

Deep Learning Algorithms for the Detection of Suspicious Pigmented Skin Lesions in Primary Care Settings: A Systematic Review and Meta-Analysis

Review began 06/03/2024
Review ended 07/20/2024
Published 07/22/2024

© Copyright 2024
Abdalla et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.65122

Ahmed R. Abdalla^{1, 2}, Ahmed W. Hageen^{3, 2}, Haneen H. Saleh^{4, 2}, Omar Al-Azzawi^{5, 2}, Mahmoud Ghalab^{6, 2}, Amani Harraz^{7, 2}, Bola S. Eldoqsh^{8, 2}, Fatma E. Elawady^{9, 2}, Ayman H. Alhammadi^{10, 2}, Hesham Hassan Elmorsy^{11, 2}, Majd Jano^{12, 2}, Mohamed Elmasry^{7, 2}, Eshak I. Bahbah^{13, 2}, Ahmed Elgebalay^{14, 2}

1. Vascular Surgery, Faculty of Medicine, Mansoura University, Mansoura, EGY 2. Artificial Intelligence Research Group, MedDots Academy, Cairo, EGY 3. Faculty of Medicine, Tanta University, Tanta, EGY 4. Faculty of Medicine, University of Jordan, Amman, JOR 5. Faculty of Pharmacy, İstinye University, İstanbul, TUR 6. Radiology Department, Kafrelsheikh University, Kafr El Sheikh, EGY 7. Faculty of Medicine, Alexandria University, Alexandria, EGY 8. Faculty of Medicine, Minia University, Minia, EGY 9. Department of Ophthalmology, Port Said Specialized Hospital of Ophthalmology, Port Said, EGY 10. Department of Radiology, Faculty of Medicine, Alexandria University, Alexandria, EGY 11. Faculty of Pharmacy, Helwan University, Helwan, EGY 12. Research Department, Syrian Society for Physicians and Pharmacists, Frankfurt, DEU 13. Faculty of Medicine, Al-Azhar University, Damietta, EGY 14. Smart Health Centre, University of East London, London, GBR

Corresponding author: Eshak I. Bahbah, e.bahbah@meddots.co.uk

Abstract

Early detection of suspicious pigmented skin lesions is crucial for improving the outcomes and survival rates of skin cancers. However, the accuracy of clinical diagnosis by primary care physicians (PCPs) is suboptimal, leading to unnecessary referrals and biopsies. In recent years, deep learning (DL) algorithms have shown promising results in the automated detection and classification of skin lesions. This systematic review and meta-analysis aimed to evaluate the diagnostic performance of DL algorithms for the detection of suspicious pigmented skin lesions in primary care settings. A comprehensive literature search was conducted using electronic databases, including PubMed, Scopus, IEEE Xplore, Cochrane Central Register of Controlled Trials (CENTRAL), and Web of Science. Data from eligible studies were extracted, including study characteristics, sample size, algorithm type, sensitivity, specificity, diagnostic odds ratio (DOR), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and receiver operating characteristic curve analysis. Three studies were included. The results showed that DL algorithms had a high sensitivity (90%, 95% CI: 90-91%) and specificity (85%, 95% CI: 84-86%) for detecting suspicious pigmented skin lesions in primary care settings. Significant heterogeneity was observed in both sensitivity ($p = 0.0062$, $I^2 = 80.3\%$) and specificity ($p < 0.001$, $I^2 = 98.8\%$). The analysis of DOR and PLR further demonstrated the strong diagnostic performance of DL algorithms. The DOR was 26.39, indicating a strong overall diagnostic performance of DL algorithms. The PLR was 4.30, highlighting the ability of these algorithms to influence diagnostic outcomes positively. The NLR was 0.16, indicating that a negative test result decreased the odds of misdiagnosis. The area under the curve of DL algorithms was 0.95, indicating excellent discriminative ability in distinguishing between benign and malignant pigmented skin lesions. DL algorithms have the potential to significantly improve the detection of suspicious pigmented skin lesions in primary care settings. Our analysis showed that DL exhibited promising performance in the early detection of suspicious pigmented skin lesions. However, further studies are needed.

Categories: Dermatology, Oncology

Keywords: screening, detection, melanoma, deep learning, artificial intelligence

Introduction And Background

Skin cancer is one of the most common cancers worldwide, with an estimated 1.5 million new cases diagnosed each year [1]. Among the different types of skin cancer, melanoma is the most aggressive tumor, accounting for approximately 75% of skin cancer-related deaths [2]. The burden of skin cancers is expected to increase further due to population growth, aging, and increased exposure to ultraviolet radiation [1]. Early detection and diagnosis of skin cancer, particularly melanoma, is crucial for improving patient outcomes and survival rates. The five-year survival rate for melanoma patients with localized disease is over 99%, but it drops dramatically to 74% for regional disease and 35% for distant metastasis [3]. A significant proportion of skin cancers are diagnosed at advanced stages, leading to worse prognoses and increased healthcare costs [4].

