THE ANALYSIS STUDY OF ARTIFICIAL INTELIGENCE FOR SKIN CANCER: A COMPREHENSIVE SYSTEMATIC REVIEW

*Tia Alviani Juwita, Indah Sari Siregar

*1Faculty of Medicine, HKBP Nommensen University, Medan, Indonesia
2Faculty of Medicine, North Sumatera University, Medan, Indonesia

Corresponding Author:
indhsarisrg@gmail.com

ABSTRACT

Background: Skin cancer diagnosis relies heavily on the interpretation of visual patterns, making it a complex task that requires extensive training in dermatology and dermatoscopy. However, AI algorithms have been shown to accurately diagnose skin cancers, even outperforming experienced dermatologists in image classification tasks in constrained settings.

The aim: The aim of this study to show about artificial intelligence for skin cancer.

Methods: By the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020, this study was able to show that it met all of the requirements. This search approach, publications that came out between 2014 and 2024 were taken into account. Several different online reference sources, like Pubmed, SagePub, and Science Direct were used to do this. It was decided not to take into account review pieces, works that had already been published, or works that were only half done.

Result: Eight publications were found to be directly related to our ongoing systematic examination after a rigorous three-level screening approach. Subsequently, a comprehensive analysis of the complete text was conducted, and additional scrutiny was given to these articles.

Conclusion: The use of AI has high potential to facilitate the way skin cancer is diagnosed. Two main branches of AI are used to detect and classify skin cancer, namely shallow and deep techniques.

Keyword: Artificial intelligence, diagnostic, skin cancer.
INTRODUCTION
Artificial intelligence (AI) stands at the forefront of technological innovation and has permeated into almost every industry and field. In dermatology, significant progress has been made toward the application of AI in skin cancer screening and diagnosis. Notably, a milestone that marked the era of modern artificial intelligence in dermatology was the demonstration of skin cancer classification abilities by deep learning convolutional neural networks (CNNs), which was on par with the performance of board-certified dermatologists. This CNN was trained on a dataset that was two orders of magnitude greater than those previously utilized. The dermatologist-level classification ability has since been experimentally validated by other papers. Recent progress in the field of AI enables models to not only analyze image data but also integrate clinical information, including patient demographics and past medical history. Advancements allow for the simultaneous evaluation and identification of multiple lesions from wide-field images.1–3

Skin cancer is the abnormal growth of skin cells. The cancerous growth may affect both the layers—dermis and epidermis, but this review is concerned primarily with epidermal skin cancer; the two types of skin cancers that can arise from the epidermis are carcinomas and melanomas, depending on their cell type keratinocytes or melanocytes, respectively. It is a challenge to estimate the incidence of skin cancer due to various reasons, such as the multiple sub-types of skin cancer. This poses as a problem while collating data, as non-melanoma is often not tracked by registries or are left incomplete because most cases are treated via surgery. As of 2020, the World Cancer Research Fund International reported a total of 300,000 cases of melanoma in skin, and a total of 1,198,073 cases of non-melanoma skin cancer. The reasons for the occurrence of skin cancer cannot be singled out, but they include and are not limited to exposure to ultraviolet rays, family history, or a poor immune system. The affected spot on the skin is called a lesion, which can be further segregated into multiple categories depending on its origin.4–6

Cancer is one of the major healthcare burdens across the world. Global statistics suggest almost 10.0 million deaths (9.9 million excluding non-melanoma skin cancer) due to cancer in the year 2020. The most commonly diagnosed cancers include breast cancer in females, lung cancer, and prostate cancers. Lung, liver, and stomach cancers are the major contributors of cancer related deaths. Skin cancer, including both malignant melanoma and non-melanoma skin cancer (NMSC), are common cancers in Caucasians and their incidence is on the rise. According to the US Skin Cancer Foundation, skin cancer affects more people in the United States each year than all other cancers combined.7

