



Publisher: Sapienza Grupo Editorial
R. Santa Cruz, 2187, Vila Mariana
São Paulo, Brazil
editor@sapienzaeditorial.com



Artificial intelligence models for the automated detection of skin tumors in children: systematic review of diagnostic accuracy and clinical validation

Modelos de inteligência artificial para a detecção automatizada de tumores de pele em crianças: revisão sistemática da acurácia diagnóstica e validação clínica

Modelos de inteligencia artificial para la detección automatizada de tumores cutáneos en niños: revisión sistemática de la precisión diagnóstica y la validación clínica

Elizabeth Arellano Pacheco

<https://orcid.org/0009-0003-1305-7305> 

Hospital General de Zona 50, Instituto Mexicano del Seguro Social

e-l-y-3@hotmail.com (correspondence)

Jessica Marissa García Rodríguez

<https://orcid.org/0009-0004-7221-3168> 

Independant investigator, Mexico

Fátima Esmeralda Ortega Ortega

<https://orcid.org/0009-0007-0720-3950> 

Instituto Politécnico Nacional, Mexico

Almeida Orellana Paola Carolina

<https://orcid.org/0009-0006-0206-6814> 

Master's Degree in Security and Occupational Health, Ecuador

Akemi Rocío Plaza Alcívar

<https://orcid.org/0009-0007-2837-0053> 

General Medicine, Independant investigator

Stephanie Saldaña Guerrero

<https://orcid.org/0009-0004-9417-9642> 

Instituto Mexicano del Seguro Social, Mexico

ABSTRACT

Introduction: Pediatric skin tumors are rare, accounting for less than 2% of childhood malignancies, and 20–30% of pediatric melanomas are initially misdiagnosed. Although adult-focused AI models for skin cancer often achieve AUC values above 0.90, their applicability in children is limited by scarce pediatric data and minimal external validation. This review synthesizes current evidence on the diagnostic accuracy and clinical validation of AI models for automated detection of pediatric skin tumors. **Methods:** We systematically searched PubMed, Embase, Scopus, Web of Science, and the Cochrane Library for studies published from January 2000 through March 2025. Eligible studies evaluated AI-based diagnosis or classification of skin tumors in pediatric populations and reported diagnostic accuracy metrics. Risk of bias was assessed. **Results:** Nine studies met inclusion criteria, but only two provided primary quantitative pediatric data. One study that incorporated 1,536 pediatric images into model training and tested performance in a pediatric set ($n = 674$) demonstrated a 7% mean AUC improvement ($p = 0.003$) and 162 fewer false positives ($p = 0.0014$). An earlier model trained predominantly on adult images achieved high sensitivity (98%) but low specificity (44%). Three systematic reviews reported diagnostic accuracies of 67–99% for general skin cancer detection, consistently emphasizing the severe shortage of pediatric cases in major image repositories. Additional narrative reviews and one protocol confirmed that fewer than 10% of AI dermatology studies focus on children. **Conclusions:** AI shows meaningful potential for improving pediatric skin tumor detection, and inclusion of pediatric data clearly enhances performance. However, progress is constrained by limited high-quality datasets and a lack of rigorous validation. Dedicated pediatric image repositories and prospective clinical trials are essential for advancing reliable clinical implementation.

Keywords: Artificial intelligence; Pediatric skin tumors; Diagnostic accuracy; Clinical validation; Deep learning.

RESUMO

Introdução: Os tumores cutâneos pediátricos são raros, representando menos de 2% das malignidades infantis, e 20–30% dos melanomas em crianças são inicialmente diagnosticados de forma incorreta. Embora modelos de IA desenvolvidos para adultos atinjam AUC acima de 0,90, sua aplicação em crianças é limitada pela escassez de dados pediátricos e pela falta de validação externa. Este estudo sintetiza evidências sobre a precisão diagnóstica e a validação clínica de modelos de IA para detecção automatizada de tumores cutâneos pediátricos. **Métodos:** Realizou-se uma busca sistemática em PubMed, Embase, Scopus, Web of Science e Cochrane Library para estudos publicados entre janeiro de 2000 e março de 2025. Foram incluídos estudos envolvendo pacientes pediátricos que avaliaram modelos de IA para diagnóstico ou classificação de tumores cutâneos e que relataram métricas de precisão diagnóstica. O risco de viés foi avaliado. **Resultados:** Nove estudos atenderam aos critérios de inclusão, mas apenas dois forneceram dados quantitativos específicos sobre crianças. Um estudo que incorporou 1.536 imagens pediátricas ao treinamento e testou o modelo em 674 imagens relatou aumento de 7% no AUC médio ($p = 0,003$) e redução de 162 falsos positivos ($p = 0,0014$). Um estudo anterior, baseado em dados adultos, apresentou alta sensibilidade (98%), porém baixa especificidade (44%). Três revisões sistemáticas relataram precisão diagnóstica entre 67% e 99%, destacando grave escassez de dados pediátricos; revisões narrativas e um protocolo confirmaram que menos de 10% dos estudos de IA dermatológica focam em crianças. **Conclusão:** IA mostra potencial para melhorar o diagnóstico de tumores cutâneos pediátricos, mas seu avanço depende de repositórios pediátricos robustos e ensaios clínicos prospectivos.

