Costs, benefits and diagnostic accuracy of machine learning applications for the autonomous detection of melanoma in high-risk individuals.

A thesis submitted in fulfilment of requirements for the degree of

Master of Philosophy

By

Angela Olsson

Faculty of Medicine and Health, The University of Sydney

Under the supervision of

Professor Rachael Morton and **Dr Mbathio Dieng**

June 2025

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purpose.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

This thesis is presented as a record of work conducted for the degree of Master of Philosophy at NHMRC Clinical Trials Centre, Faculty of Medicine and Health, The University of Sydney.

Acknowledgements

I would like to thank my supervisors Professor Rachael Morton, and Dr Mbathio Dieng for their patience, support and time. I would also like to thank my colleague Dr Jake Williams for his advice and support participating in and reviewing the manuscript for the protocol and scoping review. I would also like to acknowledge the assistance provided by the Library HDR team at The University of Sydney. Your help to structure the literature search which formed the basis of my research was much appreciated.

A special thanks goes to Professor Katherine Mills who was able to help me cross the finish line during a tumultuous final 12 months.

Last but not least, I would also like to recognise my family and friends, without who I would not have been able to complete this research. Lars, you are an unwavering support, full of light and positivity especially when I feel the need to discuss Turing's work in the early hours of the morning. Thank you.

The research reported in this thesis was supported by the award of a Research Training Program Fee Offset by the Australian Government. Portions of the text in this thesis were revised with the assistance of OpenAI's ChatGPT 4.0, which was used to improve clarity, grammar, and expression. All academic content, analysis, and interpretation are my own.

Table of Contents

List of Figures

List of Tables

Abbreviations

2D TBP	Two-dimensional total body photography
3D TBP	Three-dimensional total body photography
AI	Artificial intelligence
BCC	Basal cell carcinoma
CI	Confidence Interval
CNNs	Convolutional Neural Network
GP	General practitioner
ML	Machine Learning
NMSC	Non melanoma skin cancer
SCC	Squamous cell carcinoma
USPSTF	United States Preventative Services Task Force
UV	Ultra-violet

Abstract

Melanoma of the skin is one of the most frequently diagnosed cancers in Australia, with an estimated 18,257 new cases and over 1300 deaths in 2023. Australia continues to record the highest global incidence. Early detection of invasive melanoma is critical for survival, and although systemic therapies have advanced, timely diagnosis remains a priority. While population-level screening is not recommended, research has demonstrated that screening for melanoma in a high-risk clinic in Australia can be cost-effective.

Methods to improve melanoma detection include the use of artificial intelligence, or the practical application, machine learning. Machine learning uses algorithms to learn from images, recognising patterns within the image. Initial trials suggested dermatologists outperformed machine learning algorithms; however, more recent studies demonstrate machine learning achieving, and in some cases surpassing, dermatologist-level diagnostic accuracy.

This thesis examines the existing body of literature on the use of machine learning to detect melanoma in adults at high-risk of developing melanoma of the skin, focusing on diagnostic accuracy, costs and benefits. Chapter 1 provides an overview of melanoma and non-melanoma skin cancer; Chapter 2 is a narrative review of machine learning in healthcare. A simplified description of machine leaning as it applies to melanoma detection is provided. Chapter 3 outlines the protocol for the scoping review and Chapter 4 is the scoping review. 9,188 records were screened, of which 55 studies met the inclusion criteria. Machine learning demonstrated high diagnostic accuracy in

controlled settings with 78% of reader studies reporting ROAUC >0.8, compared to 60% of studies in a clinical setting. Only five studies reported on costs and benefits in high-risk populations, two conducted formal economic evaluations. One found no significant cost-effectiveness advantage, while the other reported cost savings and reduced clinician workload with machine learning. Patient and clinician preference studies indicated higher trust in clinician-led machine learning models (e.g., 2D/3D total body photography) compared with standalone consumer-led apps.

Evidence supporting the use of machine learning in high-risk populations remains limited, particularly regarding cost-effectiveness and studies in real-world clinical settings. Future research should prioritise prospective, real-world studies, cost benefit and cost-effectiveness analyses to better assess the utility of machine learning in melanoma surveillance and diagnosis, and to inform reimbursement policy.

Chapter 1 – Melanoma

The skin

The skin is the largest organ of the body. It has two main layers, the epidermis and dermis. There are three main cell types in the epidermal layer of the skin, including basal, squamous and melanocyte cells. Basal cells and squamous cells are keratinocytes, with basal cells originating from undifferentiated keratinocytes and squamous cells from differentiated keratinocytes.[1] Basal cells are located in the deepest part of the epidermis and squamous cells are found in the upper layers of the epidermis, above the basal layer. Squamous cells form the skins protective barrier. Melanocytes are epidermal cells that sit between the basal cells in the deeper level of the epidermis and produce melanin which is the pigment that gives skin its colour. When skin is exposed to ultraviolet (UV) radiation, melanocytes produce melanin to protect the skin. Melanocytes are also found in naevi on the skin. [2, 3] The dermis is the layer of skin that sits below the epidermis and is made up of fibrous tissue, hair follicles, sweat glands, blood vessels, lymph vessels and nerves. [2, 3]

Skin cancer

Skin cancer may be broadly categorised into two distinct types, melanoma and non-melanoma skin carcinoma. Non-melanoma skin cancer or basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) may also be referred to as keratinocytic carcinomas as they originate from keratinocytes.[2] When unregulated growth occurs in the cells of the epidermis, it may lead to skin cancer.[3]

The three main types of skin cancer include[2, 4-7]

- Basal Cell Carcinoma (BCC) which originate from the basal cells in the epidermis and is the most common type of skin cancer.
- Squamous Cell Carcinoma (SCC) which arise from squamous cells and is the second most common type of skin cancer; and
- Melanoma which originates in the melanocyte cells of the epidermis.

Even though melanoma is less common than the non-melanoma skin cancers BCC and SCC, it has the highest mortality due to its propensity to metastasise and spread.(6) Melanoma represents only 2-3% of all skin cancers detected, however, it leads to 65% of all skin cancer related deaths.(8)

Characteristics of people with higher risk for developing melanoma

Characteristics of people with higher risk for developing melanoma include fair skin, light hair and eye colour, skin sensitivity, exposure to UV radiation, naevus (mole) count, personal or family history, exposure to certain chemicals and weak immune systems.(9)

Incidence and mortality risk

Melanoma is one of the leading causes of cancer in Australia.(6) The incidence of melanoma in Australia is the highest in the world with 18,257 cases of melanoma diagnosed in 2023.(6, 10) The incidence of melanoma in Australia has steadily increased over the past few decades, however the lifetime mortality risk for melanoma, or risk of dying from melanoma, is starting to decrease.(6) This reduction in lifetime mortality risk may correspond to improvements in the treatment of melanoma or from the result of risk minimisation programs including the 'slip slop slap' disease awareness campaign. (6, 11-14) The 5-year survival rate for melanoma of the skin in 2016–2020 was 94% and is the highest rate recorded.[6]

The earlier invasive melanoma is detected, the greater chance of survival. This increase in survival with early detection is illustrated by the difference in survival rates between early and late-stage melanoma. The 5-year relative survival for Stage 1 melanoma is 99.2% (CI 98.5-99.9), however for Stage IV this is reduced to 26.2% (CI 20.3-32.4%),[15] see Figure 1. Despite recent advances in systemic therapies for the treatment of melanoma,[11-14, 16, 17] early detection remains a priority. This is recognised by the Federal Government in Australia, with a recent $10million investment in a roadmap for a national targeted skin cancer screening program.[18]

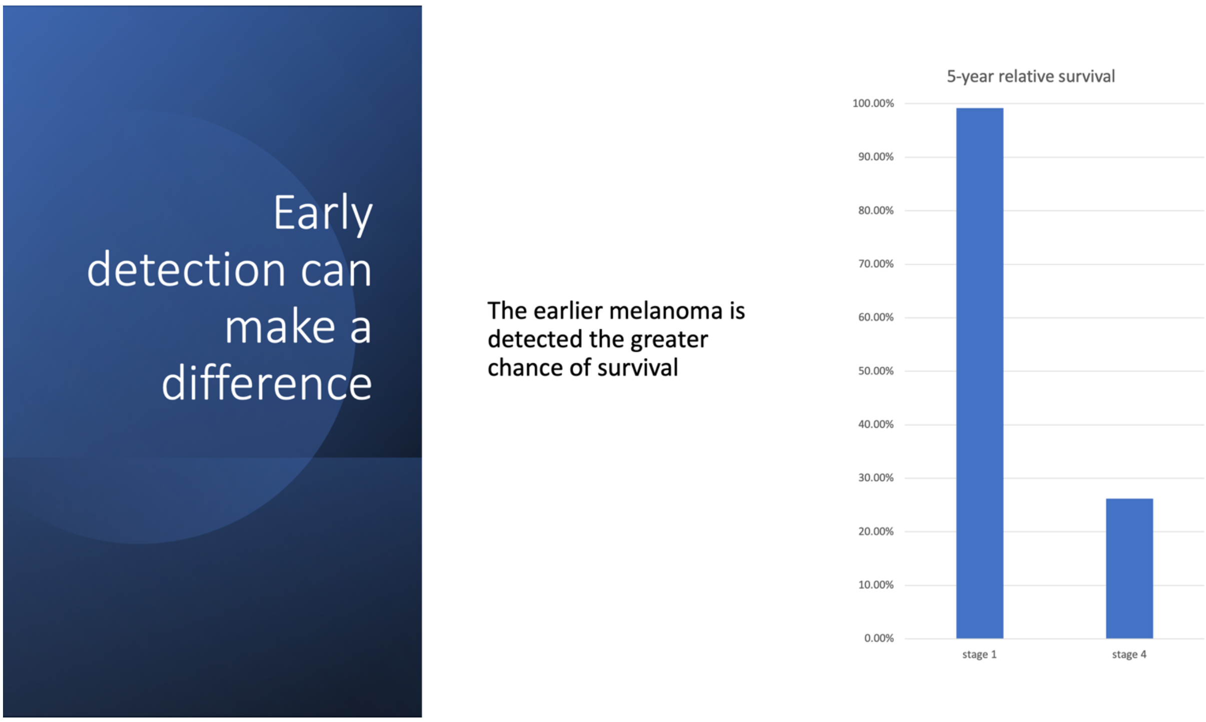


Figure 1 Early detection of melanoma can impact survival.

Source: AIHW 2023

There is currently a concern that heightened surveillance for melanoma may lead to overdiagnosis, where a proportion of melanomas are diagnosed that otherwise may not have warranted clinical attention as they were not destined to cause morbidity or death in a patient's lifetime.[19, 20] The challenge of over-identifying melanoma in-situ remains significant, as currently harmless melanomas cannot be prospectively identified, and for the moment melanoma survival is significantly higher in Australian regions with increased diagnostic rates.[21-23]

Screening

In Australia, there is no standardised screening program for the early detection of melanoma. Current Cancer Council Australia guidelines recommend people at very high-risk of new primary melanoma and their partners and carers are educated to recognise and document lesions suspicious of melanoma. In addition they should also be checked regularly by a clinician with six-monthly full skin examination supported by total body photography and dermoscopy.[24]

The US Preventative Services Task Force (USPSTF) recently found that there was insufficient evidence to determine the balance of benefits and harms for visual skin examination by a clinician to screen for skin cancer in asymptomatic adolescents and adults (a general population). They did note though that there was a need for validated risk assessment tools to identify persons at highest risk for skin cancer, who might benefit from screening.[25]

This is supported when considering the accuracy of a test, and when the same test is used in a population with low prevalence versus high prevalence. The predictive value of the test such as a Positive Predictive Value (PPV) or Negative Predictive value (NPV) depends on the prevalence of the condition in the population being tested. If a condition is rare, even an accurate test (high sensitivity and specificity) will yield a low positive predictive value (PPV),[26] see Figure 2.

The need for a standardised screening program in Australia has been recognised and the Melanoma Institute Australia is working with other stakeholders to research risk-based and cost-effective national screening approaches to provide an evidence-based roadmap for a targeted skin cancer screening program in Australia.[18]

The potential benefit of a screening program for those at the highest risk for skin cancer is supported by studies demonstrating that screening for melanoma in a high-risk clinic in Australia was cost-effective,[27] and targeted interventions that provide individuals with genomic risk of melanoma, are cost effective.[28] As such, there is a place for further research investigating the early detection of melanoma in adults at *higher risk* of developing melanoma rather than the general population. This is especially pertinent to Australia.

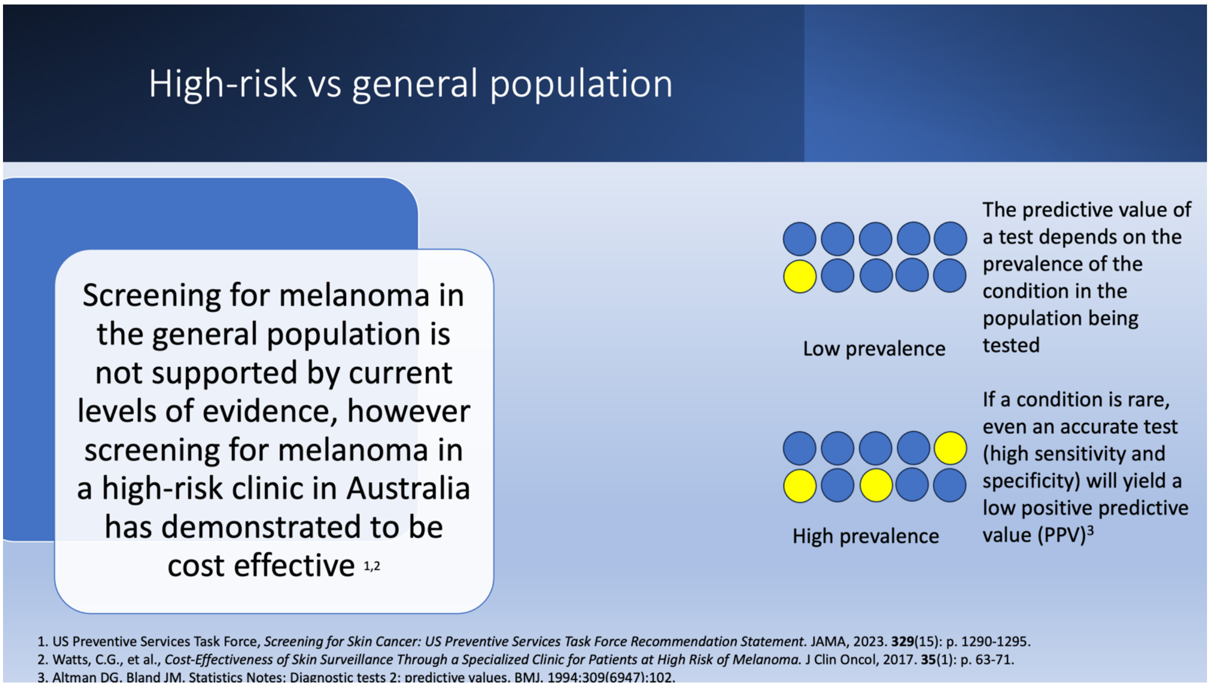


Figure 2 Targeted screening for those at high-risk, versus population-based screening

In the UK, there is no routine screening for skin cancer, and the first assessment of a skin condition is completed by a GP in primary care to determine if urgent referral for suspected skin cancer is required. People should be given a diagnosis or rule out cancer within 28 days of being referred by their GP. Referral to a teledermatology pathway has been added as an alternative to face-to-face dermatologist appointments in order to reduce the increasing demand on dermatology services within the NHS. Artificial intelligence (AI) technologies have now also been included in this pathway as an early value assessment to determine if they can be an alternative to teledermatology or face-to-face dermatologist appointments.[29] This represents an important step for using AI technologies to reduce referrals to in-person dermatology appointments.

Prevention strategies

As sun exposure leads to an increased risk of skin cancers including melanoma, preventive strategies in Australia have been aimed at reducing the cumulative risk of sun exposure.