Melanoma is the skin cancer with the worst prognosis. If diagnosed early, it can be treated successfully with surgical procedures. However, once there is metastasis, rates of survival are reduced significantly. Diagnosis of melanoma depends on the clinical examination and classic findings on the lesion biopsy. Examples of NMSC include basal cell carcinoma (NMSC) and squamous cell carcinoma. The success of skin cancer depends on early diagnosis and appropriate treatment. Visual inspection may not be sufficient to differentiate benign lesions from malignant tumors. The gold standard procedure is histopathology examination of the skin biopsy. The invasive nature of the procedure, associated pain, and the need for repeated samples in suspected lesions with varied presentations are some of the limitations for skin biopsy. Non-invasive tools can also assist in clinical diagnosis.7–9

METHODS

PROTOCOL
By following the rules provided by Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020, the author of this study made certain that it was up to par with the requirements. This is done to ensure that the conclusions drawn from the inquiry are accurate.

CRITERIA FOR ELIGIBILITY
For the purpose of this literature review, we compare and contrast artificial intelligence for skin cancer. It is possible to accomplish this by researching of artificial intelligence for skin cancer. As the primary purpose of this piece of writing, demonstrating the relevance of the difficulties that have been identified will take place throughout its entirety.

In order for researchers to take part in the study, it was necessary for them to fulfil the following requirements: 1) The paper needs to be written in English, and it needs to determine about artificial intelligence for skin cancer. In order for the manuscript to be considered for publication, it needs to meet both of these requirements. 2) The studied papers include several that were published after 2014, but before the time period that this systematic review deems to be relevant. Examples of studies that are not permitted include editorials, submissions that do not have a DOI, review articles that have already been published, and entries that are essentially identical to journal papers that have already been published.

SEARCH STRATEGY
We used " Artificial intelligence for skin cancer." as keywords. The search for studies to be included in the systematic review was carried out using the PubMed, SagePub, and Science Direct databases by inputting the words: {"Skin cancer"[MeSH Subheading] OR " Artificial intelligence"[All Fields] OR "Diagnostic" [All Fields]} AND
DATA RETRIEVAL
After reading the abstract and the title of each study, the writers performed an examination to determine whether or not the study satisfied the inclusion criteria. The writers then decided which previous research they wanted to utilise as sources for their article and selected those studies. After looking at a number of different research, which all seemed to point to the same trend, this conclusion was drawn. All submissions need to be written in English and cannot have been seen anywhere else.

**Identification of studies via databases and registers**

Records identified from*:
- PubMed (n: 227)
- SageJournal (n: 329)
- Sciencedirect (n: 1146)

Records screened (418)

Reports sought for retrieval (10)

Reports assessed for eligibility (10)

Studies include in systematic review (8)

Records remove before screening:
- Duplicate records removed (986)
- Records marked as ineligible by automation tools (297)
- Records remove for other reasons (1)

Records exclude*:
- Wrong population (382)
- Wrong study design (12)
- Wrong intervention (13)
- Wrong publication type (1)

Reports not retrieved (0)

Reports exclude (2) due to:
- No comparison (1)
- Wrong intervention (1)

**Figure 1. Article search flowchart**

Only those papers that were able to satisfy all of the inclusion criteria were taken into consideration for the systematic review. This reduces the number of results to only those that are pertinent to the search. We do not take into consideration the conclusions of any study that does not satisfy our requirements. After this, the findings of the research will be analysed in great detail. The following pieces of information were uncovered as a result of the inquiry that was carried out for the purpose of this study: names, authors, publication dates, location, study activities, and parameters.

QUALITY ASSESSMENT AND DATA SYNTHESIS
Each author did their own study on the research that was included in the publication’s title and abstract before making a decision about which publications to explore further. The next step will be to evaluate all of the articles that are suitable for inclusion in the review because they match the criteria set forth for that purpose in the review. After that, we'll
determine which articles to include in the review depending on the findings that we've uncovered. This criteria is utilised in the process of selecting papers for further assessment. in order to simplify the process as much as feasible when selecting papers to evaluate. Which earlier investigations were carried out, and what elements of those studies made it appropriate to include them in the review, are being discussed here.