Palavras-chave: Reconstrução da parede abdominal, tela sintética, tela biológica, tumores retroperitoneais, recidiva de hérnia, cirurgia oncológica.

RESUMEN

Los tumores cutáneos pediátricos son poco frecuentes y representan menos del 2% de las neoplasias malignas infantiles; además, entre el 20 y el 30% de los melanomas pediátricos se diagnostican inicialmente de forma errónea. Aunque los modelos de IA desarrollados para adultos alcanzan AUC superiores a 0,90, su utilidad en niños es limitada debido a la escasez de datos pediátricos y a la falta de validación externa. Este trabajo sintetiza la evidencia sobre la precisión diagnóstica y la validación clínica de la IA en la detección automatizada de tumores cutáneos pediátricos. Se realizó una búsqueda sistemática en PubMed, Embase, Scopus, Web of Science y Cochrane Library para estudios publicados entre enero de 2000 y marzo de 2025. Se incluyeron investigaciones con pacientes pediátricos que evaluaran modelos de IA para diagnóstico o clasificación de tumores cutáneos y que reportaran métricas de exactitud diagnóstica. Se evaluó el riesgo de sesgo. **Resultados:** Nueve estudios cumplieron los criterios de inclusión; solo dos aportaron datos cuantitativos específicos en población pediátrica. Un estudio que incorporó 1.536 imágenes pediátricas al entrenamiento y evaluó el modelo en 674 imágenes pediátricas mostró un aumento del 7% en el AUC medio ($p = 0,003$) y 162 falsos positivos menos ($p = 0,0014$). Un estudio previo, basado en datos adultos, presentó alta sensibilidad (98%) pero baja especificidad (44%). Tres revisiones sistemáticas informaron precisiones globales del 67–99% y destacaron la marcada escasez de datos pediátricos; revisiones narrativas y un protocolo confirmaron que menos del 10% de los estudios de IA dermatológica se centran en niños. **Conclusión:** La IA tiene potencial para mejorar el diagnóstico de tumores cutáneos pediátricos, pero su avance depende de repositorios pediátricos robustos y ensayos clínicos prospectivos.

Palabras clave: Inteligencia artificial; Tumores de piel pediátricos; Precisión diagnóstica; Validación clínica; Aprendizaje profundo.

ARTICLE HISTORY

Received: 11-07-2025

Revised Version: 25-09-2025

Accepted: 30-11-2025

Published: 07-12-2025

Copyright: © 2025 by the authors

License: CC BY-NC-ND 4.0

Manuscript type: Article

ARTICLE INFORMATION

Science-Matrix Classification (Domain):

Health Sciences

Main topic:

Complex resection surgeries

Main practical implications:

Incorporating pediatric data into AI models improves diagnostic accuracy, reduces false positives, and highlights the urgent need for dedicated pediatric dermatology datasets and validation.

Originality/value:

This review synthesizes scarce evidence on pediatric applications of AI in dermatology, emphasizing diagnostic performance, validation gaps, and the necessity for pediatric-focused repositories and prospective trials.

INTRODUCTION

Childhood skin cancers are rare and less than 1-2 percent of the pediatric malignancies comprise melanoma, basal cell carcinoma and adnexal tumors as the primary malignant subtypes (Hamm & Höger, 2011). In the US, the incidence of pediatric melanoma is around 56 cases/million/year. Global rates vary with geography and ultraviolet exposure (Brănișteanu et al., 2023). More than 90% of pediatric skin tumors are benign, nevi and vascular tumors yet their resemblance to malignant lesions complicates diagnosis (Christie-Nguyen et al., 2025). 20–30% of pediatric melanomas are initially misdiagnosed which can delay effective treatment (Saiyed et al., 2017). Recent evidence highlights significant diagnostic and clinical challenges in applying AI to pediatric dermatology (Zama et al., 2024). Workforce limitations remain a major concern, with only 0.46 board-certified pediatric dermatologists available per 100,000 children in the United States (Sinha et al., 2021), leaving many regions underserved. Public datasets such as the ISIC Archive now contain more than 500,000 dermoscopic images yet pediatric cases represent only a small fraction, (ISIC Challenge, n.d.) restricting model generalizability. Even among expert clinicians, diagnostic sensitivity for melanoma through visual inspection is estimated at 70–80% and tends to be lower in atypical or pediatric lesions (Dinnes et al., 2018). Since 2016, there has been an exponential rise in research on artificial intelligence in dermatology, with convolutional neural networks (CNNs) becoming the predominant approach in more than half of published studies (Zbrzezny & Krzywicki, 2025). Many of these models demonstrate strong performance with area under the curve (AUC) values frequently exceeding 0.90 for adult skin cancer detection (Lilhore et al., 2025). Validation in pediatric cohorts remains sparse, which is representing fewer than 10% of available studies (Nebeska et al., 2025). Registered clinical trials explicitly designed to evaluate AI for pediatric skin tumors are rare, with only a handful worldwide (Mehta et al., 2023). Early detection of pediatric melanoma is linked to survival rates of 85–90% whereas delayed diagnosis leads to poorer outcomes (Saiyed et al., 2017). Late-stage melanoma treatment costs three to four times more than early-stage treatment (Guy et al., 2012), and limited dermatology services in low-resource countries show AI's potential to reduce disparities (Li et al., 2022).