Primary care physicians (PCPs) are often the first point of contact for patients with suspicious skin lesions. In primary care settings, visual inspection of patients to identify suspicious pigmented lesions is a well-

How to cite this article

Abdalla A R, Hageen A W, Saleh H H, et al. (July 22, 2024) Deep Learning Algorithms for the Detection of Suspicious Pigmented Skin Lesions in Primary Care Settings: A Systematic Review and Meta-Analysis. Cureus 16(7): e65122. DOI 10.7759/cureus.65122

established dermatological practice. The ABCDE criteria, which include asymmetry, border unevenness, color distribution, diameter, and evolution, have been traditionally used for early-stage melanoma screening by PCPs [5]. However, recent research suggests that highly accurate and skilled clinical detection of melanoma relies more on unconscious visual pattern recognition and "ugly duckling" comparisons than on the simplified ABCDE criteria. However, the accuracy of clinical diagnosis by PCPs is suboptimal, with studies reporting sensitivities ranging from 55% to 73% [6,7]. This can result in unnecessary referrals and biopsies, causing anxiety and stress for patients and increasing the burden on healthcare systems [8]. Therefore, there is a pressing need for more accurate and efficient diagnostic tools to aid PCPs in the early detection and diagnosis of skin cancers.

In recent years, artificial intelligence (AI) has shown great promise in improving the diagnosis and management of various medical conditions, including skin cancers [9]. Deep learning (DL) is a subset of machine learning (ML) that uses artificial neural networks to model and solve complex problems. These algorithms can automatically learn and extract relevant features from large datasets, enabling them to identify patterns and make predictions with high accuracy [10]. In the context of skin cancer, DL algorithms have been applied to the analysis of clinical images to improve the diagnostic accuracy of skin lesions compared to naked-eye examination alone by providing automated and objective analysis of dermoscopic images. Several studies have demonstrated the efficacy of DL algorithms in detecting and classifying skin lesions, including melanoma, with high sensitivity and specificity [11]. Nonetheless, these studies evaluated DL algorithms for dermoscopic images, which require specialized training and expertise that are not readily available in primary care settings [11].

In recent years, advances in smartphone technologies have increased access to high-quality personal cameras and robust mobile computing systems, which have been applied to dermatology using computer-aided diagnosis (CAD) systems [12]. Recent studies have evaluated the usefulness of employing DL algorithms in next-generation CAD systems for the evaluation of suspicious pigmented lesions in primary care settings. These studies have shown that DL-based models can achieve comparable or even superior diagnostic accuracy to board-certified dermatologists in visual inspection [13,14]. However, the diagnostic performance of DL algorithms for the detection of suspicious pigmented skin lesions in primary care settings has not been systematically evaluated.

Therefore, this systematic review and meta-analysis aimed to evaluate the diagnostic accuracy of DL algorithms for the detection of suspicious pigmented skin lesions in primary care settings.

Review

Methods

The present systematic review and meta-analysis were prepared in concordance with the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [15].

Search Strategy

A systematic literature search was conducted across multiple electronic databases, including MEDLINE via PubMed, Scopus, IEEE Xplore, Cochrane Central Register of Controlled Trials (CENTRAL), and Web of Science databases. The search strategy for PubMed and IEEE Xplore were as follows: (((((Learning, Deep OR Hierarchical Learning) AND (Intelligence, Artificial OR Computational Intelligence OR Machine Intelligence OR Computer Reasoning OR Computer Vision Systems OR Computer Vision System OR Knowledge Acquisition) AND (machine learning OR Transfer Learning) AND (skin lesions OR skin cancer OR Skin Neoplasm) AND (melanoma OR Melanomas OR Malignant Melanoma)))))). For the remaining databases, we employed the following search strategy: (Computer Vision Systems OR Deep learning) AND (skin cancer) AND (melanoma OR Malignant Melanoma).

Eligibility Criteria and Screening

We included all studies that utilized DL algorithms for the detection of suspicious pigmented skin lesions in primary care settings. We limited our inclusion to studies that reported sufficient data to calculate the diagnostic accuracy of DL algorithms, including true positives, false positives, true negatives, and false negatives. Studies that were non-English, case reports, case series, review articles, editorial comments, letters, and studies that focused on the use of DL algorithms outside of the primary care setting or for non-pigmented lesions were excluded.

A two-step screening process was conducted to identify studies that met the predefined inclusion criteria. First, two reviewers independently screened the titles and abstracts of all retrieved studies to identify potentially eligible studies. Then, the full texts of the potentially eligible studies were obtained and assessed independently by the same two reviewers to determine their eligibility for inclusion. Disagreements between the reviewers were resolved by discussion to reach a final decision. If a consensus could not be reached, a third reviewer was consulted to make a final decision.

