AUCTORES
Globalize your Research

Journal of Dermatology and Dermatitis
Mohammed S Al Abadie *

Open Access

Research Article

Patch Testing in the Age of Artificial Intelligence (AI): Opportunities and Risks

Shaymaa Al-tharwane [1], Mohammed S Al Abadie [2]*

[1]University Hospitals of Leicester NHS Foundation Trust.

[2]University of Wolverhampton, North Cumbria Integrated Care NHS Foundation Trust, United Kingdom.

***Corresponding Author:** Mohammed S Al Abadie, FRCP, University of Wolverhampton, North Cumbria Integrated Care NHS Foundation Trust, United Kingdom.

Received date: December 30, 2025; **Accepted date:** February 03, 2026; **Published date:** March 28, 2026

Citation: Shaymaa Al-tharwane, Mohammed S Al Abadie, (2026), Patch Testing in the Age of Artificial Intelligence (AI): Opportunities and Risks, *Dermatology and Dermatitis*, 13(1); **DOI:10.31579/2578-8949/204**

Abstract

Background: Patch testing is constrained by subjectivity, interobserver variability, and structural barriers to access. Artificial intelligence (AI), particularly machine learning and deep learning, offers a potential solution to standardise interpretation and expand diagnostic capacity. However, current gaps in validation and standardisation limit these systems to exploratory roles rather than deployable clinical instruments.

Aim: This review aims to critically evaluate the application of AI to patch testing in contact dermatitis, focusing on diagnostic performance, methodological strengths, and current limitations regarding clinical integration.

Materials and Methods: Following PRISMA 2020 guidelines, MEDLINE (PubMed) was searched for records from inception to December 2025. Eligible studies included original research involving human participants that utilised recognised AI techniques, such as convolutional neural networks (CNNs) or random forests, for patch testing or skin sensitisation assessment and reported quantifiable performance metrics.

Results: Eleven studies met the inclusion criteria from 123 records screened. Methodologies spanned automated image interpretation, molecular analysis, and sensitisation risk modelling. CNNs achieved high classification accuracy for patch test reactions, ranging from 90.1% to 99.5%. Molecular models discriminated allergic from irritant contact dermatitis with accuracies between 86% and 100%. Conversely, clinical risk prediction models showed moderate discrimination (mean AUC 0.69). Critical limitations included small, single-centre datasets, class imbalance, and inconsistent reporting of skin phototypes.

Conclusions: AI demonstrates credible technical capability in reaction detection and molecular discrimination but is currently constrained by dataset heterogeneity and a lack of external validation. Consequently, AI currently functions best as a decision-support system rather than an autonomous diagnostic tool.

Keywords: artificial intelligence; convolutional neural networks; sensitisation risk modelling

Introduction

Allergic contact dermatitis occupies an important space in dermatology: common, clinically impactful, and, despite decades of methodological refinement, still constrained by the subjectivity and labour intensity of patch testing. Interobserver variability persists even among experienced clinicians, with discrepancies reported in both reaction grading and final interpretation [1]. Patch testing also demands multiple visits and consistent expertise, creating structural barriers to timely diagnosis and equitable access. These limitations have made the field increasingly receptive to computational support systems capable of standardising interpretation, reducing human error, and expanding diagnostic capacity. Artificial intelligence, particularly in the form of machine learning and deep learning, has rapidly become the most prominent candidate for such augmentation.

Early dermatology applications of AI focused on skin cancer classification, but the visual and pattern-recognition demands of patch testing make it an equally natural target. Recent studies confirm that convolutional neural networks (CNNs) can classify patch test reactions with high accuracy: several image-based models reach performance levels between 90% and 99.5%, even when trained on heterogeneous clinical photographs [1]. Chan et al. demonstrated near-perfect accuracy using an Xception-based architecture [2], while Hall et al. showed that

CNNs can outperform expert dermatologists in sensitivity, albeit at a modest cost to specificity [3]. These findings suggest that AI could meaningfully reduce subjectivity in patch test readings, standardise assessments across centres, and potentially support remote or at-home evaluation workflows.

