monal Yuwanati, and Pratibha Ramani contributed equally to this work. Their contributions include Conceptualization, Methodology, Data Curation, Formal Analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, and Project Administration.

All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work.

Clinical trial registration details

Not applicable as it is a Systematic Review.

Statement of clinical relevance

This research highlights the potential of AI, particularly convolutional neural networks, to assist clinicians in the early and accurate diagnosis of pemphigus and pemphigoid disorders based on clinical images. By reducing diagnostic delays and minimizing reliance on invasive procedures, AI tools can enhance decision-making in dermatological and oral pathology practice. Integrating such models could support resource-limited settings and improve patient outcomes through timely intervention.

Funding

Authors have not received any financial support or funding for this systematic review and meta-analysis.

Ethical clearance declaration

This systematic review and meta-analysis does not involve

recruitment of patients and the collection or analysis of individual patient data. Instead, it has utilized data from published research studies. All data sources were appropriately cited. As this is a secondary analysis of published literature, ethical approval was not required.

Patient's/guardian's consent declaration

Consent was not required, as this study is a systematic review and meta-analysis based solely on previously published research studies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Nil.

References

1. Simionescu O, Tudorache SI. Autoimmune pemphigus: difficulties in diagnosis and the molecular mechanisms underlying the disease. *Front Immunol*. 2025;16, 1481093. https://doi.org/10.3389/fimmu.2025.1481093.
2. Rosi-Schumacher M, Baker J, Waris J, Seiffert-Sinha K, Sinha AA. Worldwide epidemiologic factors in Pemphigus vulgaris and bullous pemphigoid. *Front Immunol*. 2023;14, 1159351. https://doi.org/10.3389/fimmu.2023.1159351.
3. Amber KT, Murrell DF, Schmidt E, Joly P, Borradori L. Autoimmune subepidermal bullous diseases of the skin and mucosae: clinical features, diagnosis, and management. *Clin Rev Allergy Immunol*. 2017;54(1):26–51. https://doi.org/10.1007/s12016-017-8633-4.
4. Timóteo RP, Pessoa-Gonçalves YM, do Carmo Neto JR, Rodrigues WF, da Silva MV, Oliveira CJF. A global view of pemphigus: geographical variations. *Clin Rev Allergy Immunol*. 2024;66(1):14–29. https://doi.org/10.1007/s12016-024-08980-w.
5. Costan VV, Popa C, Hâncu MF, Porumb-Andrese E, Toader MP. Comprehensive review on the pathophysiology, clinical variants and management of pemphigus (Review) *Exp Ther Med*. 2021;22(5):1335. https://doi.org/10.3892/etm.2021.10770.
6. Daniel BS, Murrell DF. Review of autoimmune blistering diseases: the pemphigoid diseases. *J Eur Acad Dermatol Venereol*. 2019;33(9):1685–1694. https://doi.org/10.1111/jdv.15679.
7. Hammers CM, Stanley JR. Mechanisms of disease: pemphigus and bullous pemphigoid. *Annu Rev Pathol*. 2016;11:175–197. https://doi.org/10.1146/annurev-pathol-012615-044313.
8. Di Lernia V, Casanova DM, Goldust M, Ricci C. Pemphigus vulgaris and bullous pemphigoid: update on diagnosis and treatment. *Dermatol Pract Concept*. 2020;10(3). https://doi.org/10.5826/dpc.1003a50. e2020050-e2020050.
9. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). *Comput Methods Progr Biomed*. 2022;226, 107161. https://doi.org/10.1016/j.cmpb.2022.107161.
10. Hammoudi K, Benhabiles H, Melkemi M, et al. Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. *J Med Syst*. 2021;45(7):75. https://doi.org/10.1007/s10916-021-01745-4.
11. Kodera S, Akazawa H, Morita H, Komuro I. Prospects for cardiovascular medicine using artificial intelligence. *J Cardiol*. 2022;79(3):319–325. https://doi.org/10.1016/j.jjcc.2021.10.016.
12. Balasamy S, Somasundaram J, Sundramoorthy AK. MRI-Based deep learning and radiomics for occult cervical lymph node metastasis (OCLNM) prediction. *Oral Oncol*. 2024;159, 107019. https://doi.org/10.1016/j.oraloncology.2024.107019.
13. Pereira-Prado V, Martins-Silveira F, Sicco E, et al. Artificial intelligence for image analysis in oral squamous cell carcinoma: a review. *Diagnostics*. 2023;13(14):2416. https://doi.org/10.3390/diagnostics13142416.
14. Murugan R. Leveraging optogenetics and machine learning for precision dentistry. *Br Dent J*. 2025;238(4). https://doi.org/10.1038/s41415-025-8447-3, 210-210.
15. Maheswari TNU, Bohra A, Sharuniveda S. Pioneering a new frontier: artificial intelligence (AI)-Driven lip print pattern Analysis—A systematic review. *J Indian Acad Oral Med Radiol*. 2024;36(3):206–212. https://doi.org/10.4103/jiaomr.jiaomr_132_24.
16. Manuelyan K, Dragolov M, Drenovska K, Shahid M, Vassileva S. Artificial intelligence in autoimmune bullous dermatoses. *Clin Dermatol*. 2024;42(5):426–433. https://doi.org/10.1016/j.clindermatol.2024.06.008.
17. Capurro N, Pastore VP, Touijer L, et al. A deep learning approach to direct immunofluorescence pattern recognition in autoimmune bullous diseases. *Br J Dermatol*. 2024;191(2):261–266. https://doi.org/10.1093/bjd/ljae142.