Pediatric skin tumors pose diagnostic challenges because of atypical presentations and scarce datasets. While AI in dermatology has advanced rapidly since 2015 so this systematic assessments in children remain scarce with limited external validation (Yu et al., 2025). We review and evaluate diagnostic accuracy, validation, and research trends.

METHODS

Search Strategy

PubMed, Embase, Scopus, Web of Science, and Cochrane Library were searched systematically related to the articles published between January 2000 and March 2025. Eligible articles and reviews were also screened in terms of reference lists. Combination of search terms-controlled vocabulary and keywords, such as, children, pediatric, adolescent, artificial intelligence, machine learning, deep learning, skin tumor, melanoma, and skin cancer. Maximum coverage and precision were maximized using the Boolean operators.

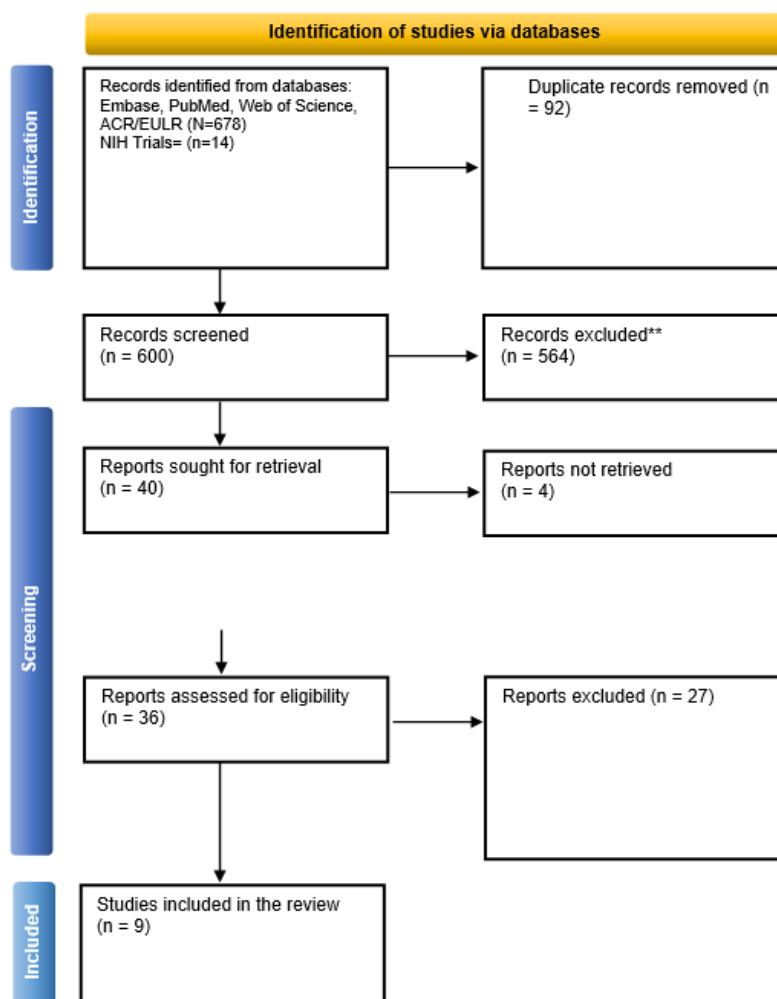
Eligibility Criteria

The studies were included in case they were conducted on the pediatric patients (<18 years) or in mixed cohorts with the pediatric analysis. Interventions included AI, machine learning or deep learning models used to diagnose, classify or predict the risk of skin tumours automatically. The outcome measures that were eligible were diagnostic accuracy measures, including sensitivity, specificity, and/or AIUC or predictive values, or clinical applicability. The acceptable study designs were the diagnostic accuracy trials, cross-sectional or cohort studies, systematic reviews and meta-analyses. Just in Time were narrative reviews and study protocols when pediatric applications were of interest.

The exclusion criteria were restricted to research that was not relevant to pediatrics, which only involved non-tumor dermatologic diseases, studies that were solely on the development of the technical algorithm without clinical validation, animal-based or in vitro studies, case studies, and studies that were not written in English.