Data Extraction and Quality Assessment

Data extraction was performed systematically to extract relevant information from the included studies. The data extraction process was conducted by three reviewers who independently extracted data from each included study and recorded it in a predesigned Excel sheet (Microsoft Corporation, Redmond, WA). The extracted data included authors, year of publication, country, study design, demographic characteristics of the population, type of the lesion investigated, DL algorithms characteristics, and their sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic (ROC) curve.

The risk of bias assessment of the included studies was conducted using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) score to evaluate the methodological quality and applicability of primary diagnostic accuracy in the included studies [16]. The risk of bias was assessed using four domains: patient selection, index test, reference standard, flow, and timing. Studies were rated as "low risk of bias," "high risk of bias," or "unclear risk of bias."

Data Analysis

The statistical analysis was conducted using the meta package in R software version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria). We performed a bivariate random-effects meta-analysis model to synthesize sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratios (DOR) from the included studies. The diagnostic performance across studies was summarized using the summary receiver operating characteristic (SROC) curve, along with the calculation of the area under the curve (AUC). Heterogeneity among the included studies was assessed using the I^2 statistic, where values of 25%, 50%, and 75% represent low, moderate, and high heterogeneity, respectively. Cochran's Q test was also used to evaluate the statistical significance of observed heterogeneities. A p-value of less than 10% represented significant heterogeneity.

Results

Literature Search Results

A total of 473 unique records were identified through the literature search. A total of 202 records were excluded during the initial screening phase, and 271 full-text articles were assessed for eligibility. Of them, three studies were included in the present systematic review and meta-analysis (Figure 1) [12,14,17].

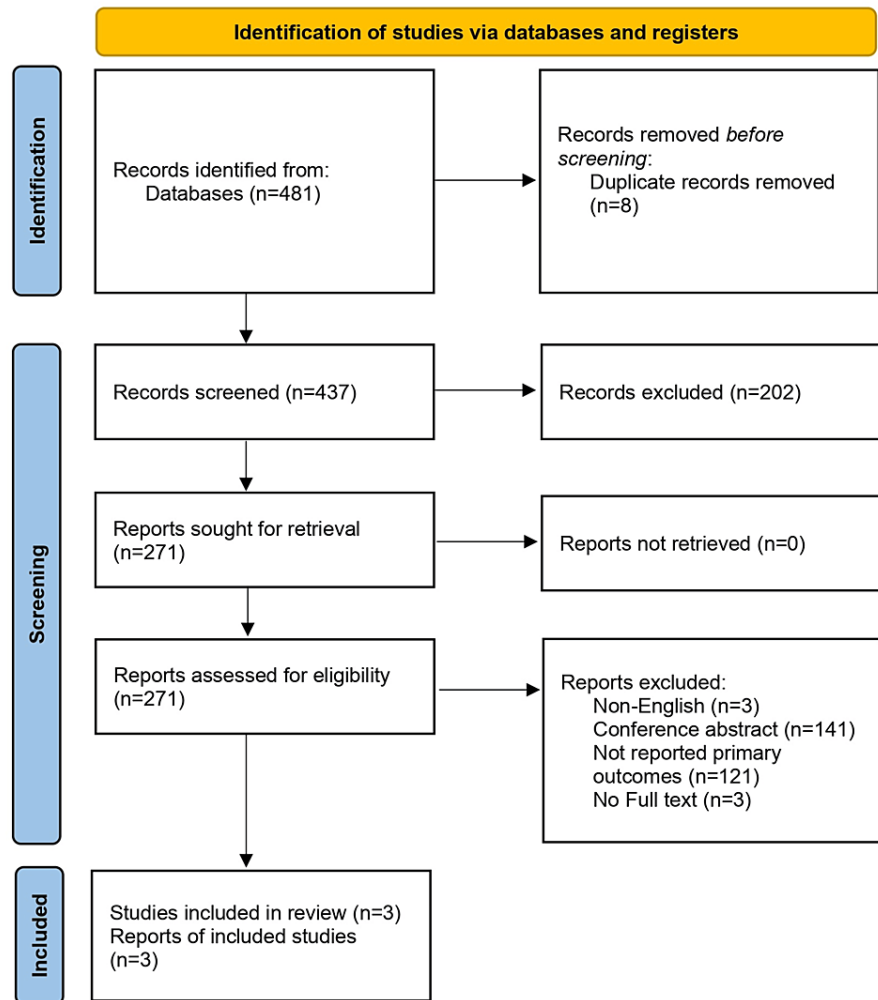


FIGURE 1: PRISMA 2020 flow diagram.

