Efficacy of Artificial Intelligence in Diagnosing Melanoma with Dermoscopy

Nima Aminzadeh, Heejee (Julia) Yoo, Jaisleen Kaur, Mario Almonte
Idaho College of Osteopathic Medicine, Meridian, ID

Introduction

Melanoma is a skin cancer that is derived from melanocytes, the cells in the body responsible for making pigment of the skin. The American Cancer Society estimated 104,960 new cases of melanoma and 8,430 deaths in 2025[1]. It is the most common cancer in adults aged 25-29; however, the highest incidence occurs in older men[1].

Melanomas are a multifactorial disease with genetic and environmental risk factors[2]. Skin screening is essential in early melanoma diagnosis, and dermoscopy is commonly used to detect melanoma despite its questionable cost-effectiveness and necessity for training[3].

Artificial Intelligence (AI) has been increasingly popular since its implementation in numerous fields, but its use in medicine has been debated and is a highly researched area. AI can change the current skill required to make an accurate dermatological diagnosis from years of clinical experience to computer algorithms[4]. Physicians can utilize AI to distinguish between diagnoses and improve clinical accuracy.

There has been significant research on the efficacy of AI in the diagnosis of dermatological conditions and skin cancer. However, AI is a rapidly developing field, and there are few recent reports on the accuracy of AI vs dermatologists in dermoscopic imaging. **This review aims to explore the efficacy of AI versus dermatologists alone and AI with dermatologists combined in melanoma diagnosis.**

Methodology

This literature review analyzed 36 published papers in English from 2020 to May 2025 regarding the accuracy of AI in the diagnosis of melanoma. Figure 1 displays the selection of the studies for this poster. Data on AI and human accuracy, sensitivity, specificity, and area under the curve (AUC) was tabulated in table 1 and metrics were averaged in table 2.

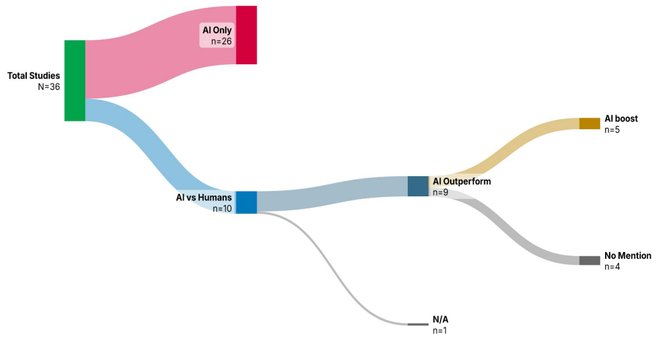

Figure 1: Overview of Total Studies Reviewed and Type of Studies

Results

Table 1: Review of 10 Studies Utilizing AI vs Human for Melanoma Diagnosis

Citation	AI technique	Dataset	AI Diagnostic Metrics	Human Metrics	AI + Human
Marchetti et al. (2023)	All Dat Are Ext (ADAE)	10,982 public dermoscopy images	AUC: 0.857 Sensitivity: 96.8% Specificity: 37.4%	AUC: 0.780 Sensitivity:89-100 % (94.5%) Specificity:0-45% (22.5%)	AI boosts human accuracy With another dataset AUC improved from 0.7798 to 0.8161 after exposure to ADAE (p = 0.042)
Heinlein et al. (2024)	ADAE	58,457 public dermoscopy images	Sensitivity: 92.1% Specificity: 67.3% Accuracy:79.8%	Sensitivity: 73.4% Specificity: 82.8% Accuracy: 78.1%	AI is more sensitive but humans are more specific
Winkler et al. (2023)	Convolutional Neural Networks (CNN): Moleanalyzer-Pro (FotoFinder Systems)	CNN: 100 dermoscopy images Dermatologist: 228 suspected melanocytic lesions from 188 patients.	AUC: 0.904 Sensitivity: 81.6% Specificity: 88.9% Accuracy: 87.7%	AUC: 0.895 Sensitivity: 84.2% Specificity: 72.1% Accuracy: 74.1%	AI + human outperformed humans AUC: 0.968 Sensitivity: 100% Specificity: 83.7% Accuracy 86.4%
Ertürk Zararsız et al. (2025)	Hybrid, DL, ML	Web of Science (WOS), Scopus, and MEDLINE	Sensitivity: 89% Specificity: 92%	Sensitivity: 79.8% Specificity: 73.6%	AI+ human outperformed both AI and Human diagnosis AUC: 0.98 Sensitivity: 94% Specificity: 95%
Maron et al. (2020)	CNN	Open source images	Sensitivity: 84.7% Specificity: 72.4% Accuracy: 73.6%	Sensitivity: 59.4% Specificity: 70.6% Accuracy: 65%	AI boosts dermatologist accuracy Sensitivity: 67.4% Specificity: 85.2% Accuracy: 86.3%
Yu et al. (2022)	Sequential Images Modelling	179 Skin lesions	Accuracy: 63.69%	Accuracy: 54.3%	AI better than human especially for early melanoma detection
Marchetti et al. (2020)	Deep Learning	ISIC 2017	AUC: ~0.87	Accuracy: 60-84% (72%)	AI improves human performance (direct statistics not mentioned in paper)
Pham et al. (2021)	Deep Learning	17,302 dermoscopic images	AUC: 0.944 Sensitivity: 85.0% Specificity: 95.0%	N/A	AI outperformed all 157 dermatologist (data not provided in paper)
Tognetti et al. (2021)	Deep Convolutional Neural Network (DCNN)	630 dermoscopic images	AUC: 0.903 Sensitivity: 86.5% Specificity: 73.6%	Sensitivity: 74.1% Specificity: 60%	AI outperformed dermatologist
Minagawa et al. (2021)	Deep Learning	ISIC and Shinsu	Avg Sensitivity: 73.05% Avg Specificity: 98.1%	Dermatologist performance varied	AI improved diagnostic equity by closing performance gap of dermatologists