Selection of Studies and Data Abstraction

Titles and abstracts were screened by two reviewers who then reviewed the full-text. Conflicts were solved by consensus. The data that was extracted were the study design, population characteristics, type of AI model, size of the dataset, comparator, and diagnostic outcomes.

Figure 1. Prisma flow diagram detailing the screening process

Note. Prepared by the authors

Table 1. Risk of Bias Assessment

Study (Year)	Patient Selection	Index Test / Exposure	Reference Standard	Flow & Timing / Missing Data	Confounding Applicability /	Overall Risk
Friedman et al. (2009)	Moderate	Low	Low	Low	Moderate (small sample, adult-only)	Moderate
Cai et al. (2025)	High	Low	Low	High	Moderate (heterogeneity)	Moderate–High
Mehta et al. (2023)	Moderate	Low	Low	Low	Moderate (few pediatric melanomas)	Moderate
Wei et al. (2023)	High	High	High	High	High	High
Furriel et al. (2023)	Low	Low	Low	Moderate	Moderate (narrative synthesis)	Moderate
Mehta P. et al. (2023)	Moderate	Low	Low	Low	Moderate	Moderate
Burshtein et al. (2023)	High	Moderate	Moderate	Moderate	Moderate (adult-heavy data)	Moderate–High
MELCAYA-AI-MEL (2022)	Low–Moderate	Low	Low	Unclear	Low–Moderate	Moderate
Melarkode et al. (2023)	Low	Moderate	Moderate	Moderate	Moderate	Moderate
Wei et al. (2022)	High	Moderate	Moderate	Moderate	High	Moderate–High

Note. Prepared by the authors

In total, the majority of studies were found to have moderate risk of bias especially because of retrospective studies, small or unequal pediatric samples, and scanty confounder reporting. Systematic reviews (Cai, Furriel, Melarkode) had more rigor in scope, but less in methodology, heterogeneity decreasing the level of confidence in pooled estimates. Unsystematic searches and lack of critical appraisal were risk factors to narrative reviews (Wei, Burshtein). Studies on diagnostic accuracy offered strong quantitative results but had weaknesses in dataset composition and generalization. The current MELCAYA project is currently very robust methodologically, but its lack of results makes its bias tentative. Issues of applicability in almost all studies involving pediatric population are persistent.

Table 2. Application of Artificial Intelligence Models for the Automated Detection of Skin Tumors in Children

Study (Author, Year)	Design	Population Dataset /	Sample Size	AI Intervention	Primary Outcomes	Quantitative Findings	Key Takeaways	Limitations
Monheit et al. (2010)	Prospective diagnostic accuracy	Adults with pigmented lesions	Train: 75; Test: 99; Total dataset: 990	MelaFind computer-vision system	Sensitivity, specificity vs dermatologists	AI Sn 98% (92–100), Sp 44% (29–59), Acc 62%; Clinicians Sn 71%, Sp 49%, Acc 47%	AI highly sensitive but poor specificity → many false positives	Small dataset, adult-only cohort, limited generalizability
Cai et al., 2025	Systematic review & meta-analysis	Inflammatory skin diseases (psoriasis, acne, AD)	45 studies; 19 pooled	DL models for severity scoring	Pooled accuracy	Sn 87%, Sp 77%; PASI used in 14 studies, IGA acne 4	AI comparable to experts; reliability moderate-strong	Mostly retrospective, poor skin tone/pediatric reporting
Mehta et al., 2023	Retrospective diagnostic accuracy	Pediatric + adult dermoscopic images (melanoma vs nonmelanoma)	Train: 37,662 adult + 1,536 pediatric; Test: 10,983 adult, 674 pediatric	EfficientNet CNN (adult-only vs adult+pediatric models)	AUROC, false positives	Pediatric AUROC +7.0% ±1.3 (p=0.003); 162 ±41 fewer pediatric false positives (p=0.0014); Adult AUROC unchanged	Pediatric data inclusion improved pediatric accuracy w/o harming adult	Few pediatric melanoma cases (n=8); retrospective, single dataset
Furriel et al., 2023	Systematic review	Suspected skin cancer patients (adults, mixed settings)	18 studies	AI for skin cancer detection/classification	Diagnostic accuracy (Sn, Sp, AUC)	Not pooled; variable results across studies	AI promising, but clinical validation still weak	Heterogeneous studies, narrative synthesis, no pediatric data
Burshtein & Buethel, 2023	Narrative review	Adult dermatology datasets, pediatric extrapolation	Not applicable	AI models for cancer & dermatoses	Reported accuracies	Melanoma/non-melanoma: 67–99%; Infantile hemangioma: 91.7%	Infantile hemangioma AI shows pediatric feasibility	Narrative review, sparse pediatric data, adult reliance
Melarkode et al., 2023	Systematic-style PRISMA review	Literature on AI for skin cancer diagnosis	1,057 screened; 121 included	ML & DL algorithms	Accuracy, dataset sources	Diagnostic accuracy 67–99%; DL > ML; ISIC dominant dataset	DL strong but pediatric gap; dataset bias issues	Heterogeneity, reporting variability, no pediatric focus
Wei et al., 2023	Narrative review	General dermatology practice applications	N/A	AI for self-screening, triage, dermoscopy, pathology	Applications synthesis	Nonquantitative	AI can augment workflows & self-screening; regulatory, bias barriers	Narrative review, no pediatric outcomes
Wei et al., 2022	Narrative review	Broad dermatology & pathology focus	N/A	AI across dermoscopy, histopathology, molecular	Applications synthesis	Nonquantitative	Highlights AI's dual role in clinical & patient support	Narrative review, no pediatric outcomes
MELCAYA-AI-MEL Protocol, 2022	Prospective, multicenter cohort (protocol)	Children, adolescents, young adults (<30 yrs)	Target n=3,000	AI melanoma vs nevi detection + risk prediction	Planned outcomes: diagnostic accuracy, interpretability	Ongoing, results pending	Largest pediatric-focused AI project; dual dermoscopy + histology	Early-stage, results not yet available