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Characteristics of the Included Studies

Three studies employed DL techniques to detect pigmented skin lesions in primary settings. Sangers et al. conducted a prospective diagnostic accuracy study using a convolutional neural network (CNN, version RD-174). The study involved 372 participants with a median age of 71 years (range = 58-78), and approximately half were males (49.2%). Out of the 785 skin lesions assessed, 275 were identified as premalignant or malignant and 510 as benign. The main finding was that the AI-powered, commercially available application exhibited high sensitivity and specificity in detecting suspicious lesions [17]. Birkenfeld et al. also carried out a prospective diagnostic accuracy study but used logistic regression combined with principal component analysis (PCA). The sample included 133 individuals, predominantly male (54.13%), with ages ranging from 16 to 76 years. The study analyzed a total of 1759 lesions, distributed between a training set of 1187 lesions and a test set of 572 lesions. The results highlighted that the DL-powered CAD system could efficiently differentiate between suspicious and non-suspicious lesions [12]. Soenksen et al. utilized a deep convolutional neural network (DCNN) in a retrospective diagnostic accuracy study involving a large dataset of 33,980 images, which included 4,063 suspicious pigmented lesions (SPLs). Although specific demographic details were not provided, the study concluded that the DL-powered tool enabled accurate detection of SPLs within a primary care setting [14], as shown in Table 1.

Authors, year	Country	Study design	DL algorithm	Minimum specifications	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	AUC, % (95% CI)	Accuracy, % (95% CI)
Sangers et al. (2022) [17]	Netherlands	Prospective diagnostic accuracy study	CNN (version RD-174)	12-megapixel camera running either an Android 10 (Galaxy S9, Samsung, Seoul, South Korea) or iOS 13 (iPhone XR, Apple Inc., Cupertino, CA)	86.9% (82.3, 90.7)	70.4% (66.2, 74.3)	61.3% (57.9, 64.6)	90.9% (88, 93.2)	NR	76.2% (73.0, 79.1)
Birkenfeld et al. (2020) [12]	Spain	Prospective diagnostic accuracy study	Logistic regression and PCA	10-megapixel camera with camera-specific software (OLYMPUS Viewer 3, OM Digital Solutions, Tokyo, Japan)	84.0% (NR)	72.1% (NR)	58.2% (NR)	90.7% (NR)	0.89 (0.85, 0.92)	75.9% (NR)
Soenksen et al. (2021) [14]	Spain	Retrospective diagnostic accuracy study	DCNN	10-megapixel camera with camera-specific software (Olympus E-420, 14- to 42-mm lens, Olympus Corporation, Tokyo, Japan)	90.3% (90, 90.6)	89.9% (89.6, 90.2)	NR	NR	0.97 (0.969, 0.971)	86.6% (86.3, 86.8)

TABLE 1: Summary characteristics of the included studies.

DCNN: deep convolutional neural network; PCA: principal component analysis; CNN: convolutional neural network; DL: deep learning; PPV: positive predictive value; NPV: negative predictive value; NR: not reported.

Quality Assessment

Birkenfeld et al. [12] showed high patient selection bias, while the subsequent two had a low risk of bias. However, all studies demonstrated low bias in index tests, reference standards, flow, and timing (Figure 2).

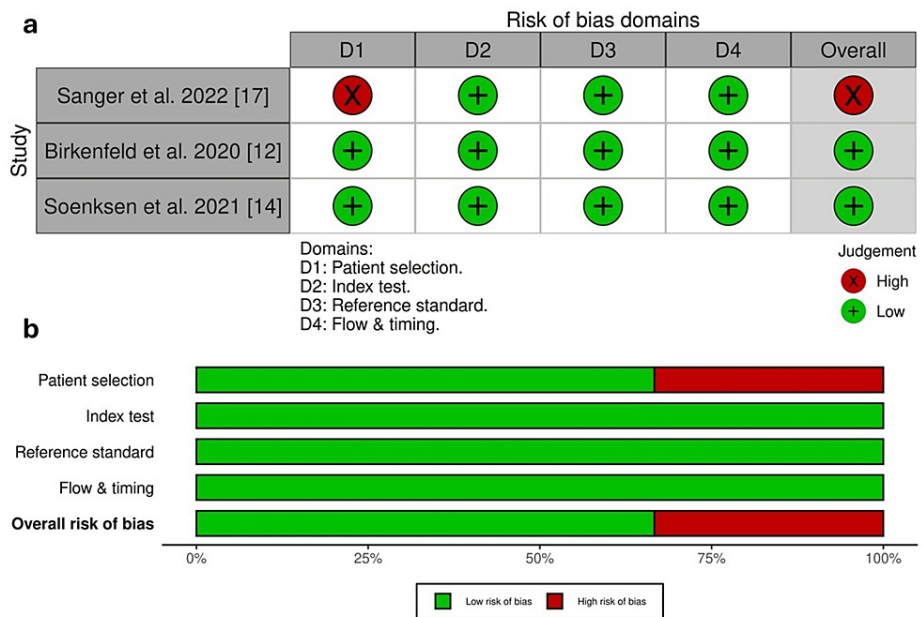


FIGURE 2: The risk of bias assessment was performed using the QUADAS-2 score to evaluate the quality of the included observational studies.