The 'Slip, Slop, Slap' public health campaign was a very large skin cancer awareness and prevention campaign commencing from the early 1980s. In 2024, the population aged under 40 were born after or around the 'Slip, Slop, Slap' campaign and have spent their lives in an environment where skin cancer awareness has been greater. Skin cancer awareness and prevention advice continues today. While populations over 40 have increasing incidence rates, the rate increases are greater for the oldest populations who are likely to have spent more of their lives in times when there was less skin cancer awareness.[15]

Other strategies include regulation to ban the use of commercial solariums, school policies including 'No Hat, No play' in schoolyards, workplace occupational health and safety regulations including the provision of sun protection at worksites and the National Skin Cancer Prevention Strategy which prioritises reducing UV exposure, increasing sun protective behaviours, improving early detection, supporting research and importantly addressing inequities in rural, remote and indigenous communities.[8]

Chapter 2 - Machine learning and image analysis in healthcare

Artificial intelligence

Artificial intelligence or AI, refers to computer systems performing tasks that typically require human intelligence such as making decisions, solving problems and visual perception.[30] AI, as a formal field of study started in the early 1950's with Alan Turing introducing the concept of a Turing test, which assessed a machine's ability to display human-like intelligence through consideration of the question 'Can machines think?' and the development of the 'imitation game'.[31] Since this time, advances in computing power, algorithm development, cloud storage and the rise of deep learning have led to the widespread use of AI, including within healthcare.[30]

The integration of AI in healthcare has the potential to enhance patient and clinical team outcomes, decrease costs, and positively impact population health. AI is expected to play a significant and complementary role alongside human cognition in supporting the delivery of personalised healthcare services however currently, there is limited data regarding the actual effect on patient outcomes and cost of care.[32]

Examples of the use of AI in healthcare include medical imaging and diagnostics, predictive analytics, clinical decision support, personalised medicine, resource and workflow optimisation and remote monitoring and wearable data analysis.

Machine learning

Machine learning is a subfield of AI that allows computers to learn from data without being explicitly programmed. Machine learning uses statistical and mathematical modelling techniques that use different approaches to automatically learn and improve the prediction of a target state without explicit programming.[33] Machine learning differs from traditional computer programming in that the function is learnt based on the data, rather than traditional computer programming where functions are written using manually coded rules.[33]

The three fundamental components of machine learning can be expressed by the formula $Y=f(X)$, where the input (X) is the data that is entered into the program; the function (f) is the model, and the output (Y) is true or false. The function is a mathematical model that predicts the output from the input data. Examples of functions used in machine learning include Convolutional Neural Networks (CNNs), which map image pixels to diagnostic labels, or logistic regression which learns the probability of an outcome i.e. risk prediction.[34]

Types of machine learning

Machine learning can be broadly categorised into three types: supervised learning, unsupervised learning and reinforcement learning.

Supervised learning involves a program learning the function that connects input to output. For example, if the input is a dermoscopic image of a melanoma and the output is a label that is yes or no to the image being a melanoma then the program will learn a function that links the labelled data. If this is repeated numerous times, the model will continuously refine its function to accurately map the inputs to labels. This process known as fitting or training the model,[33] see Figure 3.

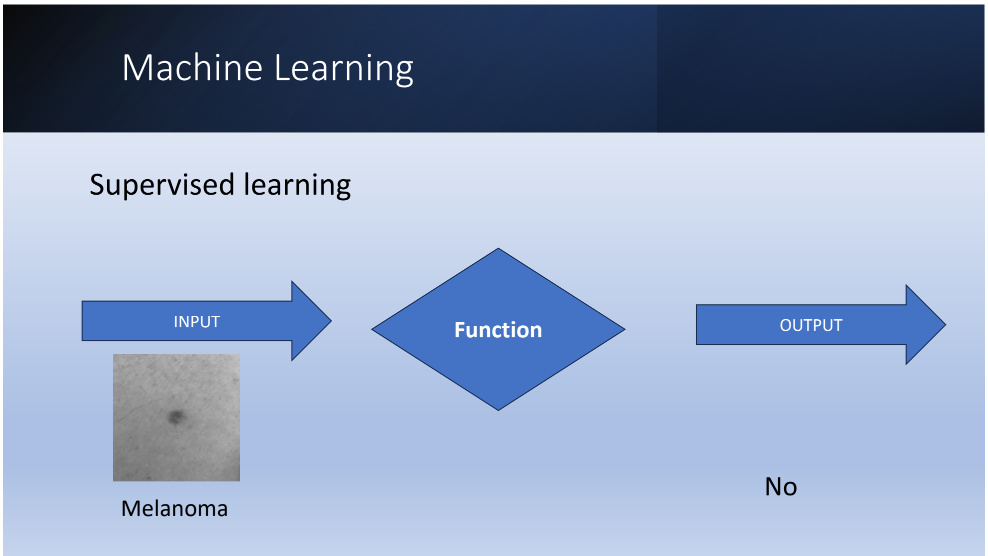


Figure 3. Supervised learning schematic

The model is given a training set of examples to learn the function that links the input and outputs of a given question, the model is then validated by using a separate set of examples and finally it is tested to check if the model can perform using test examples that have not been seen before during the model development. This training loop allows the model to be developed, validated and tested on large amounts of data. Training minimises the difference between the output of the models function and the true label. Loss is the difference between the function output and the true label. In order to fit the model, the aim is to minimise loss.[33] The model can also be adjusted to accommodate a set operating point at which more examples are determined as true rather than false.[33] It should be noted that the use of AI beyond the populations represented in the training and validation datasets may affect the accuracy of the test.[32]

Unsupervised learning is used with very large datasets, with no labels. Unsupervised learning examines the data and groups the data together using a common feature. This clustering is then used to define or label the group. The third example of machine learning is **reinforcement learning** which is often used for game playing where the model receives feedback for decisions made and learns over time which actions maximise the positive feedback.(35)

Machine learning in healthcare

Machine learning research in healthcare has focused on image classification or the ability to recognise patterns within an image. Patterns are recognised using neural networks, where data is filtered by a neuron or gate and the information is processed and passed to the next layer of neurons.(36) The architecture of a neural network refers to the way the layers of neurons are organised and interconnected, see Figure 4.

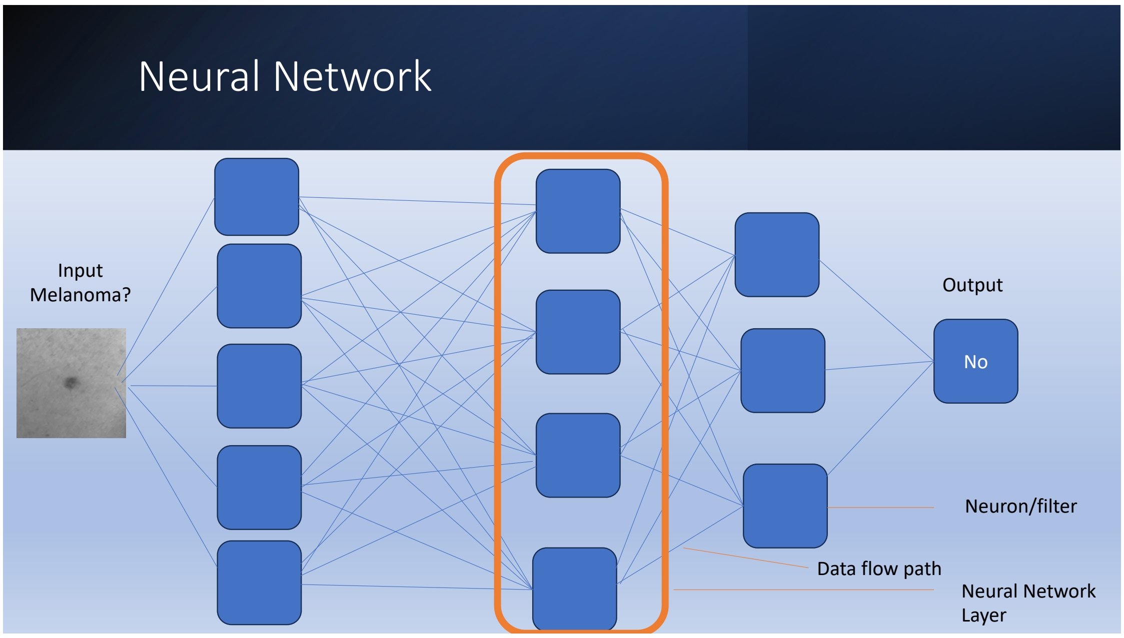


Figure 4 Neural network schematic with layers of neurons. Figure note: Blue boxes represent neurons or filters within each layer of the neural network. Blue lines indicate the data flow paths and connections between neurons across the layers.

Convolutional Neural Networks (CNNs) are a deep learning model that recognises patterns in images by learning to detect features like edges, textures, and shapes and combining them to recognise complex structures. Instead of data being processed by every neuron, each network layer will scan for a pattern e.g. a smooth edge or change in pigment and this will lead to an output. CNNs are often used for medical image analysis and facial recognition.[37] Other machine learning methods include Natural Language Processing which use Recurrent Neural Networks. Common CNN architectures include AlexNet with 8 layers (2012), VGG with 19 layers (2014), GoogLeNet Inception with 22 layers (2014), and ResNet with over 150 layers (2015). EfficientNet with over 200 layers (2019) was released by Google AI and is based on the notion of compound scaling.[33]

Accuracy of a CNN is often reported using Receiver Operating Characteristic Area Under the Curve (ROAUC) which quantifies the CNNs ability to distinguish between true and false across different classification thresholds. A value of 1.0 indicates perfect accuracy, whereas 0.5 reflects no better than chance. Higher ROAUC values suggest better diagnostic discrimination by the machine learning model.

Machine learning for the detection of melanoma

Machine learning may be a tool used by clinicians to improve the chance of detecting melanoma (clinician-led machine learning). Alternatively, machine learning can be used by patients and carers as a stand-alone, autonomous tool to detect melanoma (patient-led machine learning).[38] The use of patient-led machine learning applications, if accurate, has the potential to enable earlier detection of melanoma as it is not dependent on a clinician for skin examination.

Machine learning may improve melanoma detection when integrated with imaging techniques including total body photography (TBP), sequential digital dermoscopic imaging, reflectance confocal microscopy and smartphone applications.[39] Skin lesion tracking applications available for use on smartphones were originally a tool to track lesions, capture images of lesions/moles, compare images over time and provide reminders for skin checks and dermatologist review. [40, 41] The smartphone apps were initially inaccurate and associated with a high likelihood of missing melanomas, however this technology is rapidly improving.[42-45] Early clinical trials compared the accuracy of CNNs at detecting images of melanoma compared with dermatologists. Dermatologists outperformed the neural networks, however, more recently neural networks have demonstrated superior accuracy to dermatologists.[46-51]

Challenges

Machine learning in healthcare offers potential but also faces several key challenges. These include the risk of algorithmic and user bias; applying a tool to a dataset that has different demographics to the population in which the network was trained and validated; interpretability and transparency - the concept of a 'black box' where the decision making process is not easily interpretable rather than explainable AI; concern for patient privacy; medico-legal responsibility including legal responsibility if a cancer that can be detected by machine learning on an image is missed by a clinician, and regulatory compliance.[32]

Conclusion

As machine learning continues to evolve, its utility in healthcare will depend not only on its accuracy but also on its integration into clinical workflows, regulatory compliance, ethical implementation and acceptance by users. Continued research, particularly in diverse and high-risk populations is essential for machine learning to fulfill its promise to improve patient outcomes.

Chapter 3 - Protocol

Title: Costs and benefits of using Artificial Intelligence/Machine Learning for the early detection of melanoma in adults at high-risk of developing melanoma: A Scoping review protocol

A Olsson[a], J T W Williams[b], M Dieng[a], R L. Morton[a],

a. NHMRC Clinical Trials Centre, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia
b. Faculty of Medicine and Health, School of Public Health, The University of Sydney, Camperdown, Australia

Introduction

Melanoma of the skin is one of the most frequently diagnosed cancers in Australia. In 2023, an estimated 18,257 melanomas were diagnosed nationally, and more than 1,300 Australians died from this disease.[52] In 2022, Australia reported the highest incidence of melanoma worldwide, with an age-standardised incidence rate of 37 per 100,000 compared to the world incidence rate of 3.2 per 100,000.[53]

Early detection of melanoma improves survival outcomes. This increase in survival with early detection is highlighted by the difference in five-year relative survival rates between early and late-stage disease. Stage 1 melanoma has a five-year relative survival rate of 99.2% (CI 98.5-99.9), however for Stage IV this is reduced to 26.2% (CI 20.3-32.4%).[15] Despite advances in systemic therapies for the treatment of melanoma,[11-14, 16, 17] early detection remains a priority. While population-wide melanoma screening is not recommended,[25] research in Australia has demonstrated that targeted screening in a high-risk clinic can be cost-effective.[27] As such, there may be a place for investigating the early detection of melanoma in adults at higher risk of developing melanoma rather than the general population. High-risk individuals include those with fair skin, light hair and eye colour, skin sensitivity, exposure to UV radiation, high naevus (mole) count, a personal or family history of melanoma, exposure to certain chemicals or immunosuppression.[9]

Strategies to improve melanoma detection include Artificial Intelligence (AI), or the practical application, machine learning (ML). ML uses algorithms to learn from images, recognising patterns within the image. This is done using neural networks, where data passes through a neuron or gate and the information is processed and passed to the next layer of neurons.[54] Early clinical trials evaluated the diagnostic performance of neural networks in detecting images of melanoma compared to dermatologists. Dermatologists initially outperformed the neural networks, however, more recent studies have shown neural networks achieving, and in some cases exceeding, dermatologist-level accuracy.[47-49, 55-57]

Examples where AI/ML may improve melanoma detection include Total Body Photography (TBP), Sequential Digital Dermoscopic imaging (SDDI), reflectance confocal microscopy (RCM) and the use of smartphone technology and applications.[39] Skin tracking applications available for use on smart phones were originally a tool to track lesions, capture images of lesions/moles, compare images over time and provide reminders for skin checks and dermatologist review.[41] These tracking applications are now starting to use machine learning to track the size of a mole and to provide an alert if a mole has changed size, shape, or colour. Initially the smartphone applications using AI/ML based analysis to identify melanoma were inaccurate and associated with a high likelihood of missing melanomas, however this field is rapidly changing.[42, 43]

Hornung et al. [58] recently conducted a systematic review of the value of TBP for the early detection of melanoma. They demonstrated that TBP identified a higher proportion of melanomas with lower average Breslow thickness compared to the comparison groups without TBP. Whilst AI was not used to evaluate the captured images, the authors recommended that AI should ultimately be integrated into 3D TBP systems to support image analysis, pattern recognition, and detect changes of moles. Current trials investigating the cost-effectiveness of TBP in Australia (ACEMID) are ongoing.[59]

In Australia, current guidelines recommend people at very high-risk of new primary melanoma and their partners / carers are educated to recognise and document lesions suspicious of melanoma, in addition they should also be checked regularly by a clinician with six-monthly full skin examination supported by total body photography and dermoscopy. [24] AI/ML is used as a 'tool' to improve a clinicians chance of detecting melanoma rather than an autonomous, stand-alone tool that can be used by patients / carers.[38] The use of an autonomous, stand-alone tool for melanoma detection, if accurate, has the potential to enable earlier detection of melanoma as it is not dependent on a clinician for a full body skin examination.

A preliminary search of MEDLINE, the Cochrane Database of Systematic Reviews, PROSPERO and JBI Evidence Synthesis was conducted and identified several systematic and scoping reviews investigating the use of AI/ML for melanoma detection.[60-64] As of 9 May 2024, no systematic or scoping review were identified that were researching the autonomous use of AI/ML for the detection of melanoma of the skin in *high-risk* populations. The systematic reviews were either identifying literature comparing the accuracy of AI/ML to clinicians at classifying melanoma;[60] reviewing the use of AI/ ML to detect skin cancer in a general rather than a high-risk population;[61] or looking at the use of AI/ML as a tool to assist skin cancer diagnosis from images.[62-64]

The primary objective of this scoping review is to determine the extent of the literature describing the costs and benefits of using AI / ML as an autonomous tool for the detection of melanoma of the skin in adults at high-risk of developing melanoma compared to standard care. This will be supported by addressing secondary objectives which include determining the extent literature regarding:

- The accuracy of AI / ML to autonomously detect melanoma in adults at *high-risk* of developing melanoma of the skin;
- The effectiveness of AI/ML at detecting melanoma of the skin autonomously in adults at *high-risk* of developing melanoma;
- The time horizon in which costs and effects of AI/ML diagnosis / surveillance might be accrued; and
- The shelf life of cost-effectiveness evidence.