RESULT
Using reputable resources like Science Direct, PubMed, and SagePub, our research team first gathered 1702 publications. A thorough three-level screening strategy was used to identify only eight papers as directly relevant to our ongoing systematic evaluation. Next, a thorough study of the entire text and further examination of these articles were selected. Table 1 compiles the literature that was analyzed for this analysis in order to make it easier to view.

Table 1. The literature include in this study

<table>
<thead>
<tr>
<th>Author</th>
<th>Origin</th>
<th>Method</th>
<th>Sample</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu, Q et al., 2023</td>
<td>China</td>
<td>On July 14, 2023, articles and reviews about the application of AI in skin cancer, spanning the years from 1900 to 2023, were selected from the Web of Science Core Collection.</td>
<td>512</td>
<td>A total of 512 papers matching the search terms and inclusion/exclusion criteria were published between 1991 and 2023. The United States leads in publications with 149, followed by India with 61. Germany holds eight positions among the top 10 institutions, while the United States has two. The most prevalent journals cited were Cancer, the European Journal of Cancer, and Sensors. The most frequently cited keywords include “skin cancer”, “classification”, “artificial intelligence”, and “deep learning”.</td>
</tr>
<tr>
<td>Jagemann, I et al., 2024</td>
<td>Germany</td>
<td>A choice-based conjoint analysis was used to examine the acceptance of medical AI for a skin cancer screening from the patient’s perspective.</td>
<td>383</td>
<td>Among the 383 clicks on the survey link, a total of 126 (32.9%) respondents completed the online survey. The conjoint analysis showed that the three attributes had more or less equal importance in contributing to the participants’ choices, with provider being the most important attribute. Inspecting the individual part-worths of conjoint attributes showed that treatment by a physician was the most preferred modality, followed by electronic consultation with a physician and personalized AI; the lowest scores were found for the three AI levels. Concerning the relationship between sociodemographic characteristics and relative importance, only age showed a significant positive association to the importance of the attribute provider ($r=0.21, P=.02$), in which younger participants put less importance on the provider than older participants. All</td>
</tr>
<tr>
<td>Marsden, H et al., 2023&lt;sup&gt;12&lt;/sup&gt;</td>
<td>United Kingdom</td>
<td>The DERM-003 study was a prospective, multi-center, single-arm, masked study that aimed to demonstrate the effectiveness of an AI as a Medical Device (AIaMD) to identify Squamous Cell Carcinoma (SCC), Basal Cell Carcinoma (BCC), pre-malignant and benign lesions from dermoscopic images of suspicious skin lesions.</td>
<td>572 patients (49.5% female, mean age 68.5 years, 96.9% Fitzpatrick skin types I-III) were recruited from 4 UK NHS Trusts, providing images of 611 suspicious lesions. 395 (64.6%) lesions were biopsied; 47 (11%) were diagnosed as SCC and 184 (44%) as BCC. The AIaMD AUROC on images taken by iPhone 6S was 0.88 (95% CI: 0.83–0.93) for SCC and 0.87 (95% CI: 0.84–0.91) for BCC. For Samsung 10 the AUROCs were 0.85 (95% CI: 0.79–0.90) and 0.87 (95% CI, 0.83–0.90), and for the iPhone 11 they were 0.88 (95% CI, 0.84–0.93) and 0.89 (95% CI, 0.86–0.92) for SCC and BCC, respectively. Using pre-determined diagnostic thresholds on images taken on the iPhone 6S the AIaMD achieved a sensitivity and specificity of 98% (95% CI, 88–100%) and 38% (95% CI, 33–44%) for SCC; and 94% (95% CI, 90–97%) and 28% (95 CI, 21–35%) for BCC. All 16 lesions diagnosed as melanoma in the study were correctly classified by the AIaMD.</td>
<td></td>
</tr>
<tr>
<td>Jutzi, TB et al., 2020&lt;sup&gt;13&lt;/sup&gt;</td>
<td>Germany</td>
<td>A web-based questionnaire was designed using LimeSurvey, sent by e-mail to university hospitals and melanoma support groups and advertised on social media.</td>
<td>298 individuals (154 with a melanoma diagnosis, 143 without) responded to the questionnaire. About 94% [95% CI = 0.91–0.97] of respondents supported the use of artificial intelligence in medical approaches. 88% [95% CI = 0.85–0.92] would even make their own health data anonymously available for the further development of AI-based applications in medicine. Only 41% [95% CI = 0.35–0.46] of respondents were amenable to the use of artificial intelligence as stand-alone system, 94% [95% CI = 0.92–0.97] to its use as assistance system for physicians. In sub-group analyses, only minor differences were detectable. Respondents with a previous history of melanoma were more amenable to the use of AI applications for early detection even at home. They would</td>
<td></td>
</tr>
</tbody>
</table>
prefer an application scenario where physician and AI classify the lesions independently. With respect to AI-based applications in medicine, patients were concerned about insufficient data protection, impersonality and susceptibility to errors, but expected faster, more precise and unbiased diagnostics, less diagnostic errors and support for physicians.