Despite accelerating progress, current evidence is limited by small, single-centre datasets, inconsistent reporting of skin phototypes, and the frequent use of internally recycled training and test sets, which risks overfitting. Most studies lack external validation and do not address crucial challenges such as harmonised imaging protocols, algorithmic transparency, or real-world integration into clinical workflows. These gaps underscore why AI systems currently function as exploratory tools rather than deployable clinical instruments. Nevertheless, the trajectory of research points toward a future in which AI enriches diagnostic precision, reduces variability, and supports more scalable patch testing, provided that methodological standardisation, representative datasets, and rigorous prospective validation are achieved.

This article aims to critically evaluate how artificial intelligence is being applied to patch testing in contact dermatitis, with emphasis on methodological strengths, diagnostic performance, and current limitations. By integrating the available evidence, the review will establish a balanced, evidence-based perspective on how AI might realistically augment, not replace, clinical expertise in the diagnosis and management of allergic and irritant contact dermatitis.

Materials and Methods

Search strategy and study selection

Following PRISMA 2020, we searched MEDLINE (PubMed) for records from inception to December 2025 using controlled vocabulary and free-text terms such as “patch testing”, “allergic contact dermatitis”, “artificial intelligence” and “machine learning” (Supplementary Data 1). A PRISMA flow diagram (Supplementary Figure 1) summarises the selection.

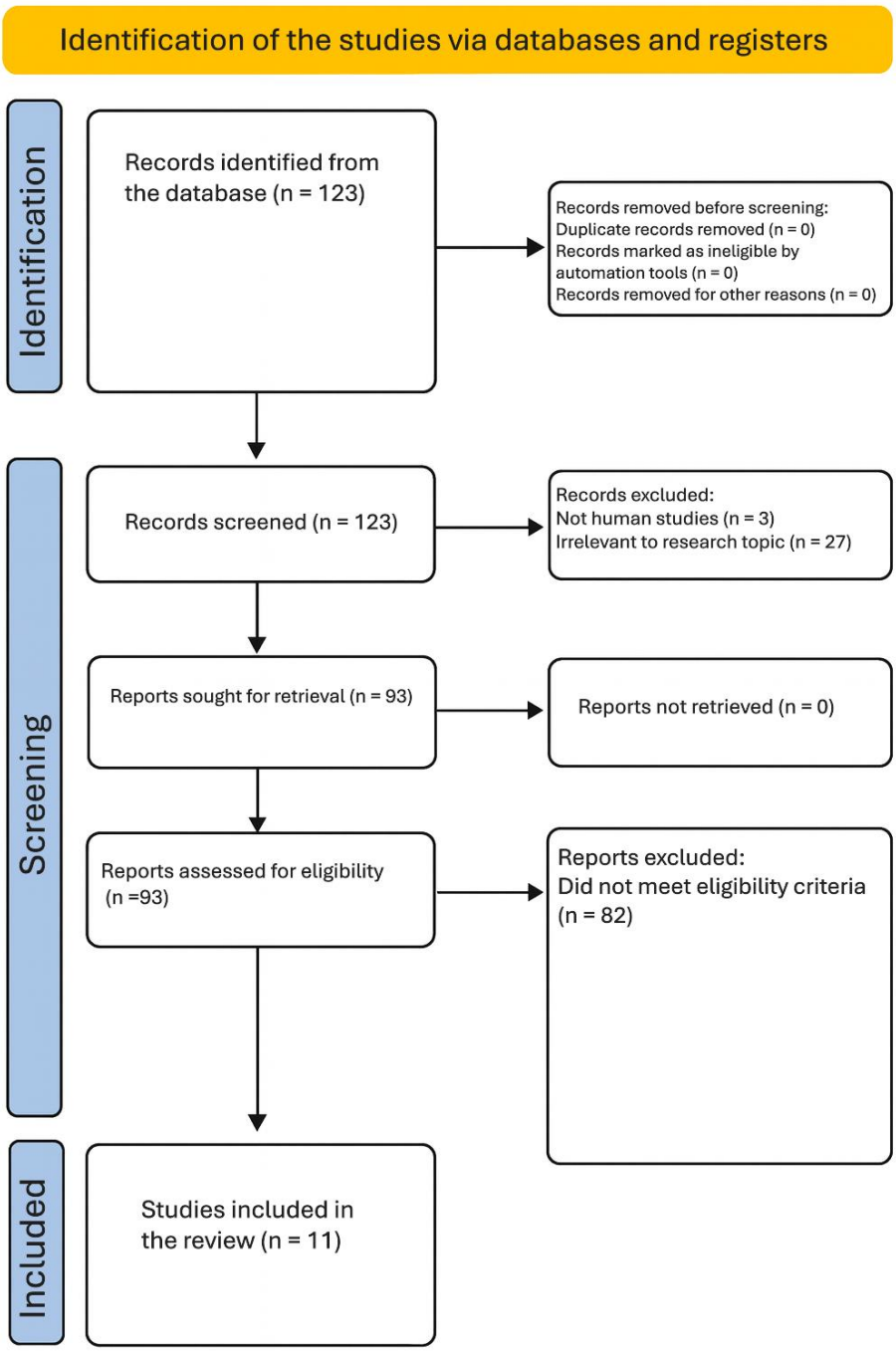
Category	Search Terms