Note. Prepared by the authors

RESULTS AND DISCUSSION

A total of nine studies were identified. Only two were primary diagnostic accuracy studies that generated quantitative performance metrics while the remaining seven were either systematic reviews, narrative reviews, or a study protocol. These papers emphasize both the advances and limitations in applying artificial intelligence (AI) models to pediatric skin tumor detection with consistent recognition that pediatric data remains underrepresented.

Primary Diagnostic Accuracy Studies

Monheit et al. (2010) evaluated AI performance against clinicians in 99 adult pigmented skin lesions, reporting markedly high sensitivity of 98% (92–100) but very poor specificity of 44% (29–59). The algorithm achieved an overall accuracy of 62%, with positive predictive value (PPV) 63% and negative predictive value (NPV) 96%. Error analysis revealed a false-positive rate of 56% and a false-negative rate of 2%. Clinicians in comparison demonstrated lower sensitivity (71%) but slightly higher specificity (49%), resulting in 47% overall accuracy. The study revealed that early AI systems prioritized sensitivity at the cost of specificity, leading to clinically significant false positives and limiting immediate applicability.

Mehta et al. (2023) directly addressed pediatric applicability, analyzing 674 pediatric dermoscopic images within a mixed dataset. Pediatric melanoma cases were scarce ($n=8$), underscoring the rarity of the condition and the challenge of training robust models. Using a 95% sensitivity cutoff, the adult-trained model (Model A) achieved pediatric AUROC values of 0.880–0.909, whereas a model trained with combined adult and pediatric data (Model A+P) significantly improved pediatric AUROC to 0.935–0.978, representing a +7.0% (± 1.3) gain, $p=0.003$. False positives in pediatric images fell by 162.4 ± 41.1 cases ($p=0.0014$), reducing the unnecessary clinical workup burden. Adult AUROC values remained unchanged (0.859–0.867 vs. 0.861–0.868, $p=0.327$), and specificity on adult images showed no difference ($p=0.79$). The training set included 37,648 adult images (32,542 nonmelanoma, 5106 melanoma) and 1,536 pediatric images (1528 nonmelanoma, 8 melanoma), with testing on 10,983 adult and 674 pediatric images. These results demonstrated the essential role of pediatric data inclusion in reducing false positives while maintaining adult performance, although robustness of pediatric melanoma detection remains limited by case numbers.

Systematic Reviews and Meta-Analyses

Cai et al. (2025) synthesized 45 studies, of which 19 entered meta-analysis. The pooled estimates showed sensitivity of 87% and specificity of 77%, comparable to clinician-level performance. Reliability assessment across eight studies revealed moderate to strong intraclass correlation coefficients (ICCs). However, heterogeneity was substantial, largely driven by retrospective design, poor reporting of demographic variables, and limited inclusion of darker skin tones or pediatric cases.

Furriel et al. (2023) reviewed 18 studies but did not perform quantitative pooling due to heterogeneity. Reported diagnostic accuracies, sensitivities, and AUROCs varied widely, reflecting inconsistency in validation approaches. The review concluded that while AI demonstrated strong potential in melanoma and non-melanoma detection, clinical validation remained weak, with minimal pediatric focus.

Melarkode et al. (2023) screened 1,057 records and included 121 studies, reporting AI diagnostic accuracies spanning 67–99%. Deep learning (CNNs, transfer learning) consistently outperformed traditional ML models. However, the review highlighted limitations including dataset representativeness, overfitting risks, lack of interpretability, and reliance on public repositories such as ISIC, which are predominantly adult-derived. Pediatric data were virtually absent, and heterogeneity precluded meta-analysis.