Birkenfeld et al. 2020 [12], Soenksen et al. 2021 [14], and Sangers et al. 2022 [17].

D1: patient selection; D2: index test; D3: reference standard; D4: flow & timing.

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies 2.

Outcomes

The pooled sensitivity was 0.90 (95% CI: 0.90-0.91), indicating a high ability to identify true positives correctly (Figure 3A). Significant heterogeneity exists among the studies ($p = 0.006$, $I^2 = 80.3\%$). Likewise, the pooled specificity was 0.85 (95% CI: 0.84-0.86), indicating a strong ability to correctly identify true negatives (Figure 3B). Significant heterogeneity is observed among the studies ($p < 0.001$, $I^2 = 98.8\%$). In terms of PLR, the pooled effect size was 4.30 (95% CI: 1.65-11.18), indicating a moderate ability to correctly identify true positives (Figure 3C); there was substantial heterogeneity ($p < 0.001$, $I^2 = 99.3\%$). The pooled NLR was 0.16 (95% CI: 0.09-0.29), indicating a moderate ability to rule out the condition (Figure 3D); there was a significant heterogeneity ($p < 0.001$, $I^2 = 94.8\%$).

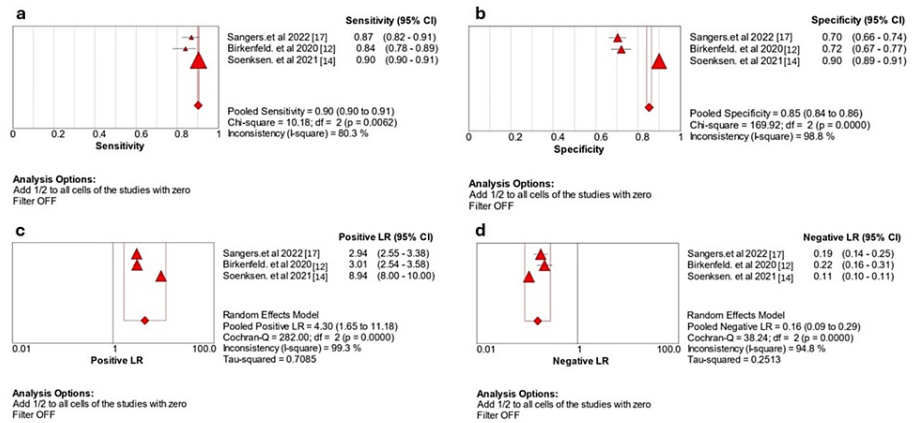
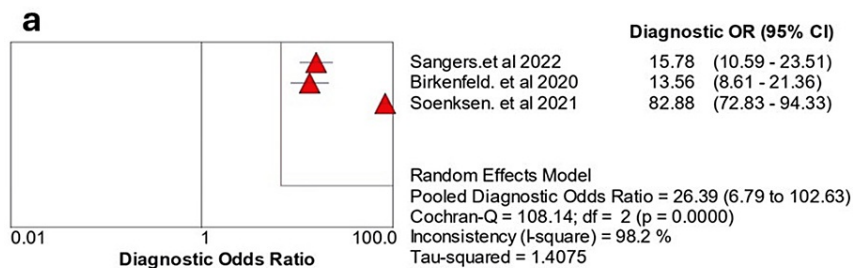


FIGURE 3: Data presents (a) a meta-analysis of sensitivity across three studies, (b) a meta-analysis of specificity across three studies, (c) a meta-analysis of positive likelihood ratios from three studies, and (d) a meta-analysis of negative likelihood ratios from three studies.

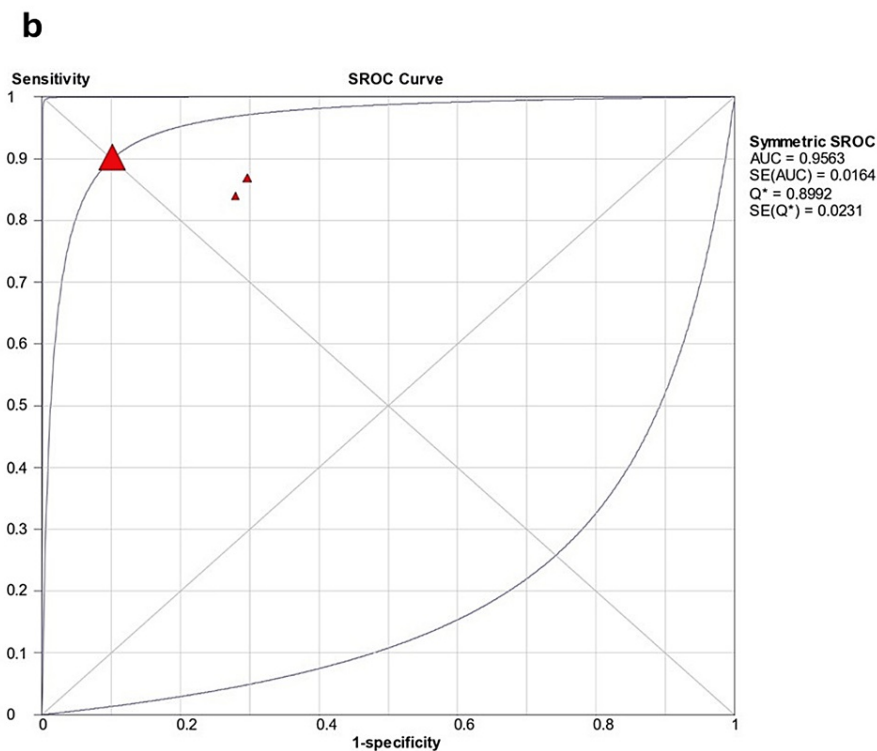
Birkenfeld et al. 2020 [12], Soenksen et al. 2021 [14], and Sangers et al. 2022 [17].