Contact Dermatitis	("Dermatitis, Contact"[Mesh] OR "Dermatitis, Allergic Contact"[Mesh] OR "Dermatitis, Photoallergic"[Mesh] OR "Dermatitis, Occupational"[Mesh] OR contact dermatitis OR Allergic Contact Dermatitis OR Allergic Eczematous Dermatitis OR allergic dermatitis OR Photocontact Dermatitis OR Photosensitive Dermatitis OR photoallergic OR photoallergy OR "Contact Allergy" OR "Contact Sensitivity" OR "Contact Hypersensitivity" OR Dermatitis Allergic Occupational OR Industrial Dermatosis OR "occupational dermatitis" OR dermatitis)
Machine learning	("Artificial Intelligence"[Mesh] OR "Machine Learning"[Mesh] OR "Deep Learning"[Mesh] OR "Supervised Machine Learning"[Mesh] OR "Unsupervised Machine Learning"[Mesh] OR "Support Vector Machine"[Mesh] OR "Neural Networks, Computer"[Mesh] OR "Natural Language Processing"[Mesh] OR "Computer Heuristics"[Mesh] OR "Expert Systems"[Mesh] OR "Fuzzy Logic"[Mesh] OR "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Networks" OR "Convolutional Neural Networks" OR "Supervised Learning" OR "Unsupervised Learning" OR "Natural Language Processing" OR "Computer Vision" OR "Image Analysis" OR "Image Recognition" OR "Pattern Recognition" OR "Data Mining" OR "Predictive Modeling" OR "Automated Diagnosis" OR "Algorithmic Diagnosis" OR "Decision Support Systems" OR "Diagnostic Decision Support" OR "Computer-Aided Diagnosis" OR "Machine Intelligence" OR "Cognitive Computing" OR "Health Informatics" OR "Ensemble Learning" OR "Augmented Intelligence" OR "Large Language Models" OR "Foundation Models" OR "reinforcement learning" OR "Generative Adversarial Network")
Patch Test	("Patch Tests"[Mesh] OR "Patch Test Allergens" OR "Drug Patch Test" OR "patch test" OR "patch testing" OR "patch tests" OR epicutaneous test OR Contact Sensitization Testing)
Skin	(epidermis OR skin OR cutaneous OR epicutaneous OR "skin barrier" OR skin inflammation OR skin sensitization)