| Marchetti, MA et al., 2023¹⁴ | USA | We conducted a prospective, observational clinical study to assess the diagnostic accuracy of the AI algorithm (ADAE) in predicting melanoma from dermoscopy skin lesion images. | 435 | Four hundred thirty-five participants were enrolled and contributed 603 lesions (95 melanomas). Participants had a mean age of 59 years, 54% were female, and 96% were White individuals. At the predetermined 95% sensitivity threshold, ADAE had a sensitivity of 96.8% (95% CI: 91.1–98.9%) and specificity of 37.4% (95% CI: 33.3–41.7%). The dermatologists’ ability to assess melanoma risk significantly improved after ADAE exposure (AUC 0.7798 vs. 0.8161, \( p = 0.042 \)). Post-ADAE dermatologist decisions also had equivalent or higher net benefit compared to biopsying all lesions. We validated the accuracy of an open-source melanoma AI algorithm and showed its theoretical potential for improving dermatology experts’ ability to evaluate lesions suspicious of melanoma. Larger randomized trials are needed to fully evaluate the potential of adopting this AI algorithm into clinical workflows. |
| Willingham, ML et al., 2021¹⁵ | Hawaii | To validate our AI’s efficiency in distinguishing between images, we utilized 50 images obtained from our ISIC dataset (n=25) and pathologically confirmed lesions (n=25). We compared the ability of our AI to The AI model better differentiated between melanoma vs non-melanoma with an AUC of 0.948. The three panel member dermatologists correctly diagnosed a similar number of images (n=35) as the AI program (n=34). Fleiss’ Kappa (\( \kappa \)) score for the raters and AI indicated fair (0.247) agreement. However, the combined result of the dermatologists panel with the AI assessments correctly identified 100% of the images from the test data set. | 50 |  |
visually diagnose these 50 skin cancer lesions with a panel of three dermatologists.

**Sangers, TE et al., 2023**

Netherlands

A qualitative focus group study, consisting of six focus groups with 16 dermatologists and 17 GPs, varying in prior knowledge and experience with AI, gender, and age.

33

Dermatologists and GPs perceive substantial benefits of AI, particularly an improved health outcome and care pathway between primary and secondary care. Doubts about accuracy, risk of health inequalities, and fear of replacement were among the most stressed barriers. Essential preconditions included adequate algorithm content, sufficient usability, and accessibility of AI. In conclusion, dermatologists and GPs perceive significant benefits from implementing AI in skin cancer care. However, to successfully implement AI, key barriers need to be addressed. Efforts should focus on ensuring algorithm transparency, validation, accessibility for all skin types, and adequate regulation of algorithms. Simultaneously, improving knowledge about AI could reduce the fear of replacement.

**Chevrier, MJ et al., 2014**

Canada

AI skin cancer publications were retrieved in June 2022 from the Web of Science. Publications were screened by title, abstract, and keywords to assess eligibility. Publications were fully reviewed.