LR: likelihood ratio.

Figure 4A shows the findings of the DOR from three studies. The individual diagnostic ORs ranged from 15.78 to 82.88. When combined using a random effects model, the pooled DOR was 26.39 (95% CI: 6.79-102.63), indicating a substantial overall diagnostic performance. However, the studies had significant heterogeneity ($p < 0.001$, $I^2 = 98.2\%$). The SROC curve showed that the pooled AUC was 0.9563, indicating high discriminatory power (Figure 4B).



Analysis Options:
Add 1/2 to all cells of the studies with zero
Filter OFF



Analysis Options:
Add 1/2 to all cells of the studies with zero
Filter OFF
Symmetric SROC curve fitted using Moses' Model [Weighted regression (Inverse Variance)]
Defined relevant region: All ROC spac

FIGURE 4: The findings of the meta-analysis of diagnostic odds ratios (a) and summary receiver operating characteristic curve (b).

Birkenfeld et al. 2020 [12], Soenksen et al. 2021 [14], and Sangers et al. 2022 [17].

OR: odds ratio; SROC: summary receiver operating characteristic curve; ROC: receiver operating characteristic; AUC: area under the curve.

Discussion

Our research aimed to investigate and shed light on the effectiveness of DL algorithms in the early diagnosis of patients suffering from suspicious pigmented skin lesions in primary care settings. We compared the performance of these advanced technologies with the usual diagnostic methods followed by doctors to gain a comprehensive understanding of the potential benefits of DL algorithms. The sensitivity analysis revealed the high sensitivity and specificity of the DL algorithms for detecting suspicious pigmented skin lesions in primary care settings. Our data also suggested that the DL algorithms had increased the odds of correctly diagnosing suspicious pigmented skin lesions, with an excellent discriminative ability to distinguish between benign and malignant pigmented skin lesions. Consequently, they can serve as a reliable tool for primary care physicians to detect skin cancers such as melanoma early. However, our results

should be interpreted cautiously as significant heterogeneities in the pooled analyses suggest variations across the studies. This could be attributed to differences in dataset characteristics, algorithm design, lesion type variations, the doctors' expertise, or variations in the reference standard used for comparison.

Our findings align with previous studies by Birkenfeld et al. [12] and Soenksen et al. [14], who also reported significant improvements in suspicious pigmented skin lesion outcomes following similar interventions in different populations. Furthermore, Tschandl et al. also found that when an AI algorithm supports a physician's diagnosis-making, the diagnostic accuracy improves over that of either AI or physicians alone [18]. These findings have also been confirmed by other researchers [19].

Despite the promising results, there are several challenges and limitations that need to be acknowledged. Publication bias may exist, as studies with significant positive results are more likely to be published. Additionally, the quality of the included studies varied, and some may have introduced biases that could affect the overall findings. Moreover, one major limitation is the lack of interpretability of these algorithms. Hence, the accuracy of these algorithms is hard to detect when they are applied without any physician input. Besides, it is hard to explain how these algorithms come to their results. ML algorithm often functions as a black box [17] that takes in inputs and produces outputs with no interpretation of how it produced the conclusions and results. This lack of transparency can pose limitations in gaining the trust and acceptance of healthcare physicians [20]. Nevertheless, our study highlights the potential benefits of integrating AI and ML technologies, specifically DL algorithms, into routine primary care practice. These technologies can potentially enhance early detection, improve patient outcomes, and alleviate the burden on dermatologists [18]. Recently, DL methods have also been explored in the non-invasive diagnosis of skin lesions and demonstrated their ability to classify skin lesions with high accuracy. However, some other opinions believe that it is difficult to enable AI and ML technologies in daily dermatological examinations [21]. From a policy perspective, implementing large-scale skin cancer screening programs is not only likely to be a complex task but will also be infeasible in most resource-limited healthcare systems worldwide. In the United States, for example, there are fewer than 12,000 practicing dermatologists [22], and with fewer than 15 visits per 100 individuals per year [22], it is expected that most dermatology practices across the world are already too saturated and time-constrained to provide additional screening services.