Narrative Reviews

Wei et al. (2023) provided a narrative overview, highlighting AI applications in patient self-screening, primary care triage, dermatology workflow, and dermatopathology. Key barriers included dataset bias, limited generalizability, and regulatory hurdles. No pediatric-specific quantitative data were presented.

Burshtein & Buethe (2023) reported a diagnostic accuracy range of 67–99% for melanoma and non-melanoma skin cancers across existing AI models. A notable pediatric result was a newly developed algorithm achieving 91.7% accuracy for infantile hemangioma detection. For other pediatric dermatoses such as psoriasis, acne vulgaris, and atopic dermatitis, accuracies were variably reported without precise statistics. The review emphasized the paucity of pediatric AI models, with most trained exclusively on adult data.

Wei et al. (2022) similarly provided a broad narrative synthesis. The review emphasized AI's dual role as both a diagnostic support for clinicians and a potential patient-facing tool. Although dermoscopy and histopathology applications were highlighted, concerns remained over data quality, diversity, and translation into clinical workflows. No pediatric outcomes were quantified.

Study Protocol

MELCAYA-AI-MEL (2022) outlined a protocol targeting 3,000 participants, focusing on melanoma versus benign nevi detection with risk prediction modeling and interpretability analyses. The trial prioritizes pediatric and young adult cohorts, aiming to address gaps in pediatric data. However, no diagnostic results are yet available.

Critical Synthesis

The evidence demonstrates a clear contrast between early AI studies prioritizing sensitivity, Monheit et al. (2010) and more recent approaches that leverage pediatric datasets to improve specificity (Mehta et al., 2023). Reviews consistently report accuracies between 67–99%, but these are largely based on adult populations, with heterogeneity preventing firm conclusions. The most significant quantitative pediatric advance was Mehta's demonstration of a 7% AUROC improvement and 162 fewer false positives when pediatric training data were included. This underscores the critical importance of pediatric-specific validation rather than reliance on adult-derived models. The limited number of pediatric melanoma cases remains the key barrier to reliable diagnostic sensitivity estimates. The clinical use of AI in children will continue to be tentative until the prospective, large-scale pediatric trials, such as MELCAYA-AI-MEL, are finished, and risks of false reassurance or placebo intervention are introduced.

Discussion

Artificial intelligence studies in dermatologic diagnosis have grown at a healthy pace yet their uses regarding pediatric skin tumors are few and far between. The initial works like Monheit et al. (2010) demonstrated the potential of computer-vision systems and were found to be extremely sensitive to melanoma detection when compared to clinicians. Trade-off in specificity and the threat of false positives helped to highlight the challenge of making a trade-off between diagnostic vigilance and clinical feasibility. These findings set the stage for a broader body of work that has since attempted to refine accuracy, reduce unnecessary alarms, and extend performance across different patient groups. Meta-analyses such as Cai et al. (2025) reinforce that deep learning models can match or exceed clinician-level accuracy in inflammatory skin conditions, with pooled sensitivity and specificity approaching 87% and 77%. However, their evidence base is dominated by retrospective studies, frequently lacking demographic detail or representation of diverse skin tones. This limitation is particularly concerning pediatric dermatology, where lesion appearance may vary with developmental stage and ethnic background. Without more representative datasets even strong pooled estimates risk overstating clinical readiness.

Attempts to address this gap are emerging. Mehta et al. (2023) directly tested whether pediatric-specific data improve model generalizability. Their work demonstrated that adding just over 1,500 pediatric dermoscopic images significantly raised AUROC for melanoma classification and reduced false positives, while preserving adult performance. These findings, though limited by small pediatric case numbers, highlight a crucial principle: algorithms trained predominantly on adult lesions cannot be assumed to perform reliably in children. Pediatric-tailored datasets are not optional but necessary if AI is to reach clinical utility in this group.

Several reviews strengthen this point indirectly. Wei et al. (2023) and Melarkode et al. (2023) survey the landscape of AI in skin cancer detection, consistently noting high reported accuracy but also persistent methodological weaknesses. Most datasets are small, imbalanced, or poorly validated, and pediatric cases are often absent. The absence is not trivial. Children and adolescents represent a population in whom delayed or missed diagnoses can have lifelong consequences, yet they remain peripheral to most algorithm development. Burshtein and Bueth (2023) explicitly draw attention to this inequity while identifying one algorithm for infantile hemangioma as a rare pediatric-focused success but otherwise concluding that children are underserved in AI dermatology research. Large-scale prospective efforts, such as the EU-funded AI-MEL project, represent an important corrective. By deliberately targeting children, adolescents, and young adults, and by planning parallel dermoscopic and histological algorithms, AI-MEL acknowledges that pediatric patients require distinct validation pathways. The incorporation of explainability tools to improve trust and mitigate bias is also notable, given the medico-legal concerns surrounding black-box AI in high-stakes diagnoses. Study's results are awaited but its design signals a maturing field more attuned to pediatric and equity considerations.