Inclusion and exclusion criteria:

Studies were eligible for inclusion if they presented original research applying artificial intelligence methods to contact dermatitis or patch testing. Eligible work had to involve human participants or human-derived data, including clinical patch test images, demographic and clinical datasets, or molecular profiles such as transcriptomics. Only studies that implemented recognised AI techniques—such as machine learning or deep learning models, including convolutional neural networks, random forests, gradient boosting, support vector machines, or ensemble approaches—were included. To ensure evaluability, studies were required to report quantifiable performance metrics, such as accuracy, sensitivity, specificity, area under the curve, or biomarker classification performance. Full-text articles published in English and providing sufficient methodological detail to appraise algorithm development and validation were considered.

Studies were excluded if they did not apply AI methodologies, relied solely on traditional statistical analyses, or failed to report measurable algorithm outputs. Non-original research, including reviews, commentaries, editorials, conference abstracts without primary data, and expert opinion pieces, was excluded. Work based on animal models, in vitro systems, or synthetic datasets without clinical linkage was not considered. Studies were also excluded if they lacked relevance to patch testing or contact dermatitis despite incorporating AI, if they did not involve human-derived data, or if the full text was unavailable in English.

Supplementary Figure 1



Results

From the 123 records screened, 11 studies met the inclusion criteria and reported the application of AI methods to patch testing and skin sensitisation assessment (Table 1). The included studies spanned multiple methodological domains, including automated image-based interpretation of patch test reactions, molecular and transcriptomic analyses, and predictive models for sensitisation risk.

The studies utilised AI and machine learning (ML) to address several specific stages of the patch testing workflow, spanning risk stratification, reaction detection, and molecular discrimination.

A primary objective was to predict the risk of a clinically relevant positive patch test result using established clinical and demographic parameters. Cunningham et al. noted that this quantification of risk is intended to assist clinical decision-making and inform patient discussions regarding the utility of patch testing.[10] Adler et al. pursued another key goal: investigating factors associated with polysensitization (PS)—defined as three or more positive reactions—by identifying combinations of positive reactions to pairs of allergens.[9] Kyritsi et al. also aimed to investigate contact allergy patterns related to specific agents, such as preservatives, by uncovering the linkage among hypersensitivity to preservatives, occupational profiles, and patients' clinical and demographic characteristics [18,19].

Chan et al., Ravishankar et al., and Hall et al. aimed to develop deep learning models for the automated detection and classification of patch test reactions from photographs. [2,3,12] This task focused on discriminating positive reactions from non-reactions (negative or irritant). Vezakis et al. set specific objectives, including evaluating the impact of different image modalities (e.g., Colour, Redness) and novel pre-

processing techniques (e.g., context-retaining masks) on classification accuracy [7].

Fortino et al. and Lefevre et al. focused on distinguishing the clinical phenotype of irritant and allergic contact dermatitis (ACD/ICD) using integrative transcriptomic analysis and machine learning to decipher unique molecular signatures and find robust biomarker sets. [4,5]

The utilised AI techniques included traditional supervised and unsupervised machine learning models, deep learning architectures, and hybrid ensemble approaches, depending on the data modality.

Cunningham et al. utilised supervised models, including gradient boosting, logistic regression, random forest, and AdaBoost, for clinical risk prediction from demographic and patch-test variables [10]. Cunningham et al. found that gradient boosting outperformed linear methods by capturing non-linear feature interactions [10]. Kyritsi et al. employed unsupervised dimensionality reduction techniques, specifically multiple correspondence analysis and categorical principal components analysis, to explore latent sensitisation patterns within categorical patient registries.[18] For biomarker discovery, Fortino et al. and Lefevre et al. leveraged hybrid approaches, combining genetic algorithms with random forest classifiers to optimise minimal gene sets for discriminating baseline skin from allergic and irritant contact dermatitis, alongside complementary feature selection using Boruta.[4,5] Panagiotidis et al. applied classical classifiers including random forest, support vector machines, k-nearest neighbours, decision trees, and XGBoost to engineered feature representations for radiomics-based image analysis [21].

Chan et al., Ravishankar et al., Hall et al., and Vezakis et al. established Convolutional Neural Networks (CNNs) as the standard architecture for image analysis [2,3,7,12], Chan et al. used the Xception CNN architecture, pretrained on ImageNet, via transfer learning.[2] Vezakis et al. and Panagiotidis et al. selected the EfficientNetB0 architecture.[7,21] Vezakis et al. specifically found that this architecture achieved better outcomes compared to ResNet50 and MobileNetV2 in a study comparing image modalities [7]. Hall et al. and Ravishankar et al. also designed custom 5-layer fully CNNs for classification [3,12].

Panagiotidis et al. evaluated a Fusion algorithm that combined features extracted by deep learning (EfficientNetB0 CNN) and traditional radiomics features (used in Random Forest) [21]. Panagiotidis et al. intended this combined approach to benefit from the information captured by both methods [21].

The studies analysed employed three principal data modalities: large clinical registries for risk modelling, biopsy-derived molecular datasets for mechanistic discrimination, and image repositories acquired using both standard photography and specialised multi-modal imaging systems.

Population-level risk prediction and sensitisation analyses relied on extensive retrospective registries. Cunningham et al. examined 49,217 patients from a single UK centre over four decades using an extended European baseline series. [10], Adler et al. analysed 105,325 patients from the multicentre IVDK network across Central Europe, focusing on a harmonised subset of baseline allergens [9], Kyritsi et al. performed more targeted pattern analysis in a Greek cohort of 800 monosensitised patients using a restricted preservative panel [18].