18. Hocke J, Krauth J, Krause C, et al. Computer-aided classification of indirect immunofluorescence patterns on esophagus and split skin for the detection of autoimmune dermatoses. *Front Immunol*. 2023;14. https://doi.org/10.3389/fimmu.2023.1111172.
19. Arockiam S, Raj B, Prabha N, Dharani M. Diagnostic utility of immunofluorescence in oral lesions: a systematic review. *J Oral Maxillofac Res*. 2024;15(3). https://doi.org/10.5037/jomr.2024.15302.
20. Doolan BJ, Thomas BR. Bursting the bubble on diagnostics: artificial intelligence in autoimmune bullous disease. *Br J Dermatol*. 2024;191(2):160–161. https://doi.org/10.1093/bjd/ljae197.
21. Ehtesham H, Safdari R, Mansourian A, Tahmasebian S, Mohammadzadeh N, Pourshahidi S. Developing a new intelligent system for the diagnosis of oral medicine with case-based reasoning approach. *Oral Dis*. 2019;25(6):1555–1563. https://doi.org/10.1111/odi.13108.
22. Wells A, Patel S, Lee JB, Motaparthi K. Artificial intelligence in dermatopathology: diagnosis, education, and research. *J Cutan Pathol*. 2021;48(8):1061–1068. https://doi.org/10.1111/cup.13954.
23. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*. 2022;18(2). https://doi.org/10.1002/cl2.1230.
24. Ahmed M, Islam SB, Alif AU, Islam M, Saima SM. AI empowered diagnosis of pemphigus: a machine learning approach for automated skin lesion detection. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*. 2023;13(4):21–26. https://doi.org/10.35784/iapgos.5366.
25. Huang K, Jiang Z, Li Y, et al. The classification of six common skin diseases based on Xiangya-Derm: development of a Chinese database for artificial intelligence. *J Med Internet Res*. 2021;23(9). https://doi.org/10.2196/26025. e26025-e26025.
26. Noyan MA, Durdu M, Eskiocak AH. TzanckNet: a convolutional neural network to identify cells in the cytology of erosive-vesiculobullous diseases. *Sci Rep*. 2020;10(1), 18314. https://doi.org/10.1038/s41598-020-75546-z.
27. Schielein MC, Christl J, Sitaru S, et al. Outlier detection in dermatology: performance of different convolutional neural networks for binary classification of inflammatory skin diseases. *J Eur Acad Dermatol Venereol*. 2023;37(5):1071–1079. https://doi.org/10.1111/jdv.18853.
28. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br Med J*. 2021;372, n71. https://doi.org/10.1136/bmj.n71.
29. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—A web and Mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210. https://doi.org/10.1186/s13643-016-0384-4.
30. Whiting PF, Rutjes AWS, Westwood ME, et al. Quadas-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.
31. Plana MN, Arevalo-Rodriguez I, Fernández-García S, et al. Meta-DiSc 2.0: a web application for meta-analysis of diagnostic test accuracy data. *BMC Med Res Methodol*. 2022;22(1):306. https://doi.org/10.1186/s12874-022-01788-2.
32. Dubuc A, Zitouni A, Thomas C, et al. Improvement of Mucosal Lesion Diagnosis with Machine Learning Based on Medical and Semiological Data: An Observational Study. *J Clin Med*. 2022;11(21):6596. https://doi.org/10.3390/jcm11216596.
33. Amani T, Priya AH, Sowmya S. Artificial intelligence (AI) in the identification of Pemphigus vulgaris as compared with healthy mucosa using clinical images: a retrospective study. *J Indian Acad Oral Med Radiol*. 2025;37(1):50–55. https://doi.org/10.4103/jiaomr.jiaomr_322_24.
34. Ramesh E, Ganesan A, Lakshmi KC, Natarajan PM. Artificial Intelligence—Based diagnosis of oral leukoplakia using deep convolutional neural networks Xception and MobileNet-v2. *Front Oral Health*. 2025;6. https://doi.org/10.3389/froh.2025.1414524.
35. Achararit P, Manaspon C, Jongwannasiri C, Phattarataratip E, Osathanon T, Sappayatosok K. Artificial intelligence-based diagnosis of oral Lichen Planus using deep convolutional neural networks. *Eur J Dermatol*. 2023;17(4):1275–1282. https://doi.org/10.1055/s-0042-1760300.
36. Rahmani E, Bayat Z, Farrokhi M, et al. Monkeypox: a comprehensive review of virology, epidemiology, transmission, diagnosis, prevention, treatment, and artificial intelligence applications. *Arch Acad Emerg Med*. 2024;12(1):e70. https://doi.org/10.22037/aaem.v12i1.2491.
37. Hurvitz N, Azmanov H, Kesler A, Ilan Y. Establishing a second-generation artificial intelligence-based system for improving diagnosis, treatment, and monitoring of patients with rare diseases. *Eur J Hum Genet*. 2021;29(10):1485–1490. https://doi.org/10.1038/s41431-021-00928-4.
38. Cirillo D, Catuara-Solarz S, Morey C, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digit Med*. 2020;3:81. https://doi.org/10.1038/s41746-020-0288-5.
39. Alshami ML, Aswad F, Abdullah B. A clinical and demographic analysis of oral pemphigus vulgaris: a retrospective cross-sectional study from 2001 to 2021. *Health Sci Rep*. 2022;5(5). https://doi.org/10.1002/hsr2.832. e832-e832.
40. Breger A, Biguri A, Landman MS, et al. A study of why we need to reassess full reference image quality assessment with medical images. *Journal of Imaging Informatics in Medicine*. March 24, 2025. https://doi.org/10.1007/s10278-025-01462-1.
41. Čartolovni A, Tomičić A, Lazić Mosler E. Ethical, legal, and social considerations of AI-based medical decision-support tools: a scoping review. *Int J Med Inf*. 2022;161, 104738. https://doi.org/10.1016/j.ijmedinf.2022.104738.