Review question

The primary question of this scoping review is what are the *costs* and *benefits* of using AI/ML as an autonomous tool for the detection of melanoma in adults at *high-risk* for developing melanoma of the skin compared to standard care?

Costs may include direct and indirect costs of hardware, software, resource use including physician time, societal costs, and potential for misdiagnosis / non-detection. Benefits may include any averted financial costs such as benefit derived from a reduction in the use of healthcare services, patient satisfaction, increased adherence to a surveillance regimen or increased detection of early stage rather than later stage melanoma.

The primary question will be supported by the following sub-questions.

i. What is the accuracy of AI/ML detecting melanoma in adults at *high-risk* of developing melanoma of the skin? Diagnostic accuracy includes specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV).
ii. What is the effectiveness of AI/ML at detecting melanoma in *high-risk* groups?
iii. What economic analysis has been conducted to determine, costs, benefits and performance of AI/ML to detect melanoma in adults at *high-risk*?
    - What is the time horizon in which costs and effects of AI diagnosis / surveillance may be accrued?
    - What is the shelf life of cost-effective evidence for the use of AI/ML for melanoma diagnosis /surveillance?

Eligibility criteria

Participants

Adults at high-risk for developing melanoma of the skin.

Concept

The earlier melanoma is detected, the greater chance of survival. AI/ML is one way to detect melanoma, however, it is often used as an additional tool to assist clinicians to detect melanoma rather than autonomously. This scoping review will look at the evidence for the use of AI/ML alone to detect or screen for melanoma in patients at high risk of developing melanoma of the skin. The accuracy, effectiveness, costs, and benefits of using AI/ML alone to detect melanoma in adults at high-risk of developing melanoma of the skin versus standard care will be determined.

Context

This scoping review will look at published evidence including Randomised Control Trials, prospective observational studies, and other evidence-based documents. Cost-effectiveness, cost-utility and cost-benefit studies will also be included.

Exclusion criteria

The following exclusion criteria will be used for this scoping review:

- Studies that were case studies, conference abstracts, comment papers, systematic reviews or meta-analyses.
- Non-melanoma skin cancers (e.g., basal cell carcinoma or squamous cell carcinoma).
- Ocular, mucosal and acral melanoma.

Methods

This scoping review will follow the JBI methodology for scoping reviews and the Preferred Reporting Items for Systematic review and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.[(65)] The protocol for this scoping review will be registered and prospectively made available on the Open Science Framework.[(66)]

Information sources

EMBASE, Medline, SCOPUS, Web of Science, Cinahl, Econlit, INAHTA and the CEA Registry. Non-indexed and grey literature will be searched including government websites.

Search Strategy

The search strategy in Appendix A will be used to search EMBASE, Medline, SCOPUS, Web of Science, Cinahl, Econlit, INAHTA and the CEA Registry. Non-indexed and grey literature will be searched including government websites. We will review the reference list of identified studies to ensure we have captured all key studies and contact authors as required.

The search term 'high-risk' will be used as an eligibility criterion rather than a search term to ensure the literature search is broad enough to include all relevant results. Country specific guidelines for determining those at high risk for developing melanoma will be followed.

Limitations

As most publications have occurred in the last decade literature will be limited to years 2012 and onwards.

Source of Evidence Selection

Literature searches will be conducted according to the Search Strategy in Appendix A.

References will be transferred to the reference manager Endnote 20 (Clarivate Analytics, PA, USA).[67]

References will be uploaded into Covidence (Veritas Health Innovation, Australia; https://www.covidence.org) and duplicates will be removed.[68]

Non-English articles will be translated using DeepL translator.

Study Selection

Title and abstract screening will be completed by a single reviewer. Full text articles will then be reviewed by two independent reviewers with any non-conformance to be discussed between the reviewers and referred to a third reviewer as required.

Data Extraction, analysis and presentation

Data will be extracted from the articles by a single reviewer using a data extraction tool. Data extracted will include study characteristics and criteria reported in health economic and AI in healthcare reporting guidelines including the Consolidated Health Economic Evaluation Reporting Standards (CHEERS)[69] and the SPIRIT-AI[70] and Consort-AI reporting guidelines.[71] The data extraction tool, developed by the reviewers will be modified and revised as necessary during the process of extracting data. Any modifications will be detailed in the scoping review.

Data analysis will focus on the scope of published information and will include a descriptive numerical summary and qualitative thematic analysis. Results will be presented in tables, graphically and with a supporting narrative description.

Acknowledgements

This scoping review will contribute to a MPhil (Medicine and Health) The University of Sydney by Angela Olsson.

Funding

Jake Williams is supported by a University of Sydney Faculty of Medicine and Health Research Centres Stipend Scholarship. Rachael Morton is supported by a NHMRC Investigator grant #1194703.

Conflicts of interest

Nil declared

Appendix

Appendix A

Search Strategy

Medline

Ovid MEDLINE(R) ALL <1946 to Month XX, 202X>
1 Skin Neoplasms/
2 skin cancer.mp.
3 Melanoma/
4 melanoma.mp.
5 1 or 2 or 3 or 4
6 exp Artificial Intelligence/
7 AI.mp.
8 AI.tw.
9 machine learning.mp.
10 deep learning.mp.
11 neural network*.mp.
12 computational intelligence.mp.
13 6 or 7 or 8 or 9 or 10 or 11 or 12
14 5 and 13
15 limit 14 to yr="2012 -Current"

Embase
Embase Classic+Embase <1947 to 202X Month XX>
1 Skin Neoplasms/
2 skin cancer.mp.
3 Melanoma/
4 melanoma.mp.
5 1 or 2 or 3 or 4
6 artificial intelligence.mp. or artificial intelligence/
7 AI.mp.
8 AI.tw.
9 machine learning.mp. or machine learning/
10 deep learning.mp. or deep learning/
11 artificial neural network/ or neural network*.mp.
12 computational intelligence.mp.
13 6 or 7 or 8 or 9 or 10 or 11 or 12
14 5 and 13
15 limit 14 to yr="2012 -Current"

Scopus
(TITLE-ABS-KEY (melanoma) OR TITLE-ABS-KEY ("skin neoplasm") OR TITLE-ABS-KEY ("skin cancer")

AND
TITLE-ABS-KEY ("artificial intelligence") OR TITLE-ABS-KEY ("machine learning") OR TITLE-ABS-KEY ("deep learning") OR TITLE-ABS-KEY ("neural network*") OR TITLE-ABS-KEY ("computational intelligence"))
AND
(LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012))
AND
(EXCLUDE (DOCTYPE , "cp") OR EXCLUDE (DOCTYPE , "cr"))

Web of science

(TI=(Melanoma OR melanoma OR "skin cancer" OR "skin neoplasm") OR AB=(Melanoma OR melanoma OR "skin cancer" OR "skin neoplasm"))
AND
AB=("Artificial intelligence" OR “AI” OR "machine learning" OR "deep learning" OR "artificial neural network" OR "neural network*")

Cinahl

#	**Query**
S9	S4 AND S8
S8	S5 OR S6 OR S7
S7	"computational intelligence"
S6	(MH "Neural Networks (Computer)") OR (MH "Machine Learning+") OR (MH "Deep Learning") OR (MH "Artificial Intelligence+") OR "machine learning or artificial intelligence or deep learning or neural network"
S5	AB artificial intelligence or ai or a.i.
S4	S1 OR S2 OR S3
S3	AB melanoma
S2	(MH "Skin Neoplasm+") OR "skin cancer"
S1	TX skin neoplasm

Econlit

#	**Query**
S8	S3 AND S7
S7	S4 OR S5 OR S6
S6	AB computational intelligence
S5	AB machine learning or artificial intelligence or deep learning or neural network
S4	AB neural networks and machine learning
S3	S1 OR S2
S2	AB skin cancer or skin neoplasm or skin carcinoma
S1	AB melanoma or skin cancer

Chapter 4 – The costs, benefits, and diagnostic accuracy of machine learning for the autonomous detection of melanoma in high-risk adults: a scoping review

Authors and institutions:

Angela Olsson[a], Jake T. W. Williams[b], Mbathio Dieng[a], Rachael L. Morton[a]

a. NHMRC Clinical Trials Centre, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia 2006
b. Sydney School of Public Health, The University of Sydney, Camperdown, Australia, 2006
Keywords: Artificial intelligence, Machine learning, Melanoma, Diagnostic accuracy

Figures: 6

Tables: 4

Box: 1

Funding: RL Morton is supported by an Australian NHMRC Investigator grant #1194703.

JW is supported by a Faculty of Medicine and Health Research Centre Stipend Scholarship (SC4167) from the University of Sydney.

AO was supported by the award of a Research Training Program Fee Offset by the Australian Government.

Abstract:

Background

Melanoma of the skin is one of the most frequently diagnosed cancers in Australia. It is estimated that approximately 18,257 melanomas were identified in Australia during 2023 and more than 1,300 Australians died from this disease. In 2022, Australia had the highest incidence of melanoma in the world.[(53)] As such the prevention, early detection, and treatment for melanoma in Australia remains a high priority. The earlier invasive melanoma is detected, the greater chance of survival. Screening for melanoma in the general population is not currently recommended in Australia or the USA, however, there may be scope for screening of higher-risk populations.

Machine learning, in particular the use of convolutional neural networks (CNNs) has shown increasing potential to autonomously detecting melanoma from medical images. However, the evidence base supporting its use in a clinical setting in the high-risk population remain unclear. This scoping review aimed to explore the extent, accuracy, benefits and costs of using machine learning autonomously, either consumer-led or clinician-led for detecting melanoma in adults at high-risk.

Methods

A scoping review was conducted in accordance with the Joanna Briggs Institute methodology and the PRISMA-ScR guidelines. A comprehensive literature search was undertaken for studies published between 1 Jan 2012 and 1 May 2024. Eligible studies include randomised controlled trials, prospective observational studies, cohort studies and economic evaluations that assessed the accuracy, costs and benefits of autonomous machine learning use in high-risk populations. Data was extracted, tabulated, and synthesised narratively.

Results

From 9,188 records, 55 studies met the inclusion criteria. Most studies were conducted in Germany, the USA or Australia and involved either diagnostic reader studies (n=26) or real-world clinical settings (n=29). Machine learning demonstrated high diagnostic accuracy in controlled settings with 78% of reader studies reporting ROAUC >0.8, compared with 60% of clinical studies. Only five studies reported costs and benefits in high-risk populations; and two conducted formal economic evaluations. One of these did not find machine learning to be cost-effective, while the other reported cost savings and reduced clinician workload. Patient and clinician preference studies indicated higher trust in clinician-led machine learning models (e.g., 2D/3D total body photography) compared with standalone consumer apps.

Conclusion

This review highlights growing interest in machine learning for melanoma detection, but reveals differences in study design, comparator standards, image modalities and outcome reporting. Evidence supporting the use of machine learning in high-risk populations remains limited, particularly regarding cost-effectiveness and studies in real-world clinical settings. Future research should prioritise prospective, real-world studies, cost benefit and cost effectiveness analysis to better assess the utility of machine learning in melanoma surveillance and diagnosis.

Introduction

Melanoma of the skin is one of the most commonly diagnosed cancers in Australia. In 2023, an estimated 18,257 new melanoma cases were identified, and over 1300 Australians died from the disease.[(52)] In 2022, Australia had the highest incidence of melanoma in the world with an age-standardised incidence rate of 37 per 100,000 compared to the world incidence rate of 3.2 per 100,000.[(53)] As such the prevention, early detection, and treatment for melanoma remains a national health priority.

The earlier invasive melanoma is detected, the greater chance of survival. This increase in survival with early detection is illustrated by the difference in survival rates between early and late-stage melanoma. The 5-year relative survival for Stage 1 melanoma is 99.2% (CI 98.5-99.9), however for Stage IV this is reduced to 26.2% (CI 20.3-32.4%).[(15)] Although systemic therapies for the treatment of melanoma have improved, early detection continues to be the most effective means of improving survival. [(11-14, 16, 17)]

At present, routine melanoma screening in not recommended for the general population. [(25)] However, targeted screening of high-risk individuals in Australian has been shown to be cost-effective.[(27)] As such, there may be a place for investigating the early detection of melanoma in adults at higher risk of developing melanoma rather than the general population. Characteristics of people with higher risk for developing melanoma include fair skin, light hair and eye colour, skin sensitivity, exposure to UV radiation, naevus (mole) count, personal or family history, exposure to certain chemicals and weak immune systems.[(9)]

Methods to improve melanoma detection include the use of artificial intelligence (AI) or the practical application machine learning (ML). AI refers to computer systems that are able to perform tasks that typically require human intelligence such as making decisions, solving problems and visual perception[30]. ML is a subfield of AI that allows computers to learn from data without being explicitly programmed. ML uses algorithms to learn from images and to recognise patterns within the image. Patterns are recognised using neural networks, where data is filtered by a neuron or gate and the information is processed and passed to the next layer of neurons. [54] Convolutional Neural Networks (CNNs) is a deep learning model that recognises patterns in images by learning to detect features like edges, textures, and shapes and combining them to recognise complex structures. CNNs are often used for medical image analysis and facial recognition.[37] Early clinical trials compared the accuracy of CNNs at detecting images of melanoma compared with dermatologists. Dermatologists outperformed the neural networks, however, more recently neural networks have demonstrated superior accuracy to dermatologists.[47-49, 55-57]

Accuracy of the CNN is often reported using Receiver Operating Characteristic Area Under the Curve (ROAUC) which quantifies the CNNs ability to distinguish between melanoma and non-melanoma cases across different classification thresholds. A value of 1.0 indicates perfect accuracy, whereas 0.5 reflects no better than chance. Higher ROAUC values suggest better diagnostic discrimination by the ML model.

Examples where ML may improve melanoma detection include total body photography (TBP), sequential digital dermoscopic imaging (SDDI), reflectance confocal microscopy (RCM) and the use of smartphone technology and applications.[39] Skin tracking applications[40] available for use on smartphones were originally a tool to track lesions, capture images of lesions/moles, compare images over time and provide reminders for skin checks and dermatologist review.[41] The smartphone apps were initially inaccurate and associated with a high likelihood of missing melanomas, however this technology is rapidly improving.[42-45]

In Australia, current Cancer Council Australia guidelines recommend people at very high-risk of new primary melanoma and their partners and carers are educated to recognise and document lesions suspicious of melanoma. In addition they should also be checked regularly by a clinician with six-monthly full skin examination supported by total body photography and dermoscopy.[24]

ML may be a tool used by clinicians to improve the chance of detecting melanoma (clinician-led ML applications). Alternatively, ML can be used by patients and carers as a stand-alone, autonomous tool to detect melanoma (patient-led ML applications).[38] The use of patient-led ML applications, if accurate, has the potential to enable earlier detection of melanoma as it is not dependent on a clinician for skin examination.

A preliminary search of MEDLINE, the Cochrane Database of Systematic Reviews, PROSPERO and JBI Evidence Synthesis was conducted and identified several systematic and scoping reviews investigating the use of ML for melanoma detection.[60-64] As of 9 May 2024, no systematic or scoping reviews were identified that were researching the autonomous use of ML for the detection of melanoma of the skin in *high-risk* populations. The systematic reviews were either identifying literature comparing the accuracy of ML to clinicians at classifying melanoma;[60] reviewing the use of ML to detect skin cancer in a general rather than a high-risk population;[61] or looking at the use of ML as a tool to assist skin cancer diagnosis from images.[62-64]

The primary objective of this scoping review is to determine the extent of the literature describing the costs and benefits of using ML as an autonomous tool used by clinician's (clinician-led ML applications) or patient's (consumer-led ML applications) for the detection of melanoma of the skin in adults at high-risk of developing melanoma compared with standard care.

Review question

The primary question of this scoping review is what are the *costs* and *benefits* of using ML as an autonomous tool for the detection of melanoma in adults at *high-risk* for developing melanoma of the skin compared to standard care?