Molecular discrimination between allergic and irritant contact dermatitis was investigated using small but highly characterised biopsy datasets. Fortino et al. analysed 89 patch-test biopsies from Finnish clinical cohorts using selected allergens and irritants [4]. Lefevre et al. evaluated transcriptomic profiles from 47 patients across French and Belgian centres, incorporating a broader range of allergen and irritant exposures [5].

Image-based classification studies varied widely in scale and acquisition method. Chan et al. analysed 3,695 colour images from 77 patients. [2] Hall et al. assembled a larger single-centre dataset comprising over 28,000 test sites from 201 patients. [3] Ravishankar et al. evaluated 13,622 images from 125 patients. [12] Vezakis et al. and Panagiotidis et al. employed multi-modal imaging with an Antera 3D® system to capture 1,190 annotated reactions across 200 patients, enabling analysis of colour, redness, texture, and volumetric features [7,21].

Performance Metrics Reported

Cunningham et al. reported that clinical risk prediction models achieved moderate discrimination, with gradient boosting yielding a mean AUC of 0.69 [10], Adler et al. observed higher discrimination in polysensitisation modelling, reporting AUC values exceeding 0.90 for selected allergen pairs [9]. Binary image classification models consistently reported high accuracy, ranging from 90.1% (Ravishankar et al.) to 99.5% (Chan et al.), with AUCs between 0.885 (Hall et al.) and 0.940 (Ravishankar et al.).[2,3,12], Hall et al. noted that sensitivity–specificity trade-offs reflected underlying class imbalance, as their model achieved 70.1% sensitivity but 91.7% specificity [3], Vezakis et al. demonstrated incremental gains using multi-modal image analysis, with redness-based approaches achieving mean accuracy up to 0.91 [7], Panagiotidis et al. found that feature-fusion approaches achieved accuracies up to 0.85 and improved F1 scores [21]. The highest discriminatory performance was observed in molecular models: Fortino et al. reported random forest–based classifiers achieving accuracy ranges of 86–94% and F1 scores averaging 94% for allergic and 92% for irritant reactions [4], Lefevre et al. reported accuracy ranges of 90–100%.[5]

	Author	Study aim	AI Algorithm Used	Sample size	Materials used	Accuracy
1	Adler et al, 2017[9]	To identify if certain pairs of positive reactions to allergens may be associated with polysensitization	Logistic Regression (LR). • Random Forest (RF) – secondary supervised ML model.	105325 patients	24 allergens	LR: AUC: >0.90
2	Fortino et al, 2020[4]	To identify and validate biomarkers to distinguish allergic and irritant contact dermatitis in human	Genetic Algorithm for biomarker selection + RF classifier optimisation (GARBO)	85 patients	4 allergens	Accuracy: 86%- 94%, F1 score: 94% for allergic contact dermatitis, 92% for irritant contact dermatitis

3	Cunningham et al, 2021[10]	To compare the predictive accuracy of logistic regression with more sophisticated machine learning approaches such as gradient boosting in predicting patch testing results	Random Forest (RF), AdaBoost, logistic regression (LR), Gradient boosting,	42434 patients	36 allergens	Gradient boosting: AUC mean: 0.69 (SD 0.06). RF: AUC mean: 0.60 (SD 0.052). AdaBoost: AUC mean: 0.58 (SD 0.048). LR: AUC mean: 0.65 (SD 0.068).
4	Chan et al, 2021[2]	To develop a machine learning approach for accurate classification of patch-test photographs	Convolutional neural network (CNN)	77 patients	80 allergens	AUC: 0.915, accuracy: 99.5%, F1 score: 0.89
5	Lefevre et al, 2021[5]	To characterize the molecular signatures of chemical-induced skin inflammation Through comprehensive transcriptomic analysis	Random forest (RF),Boruta,	47 patients	6 allergens, 3 irritants	RF: accuracy: 90%-100%
6	Vezakis et al, 2023[7]	To investigate the feasibility of using a deep learning classifier for automating the identification of allergens causing ACD\	CNN	240 patients	30 allergens	Accuracy:90%, F1 score: 0.83, accuracy: 90%, specificity: 95%, recall: 79%, precision: 87%
7	Kyritsi et al, 2023[18]	To investigate the patterns of contact sensitization	Multiple Correspondence Analysis MCA	240 patients	3 allergens	Not reported
8	Hall et al, 2024[3]	To develop a deep learning algorithm for the analysis of patch testing	CNN	201 patients	80 allergens	AUC: 0.885, accuracy: 90.9%, sensitivity: 70.1%, specificity: 91.7%, F1 score: 37.1
9	Ravishankar et al, 2024[12]	To evaluate the use of convolutional neural networks to determine presence of patch test reactions	Convolutional neural network (CNN)	125 patients	Not reported	Area under the curve (AUC): 0.940, accuracy: 90.1%, sensitivity: 86.0%, Specificity: 90.2%
10	Kyritsi et al, 2024[19]	To investigate the patterns of contact allergy	Multiple correspondence analysis (MCA), CATPCA (categorical principal components analysis)	800 patients	4 allergens	Not reported
11	Panagiotidis et al, 2024[21]	To develop and evaluate machine-learning models (including deep learning and ensemble approaches) for automatic detection and classification of allergic contact dermatitis reactions from patch-test images.	Random Forest, Support Vector Machines, and XGBoost, CNN	200 patients	30 allergen substances	The accuracy achieved by this model is 84%, F1 score of 79%