Table 2: Aggregate Statistical Trends (*= statistically significant at p<0.05)

Metric	AI (Range, Average, n=)	Human (Range, Average, n=)	AI + Human (Range, Average, n=)
AUC	0.857-0.944, 0.896, n=5	0.780-0.895, 0.838, n=2	0.816-0.980, 0.921, n=3
Sensitivity	67.3-96.8%, 83.05%, n=8	59.4-94.5%, 77.57%, n=6	67.4-100%, 87.13%, n=3
Specificity	37.4-98.1%, 78.09%, n=8	22.5-82.8%, 63.6%, n=6	83.7-95%, 87.97%, n=3
Accuracy	63.7-87.7%, 76.2%, n=4	54.3-84%, 68.7%, n=5 *	86.3-86.4%, 86.35%, n=2 *

Discussion/Conclusion

Limitations and Challenges

1. Underrepresentation of diverse populations: most studies relied on the International Skin Imaging Collaboration (ISIC), Human Against Machine (HAM10000), and PH2 datasets, which are heavily skewed toward lighter skin tones.
2. Implication in clinical practice: there is an absence of regulation and a lack of clinical integration, limiting the applicability of these models in dermatologic practice. Furthermore, deep learning “black-box” models, like CNNs and DCNNs, lack transparency and are hard to interpret. These systems also raise ethical concerns regarding the potential biases and accountability issues
3. Melanoma types: specific melanoma subtypes were not compared in this review; however, for early diagnosis of melanoma, AI has been proven to be more accurate than dermatologists[10]. Further research could explore AI diagnostic accuracy for specific melanoma subtypes.

Comparative Analysis

Across all studies, aggregate values of specificity, sensitivity, accuracy, and AUC showed that AI had higher averages for all metrics when compared to humans. T-test for unequal variance revealed statistical significance only between human and AI + Human for accuracy ($p = 0.0134$ at $p<0.05$). However, due to low and variable sample sizes, statistical analysis may not be accurate.

All 10 studies concluded AI was better than human in melanoma diagnosis in at least one metric. For example, the PROVE-AI study demonstrated a sensitivity of 96.8% in identifying melanoma within dermoscopic images[5]. Similarly, Pham et al. found that AI surpassed all participating dermatologists in melanoma diagnosis based on dermoscopic images, particularly in terms of sensitivity[12]. In the ISIC 2017 challenge, certain AI models achieved an AUC of 0.87, compared to an AUC of 0.66 among dermatologists[5]. However, areas where dermatologists outperform AI remain underexplored in current literature.

Dermatologist use of AI boosts diagnostic accuracy of melanoma, which could implicate AI as a useful collaborative tool for clinical practice[7,8,9,11,12].

Takeaway and Future Aims

1. AI can be utilized as a tool to improve human accuracy in dermoscopic melanoma diagnosis.
2. As AI improves, this review can be updated with newer studies to explore AI diagnostic capabilities.
3. More studies analyzing AI performance with various skin tones is essential for the future integration of AI in melanoma diagnosis for diverse populations.

References