Costs may include direct and indirect costs of hardware, software, healthcare use including physician time, societal costs, and potential for misdiagnosis / non-detection. False negatives may incur a cost due such as more intensive treatment, lower survival rates and loss of productivity and false positives may lead to unnecessary biopsies, risk of scarring, patient anxiety and increased use of services such as pathology or clinician time. Diagnostic accuracy metrics including sensitivity (true positive rate), specificity (true negative rate) and Area under the ROC curve are accuracy metrics that can be used as indicators of these costs. Benefits may include patient satisfaction, increased adherence to a surveillance regimen or increased detection of early stage rather than later stage melanoma.

The primary question will be supported by the following sub-questions.

- What is the accuracy of ML detecting melanoma in adults at high-risk of developing melanoma of the skin? Diagnostic accuracy includes specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV).
- What is the effectiveness of ML at detecting melanoma in high-risk groups?
- What economic analysis has been conducted to determine, costs, benefits and performance of ML to detect melanoma in adults at high-risk?
- What is the time horizon in which costs and effects of ML diagnosis / surveillance may be accrued?
- What is the shelf life of cost-effective evidence for the use of ML for melanoma diagnosis /surveillance?

Eligibility criteria

Participants

Adults at high-risk for developing melanoma of the skin.

Concept

The earlier melanoma is detected, the greater chance of survival. ML is one way to detect melanoma, however, it is often used as an additional tool to assist clinicians to detect melanoma rather than autonomously. This scoping review will look at the evidence for the use of ML to detect or screen for melanoma in patients at high risk of developing melanoma of the skin. The accuracy, effectiveness, costs, and benefits of using ML alone to detect melanoma in adults at high-risk of developing melanoma of the skin versus standard care will be determined.

Context

This scoping review will look at published evidence including Randomised Control Trials, prospective observational studies, and other evidence-based documents. Cost-effectiveness, cost-utility and cost-benefit studies will also be included.

Exclusion criteria

The following exclusion criteria will be used for this scoping review:

- Studies that were case studies, conference abstracts, comment papers, systematic reviews or meta-analyses.
- Non-melanoma skin cancers (e.g., basal cell carcinoma or squamous cell carcinoma).
- Ocular, mucosal and acral melanoma.

Methods

We conducted a scoping review to map the existing research and identify potential areas that may benefit from further research when looking at the use of ML to detect melanomas in high-risk patients. We were specifically interested in melanomas, rather than all skin cancers, as Australia has a higher melanoma mortality rate compared with other non-melanoma or keratinocyte skin cancers including squamous cell carcinoma (SCC) and basal cell carcinoma (BCC).(6) The scoping review adhered to the Joanna Briggs Institute (JBI) methodology for scoping reviews(72) and followed the Preferred

Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.[73]

Search Strategy

A literature search was conducted to identify studies published between 1 January 2012 and 1 May 2024, to find empiric studies about the use of ML in melanoma diagnosis. This time frame was selected due to the advances in ML over the past decade. The databases searched included Medline, EMBASE, SCOPUS, Web of Science, CINAHL, and EconLit. Keywords related to artificial intelligence (AI), machine learning (ML), deep learning, neural networks, skin cancer, and melanoma were included. Detailed search strategies can be found in **Appendix A**. To ensure completeness, reference lists of the identified studies were manually screened for additional articles.

Source of evidence, screening and selection

The scoping review included randomised controlled trials, prospective observational studies, and other economic evaluations including cost-effectiveness, cost-utility, cost-benefit analyses, and preference studies, published after 1 January 2012. Case studies, commentaries, systematic reviews, and meta-analyses were excluded. Conference abstracts were excluded unless a corresponding full-text publication was available. No language restrictions were applied, and non-English articles were translated using DeepL Translator.[74] Non-melanoma skin cancers (i.e. BCC or SCC) ocular, mucosal and acral melanomas were excluded due to differences in mortality and prevalence.

Identified publications were uploaded into Covidence software (https://www.covidence.org/), where an additional deduplication process was performed based on author, year, and title. A single reviewer (AO) performed the initial title and abstract screening. Full text review was then conducted by two independent reviewers (AO and JW). Any disagreements as to study eligibility was discussed and resolved between the reviewers or referred to a third independent reviewer.

Data extraction

Data extraction was performed by a single reviewer (AO) using a prespecified data extraction form. Data extracted included study characteristics and criteria reported in health economic and AI in healthcare reporting guidelines including the Consolidated Health Economic Evaluation Reporting Standards (CHEERS)[69] the SPIRIT-AI [70] and Consort-AI reporting guidelines.[71]

Analysis and presentation of results

Study characteristics and outcomes were identified, extracted and results were summarised and tabulated. As this was a scoping review, the quality of each study was not formally assessed. Study characteristics included aim, design, country of origin; participant characteristics included risk level for developing melanoma; clinical setting; the ML intervention including type and version, device and image type, comparator and gold standards. Study outcomes included test performance characteristics such as receiver operating characteristic area under the curve (ROAUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and reporting of costs and benefits. Descriptive statistics were used to summarise the study characteristics and outcomes. The reported ROAUC of ML used to detect melanomas in diagnostic accuracy studies, and in real-world clinical settings were charted and a narrative synthesis of results was conducted. Diagnostic accuracy studies compared images from image datasets with a predefined reference standard. Trials conducted in real-world clinical setting used images from clinical settings rather than from an image database.

Results

A total of 9,188 studies were identified in the literature search, from which 5,532 duplicates were removed. The remaining 3,656 studies were screened based on titles and abstracts, leading to the exclusion of 3,542 studies. The remaining 114 full-text studies were evaluated for eligibility resulting in a total of 55 studies included in the scoping review. Reasons for exclusion included failure to meet the prespecified inclusion criteria (e.g. studies conducted within the general population rather than among those at high-risk of melanoma). See Figure 5 for the PRISMA flow diagram.

Characteristics of included studies

Most studies (n=28; 51%) were published in 2023 or 2024. See Figure 6. Most of the studies were conducted in Germany (n= 11; 21%), the USA (n= 10; 19%), and Australia (n= 9; 17%), reflecting the high prevalence of melanoma in these countries.

Of the 55 identified studies, 26 were diagnostic accuracy or diagnostic reader studies where machine learning was used to detect melanomas from dataset images. The remaining 29 studies were performed using images taken directly from high-risk individuals reflecting real world clinical scenarios (Table 2 Study Characteristics).

A variety of image types were employed by ML to detect melanomas. Of the 55 studies identified, studies utilised dermoscopic images (n=21;38%), clinical images (n=6; 11%), or images taken by a smartphone (with or without a dermatoscope attachment) (n=10; 18%). Six (11%) studies combined dermoscopic and clinical images, and four (7%) combined clinical images with clinical information such as patient age or sex. Three studies (5%) used histopathology whole-slide digitalised images, and three (5%) employed other imaging methods including elastic scattering spectroscopy (ESS) and spectral data emitted from tissue treated with laser-induced plasma spectroscopy (LIPS). One study (2%) used 3D imaging and one used 2D total body photography images (2%). See Figure 7.

The convolutional neural networks (CNNs) used in the studies included a range of architectures, each with different structural designs and feature extraction capabilities. These included ResNet-18, ResNet 50, Inception V3 & V4, DenseNet and EfficientNet.

Diagnostic accuracy

The ROAUC for the diagnostic reader studies was plotted separately to the ROAUC for the ML used to detect melanoma in a real-world clinical setting (Figure 8 and Figure 9). The ROAUC for the diagnostic reader studies were reported to lie between 0.68[75] and 0.96[76], compared to real-world clinical studies which reported a range of 0.52[42] to 0.975[77]. 78% (18/23 reported studies) of the diagnostic accuracy studies fell in the range >0.8-1.0. This compared to only 60% (15/25) of those studies in a clinical setting with ROAUC in the range >0.8-1.0. The percentage of studies reporting ROAUC for ML in diagnostic accuracy reader studies and those in a clinical setting is presented in Figure 10.

Differences in the choice of comparator or gold standard reference was observed between studies. The majority of studies used histopathology as the gold standard comparator, however, some studies including that reporting the lowest ROAUC for ML used a dermatologist as the comparator. In this instance, ML was evaluated against the clinician's diagnosis. For benign lesions, monitoring for 3 months rather than excision was often used.

Costs and benefits

Of the 55 studies, only 5 studies reported costs and benefits for ML to detect melanoma in high-risk patients. Of these, two were health economic evaluations (see Table 1), and three were preference-based studies (see Table 4).

Of the two health economic studies identified in this scoping review, Gomez Rossi et al. (2022) used a Markov-model to calculate cost-effectiveness of ML used as a decision-support system for detecting melanoma in dermoscopic images. The study found that from the perspective of a health insurance payer and contribution model, the mean cost of AI use was USD $750 (95% CI $608-970); compared to $759 (95% CI, $618-970) for the control (standard of care by dermatologists). Quality adjusted life years (QALYs) were comparable for each group 86.6 (95%CI 84.9-88.0) with the same QALY reported for both groups, AI and standard visual recognition. The study concluded that there was no statistical difference between AI and the control group in terms of cost or outcomes, suggesting limited evidence for a clear cost-effectiveness advantage using this model.

Marsden et al. (2024) undertook a cost consequence analysis of the use of artificial intelligence as a medical device for identifying skin lesions in a clinical setting.

This study investigated the accuracy and economic implications of an AI medical device integrated into a UK-based teledermatology service. The device was used to identify premalignant and benign skin lesions in clinical settings. AI as a medical device demonstrated improved specificity by identifying more lesions that did not require biopsy or urgent referral compared to standard teledermatology care ($p = 0.001$), while maintaining similar sensitivity for skin cancer detection. This was reported as a potential cost savings of approximately £156,063.79 and 259 specialist hours per 1,000 patients, from the perspective of the UK National Health Service.

Three papers reported the patient perspective of ML, ML within a consumer-led, smart phone app; and ML using images from 2D and 3D TBP (clinician-led ML application). [(45, 78, 79)] Patients at high-risk for melanoma and patients with melanoma were more confident about a mole examination by a dermatologist rather than a smartphone app alone. Patients expected reliable results with the highest accuracy by assessment by a physician, 2D and 3D TBP devices (98%, 82% and 89% high risk patients) vs only 16% for assessment by a smartphone app alone. Both clinicians and patients were positive towards a 2D TBP clinician-led ML tool, however most patients still preferred interpretation of ML results by a specialist clinician rather than just the relying on the ML decision.

Limited cost and benefit studies were identified, and only one study reported a lifetime time horizon for their results. No studies reported the shelf-life of cost-effectiveness evidence.

Discussion

This scoping review identified 55 studies evaluating the use of ML for the autonomous detection of melanoma in adults at high-risk. While the number of publications has grown significantly over the past two years, particularly in Germany, the USA, and Australia, the evidence remains heterogeneous in terms of study design (diagnostic accuracy reader study vs clinical setting), image type (Dermoscopic, Smartphone, clinical image, WSI, 2D TBP, 3D TBP), ML application (clinician-led vs consumer-led), comparator or gold standard reference, and outcomes reported.

Only a small number of studies (n = 5) evaluated the costs or benefits of ML in high-risk populations, and just two of these were formal economic evaluations. Gomez Rossi et al. (2022) found no statistically significant cost or QALY differences between ML and standard care. Marsden et al. (2024) reported potential cost savings and reduced demand for specialist consultations with the use of a ML medical device as part of a UK-based skin cancer teledermatology service. Differences in the results of the identified health economic studies may be due to the heterogenous nature of the studies. Gomez et al is a Markov study modelled using a diagnostic accuracy study rather than real world application. Marsden et al however is derived from real-world evidence and set in a clinical situation which demonstrates a potential cost saving. Further health economic research is needed to support the use of ML to detect melanoma in high-risk populations.

Preference-based studies highlighted scepticism among patients and clinicians regarding consumer-led ML tools, particularly smartphone applications. While 2D and 3D total body photography (TBP) supported by ML was generally viewed more favourably, a preference for clinician interpretation of ML outputs remained. Dermatologists reported low confidence in standalone consumer apps, with only 8.8% of assessments indicating trust in the app's decision. These findings suggest that trust and acceptability of ML tools remain barriers to broader adoption, particularly for consumer-led surveillance.

Consumer-led ML applications have been implemented in the general-population, but studies have shown increased health system use, due to an increase in referrals for benign lesions. For example, Smak Gregoor et al. (2023)[80] found that users of an mHealth app had higher rates of claims for premalignant and benign lesions compared to matched controls. This indicates a potential risk of overdiagnosis and unnecessary healthcare utilisation when ML tools are used in low-risk populations. Targeted implementation in high-risk groups may therefore offer a more cost-effective approach and is a gap in the research, highlighted by this scoping review. There is an inherent trade-off between accessibility, diagnostic performance and cost. Highly accessible consumer-led ML applications such as smartphone-based tools, may offer equitable access, particularly in rural settings, but these may have lower diagnostic accuracy. Conversely, clinician led ML applications such as 3D TBP may be in centralised locations and require an appointment. Further economic evaluation is needed.

Future research should prioritise real-world evaluations of ML in high-risk populations and settings where incremental benefit is likely to be highest, such as in primary care, nurse-led services, or rural and remote areas where access to specialist dermatologists is limited. ML has the potential to serve as a supportive tool for clinicians, however, the incremental value of improving diagnostic accuracy among general practitioners or allied health professionals may outweigh gains achieved by marginally improving the performance of specialist dermatologists.

ResNet-18 and ResNet-50 are part of the residual network family with ResNet-18 being a relatively shallow model with only 18 layers and ResNet-50 employing 50 layers allowing for more complex features.[81] Inception V3 and Inception V4 are part of the GoogLeNet family that allow for multi-scale processing by performing convolutions in parallel within the same layer.[81] EfficientNet is a more recent architecture by Google researches which uses compound scaling to optimise performance. [81] The CNNs identified in this scoping review are opensource, however the application may be commercial such as when a CNN is trained on private datasets. Image analysis is only a single part of healthcare and diagnosis; however, machine learning, with its capacity to analyse large amounts of data and advances in processing capabilities is positioned to drive improvements. As improvements are made it is important that clinical pathways are made to utilise machine learning to save clinician time and improve patient outcomes. The use of machine learning to detect melanoma in a high-risk setting has not yet been proven to be cost effective however continued reassessment is essential given an evolving healthcare landscape.

Conclusion

The current evidence supporting ML as a decision support tool in high-risk patients from a cost-effectiveness perspective is limited. This scoping review has highlighted the heterogenous nature of the data for the accuracy and utility of machine learning for melanoma detection. Further studies are required, especially in a high-risk population, and consideration should be given to whether ML approaches to detect melanoma should be consumer-led or clinician-led.

Tables

Box 1. Scoping Review: Population, Concept, Context

Participants Adults at high-risk for developing melanoma of the skin. *Concept* The earlier melanoma is detected, the greater chance of survival. Artificial Intelligence or Machine Learning is one way to detect melanoma, however, it is often used as a tool to assist clinicians to detect melanoma rather than autonomously. This scoping review examined the extent of literature describing the use of machine learning autonomously to detect melanoma in patients at high-risk of developing melanoma of the skin. The accuracy, costs, and benefits of using machine learning autonomously to detect or screen for melanoma in adults at high risk of developing melanoma of the skin versus standard care was determined. *Context* This scoping review incorporated published evidence, including Randomised Control Trials, prospective observational studies, and other evidence-based documents. Studies examining Cost-effectiveness, cost-utility and cost-benefit were also included.