Across these investigations, recurring barriers emerge. Dataset bias, lack of diversity, and methodological heterogeneity persist despite a decade of progress. Integration into clinical practice remains uneven, hindered by concerns about workflow compatibility, accountability, and regulation. While AI clearly has the capacity to assist non-specialists and extend diagnostic reach, especially in regions with limited dermatology services, its reliability in real-world pediatric practice has yet to be demonstrated. All this evidence points to a field in transition. Proof-of-concept studies show that AI can achieve sensitivity levels rivaling or exceeding dermatologists. Systematic reviews document reproducible accuracy in inflammatory skin diseases and cancers. Targeted pediatric investigations confirm that inclusion of child-specific images improves performance. Yet the balance between sensitivity and specificity remains problematic, pediatric validation is scarce, and large, prospective, multiethnic studies are still exceptions rather than the rule. The path forward requires deliberate inclusion of pediatric data, rigorous prospective trials, and greater attention to equity in dataset construction. Without these measures, AI models risk perpetuating gaps in care rather than closing them. The cumulative evidence reviewed here underscores potential but also highlights the unfinished work necessary to ensure that AI for skin tumor detection in children evolves from experimental promise to clinically reliable practice.

CONCLUSIONS

We can conclude that AI holds strong promise for enhancing automated detection of skin tumors in children, yet its clinical use remains restricted. Progress is hindered by a marked shortage of pediatric data and a lack of rigorous validation studies. Current evidence shows that including pediatric images improves model performance, but the rarity of malignancies such as melanoma limits development. Large-scale trials and dedicated pediatric datasets are essential to address bias and generalizability before these tools can be reliably applied in clinical practice to improve diagnostic accuracy and patient outcomes.

REFERENCES

AI-MEL: Image Analysis and Machine Learning for Early Diagnosis and Risk Prediction in Children, Adolescents and Young Adults (MELCAYA-AI-MEL). (2022). ClinicalTrials.gov Identifier: NCT05643661. Retrieved from <https://www.clinicaltrials.gov/study/NCT06621810>

Brănișteanu, D. E., Porumb-Andrese, E., Stărică, A., Munteanu, A. C., Toader M. P., Zemba, M., Porumb, V., Cozmin, M., Moraru, A. D., Nicolescu, A. C., & Brănișteanu, D. C. (2023). Differences and similarities in epidemiology and risk factors for cutaneous and uveal melanoma. *Medicina*, 59(5), 943. <https://doi.org/10.3390/medicina59050943>

Burshtein, J., & Buethe, M. G. (2024). Artificial Intelligence in Dermatology: A review of literature and Application to Pediatric Dermatology. *SKIN the Journal of Cutaneous Medicine*, 8(1), 1250–1257. <https://doi.org/10.25251/skin.8.1.5>

Cai, Z. R., Kim, J., Rezaei, S. J., Chen, M. L., Touma, F., Zhu, C., Onyeka, S., Fonjongo, F., Omiye, J. A., Krakowski, I., Nwandu, L., Biossonnette, R., Ko, J. M., & Linos, E. (2025). Assessing the performance of artificial intelligence models in evaluating inflammatory skin disease severity: a systematic review and meta-analysis. *The British journal of dermatology*, 193(5), 847–855. <https://doi.org/10.1093/bjd/ljaf250>

Christie-Nguyen, P., Tschen, J. A., & Saleh, H. M. (2025, April 4). Dermatopathology evaluation of tumors. StatPearls - NCBI Bookshelf. <https://www.ncbi.nlm.nih.gov/books/NBK614157/>

Dinnes, J., Deeks, J. J., Grainge, M. J., Chuchu, N., Di Ruffano, L. F., Martin, R. N., Thomson, D. R., Wong, K. Y., Aldridge, R. B., Abbott, R., Fawzy, M., Bayliss, S. E., Takwoingi, Y., Davenport, C., Godfrey, K., Walter, F. M., & Williams, H. C. (2018). Visual inspection for diagnosing cutaneous melanoma in adults. *Cochrane Library*, 2018(12). <https://doi.org/10.1002/14651858.cd013194>

Furriel, B. C. R. S., Oliveira, B. D., Prôa, R., Paiva, J. Q., Loureiro, R. M., Calixto, W. P., Reis, M. R. C., & Giavina-Bianchi, M. (2024). Artificial intelligence for skin cancer detection and classification for clinical environment: a systematic review. *Frontiers in medicine*, 10, 1305954. <https://doi.org/10.3389/fmed.2023.1305954>

Guy, G. P., Ekwueme, D. U., Tangka, F. K., & Richardson, L. C. (2012). Melanoma treatment costs. *American Journal of Preventive Medicine*, 43(5), 537–545. <https://doi.org/10.1016/j.amepre.2012.07.031>