Table 1: Key characteristics of the included studies

Discussion

The collective evidence demonstrates a robust foundation for integrating artificial intelligence across the spectrum of contact dermatitis diagnostics, extending beyond traditional clinical observation into molecular, visual, and population-level inference. Mechanistically, AI-driven transcriptomic analyses have enabled reliable separation of allergic contact dermatitis (ACD) and irritant contact dermatitis (ICD), identifying immune activation pathways dominated by cytotoxic T-cell transcripts in ACD and proliferative or tissue renewal signatures in ICD, with high discriminatory accuracy [4,5]. This aligns with broader immunodermatology literature demonstrating that high-dimensional transcriptomic profiling can outperform conventional histopathology in inflammatory skin disease classification [6]. In parallel, convolutional neural networks (CNNs) have shown consistent capability in automating the detection of patch test reactions from images, particularly when colour-derived modalities such as redness are emphasised, achieving performance comparable to expert readers under controlled conditions [3,7]. Similar performance patterns have been observed in other dermatological imaging tasks, where CNNs demonstrate strong pattern-recognition capacity in visually constrained problems [8]. At the population level, non-linear machine learning approaches, including gradient boosting, outperform linear models in modelling sensitisation risk and allergen associations, reflecting the complex interactions between demographic, occupational, and clinical variables that are difficult to capture with traditional regression techniques [911].

Despite high reported performance metrics, the translational relevance of many image-based models remains limited. Accuracies exceeding 90% and AUCs approaching 0.94 are typically derived from simplified benchmark tasks that do not reflect routine patch test interpretation [2,3,12]. This phenomenon is well recognised across medical AI, where task simplification, curated datasets, and internal validation inflate apparent performance while obscuring fragility under real-world conditions [13,14]. In patch testing, CNNs are predominantly trained as binary classifiers, omitting multicategory grading that is fundamental to ICDRG-based diagnosis, a limitation driven largely by severe class imbalance in available datasets [3,7]. Oversampling and threshold adjustment improve sensitivity but systematically reduce specificity, resulting in false-positive rates that are poorly tolerated in clinical workflows and have been shown to erode clinician trust in decision-support systems [3,12,15]. Furthermore, complex non-linear models, while outperforming simpler approaches in controlled settings, are inherently more susceptible to overfitting, particularly when external validation is absent, a limitation repeatedly documented in clinical prediction modelling [10,11]. Crucially, benchmarking omits tactile assessment of induration and infiltration, a core diagnostic feature of patch testing, reinforcing the disconnect between image-only AI outputs and embodied clinical judgement [3,14].

Dataset heterogeneity and bias substantially constrain generalisability across all AI modalities reviewed. Image-based studies show marked variability in acquisition techniques, ranging from standardised multi-modal systems to repurposed clinical photographs, introducing inconsistencies in colour calibration, lighting, and resolution that degrade cross-domain performance [3,7]. This sensitivity to acquisition context mirrors findings from broader dermatology AI research, where distribution shift is a major driver of performance failure [16]. Demographic bias is particularly pronounced, with over-representation of Caucasian patients and Fitzpatrick skin types I–II, raising concerns regarding diminished accuracy in darker skin types and the risk of inequitable diagnostic support [2,3,12,17]. Clinical and epidemiological models further reflect geographically and occupationally specific exposure patterns, limiting transferability across regions and healthcare systems [9,18,19]. Temporal heterogeneity compounds these issues, as allergen panels and regulatory exposures evolve, meaning that models trained on historical datasets may encode patterns of diminishing contemporary relevance, a challenge increasingly recognised in longitudinal clinical AI [10,14].