Figure 5 PRISMA Flow Diagram

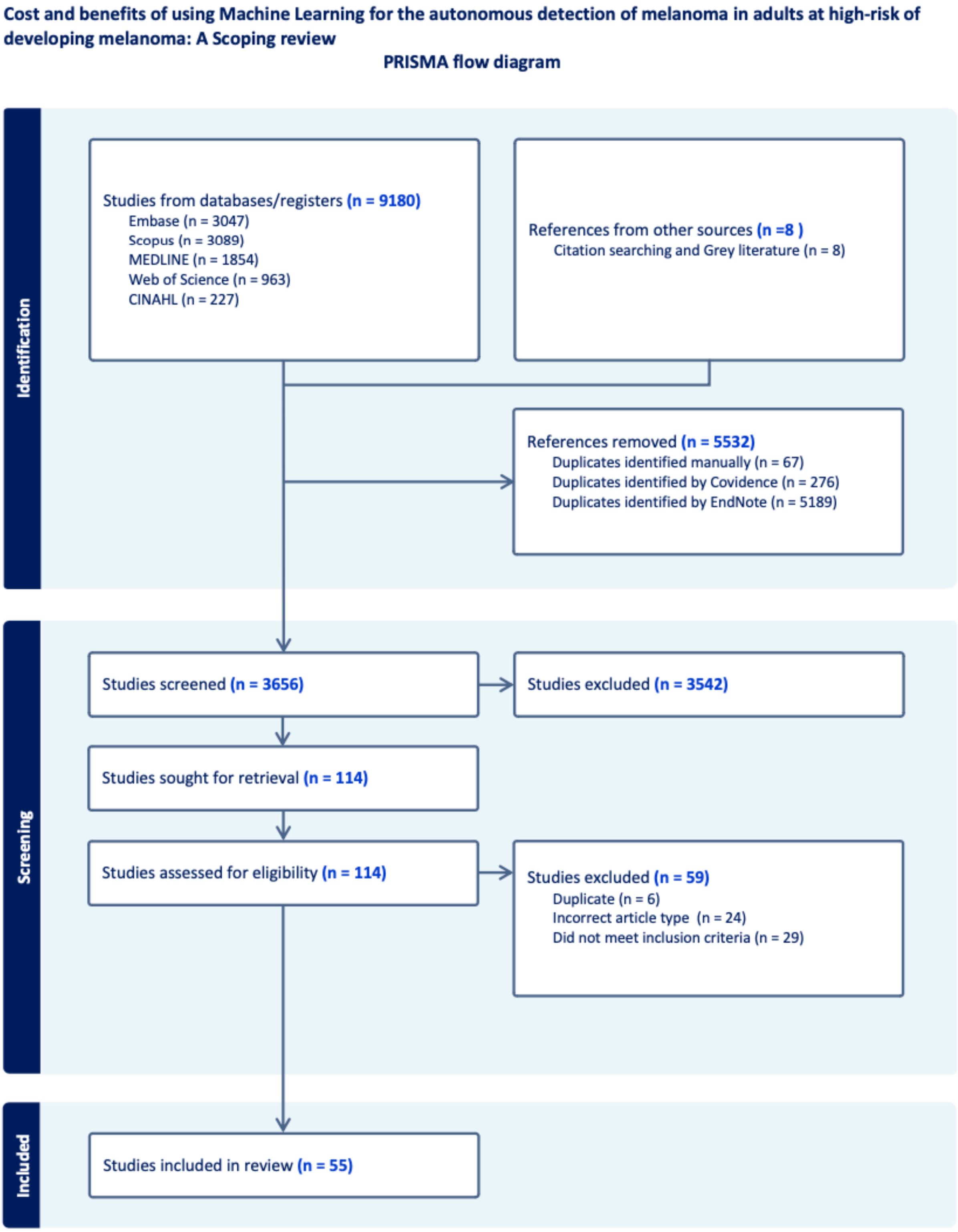


Figure 6 Number of studies identified by publication year

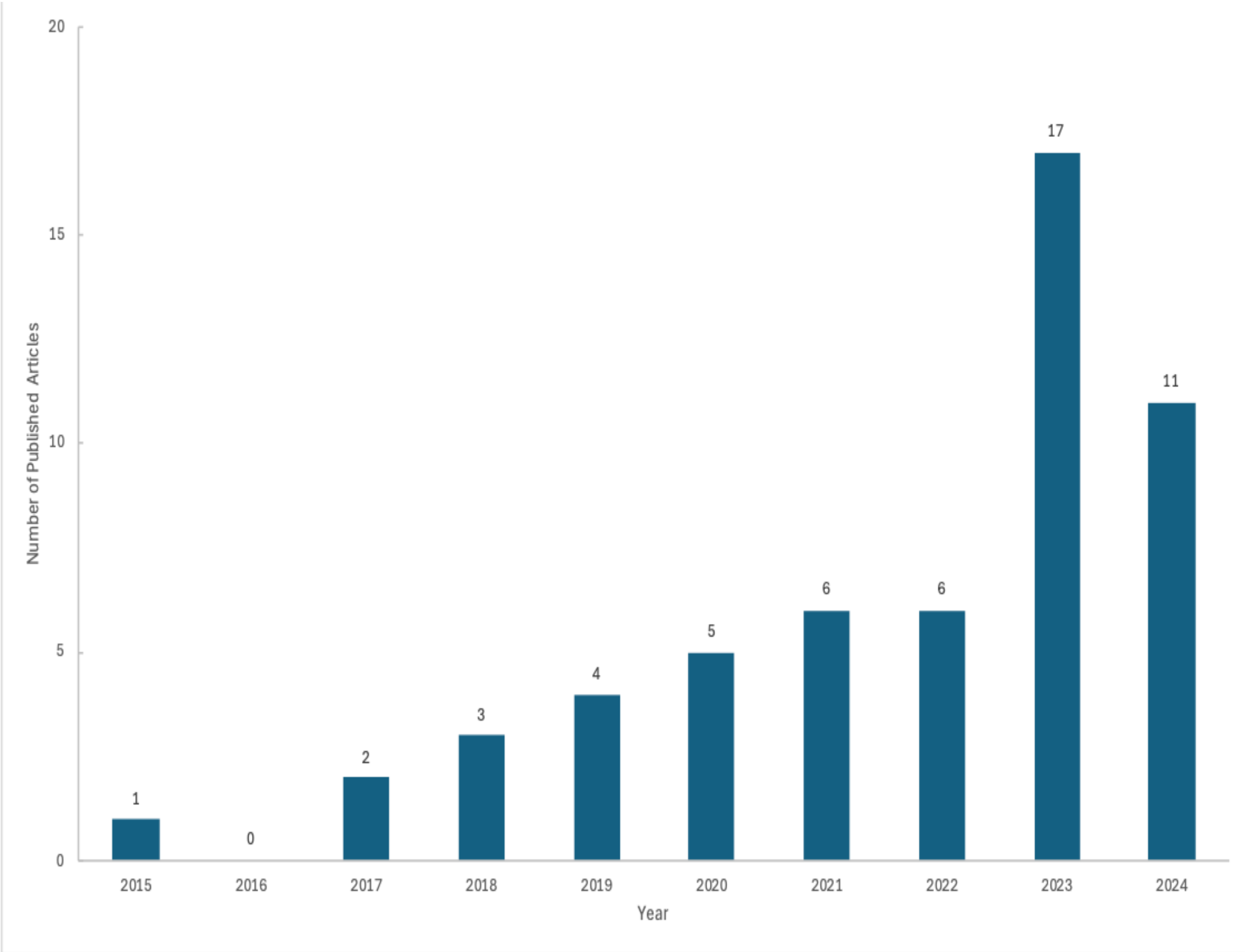
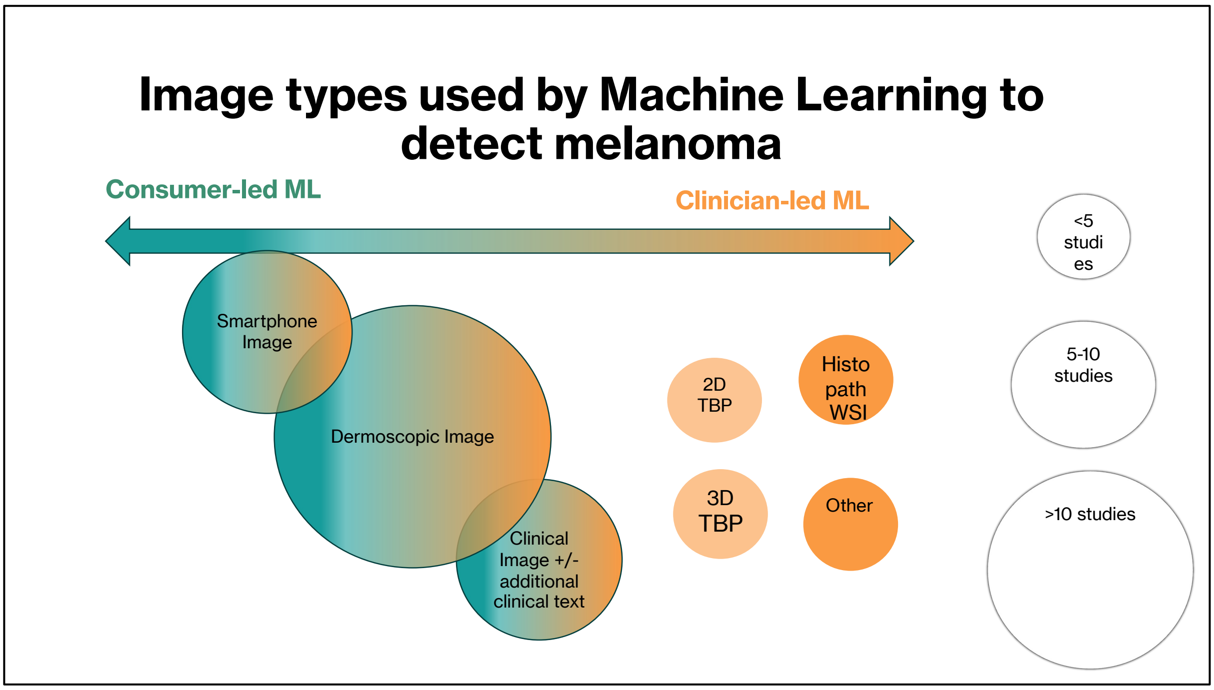


Figure 7 Types of images evaluated by Machine Learning



2D TBP – 2D Total Body Photography; 3D TBP – 3D Total Body Photography;
Histopath WSI – Histopathology Whole Slide Image

Figure 7 shows the continuum between consumer-led ML applications and clinician-led ML applications. Consumer-led machine learning can include using smartphones to collect images +/- dermoscopic lens whereby clinical images may be taken within a healthcare setting. 2D and 3D Total Body Photography (TBP) is an example of clinician-led ML, histopathology whole slide image is also an example of machine learning assisting clinicians.

Figure 8 Receiver Operating Arear Under the Curve (ROAUC) of machine learning in melanoma diagnostic accuracy studies

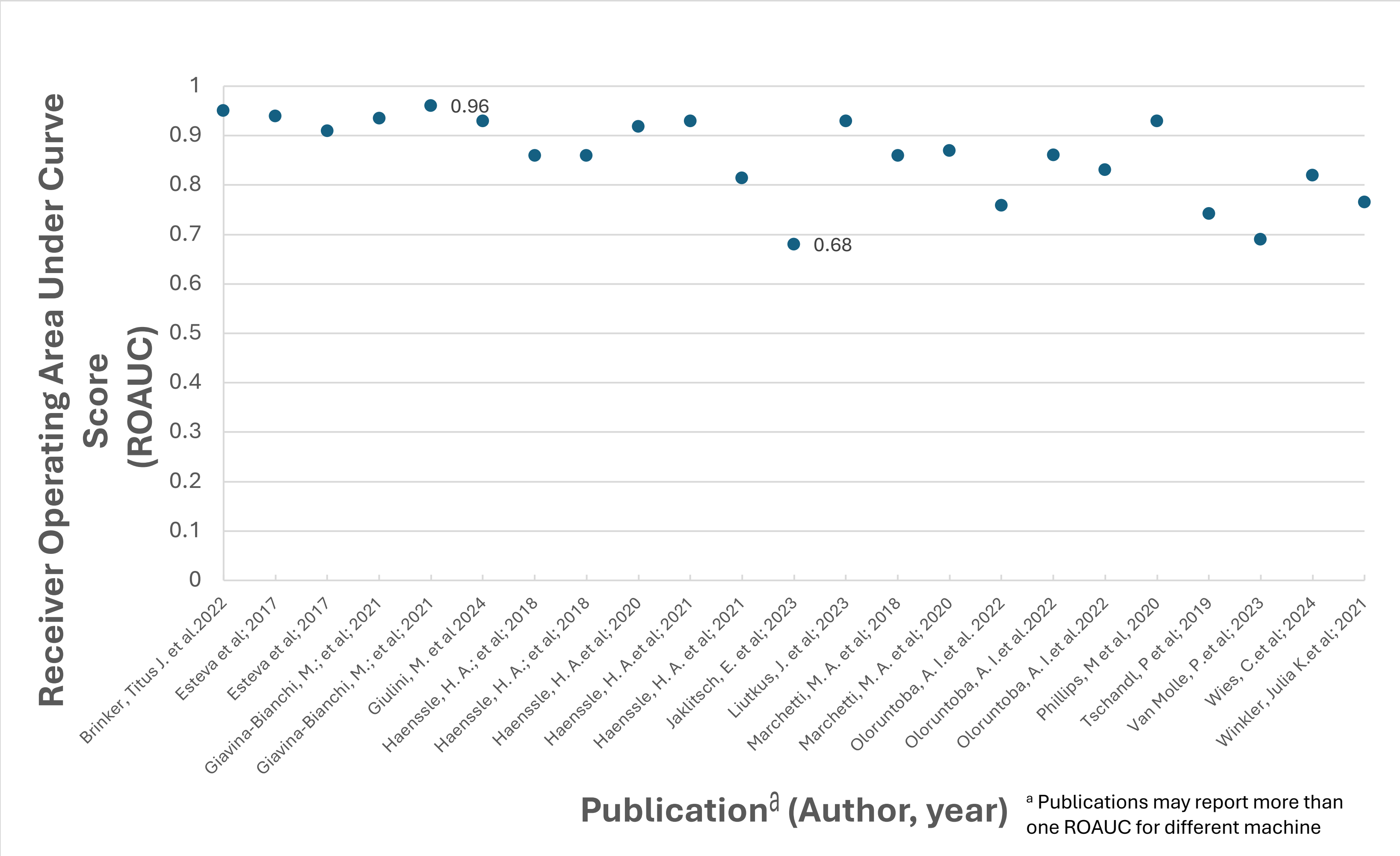


Figure 9 Reported ROAUC of Machine Learning used to detect melanoma in a clinical setting