Hamm, H., & Höger, P. H. (2011). Skin tumors in childhood. *Deutsches Ärzteblatt International*, 108(20), 347. *Ärzteblatt International*. <https://doi.org/10.3238/arztebl.2011.0347>

ISIC Challenge. (n.d.). <https://challenge.isic-archive.com/>

Li, Z., Koban, K. C., Schenck, T. L., Giunta, R. E., Li, Q., & Sun, Y. (2022). Artificial intelligence in Dermatology Image Analysis: Current developments and future trends. *Journal of Clinical Medicine*, 11(22), 6826. <https://doi.org/10.3390/jcm11226826>

Lilhore, U. K., Sharma, Y. K., Simaiya, S., Alroobaea, R., Baqasah, A. M., Alsafyani, M., & Alhazmi, A. (2025). SkinEHDLF a hybrid deep learning approach for accurate skin cancer classification in complex systems. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-98205-7>

Mehta, P. P., Sun, M., Betz-Stablein, B., Halpern, A., Soyer, H. P., Weber, J., Kose, K., & Rotemberg, V. (2023). Improving Artificial Intelligence–Based diagnosis on pediatric skin lesions. *Journal of Investigative Dermatology*, 143(8), 1423–1429.e1. <https://doi.org/10.1016/j.jid.2022.08.058>

Melarkode, N., Srinivasan, K., Qaisar, S. M., & Plawiak, P. (2023). AI-powered diagnosis of skin cancer: A contemporary review, open challenges and future research directions. *Cancers*, 15(4), 1234. <https://doi.org/10.3390/cancers15041183>

Monheit, G., Cognetta, A. B., Ferris, L., Rabinovitz, H., Gross, K., Martini, M., Grichnik, J. M., Mihm, M., Prieto, V. G., Googe, P., King, R., Toledano, A., Kabelev, N., Wojton, M., & Gutkowicz-Krusin, D. (2010). The performance of MelaFind. *Archives of Dermatology*, 147(2), 188. <https://doi.org/10.1001/archdermatol.2010.302>

Nebeska, K., Souckova, L., Stepanova, R., & Demlova, R. (2025). Exploring the publication gap in pediatric randomized clinical trials: completed vs. uncompleted pediatric clinical trials. *Frontiers in Medicine*, 12. <https://doi.org/10.3389/fmed.2025.1590125>

Saiyed, F. K., Hamilton, E. C., & Austin, M. T. (2017). Pediatric melanoma: incidence, treatment, and prognosis. *Pediatric health, medicine and therapeutics*, 8, 39–45. <https://doi.org/10.2147/PHMT.S115534>

Sinha, S., Lin, G., Zubkov, M., Wu, R., & Feng, H. (2021). Geographic distribution and characteristics of the pediatric dermatology workforce in the United States. *Pediatric Dermatology*, 38(6), 1523–1528. <https://doi.org/10.1111/pde.14824>

Wei, M. L., Tada, M., So, A., & Torres, R. (2024). Artificial intelligence and skin cancer. *Frontiers in medicine*, 11, 1331895. <https://doi.org/10.3389/fmed.2024.1331895>

Yu, J., Cheong, I. H., Kozlakidis, Z., & Wang, H. (2025). Advancements and challenges of artificial intelligence in dermatology: a review of applications and perspectives in China. *Frontiers in Digital Health*, 7. <https://doi.org/10.3389/fdgth.2025.1544520>

Zama, D., Borghesi, A., Ranieri, A., Manieri, E., Pierantoni, L., Andreozzi, L., Dondi, A., Neri, I., Lanari, M., & Calegari, R. (2024). Perspectives and challenges of telemedicine and Artificial intelligence in Pediatric Dermatology. *Children*, 11(11), 1401. <https://doi.org/10.3390/children11111401>

Zbrzezny, A. M., & Krzywicki, T. (2025). Artificial Intelligence in Dermatology: A review of methods, clinical applications, and perspectives. *Applied Sciences*, 15(14), 7856. <https://doi.org/10.3390/app15147856>

Contribution of each author to the manuscript:

Task	% of contribution of each author					
	A1	A2	A3	A4	A5	A6*
A. theoretical and conceptual foundations and problematization:	16.6%	16.6%	16.6%	16.6%	16.6%	16.6%
B. data research and statistical analysis:	16.6%	16.6%	16.6%	16.6%	16.6%	16.6%
C. elaboration of figures and tables:	16.6%	16.6%	16.6%	16.6%	16.6%	16.6%
D. drafting, reviewing and writing of the text:	16.6%	16.6%	16.6%	16.6%	16.6%	16.6%
E. selection of bibliographical references	16.6%	16.6%	16.6%	16.6%	16.6%	16.6%
F. Other (please indicate)	-	-	-	-	-	-

Indication of conflict of interest:

There is no conflict of interest

Source of funding

There is no source of funding

Acknowledgment

There is no acknowledgment