From a translational standpoint, current AI applications in patch testing are best conceptualised as decision-support systems rather than autonomous diagnostic tools. Limited interpretability of CNNs and ensemble models complicates accountability and clinician trust, consistent with concerns raised across high-stakes medical AI applications [20,3]. Integration into routine practice is further impeded by the lack of standardised imaging protocols, despite clear evidence that performance is highly sensitive to modality selection and pre-processing strategies [7,21]. Regulatory frameworks increasingly emphasise intended use and clinical impact rather than algorithmic sophistication alone, positioning image-based reaction detection and risk stratification within lower-risk decision-support categories [10,13]. In contrast, molecular AI approaches that robustly discriminate ACD from ICD align more closely with regulated in vitro diagnostics, where mechanistic grounding and analytical validity support a clearer regulatory pathway [4,5,15]. Overall, current evidence supports AI as a means of reducing observer variability and improving efficiency, rather than redefining diagnostic authority [3,12].

The evidence base is constrained by structural limitations that limit robustness and external validity. Molecular studies, while biologically informative, rely on small cohorts that restrict validation across allergens, irritants, and patient subgroups [4,5]. Conversely, large epidemiological models are constrained by incomplete clinical metadata, imposing an upper bound on predictive performance irrespective of algorithmic complexity [9,10]. Image-based research remains limited by extreme class imbalance, preventing progression from binary detection to clinically essential severity grading [2,3,7]. Across all domains, reliance on internal validation and inconsistent reporting of demographic and skin-type variables mirrors broader weaknesses in clinical AI research and precludes confident assessment of fairness and generalizability [11,14].

Conclusion

Artificial intelligence has demonstrated credible technical capability in supporting selected components of patch testing, particularly reaction detection, molecular discrimination, and population-level risk modelling. However, its translational readiness is constrained by methodological simplification, biased and heterogeneous datasets, and limited external validation, confining its current role to clinical decision support rather than autonomous diagnosis. Molecular AI approaches represent a more mature but narrowly applicable pathway, whereas image- and clinical-data models offer incremental gains in efficiency and consistency. Overall, the clinical value of AI in patch testing is determined less by headline performance metrics than by alignment between task definition, data provenance, and regulatory context.

References:

1. Tang HS, Ebriani J, Yan MJ, Wongvibulsin S. (2025). Artificial intelligence–assisted interpretation of patch test reactions: systematic review. Contact Dermatitis.
2. Chan WH, Do H, Burnett G, MacFarlane J. (2021). Automated detection of skin reactions in epicutaneous patch testing using machine learning. *Br J Dermatol*; *185*(2):456-458.
3. Hall MR, Weston AD, Wieczorek MA, Hobbs MM, Caruso MA, Siddiqui H, Pacheco-Spann LM, Lopez-Dominguez JL, Escoda-Diaz C, Carter RE, Bruce CJ. (2024). Deep learning–based assessment of patch test reactions compared with expert dermatologists: proof-of-concept automated classification of allergic contact dermatitis from clinical photographs. *J Invest Dermatol,* 144:131-138.
4. Fortino V, Wisgrill L, Werner P, Suomela S. (2020). Transcriptomic profiling differentiates allergic and irritant contact dermatitis. *J Allergy Clin Immunol*, *145*:1501-1512.
5. Lefevre MA, Nosbaum A, Rozieres A, Lenief V. (2021).

Machine learning analysis of gene expression signatures in contact dermatitis. *Allergy*, *76*:3120-3132.