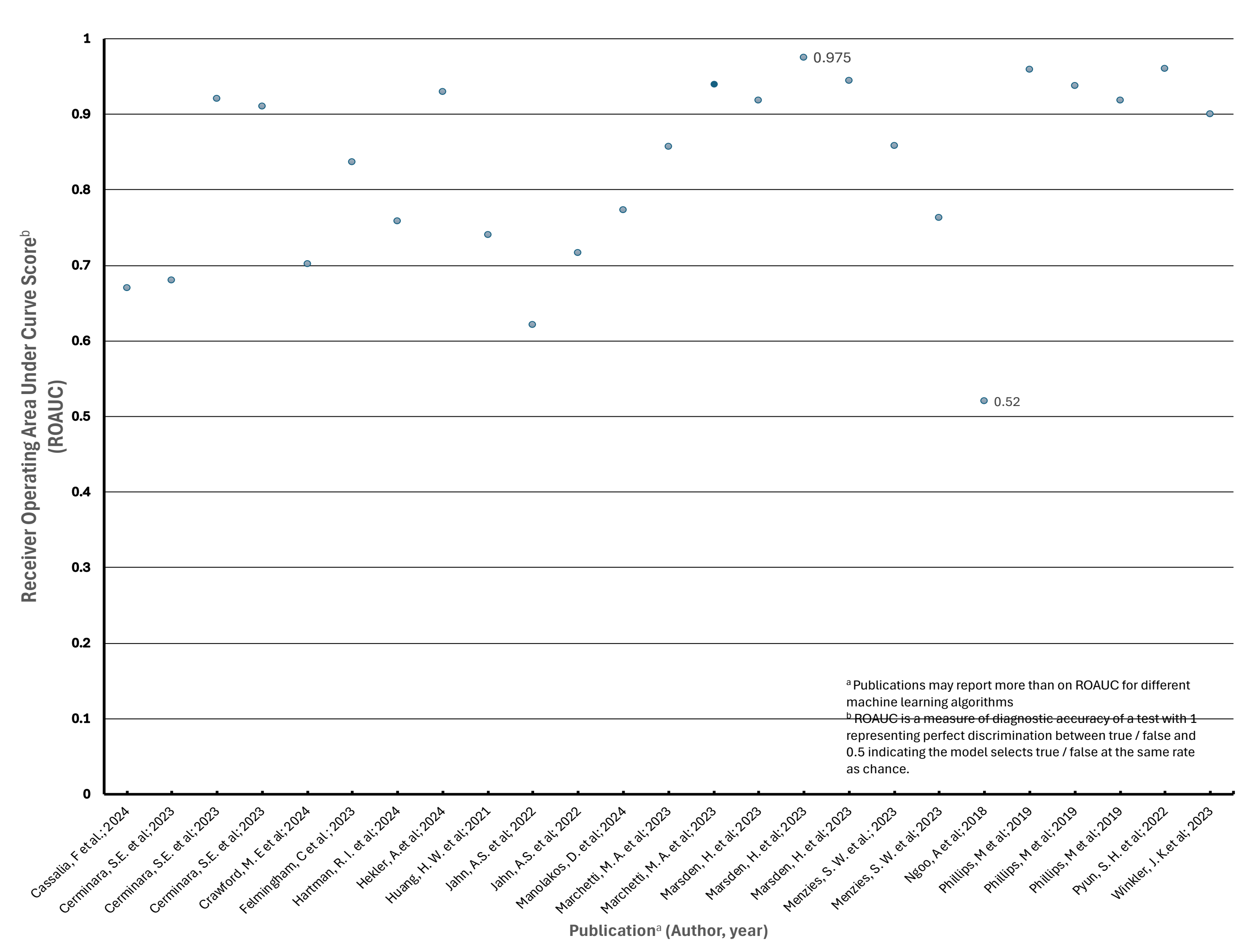
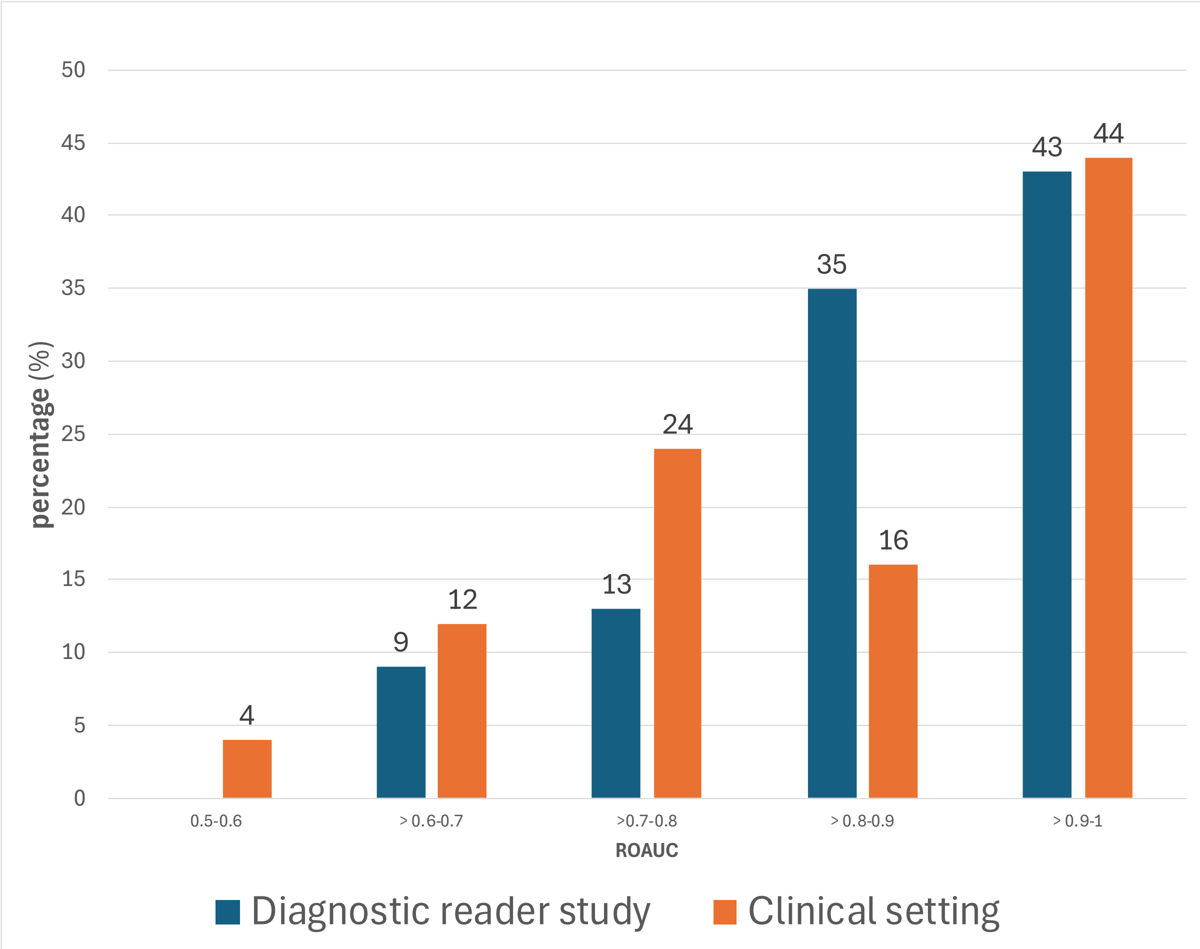


Figure 10 Reported ROAUC of Machine Learning



Appendix A

Search Strategy

Medline and Embase

Ovid MEDLINE(R) ALL <1946 to May 1, 2024>

1	Skin Neoplasms/
2	skin cancer.mp.
3	Melanoma/
4	melanoma.mp.
5	1 or 2 or 3 or 4
6	exp Artificial Intelligence/
7	AI.mp.
8	AI.tw.
9	machine learning.mp.
10	deep learning.mp.
11	neural network*.mp.
12	computational intelligence.mp.
13	6 or 7 or 8 or 9 or 10 or 11 or 12
14	5 and 13
15	limit 14 to yr="2012 -Current"

Scopus

((TITLE-ABS-KEY ("artificial intelligence") OR TITLE-ABS-KEY ("machine learning") OR TITLE-ABS-KEY ("deep learning") OR TITLE-ABS-KEY ("neural network*") OR TITLE-ABS-KEY ("computational intelligence"))) AND ((TITLE-ABS-KEY (melanoma) OR TITLE-ABS-KEY ("skin neoplasm") OR TITLE-ABS-KEY ("skin cancer"))) AND (EXCLUDE (DOCTYPE , "cp") OR EXCLUDE (DOCTYPE , "cr"))

Web of science

(TI=(Melanoma OR melanoma OR "skin cancer" OR "skin neoplasm") OR AB=(Melanoma OR melanoma OR "skin cancer" OR "skin neoplasm"))

AND

AB=("Artificial intelligence" OR “AI” OR "machine learning" OR "deep learning" OR "artificial neural network" OR "neural network*")

Publication date 2012-01-01 to 2024-05-01

Cinahl

S9	S4 AND S8
S8	S5 OR S6 OR S7
S7	"computational intelligence"
S6	(MH "Neural Networks (Computer)") OR (MH "Machine Learning+") OR (MH "Deep Learning") OR (MH "Artificial Intelligence+") OR "machine learning or artificial intelligence or deep learning or neural network"
S5	AB artificial intelligence or ai or a.i.
S4	S1 OR S2 OR S3
S3	AB melanoma
S2	(MH "Skin Neoplasm+") OR "skin cancer"
S1	TX skin neoplasm

Econlit

S8	S3 AND S7
S7	S4 OR S5 OR S6
S6	AB computational intelligence
S5	AB machine learning or artificial intelligence or deep learning or neural network
S4	AB neural networks and machine learning
S3	S1 OR S2
S2	AB skin cancer or skin neoplasm or skin carcinoma

Chapter 5 - Discussion and further research

Overview

This scoping review examined the costs, benefits and diagnostic accuracy of machine learning for the autonomous detection of melanoma in adults at high-risk.

The scoping review conducted as part of this thesis synthesised evidence from 55 studies exploring the diagnostic accuracy, clinical utility and economic implications of using machine learning for the autonomous detection of melanoma in high-risk adults. The review identified a growing number of studies especially from 2023 onwards, highlighting the current interest in the utility of machine learning to detect melanoma. While machine learning demonstrated high diagnostic accuracy in many reader studies, particularly those using dermoscopic images, this accuracy was reduced in real-world clinical settings. Only a limited number of studies evaluated economic outcomes or implementation models in high-risk populations, and trust in consumer-led machine learning applications remained low among both patients and clinicians. Although many of the studies investigating machine learning applications in high-risk populations were conducted in Australia, no specific health economic research was identified in an Australian context.

Diagnostic accuracy

Machine learning used to detect melanoma in high-risk populations demonstrated promising diagnostic accuracy in image-based reader studies with 78% reporting ROAUC values >0.8. However, in clinical settings, only 60% of studies reported ROAUC values >0.8. This discrepancy highlights the performance gap between diagnostic reader studies, which rely on images from large image databases and real-world clinical studies which are representative of clinical settings. Variability in image quality, lighting, skin types, artifacts are just some of the reasons why robust real-world validation in clinical settings are needed.

Diagnostic accuracy is impacted by prevalence of a disease. The accuracy of machine learning in a general population is different to the accuracy in a high-risk population where the prevalence of melanoma is higher. This is an important consideration when determining the generalisability of a machine learning device. Standardised trial parameter's including patient-led or clinician-led machine learning applications, comparators, outcomes, trial design and application in a clinical setting is important to allow for direct comparisons. For example, consider if it is acceptable to use a clinician's decision as the gold standard comparator or if histopathology should be used. The use of different gold standard comparators was identified as a limitation in the scoping review.

Costs and benefits

Only five studies reported economic or preference-based outcomes, with two conducting formal health economic evaluations. Of these Gomez Rossi et al (2022) found no significant cost or QALY advantage of machine learning over dermatologist-led care using a Markov model.[82] In contrast, Marsden et al (2024) reported cost savings and reduced clinician time from real-world integration of machine learning into the UK NHS teledermatology service.[83] The contrast between these findings highlights methodological challenges in economic modelling, particularly the reliance on diagnostic accuracy reader study data versus real-world clinical studies.

Few studies reported long-term cost outcomes, time horizons or the shelf-life of cost-effectiveness in a rapidly evolving technological landscape. This is a key gap in the data as without this, it is difficult to justify the investment in machine learning for melanoma detection in high-risk groups. Most studies identified in the review would mention costs and benefits of machine learning in the discussion section of the article, however empirical research was limited. Potential benefits, for example with consumer-led machine learning applications in rural settings or in areas where access to dermatologists is difficult was not reported.

Consumer vs clinician-led machine learning applications

Differences in perceived utility between consumer-led and clinician-led machine learning applications were identified. Preference studies indicated higher trust in clinician interpreted machine learning application outputs, compared to standalone smartphone apps. For example, only 16% of high-risk patients preferred a smartphone app alone for lesion assessment, compared with 89% and 98% who trusted dermatologist or TBP assessments respectively.

Clinicians also expressed caution, with dermatologists reporting low confidence in consumer-led machine learning tools. These concerns reflect broader concepts such as algorithm transparency and the 'black box'. One should consider if consumers or clinicians are more cautious of machine learning and its application to melanoma detection. Improving machine learning interpretability (i.e. understanding the reason why 'melanoma' or 'no melanoma' is determined by machine learning) and integration of machine learning into established clinical workflows is important in order to improve acceptability and uptake.

Accessibility

When evaluating machine learning for melanoma detection in high-risk populations, it is important to consider the inherent trade-off between accessibility, diagnostic performance and cost. Highly accessible consumer-led machine learning applications may offer equitable access, particularly in rural settings but these may have lower diagnostic accuracy. Alternatively, clinician-led machine learning applications such as 3D TBP may only be in centralised locations and require an appointment. These systems may achieve higher accuracy but may be less accessible and may be more costly to implement and maintain, limiting the population-level impact.

Risk of overdiagnosis

Evidence suggests that widespread use of consumer machine learning apps in general populations may lead to increased false positives and health care utilisation. One study in the Netherlands reported higher rates of claims for benign and premalignant lesions in app users versus controls, raising concerns about overdiagnosis[80]. This supports the strategy of stratifying the use of machine learning to high-risk populations and the potential for benefit and improved cost-effectiveness in this group.

Implications for future research and clinical practice

This review identifies several gaps and opportunities for future research. Clinical trials conducted in clinical settings are essential to understand machine learnings diagnostic performance and workflow integration in the population that the machine learning is validated in. Economic evaluations need to be conducted in high-risk populations including Australia and should include long-term time horizons and analysis to reflect technological change. Clinical pathways need to be designed to determine where machine learning adds the most value and governance frameworks specific for machine learning are required.

Strengths and limitations

The basis of my research was a scoping review. It used the JBI framework and PRISMA-ScR framework which provided methodological robustness, and it is a comprehensive synthesis of studies evaluating machine learning for melanoma detection in high-risk populations.

However, limitations include the lack of quality appraisal of the studies identified, consistent with scoping review methodology, heterogenous study types and the small number of economic and preference-based studies which restricts robust conclusions about costs, benefits and acceptability.

Chapter 6 – Conclusion

The current evidence supporting machine learning applications as a decision support tool in high-risk patients from a cost-effectiveness perspective is limited. This thesis has highlighted the heterogenous nature of the data for the accuracy and utility of machine learning for melanoma detection in high-risk populations. Further studies are required, and consideration should be given to whether machine learning applications to detect melanoma should be consumer-led or clinician-led.

As researchers we need to be agile and responsive to the rapid increase in research and reporting of machine learning accuracy for the detection of melanoma. As machine learning continues to evolve, its utility in healthcare will depend not only on its accuracy but also on its integration into clinical workflows, regulatory compliance, ethical implementation, accessibility, cost and acceptance by users. Continued research, particularly in diverse and high-risk populations is essential for machine learning to fulfill its promise to improve patient outcomes.

Table 1 Summary of Health Economic Evaluation

Author	Title	Intervention	Comparator	Health system perspective	Outcome
Gomez Rossi, J. et al (2022)	Cost-effectiveness of Artificial Intelligence as a Decision-Support System Applied to the Detection and Grading of Melanoma, Dental Caries, and Diabetic Retinopathy	AI to detect melanoma in dermoscopic images	Dermatologist, Standard of care	Health insurance payer and patient contribution	The use of AI was associated with mean costs of USD $750 (95% CI $608-970); vs control $759 (95% CI, $618-970) and QALY 86.6 (95%CI 84.9-88.0) for both AI and control. There was no statistical difference between the intervention (AI) or the control using this model.
Marsden, H. et al (2024)	Accuracy of artificial intelligence as a medical device as part of a UK-based skin cancer teledermatology service	Artificial Intelligence as a Medical Device (AIaMD) for identifying premalignant and benign skin lesions in a clinical setting	Teledermatology, Standard of Care	UK Health system perspective	AIaMD identified more skin lesions that did not need a biopsy or urgent referral compared to Standard of Care (p-value = 0.001) whilst retaining comparable sensitivity for identifying skin cancer. Cost savings arose from a reduction in patient reviews, face-to-face assessments, biopsies and potential specialist hours saved. For each 1,000 patients entering the pathway, using AIaMD was reported to save approx £156,063.79 and 259 specialist hours compared to teledermatology standard of care.