168

A total of 168 articles were included: 25 on NMSC, 77 on melanoma, and 66 on skin cancer. The most common types of skin cancers were melanoma (134, 79.8%), basal cell carcinoma (61, 36.3%), and squamous cell carcinoma (45, 26.9%). All articles were published between 2000 and 2022, with 49 (29.2%) of them being published in 2021. Original studies that developed or assessed an algorithm predominantly used supervised learning (66, 97.0%) and deep neural networks (42, 67.7%). The most used imaging modalities were standard dermoscopy (76, 45.2%) and clinical images (39, 23.2%).

**DISCUSSION**

Artificial intelligence (AI) is transforming health care. Deep learning (DL) has become the dominant AI technology for high-dimensional complex data, such as images. In brief, DL leverages artificial neural networks, which learn complex mappings between inputs (e.g., images) and outputs (e.g., diagnoses) without explicit human engineering. Inspired by the brain, artificial neurons arranged in deep layers adapt the strength of their connections to one another as the model self-learns features from the input, such as visual patterns, that are most relevant for predicting the output. In experimental
settings across multiple specialties, DL performs equivalently to health-care professionals for detecting disease from medical imaging.\cite{18,19}

Modern machine learning techniques rely on vast datasets to identify patterns useful for classification. Diagnostic imaging represents one of the most promising arenas for AI research. Skin cancer detection in particular serves as an appealing application for AI, given that diagnoses often hinge on the subjective visual interpretation of clinical and dermoscopic images. AI-assisted diagnosis promises several advantages. For instance, AI could improve access to specialist-level expertise. The scarcity of dermatologists is a serious problem in many regions, often leading to protracted waiting times for specialist appointments. In addition, there is growing optimism that AI-based systems might offer greater consistency and higher accuracy than human experts. First demonstrated the efficacy of convolutional neural networks (CNNs) for the task of image-based classification in dermatology. CNNs are specialized types of neural networks that are optimally suited for image analysis and are predominantly trained using supervised learning techniques.\cite{20,21}

Skin cancer is the most common form of cancer worldwide. Over the past decade, there has been a concerning 27% increase in the annual diagnosis of invasive melanoma cases. Alarmingly, more than 5,400 people die from non-melanoma skin cancer every month. In the United States alone, the annual financial burden of treating skin cancer is estimated at a staggering US$8.1 billion, with approximately US$4.8 billion allocated to non-melanoma skin cancer and US$3.3 billion to melanoma. Among skin cancer types, basal cell carcinoma ranks as the most common, followed by squamous cell carcinoma and melanoma, which stands out as the most aggressive and lethal type of skin cancer. Merkel cell carcinoma also stands out among aggressive tumors. These tumors can arise anywhere on the body but are frequently observed in regions more exposed to the sun, including the face, neck, arms, and hands. Thus, there is an imperative need for sustained efforts to promote awareness and prevention of skin cancer.\cite{22,23,24}

The deployment of this AI skin app at a broad scale shows the real-world costs of more false positives (benign lesion claims) and fewer true positives (malignant lesion claims) compared to the management of non-app users. More false positives and fewer true positives compared with conventional care can take an emotional and financial toll on patients and the healthcare system.\cite{25}

The overall cost-effectiveness of the screening may be comparable to that of a dermatologist. A recent study in the US found that the cost of detecting an additional skin premalignancy or malignancy through total body exams was $2346. Depending on the assumptions of these calculations, the skin app performed at a comparable cost per new positive identification. In context, increased total costs per app user at a comparable cost-benefit ratio suggests that the app users are enjoying more of the “benefits”—i.e., they had more skin lesions diagnosed than non-app users, likely due to increased access. This supports using AI skin apps insofar as access is the limiting determinant of diagnosis.\cite{25}

**CONCLUSION**

The use of AI has high potential to facilitate the way skin cancer is diagnosed. Two main branches of AI are used to detect and classify skin cancer, namely shallow and deep techniques. However, the reliability of such AI tools is questionable since different data set sizes, image types, and number of diagnostic classes are being used and evaluated with different evaluation metrics.

**REFERENCES**