6. Tsoi LC, Rodriguez E, Degenhardt F, Baurecht H, Wehkamp U, Volks N, Szymczak S, Swindell WR, Sarkar MK, Raja K, Shao S, Patrick M, Gao Y, Uppala R, Perez White BE, Getsios S, Harms PW, Maverakis E, Elder JT, Franke A, Gudjonsson JE, Weidinger S. (2019). *Atopic Dermatitis Is an IL-13-Dominant Disease with Greater Molecular Heterogeneity Compared to Psoriasis*. *J Invest Dermatol,* 139(7):1480-1489.
7. Vezakis IA, Lambrou GI, Kyritsi A, Tagka A, Chatziioannou A, Matsopoulos GK. (2023). *Detecting Skin Reactions in Epicutaneous Patch Testing with Deep Learning: An Evaluation of Pre-Processing and Modality Performance*. *Bioengineering,* 10(8):924.
8. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature,* 542:115-118.
9. Adler NR, et al. (2017). Predictive modelling of contact sensitisation using registry data. Contact Dermatitis, 77:307-315.
10. Cunningham L, Ganier C, Ferguson F, White IR, Watt MF, McFadden J, Lynch DM. (2022). Gradient boosting approaches can outperform logistic regression for risk prediction in cutaneous allergy. Contact Dermatitis, 86(3):165-174.
11. Beam AL, Kohane IS. (2018). Big data and machine learning in health care. JAMA, 319:1317-1318.
12. Ravishankar A, Heller N, Bigliardi PL. (2024). Demonstration of Convolutional Neural Networks to Determine Patch Test Reactivity. *Dermatitis,* 35(2):144-148.
13. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17:195.
14. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, Aviles-Rivero AI, Etmann C, McCague C, Beer L, Weir-McCall JR, Teng Z, Goh V, Rudd JHF, Sala E, Schönlieb C-B, Aerts HJWL, Higgins GS. (2021). Common pitfalls and recommendations for artificial intelligence in medical imaging. *Nature Machine Intelligence*, 3:199-217.
15. Topol EJ. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nat Med, 25:44-56.
16. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, Ko J, Swetter SM, Bailey EE, Gevaert O, Mullainathan S, Zou J. (2022). Disparities in dermatology artificial intelligence performance by skin type. *The Lancet Digital Health*, 4:726-734
17. Adamson AS, Smith A. (2018). Machine learning and health care disparities in dermatology. JAMA Dermatol, 154:1247-1248.
18. Kyritsi E, Rustemeyer T, Brans RJ, Uter W, Kezic S, John SM, Giménez-Arnau AM, Larese Filon F, Bruze M, Johansen JD, Chomiczewska-Skóra J, Aalto-Korte K, Schuttelaar MLA. (2023). Occupational determinants of contact sensitisation: machine learning analysis. *Contact Dermatitis*, 89(1):45-54.
19. Kyritsi E, Rustemeyer T, Brans RJ, Uter W, Kezic S, John SM, Giménez-Arnau AM, Larese Filon F, Bruze M, Johansen JD, Chomiczewska-Skóra J, Aalto-Korte K, Schuttelaar MLA. (2024). Temporal trends in allergen exposure and model drift in contact dermatitis. *Contact Dermatitis*, 90(2):101-110.
20. Rudin C. (2019). Stop explaining black box machine learning models for high stakes decisions. Nat Mach Intell, 1:206-215.
21. Panagiotidis G, Kyritsi E, Schuttelaar MLA. (2024). Imaging protocol standardisation for AI-based dermatology diagnostics. Eur J Dermatol, 34:12-20.
22. Tang HS, Ebriani J, Yan MJ, Wongvibulsin S, Farshchian M. (2025). Artificial Intelligence in Patch Testing: Comprehensive Review of Current Applications and Future Prospects in Dermatology. JMIR Dermatol, 8:67154.
23. Hall MR, Weston AD, Wieczorek MA, Hobbs MM, Caruso MA, Siddiqui H, Pacheco-Spann LM, Lopez-Dominguez JL, Escoda-Diaz C, Carter RE, Bruce CJ. (2024). An Automated Approach for Diagnosing Allergic Contact Dermatitis Using Deep Learning to Support Democratization of Patch Testing. Mayo Clin Proc Digit Health, 2(1):131-138.
24. Chan WH, Srivastava R, Damaraju N, Do H, Burnett G, MacFarlane J, Xie SM, Chen JK, Honari G, Sarin KY. (2021). Automated detection of skin reactions in epicutaneous patch testing using machine learning. Br J Dermatol, 185(2):456-458.
25. Andre N, Usher A, Ofri M, Horev A. (2024). Exploring the relationship between allergic contact dermatitis and atopic dermatitis in children: insights from a retrospective patch testing analysis. Int J Dermatol, 63:795-798.

DOI:10.31579/2578-8949/204