Table 2 Study Characteristics

Study Characteristics						**Participant characteristics**	**Intervention**				**Image**	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Agarwala, S et al	2021	Accuracy of a convolutional neural network for dermatologic diagnosis of tumors and skin lesions in a clinical setting	USA	To examine the real-world effectiveness of AI technology by comparing its accuracy for predicting the final clinicopathological diagnosis of a skin lesion against the accuracy of the initial clinical impression of 5 practicing dermatologists.	cohort study	178	CNN	Not reported	www.triage.com	Not reported	clinical image	353
Anderson, J. M. et al	2023	Artificial Intelligence vs Medical Providers in the Dermoscopic Diagnosis of Melanoma	USA	To evaluate the performance of Triage's dermoscopic classifier in identifying lesions as benign or malignant to determine whether AI could assist in the triage of skin cancer cases to shorten time to diagnosis.	diagnostic accuracy study	Not reported (image dataset)	CNN	Not reported	Triage (Triage Technologies Inc)	Not reported	dermoscopic image	100
Brinker, T. J. et al.	2019	A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task	Germany	To compare the performance of a CNN trained using dermoscopic images, with 145 dermatologists in a clinical image classification task to identify melanoma in clinical photographs.	diagnostic accuracy study	Not reported	CNN ResNet50	Not reported	Not reported	Not reported	clinical image	100
Brinker, T. J. et al.	2019	Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task	Germany	To compare the performance of a deep-learning trained algorithm trained by open- source dermoscopic images to dermatologists.	diagnostic accuracy study	Not reported	CNN ResNet50	Not reported	Not reported	Not reported	dermoscopic image	100
Brinker, Titus J. et al.	2022	Diagnostic performance of artificial intelligence for histologic melanoma recognition compared to 18 international expert pathologists	International (8 different countries)	To compare the performance of CNN with 18 international expert pathologists at diagnosing melanoma and nevi in hematoxylin-eosin-stained whole-slide images	diagnostic accuracy study	Not reported	CNN	Not reported	-	-	hematoxylin-eosin-stained whole slide image	100

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Brodsky, V. et al	2023	Performance of Automated Classification of Diagnostic Entities in Dermatopathology Validated on Multisite Data Representing the Real-World Variability of Pathology Workload	USA and Tanzania	To assess the value of machine learning for microscopic tissue evaluation in dermatopathology. Comparing the diagnoses of hematoxylin and eosin-stained glass slides by 2 senior board-certified pathologists vs ML.	diagnostic accuracy study	300	CNN, Resnet 34 / Resnet 18	Not reported	Mechanomind	hematoxylin eosin-stained glass slides scanned using a high- resolution Motic Digital Pathology EasyScan Pro6 scanner at 0.26 lm per pixel and the Ventana iScan Coreo scanner, both set at 340 magnification.	whole slide image	300
Cassalia, F et al.	2024	Melanoma detection: Evaluating the classification performance of a deep convolutional neural network and dermatologist assessment via a mobile app in an Italian real-world setting	Italy	To evaluate the performance of a pre-trained CNN model and online assessments by seven dermatologists reviewing images sent through the 'Clicca il Neo' app	cohort study	1813	CNN	Build 2021	ModelDerm	Model Dermatology app; Smartphone vs 'Clicca il Neo' app and smartphone	smartphone image	2226
Cerminara, S.E. et al	2023	Diagnostic performance of augmented intelligence with 2D and 3D total body photography and convolutional neural networks in a high-risk population for melanoma under real-world conditions: A new era of skin cancer screening?	Switzerland	The aim of this study was to investigate the clinical performance of 3D TBP, 2D TBP and dermatologist alone and augmented intelligence (collaboration between AI and physician in decision-making process) in the early detection of melanoma in individuals at high risk of melanoma in a real-world setting	cohort study	143	CNN	Vectra WB360 (3D CNN) ATBM Master FotoFinder (2D CNN)	Vectra WB360 (3D CNN) ATBM Master FotoFinder (2D CNN)	Vectra WB360; ATMB (FotoFinder)	dermoscopic and clinical image	1690 lesions
Crawford, M. E et al	2024	Using Artificial Intelligence as a Melanoma Screening Tool in Self-Referred Patients	Canada	The aim of this study was to assess the potential use of AI as a diagnostic tool for individuals concerned that a pigmented lesion may be cancerous.	cohort study	318	CNN	Version 6	FotoFinder Moleanalyzer Pro®	FotoFinder Moleanalyzer Pro, FotoFinder Medicam 800HD	dermoscopic image	381 lesions
Diaz-Ramon, J. L. et al	2023	Melanoma Clinical Decision Support System: An Artificial Intelligence-Based Tool to Diagnose and Predict Disease Outcome in Early-Stage Melanoma Patients	Spain	The aim of this study was to assess the performance of an AI algorithm based on clinical data and dermatoscopic imaging for the early diagnosis of melanoma, and its capacity to define the metastatic progression of melanoma through serological and histopathological biomarkers.	cohort study	196	Diagnostic module - CNN Prognostic module - random forest DT and SVM	Not reported	The Melanoma Clinical Decision Support System (CDSS)	The melanoma CDSS application	dermoscopic and clinical	Not reported

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Esteva et al	2017	Dermatologist-level classification of skin cancer with deep neural networks	USA	To test the performance of a CNN trained using a dataset of 129,450 clinical images against at least 21 dermatologist on diagnosing biopsy proven clinical images with two binary classification use cases: keratinocyte carcinomas versus benign seborrheic keratoses; and malignant melanomas versus benign nevi.	diagnostic accuracy study	Not reported	Inception v3	Inception v3 CNN architecture	-	Not reported	clinical image	1942
Felmingham, C et al	2023	improving skin cancer management with ARTificial intelligence: A pre- post intervention trial of an artificial intelligence system used as a diagnostic aid for skin cancer management in a real-world specialist dermatology setting	Australia	To assess the performance of a CNN developed by MoleMap and Monash eResearch for skin cancer diagnosis compared with Teledermatologist assessment	cohort study	214	CNN	Not reported	MoleMap	proprietary dermoscopic camera	dermoscopic image	743
Fink, C. et al	2020	Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined nevi and melanomas	Germany	To investigate the diagnostic performance of a CNN in differentiating combined naevi and melanoma in comparison with a group of trained dermatologists with different levels of experience.	diagnostic accuracy study	72	CNN; based on GoogleNet Inception_v4	Jun-19	Moleanalyzer-Pro	Moleanalyzer-Pro	dermoscopic image	72
Giavina-Bianchi, M.; et al	2021	Implementation of artificial intelligence algorithms for melanoma screening in a primary care setting	Brazil	To investigate the performance of an AI system used by primary care physicians to detect melanoma.	diagnostic accuracy study	Not reported	DenseNet	Not reported	Not reported	smartphone device	dermoscopic and clinical	308
Giavina-Bianchi, M.; et al	2021	Implementation of artificial intelligence algorithms for melanoma screening in a primary care setting	Brazil	To investigate the performance of an AI system used by primary care physicians to detect melanoma.	diagnostic accuracy study	Not reported	EfficintNetB6	Not reported	Not reported	dermatoscope	dermoscopic and clinical	2633
Giulini, M. et al	2024	Combining artificial intelligence and human expertise for more accurate dermoscopic melanoma diagnosis: A 2- session retrospective reader study	Germany	To assess the performance of clinicians using CNN to detect melanoma from dermoscopic images.	diagnostic reader study	100	CNN	Not reported	Not reported	Not reported	dermoscopic image	100

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Goessinger, E. V. et al	2024	Consistency of convolutional neural networks in dermoscopic melanoma recognition: A prospective real- world study about the pitfalls of augmented intelligence	Switzerland	To assess and compare the repeatability of computer-guided risk scores when classifying 5 sequential dermoscopic images of identical lesions using two commercially available CNNs in indirect comparison.	cohort study	66	CNN-1: Dermoscopy explainable intelligence (DEXI) score, CNN-2: MoleAnalyzer-Pro, FotoFinder ATBM® Systems GmbH	CNN-1: Version 1 CNN-2: Version 6.0.5	DEXI / Visiomed Moleanalyzer-Pro, FotoFinder ATBM / FotoFinder medicam 1000)	digital dermoscope	dermoscopic image	17 sets of 5 images
Goessinger, E. V. et al	2024	Patient and dermatologists' perspectives on augmented intelligence for melanoma screening: A prospective study	Switzerland	To investigate the perspectives of patients and dermatologists after skin cancer screening by human, artificial and augmented intelligence.	cohort study	205 patients / 8 dermatologists	CNN-1: Dermoscopy explainable intelligence (DEXI) score, Canfield Scientific, , CNN-2: MoleAnalyzer-Pro, FotoFinder ATBM® Systems GmbH,	CNN-1: Version 1 CNN-2: Version 6.0.5	DEXI / Visiomed MoleAnalyzer-Pro, FotoFinder ATBM / FotoFinder medicam 1000)	digital dermoscope	dermoscopic image	-
Haenssle, H. A.; et al	2018	Man against Machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists	International (17 countries)	To train, validate and test a deep learning CNN for the diagnostic classification of dermoscopic images of lesions of melanocytic origin (melanoma, benign nevi) and to compare the results to a large group of 58 dermatologists.	diagnostic accuracy study	300	Google's inception v4 CNN additionally trained with more than 100,000 labeled digital images	Inception v4 CNN	Not reported	dermatoscope	dermoscopic image	100

Table 2 Study Characteristics

Study Characteristics						**Participant characteristics**	**Intervention**				**Image**	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Haenssle, H. A.; et al	2020	Man against machine reloaded: performance of a market- approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions	international	To compare the CNNs sensitivity and specificity in comparison with dermatologists' management decisions based in the combination of clinical close-up images, dermoscopic images and textual case information.	diagnostic accuracy study	100	Jun-19		Moleanalyzer Pro (FotoFinder Systems GmbH Germany)	Moleanalyzer Pro	clinical close-up image, dermoscopic image, textual case information (patient age, sex, location of lesion)	100
Haenssle, H. A.; et al	2021	Skin lesions of face and scalp - Classification by a market-approved convolutional neural network in comparison with 64 dermatologists		To investigate a broad spectrum of skin lesions of the face and scalp by a market-approved CNN in comparison with the management decisions of dermatologists after reviewing clinical information in addition to images (similar to a classical store-and- forward teledermatology consultation).	diagnostic accuracy study	100	modified version of Google's inception v4 CNN trained with dermoscopic images	Not reported	Moleanalyzer Pro (FotoFinder Systems GmbH Germany)	Moleanalyzer Pro	dermoscopic image, textual information including age, sex, anatomic location.	100
Hartman, R. I. et al	2024	Multicenter prospective blinded melanoma detection study with a handheld elastic scattering spectroscopy device	USA and Australia	To investigate the performance of the ESS device in the detection of melanoma.	cohort study	311	DermaSensor	Not reported	DermaSensor	Handheld ESS device	Elastic scattering spectroscopy (ESS)	440 lesions
Hekler, A.et al	2024	Using multiple real-world dermoscopic photographs of one lesion improves melanoma classification via deep learning	Germany	To evaluate the performance of a dermoscopic image classifier to detect melanoma using multiple real-world images per lesion (MV-Real)	cohort study	617	CNN with ConveNeXT	Not reported	Not reported	dermatoscope with smartphone	dermoscopic images	656 lesions / 6 photos per lesion
Huang, H. W. et al	2021	Development of a lightweight deep learning model for cloud applications and remote diagnosis of skin cancers	Taiwan	To build a light-weight skin cancer classification model based on deep learning methods for aiding first-line medical care	diagnostic accuracy study	1222	CNN, DenseNet and EfficientNet	Not reported		mage	clinical image	1287

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Jahn, AS et al	2022	Over-Detection of Melanoma-Suspect Lesions by a CE- Certified Smartphone App: Performance in Comparison to Dermatologists, 2D and 3D Convolutional Neural Networks in a Prospective Data Set of 1204 Pigmented Skin Lesions Involving Patients' Perception	Switzerland	To assess the diagnostic accuracy of the CE-certified smartphone app SkinVision in melanoma detection in a real-world setting and to provide an insight into HCP and layperson evaluation of mHealth apps for melanoma screening.	cohort study	114	CNN	App version 6.8.1	SkinVision	iOS based iPhone SE smartphone equipped with a 12- megapixel camera and the SkinVision app version 6.8.1	smartphone image	1204
Jaklitsch, E. et al	2023	Clinical Utility of an AI-powered, Handheld Elastic Scattering Spectroscopy Device on the Diagnosis and Management of Skin Cancer by Primary Care Physicians	USA	To determine whether the use of an ESS device can help PCPs improve their diagnostic accuracy of skin malignancies.	diagnostic reader study	57 primary care physicians	CNN	Not reported	DermaSensor	EES	dermoscopic image	50
Kommoss, K. S. et al	2023	Observational study investigating the level of support from a convolutional neural network in face and scalp lesions deemed diagnostically 'unclear' by dermatologists	Germany	To investigate the level of support that might be expected from a CNN-based AI-support system when applied to skin lesions deemed diagnostically ‘unclear’ by dermatologists.	diagnostic accuracy study	100	CNN - Google’s Inception_v4	Not reported	Moleanalyzer Pro (FotoFinder Systems GmbH Germany)	Dermoscopic images	dermoscopic and clinical images	100
Kränke, T. et al	2023	New AI-algorithms on smartphones to detect skin cancer in a clinical setting—A validation study	Austria	To evaluate the diagnostic and risk-assessment accuracy of 2 CNNs in comparison to the histopathological and clinical diagnosis.	cohort study	238	2 CNNs Analyse (n - 238 patients and 1171 lesions) Detect (n - 92 patients and 552 lesions)	Not reported	Not reported	Android or IoS mobile devices with Analyze and detect CNN apps installed.	clinical image	1171 lesions

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Liutkus, J. et al	2023	Accuracy of a Smartphone-Based Artificial Intelligence Application for Classification of Melanomas, Melanocytic Nevi, and Seborrheic Keratoses	Lithuania	To assess the accuracy of a smartphone based “You Only Look Once” neural network model (NNM) for the classification of melanomas, melanocytic nevi, and seborrheic keratoses.in addition to evaluating the NNM’s performance against histopathologically confirmed diagnoses, this study compared its accuracy with that of skilled dermatologists and beginner raters	diagnostic accuracy study	Not reported	YOLO NNM - BottleNeckCSP CNN	Not reported	SmartVisSolution©	smartphone	dermoscopic image	100
MacLellan, A. N. et al	2021	The use of noninvasive imaging techniques in the diagnosis of melanoma: a prospective diagnostic accuracy study	Canada	To compare a dermatologists diagnosis using clinical examination at the bedside and remote diagnosis using clinical and dermoscopic images (teledermoscopy) with 3 non-invasive smart-imaging devices to determine the relative accuracy of each in detecting melanoma	cohort study	184	CNN Not reported however tradenames provided	Not reported	MelaFind, FotoFinder Tuebinger, Moleanalyzer Pro and Verisante Aura	DermLite Cam V2, Verisante Aura, MelaFind, FotoFinder Tuebinger and Moleanalyzer Pro	dermoscopic and clinical images	209
Maier, T. et al	2015	Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result	Germany	To evaluate prospectively the sensitivity and specificity of a recently developed smartphone application using fractal image analysis for the risk evaluation algorithm in the diagnosis of malignant melanoma compared to clinical diagnosis and histopathological result.	cohort study	195	CNN	Not reported	SkinVision	Phone 4S with 8MP camera.	smartphone image	636 (minimum 3 images per lesion)
Manolakos, D. et al	2024	Use of an elastic-scattering spectroscopy and artificial intelligence device in the assessment of lesions suggestive of skin cancer: A comparative effectiveness study	USA	To assess the safety and effectiveness of an elastic-scattering spectroscopy (ESS) device in evaluating lesions suggestive of skin cancer.	cohort study, multicenter	394	CNN, Resnet 18	version 1 / version 3	DermaSensor	DermaSensor - handheld EES device for use by GP	Elastic scattering spectroscopy (ESS)	614 lesions

Table 2 Study Characteristics

Study Characteristics						**Participant characteristics**	**Intervention**				**Image**	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Marchetti, M. A. et al	2018	Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images	USA	To compare 8 dermatologists' performance at melanoma classification from dermoscopic images to the top 5 automated algorithms at the 2016 international Symposium on Biomedical Imaging.	diagnostic reader study	Not reported	Deep Learning	Not reported	Not reported	dermoscopic image	dermoscopic image	100
Marchetti, M. A. et al	2020	Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: Results of the International Skin Imaging Collaboration 2017	USA	To determine if computer algorithms from an international melanoma detection challenge can improve dermatologists' accuracy in diagnosing melanoma	diagnostic reader study	Not reported	CNN; Depp Learning	Not reported	Not reported	dermoscopic image	dermoscopic image	150
Marchetti, M. A. et al	2023	Prospective validation of dermoscopy-based open-source artificial intelligence for melanoma diagnosis (PROVE-AI study)	USA	To prospectively validate the accuracy of ADAE for melanoma diagnosis and to assess its potential impact on dermatologist clinical decision making.	cohort study	435	All Data are Ext (ADAE) open source, non-commercial AI algorithm based on EfficientNet and ResNet architecture	Not reported	ADAE	dermoscopy; CNN via web interface	dermoscopic image and clinical metadata	603 lesions
Marchetti, M. A. et al	2023	3D Whole-body skin imaging for automated melanoma detection	USA	The objective of this pilot study was to deter- mine if a prediction model using automated analysis of 3D- images could accurately discriminate between melanoma and non-melanoma lesions.	diagnostic accuracy study	35	VECTRA DermaGraphix research software	Not reported	VECTRA WB360 / VECTRA DermaGraphix research software	Vectra WB360 TBP	3D avatar	23538