Integrating DL algorithms in dermatology highlights the importance of ongoing education and training for dermatologists and primary care physicians. Continued professional development programs can help physicians stay updated with the latest technological advancements and ensure their competent use in clinical practice [20]. Collaboration between medical schools, technology companies, and hospitals can facilitate the development of specialized training programs that equip healthcare professionals with the essential skills to effectively utilize DL algorithms. A recent study by Hekler et al. found that combining humans and AI achieves a better classification of images than only dermatologists or only classification by CNN [22]. The mean accuracy increased by 1.36% when dermatologists worked together with ML and AI.

Dermatologists can contribute their expertise in curating high-quality datasets, training the algorithms, and validating their performance to optimize the use of DL algorithms in primary care settings. Primary care physicians can provide valuable insights into the practical implementation of these algorithms and ensure their integration into existing healthcare workflows. Additionally, exploring potential combination therapies involving AI and other interventions may yield synergistic effects and further enhance skin cancer outcomes.

Future research in this field should focus on refining DL algorithms to enhance their performance, reliability, and interpretability. Prospective studies with larger sample sizes and diverse patient populations are needed to validate the findings of this meta-analysis. Additionally, long-term follow-up studies can assess the impact of AI and ML algorithms on patient outcomes, including detecting early-stage skin cancers and reducing mortality rates. Human-machine collaboration has revealed promising results for future applications. For further expansion in this field, machines could assist physicians in time-consuming practices that usually are not being applied.

Conclusions

DL algorithms have the potential to significantly improve the detection of suspicious pigmented skin lesions in primary care settings. Our analysis showed that DL exhibited promising performance in the early detection of suspicious pigmented skin lesions. However, further studies are needed. Furthermore, the integration of DL algorithms into primary care could potentially reduce the number of unnecessary biopsies and referrals, thereby optimizing resource allocation and improving patient outcomes. It is also crucial to investigate the potential challenges and limitations of DL implementation in real-world clinical settings, such as data privacy concerns and the need for standardized training datasets, to ensure its safe and effective use in dermatological practice.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Eshak I. Bahbah, Ahmed R. Abdalla, Ahmed Elgebaly

Acquisition, analysis, or interpretation of data: Eshak I. Bahbah, Ahmed R. Abdalla, Ahmed W. Hageen, Haneen H. Saleh, Omar Al-Azzawi, Amani Harraz, Bola S. Eldoqsh, Fatma E. Elawady, Ayman H. Alhammadi, Hesham Hassan Elmorsy, Ahmed Elgebaly, Majd Jano, Mohamed Elmasry, Mahmoud Ghalab

Drafting of the manuscript: Eshak I. Bahbah, Ahmed R. Abdalla, Ahmed W. Hageen, Haneen H. Saleh, Omar Al-Azzawi, Amani Harraz, Bola S. Eldoqsh, Fatma E. Elawady, Ayman H. Alhammadi, Hesham Hassan Elmorsy, Ahmed Elgebaly, Majd Jano, Mohamed Elmasry, Mahmoud Ghalab

Critical review of the manuscript for important intellectual content: Eshak I. Bahbah, Ahmed R. Abdalla, Ahmed W. Hageen, Haneen H. Saleh, Omar Al-Azzawi, Amani Harraz, Bola S. Eldoqsh, Fatma E. Elawady, Ayman H. Alhammadi, Hesham Hassan Elmorsy, Ahmed Elgebaly, Majd Jano, Mohamed Elmasry, Mahmoud Ghalab

Supervision: Eshak I. Bahbah, Ahmed R. Abdalla, Ahmed Elgebaly, Mohamed Elmasry

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Arnold M, Singh D, Laversanne M, et al.: Global burden of cutaneous melanoma in 2020 and projections to 2040. *JAMA Dermatol.* 2022, 158:495-503. [10.1001/jamadermatol.2022.0160](https://doi.org/10.1001/jamadermatol.2022.0160)
2. Saginala K, Barsouk A, Aluru JS, Rawla P, Barsouk A: Epidemiology of melanoma. *Med Sci (Basel).* 2021, 9:63. [10.3390/medsci9040063](https://doi.org/10.3390/medsci9040063)
3. American Cancer Society. Survival rates for melanoma skin cancer . (2024). Accessed: February 10, 2024: <https://www.cancer.org/cancer/types/melanoma-skin-cancer/detection-diagnosis-staging/survival-rates-for-melanoma-skin...>
4. Leiter U, Keim U, Garbe C: Epidemiology of skin cancer: update 2019. *Adv Exp Med Biol.* 2020, 1268:123-39. [10.1007/978-3-030-46227-7_6](https://doi.org/10.1007/978-3-030-46227-7_6)
5. Gaudy-Marqueste C, Wazaefi Y, Bruneu Y, et al.: Ugly duckling sign as a major factor of efficiency in melanoma detection. *JAMA Dermatol.* 2017, 153:279-84. [10.1001/jamadermatol.2016.5500](https://doi.org/10.1001/jamadermatol.2016.5500)
6. Westerhoff K, McCarthy WH, Menzies SW: Increase in the sensitivity for melanoma diagnosis by primary care physicians using skin surface microscopy. *Br J Dermatol.* 2000, 143:1016-20. [10.1046/j.1365-2133.2000.03836.x](https://doi.org/10.1046/j.1365-2133.2000.03836.x)
7. Burton RC, Howe C, Adamson L, et al.: General practitioner screening for melanoma: sensitivity, specificity, and effect of training. *J Med Screen.* 1998, 5:156-61. [10.1136/jms.5.3.156](https://doi.org/10.1136/jms.5.3.156)
8. Brown AE, Najmi M, Duke T, Grabell DA, Koshelev MV, Nelson KC: Skin cancer education interventions for primary care providers: a scoping review. *J Gen Intern Med.* 2022, 37:2267-79. [10.1007/s11606-022-07501-9](https://doi.org/10.1007/s11606-022-07501-9)
9. Das K, Cockerell CJ, Patil A, Pietkiewicz P, Giulini M, Grabbe S, Goldust M: Machine learning and its application in skin cancer. *Int J Environ Res Public Health.* 2021, 18:13409. [10.3390/ijerph182413409](https://doi.org/10.3390/ijerph182413409)
10. Liu Y, Jain A, Eng C, et al.: A deep learning system for differential diagnosis of skin diseases . *Nat Med.* 2020, 26:900-8. [10.1038/s41591-020-0842-3](https://doi.org/10.1038/s41591-020-0842-3)
11. Patel RH, Foltz EA, Witkowski A, Ludzik J: Analysis of artificial intelligence-based approaches applied to non-invasive imaging for early detection of melanoma: a systematic review. *Cancers (Basel).* 2023, 15:4694. [10.3390/cancers15194694](https://doi.org/10.3390/cancers15194694)
12. Birkenfeld JS, Tucker-Schwartz JM, Soenksen LR, Avilés-Izquierdo JA, Marti-Fuster B: Computer-aided classification of suspicious pigmented lesions using wide-field images. *Comput Methods Programs Biomed.* 2020, 195:105631. [10.1016/j.cmpb.2020.105631](https://doi.org/10.1016/j.cmpb.2020.105631)
13. Ding H, Zhang E, Fang F, et al.: Automatic identification of benign pigmented skin lesions from clinical images using deep convolutional neural network. *BMC Biotechnol.* 2022, 22:28. [10.1186/s12896-022-00755-5](https://doi.org/10.1186/s12896-022-00755-5)
14. Soenksen LR, Kassis T, Conover ST, et al.: Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. *Sci Transl Med.* 2021, 13:eabb3652. [10.1126/scitranslmed.abb3652](https://doi.org/10.1126/scitranslmed.abb3652)
15. Page MJ, McKenzie JE, Bossuyt PM, et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021, 372:n71. [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)
16. Whiting PF, Rutjes AW, Westwood ME, et al.: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011, 155:529-36. [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)
17. Sangers T, Reeder S, van der Vet S, et al.: Validation of a market-approved artificial intelligence mobile health app for skin cancer screening: a prospective multicenter diagnostic accuracy study. *Dermatology.*

- 2022, 258:649-56. [10.1159/000520474](https://doi.org/10.1159/000520474)
18. Tschandl P, Rinner C, Apalla Z, et al.: Human-computer collaboration for skin cancer recognition. *Nat Med.* 2020, 26:1229-34. [10.1038/s41591-020-0942-0](https://doi.org/10.1038/s41591-020-0942-0)
 19. Han SS, Park I, Eun Chang S, et al.: Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. *J Invest Dermatol.* 2020, 140:1753-61. [10.1016/j.jid.2020.01.019](https://doi.org/10.1016/j.jid.2020.01.019)
 20. Chan S, Reddy V, Myers B, Thibodeaux Q, Brownstone N, Liao W: Machine learning in dermatology: current applications, opportunities, and limitations. *Dermatol Ther (Heidelb).* 2020, 10:365-86. [10.1007/s13555-020-00372-0](https://doi.org/10.1007/s13555-020-00372-0)
 21. Yu L, Chen H, Dou Q, Qin J, Heng PA: Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans Med Imaging.* 2017, 36:994-1004. [10.1109/TMI.2016.2642839](https://doi.org/10.1109/TMI.2016.2642839)
 22. Hekler A, Utikal JS, Enk AH, et al.: Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer.* 2019, 120:114-21. [10.1016/j.ejca.2019.07.019](https://doi.org/10.1016/j.ejca.2019.07.019)