Table 2 Study Characteristics

Study Characteristics						**Participant characteristics**	**Intervention**				**Image**	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Marsden, H. et al	2024	Accuracy of an artificial intelligence as a medical device as part of a UK-based skin cancer teledermatology service	UK	To demonstrate that the AIaMD has a higher rate of correctly classifying lesions that did not need to be referred for biopsy or urgent face-to-face dermatologist review, compared to teledermatology standard of care, while achieving the same sensitivity to detect malignancy. Secondary endpoints included the sensitivity, specificity, positive and negative predictive values, and number needed to biopsy to identify one case of melanoma or SCC by both the AIaMD and standard of care.	cohort study, cost impact assessment	622	Not reported	Not reported	Deep Ensemble for the Recognition of Malignancy (DERM)	Smartphone and DermLite DL1 dermoscopic lens attachment.	clinical and dermoscopic image	789 lesions
Marsden, H. et al	2023	Effectiveness of an image analyzing AI-based Digital Health Technology to identify Non-Melanoma Skin Cancer and other skin lesions: results of the DERM-003 study	UK	To demonstrate the effectiveness of the AIaMD to identify SCC and BCC. Secondary objectives included demonstrating the effectiveness of the AIaMD to identify premalignant and benign conditions, comparing the AIaMD performance to dermatologists, and demonstrating the feasibility of image capture in a clinic setting.	cross-sectional study	544	Not reported	version 3	Deep Ensemble for the Recognition of Malignancy (DERM)	smartphone cameras (iPhone 6S, iPhone 11, Samsung 10) with a DL1 dermoscopic lens attachment	smartphone / digital image with dermoscopic lens	611 lesions
Menzies, S. W. et al.	2023	Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial	Australia / Austria	To test in the clinic whether there was equivalence between AI algorithms and clinicians for the diagnosis and management of pigmented skin lesions.	cross-sectional study, prospective	diagnostic n=124 management n=66	CNN - efficientNet-B3 or ResNet-10	v2.3 and v4.4	DermEngine / MetaOptima Technology	smartphone and dermatoscope / TBP	smartphone / digital image with dermoscopic lens	diagnostic - 172 lesions management - 5696 lesions

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Miller, I. J.; et al	2023	Implementation of artificial intelligence for the detection of cutaneous melanoma within a primary care setting: prevalence and types of skin cancer in outdoor enthusiasts	Australia	To assess the prevalence of skin cancer in individuals who regularly participate in activities outdoors and to report the performance parameters of a commercially available AI-powered software to assess the predictive risk of MM development.	cross-sectional study	423	CNN	3.4.1.0_(x64)	Moleanalyzer-Pro	Moleanalyzer-Pro with Medicam 1000 attachment.	dermoscopic image	Not reported - 423 full body scans
Ngoo, A et al	2018	Efficacy of smartphone applications in high-risk pigmented lesions	Australia	To assess the capability of melanoma smartphone applications in making clinical decisions about risk, compared with lesion assessment by specialist trained dermatologists.	prospective cohort study	30	SkinVision iOS; SkinVision android; SpotMole direct; SpotMole indirect; Dr Mole android	2015/2016	SkinVision SpotMole Dr Mole	Smartphone	smartphone image	57 pigmented lesions
Oloruntoba, A. I.et al.	2022	Assessing the Generalizability of Deep Learning Models Trained on Standardized and Nonstandardized Images and Their Performance Against Teledermatologists: Retrospective Comparative Study	Australia / Denmark	To assess the generalizability of CNN models trained on standardised and nonstandardised images	diagnostic accuracy study	Test set 1 n-519. test set 2 & 3 (image data set) Not reported	ResNet - 50 / ResNet-18	Not reported	Not reported	Test set 1 - iPhone 6 with a handy scope (FotoFinder Systems GmbH) Test set 2 - EOS Rebel T6i camera and ATBM mole-mapping system (FotoFinder Systems) Test set 3 - ISIC 2020 data set	dermoscopic images	569/ 422/ 33126
Phillips, M et al	2019	Assessment of Accuracy of an Artificial Intelligence Algorithm to Detect Melanoma in Images of Skin Lesions	UK	The aim of this study was to evaluate the ability of the Deep Ensemble for Recognition of Malignancy algorithm to detect melanoma from images of both biopsied and non-biopsied pigmented skin lesions, prospectively captured in dermatology and plastic surgery clinics, and to compare this with clinical diagnoses made by specialists.	prospective, cross-sectional study	514	DNN	Sep 2018-Feb2019	DERM (Skin Analytics)	iPhone 6s Galaxy s6 DLSR Dermatoscopic lens attachment DermLite DL1 and DermLite Foto II Pro	smartphone / digital image with dermoscopic lens	1550

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Phillips, M et al,	2020	Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy	UK	"The primary aim of this study was to evaluate the diagnostic accuracy of an AI algorithm (Deep Ensemble for Recognition of Melanoma [DERM]). The secondary aim was to improve the methodology for evaluating an AI diagnostic tool by comparing DERM's performance with clinical examination by physicians and stratification based on level of expertise and use of dermoscopy using a meta-analysis of diagnostic studies. "	diagnostic accuracy study	Not reported (image dataset)	DNN	Not reported	DERM (Skin Analytics)	dermatoscope	dermoscopic image	7102
Pyun, S. H. et al.	2022	Real-time, in vivo skin cancer triage by laser-induced plasma spectroscopy combined with a deep learning–based diagnostic algorithm	Australia	To investigate the diagnostic accuracy and safety of real-time noninvasive in vivo skin cancer diagnostics utilizing nondiscrete molecular LIPS combined with a deep neural network (DNN) ebased diagnostic algorithm.	cohort study	353	DNN	Not reported	LIPS	Q-switched neodymium-doped yttrium aluminium giant laser with wavelength of 1064nm and pulse duration of 4ns that spectral data from the tissue was analyzed by the DNN.	spectral data emitted from tissue treated with LIPS laser	612 lesions 8590 spectra
Gomez Rossi, J. et al	2022	Cost-effectiveness of Artificial Intelligence as a Decision- Support System Applied to the Detection and Grading of Melanoma, Dental Caries, and Diabetic Retinopathy	USA	To evaluate AI's cost-effectiveness as a diagnostic support system in dermatology, dentistry, and ophthalmology in different countries using health economic modeling via Markov models with a lifetime horizon. We decided to account for AI as fee-for-service and explored how it factored into cost-effectiveness (per-person) through sensitivity analysis. Our research goal was to test the assumption that an AI with superior diagnostic accuracy used as a decision-support system would always clearly reduce costs and improve outcomes.	Markov-model based cost effectiveness analysis; based on a diagnostic reader study (Brinker 2019)	Not reported	CNN ResNet50	Not reported	-	dermatoscope	dermoscopic image	100

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Sangers, T et al	2022	Validation of a Market-Approved Artificial Intelligence Mobile Health App for Skin Cancer Screening: A Prospective Multicenter Diagnostic Accuracy Study	The Netherlands	To validate an mHealth app currently approved for consumers in Europe, Australia, and New Zealand, which uses a deep-learning convolutional neural network (CNN) for skin premalignancy and malignancy detection in the setting of a dermatology department. The primary outcome of this study was the sensitivity and specificity of the app to detect premalignant and malignant skin lesions.	Prospective diagnostic accuracy study	372	CNN	CNN version RD-174	SkinVision	iPhone iOS 13 or Android 10 operating system with 12MP camera	smartphone image	785
Thissen,M et al	2017	mHealth App for Risk Assessment of Pigmented and Nonpigmented Skin Lesions—A Study on Sensitivity and Specificity in Detecting Malignancy	The Netherlands	To evaluate a smartphone app using image analysis for the risk assessment of skin lesions to determine its sensitivity and specificity in the diagnosis of melanoma and nonmelanoma skin cancer along with actinic keratosis and Bowen's disease.	prospective diagnostic accuracy trial	256	CNN	Not reported	SkinVision	iPhone 5 with 8 megapixel autofocus camera	smartphone image	341
Thomas, L. et al	2023	Real-world post-deployment performance of a novel machine learning-based digital health technology for skin lesion assessment and suggestions for post-market surveillance	UK	To report the prospective performance of a Deep Ensemble for Recognition of Malignancy (DERM) from its deployment within skin cancer pathways at two National Health Service hospitals in the UK.	cohort study	8571	CNN	DERM-version A (DERM- vA) (July 2021 to April 2022), and version B (DERM-vB) (April 2022 to October 2022)	DERM	smartphone and polarised dermoscopic lens attachment (Dermlite DL1 basic) or DLSR with dermoscopic lens attachment	dermoscopic image	8571
Tschandl, Philipp et al	2019	Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks	Austria and Australia	To compare the accuracy of a CNN-based classifier with physicians with different levels of experience to detect non-pigmented skin cancer including SCC, BCC and melanoma.	diagnostic accuracy study	2072 non-pigmented lesions	CNN Inception V3 / ResNet50	Not reported		Dermoscopic images and clinical close-up images were taken with different cameras and dermatoscopes at different resolutions in polarizing or nonpolarizing mode.	dermoscopic and clinical	2072
Udrea, A et al	2020	Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms	Germany	To evaluate the accuracy of the newest version of a smartphone application for risk assessment of skin lesions.	diagnostic accuracy study	5603	CNN	Jan16-July17	SkinVision	smartphone and Android devices with camera function.	smartphone image	6285

Table 2 Study Characteristics

Study Characteristics						Participant characteristics	Intervention				Image	
Author	**Year**	**Title**	**Country**	**Aim**	**Study design**	**No. subjects**	**Machine Learning**	**Version**	**Trade name**	**Device**	**Type**	**Number**
Van Molle, P.et al	2023	Dermatologist versus artificial intelligence confidence in dermoscopy diagnosis: Complementary information that may affect decision-making	Belgium	To train a CNN for skin lesion classification and evaluate its diagnostic performance and uncertainty and compare the results to the assessments by a group of dermatologists.	diagnostic accuracy study	501 (test set)	stochastic neural network - ResNet5	Not reported	Not reported	dermoscopic image	dermoscopic image	501
Wies, C.et al	2024	Evaluating deep learning-based melanoma classification using immunohistochemistry and routine histology: A three-center study	Germany	To investigate the use of deep learning (DL)-based image analysis models on MelanA- stained tissue for melanoma classification in comparison and in addition to the standard H&E-based diagnosis.	diagnostic accuracy study	Dresden 126 / Erlangen 81 / Naples 50	CNN - ResNet	Not reported	-	whole slide image	whole slide image (IHC and H&E slides)	257
Winkler, J. K.et al	2023	Assessment of Diagnostic Performance of Dermatologists Cooperating with a Convolutional Neural Network in a Prospective Clinical Study: Human with Machine	Germany	Aim to investigate the cooperation of dermatologists with a market approved CNN and to measure changes in diagnostic performance when classifying melanocytic lesions and to determine patient acceptance and trust towards the CNN.	prospective, cross-sectional study, cost impact assessment	188	CNN	Not reported	Moleanalyzer Pro (FotoFinder Systems)	Dermatoscope / Moleanalyzer Pro	dermoscopic image	228 lesions
Winkler, J. K.et al	2024	Performance of an automated total body mapping algorithm to detect melanocytic lesions of clinical relevance	Argentina, Australia, Germany and Switzerland	The aim of the study was to investigate the performance of an automated total body mapping algorithm to detect clinically relevant melanocytic lesions in a first international real- world analysis.	prospective cross-sectional study	236	Not reported	version 3.3.1.0	ATMB master	2D automated total body mapping with automated lesions detection (ATMB, master, FotoFinder Systems GmbH)	2D TBP, dermoscopic images	Not reported, 300 TBP scans
Winkler, Julia K.et al	2021	Collective human intelligence outperforms artificial intelligence in a skin lesion classification task	Germany	To investigate the diagnostic performance of a binary and multiclass CNN vs individual dermatologist's vs Collective human intelligence CoHI (120 dermatologists) at detecting malignancy of skin lesions.	diagnostic accuracy study	30	CNN - GoogleNet Inception v4	Not reported	binary CNN: Moleanalyzer Pro (FotoFinder Systems GmbH Germany)	Moleanalyzer Pro	Clinical and dermoscopic images Accompanying case data included age, sex, brief case history.	up to 2 images per lesion (30 lesions)

Table 3 Study Outcomes

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Agarwala, S et al , 2021[84]	cohort study	178	CNN	Dermatologist (pre-biopsy clinical impression)	clinicopathologic diagnosis reference standard. Biopsy results available in 339 cases.					NR	NR	NR	NR	NR	47.4%(42.0-52.8) *strict criteria 63.3% (58-68.4)** permissive criteria
Anderson, J. M. et al , 2023[85]	diagnostic accuracy study	NR (image data set)	CNN	dermatologists (n=23); family physicians (n=7), primary care mid-level providers e.g. nurse practitioners (n=12)	biopsy verified melanomas (ISIC test set)		All primary care providers: 67% (p<0.5) Dermatologists: 77% (NS)	All primary care providers: 48% (p<0.5) Dermatologists:57% (p<0.5)	All primary care providers:52% (p<0.5) Dermatologists: 61% (p<0.5)		80	95	80	95	92
Brinker, T. J. et al., 2019[86]	diagnostic accuracy study	NR	CNN ResNet50 trained on dermoscopic images	145 dermatologists	Melanoma's biopsy verified, benign nevi via expert consensus.	0.769 (all participants)	89.4% (range: 55.0%-100%)	64.4% (range: 22.5%e92.5%)		value NR	-	-	89.40	68.2 (47.5-86.25)	-

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Brinker, T. J. et al., 2019[87]	diagnostic accuracy study	NR	CNN ResNet50 trained on open-source images	157 dermatologists	histopathology	0.671	74.1 (40-100)	60 (21.3-91.3)		value NR	-	-	74.1	86.5 (70.8-91.3)	-
Brinker, T. J. et al., 2022[88]	diagnostic accuracy study	NR	CNN	18 pathologists	2 experienced dermatopathologists		88.88	91.77	90.33	0.95 (unannotated) 0.97 (annotated)			88.0 (unannotated) 94.0 (annotated)	88.0 (unannotated) 90.0 (annotated)	88.0 (unannotated) 92.0 (annotated)
Brodsky, V. et al , 2023[89]	diagnostic accuracy study	300	CNN, Resnet 34 / Resnet 18	two senior board-certified pathologists.	histopathologist	-	-	-	-	-	-	-	melanoma 97.8% BCC: 99% naevus:100%	melanoma 97.6% BCC: 100% naevus: 97.9%	-
Cassalia, F et al., 2024[90]	cohort study	1813	CNN	Model Dermatology app; Smartphone vs 'Clicca il Neo' app and smartphone	CNN compared to Clicca il Neo app and online assessments by seven dermatologists.					0.67 (95% CI: 0.63, 0.70)					

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Cerminara, S.E. et al , 2023[91]	cohort study	143	CNN	Dermatologist	histopathology or for benign images review by a dermatologist / 3D 2D TBP CNN	dermatologist - 0.88	Dermatologists - 90%	Dermatologists - 92.31%	-	2D TBP CNN - 0.68 3D TBP CNN - 0.92 Dermatologist plus AI - 0.91	2D CNN - 15.22 3D CNN - 28.12 Dermatologists plus AI - 50.00	2D CNN - 89.55 3D CNN - 97.67 Dermatologists plus AI - 98.25	2D CNN - 70% 3D CNN - 90% Dermatologists plus AI - 90%	2D CNN - 40% 3D CNN - 64.62% Dermatologists plus AI - 86.15%	-
Crawford, M. E et al, 2024[92]	cohort study	318	CNN	Dermatologist, external dermatologist and resident	2 experienced dermatopathologists, if disagreement a third dermatopathologist examined the case.	Derm 1 - 0.628 Derm 2 - 0.591 Derm 3 - 0.648 Derm4 - 0.571 Resident - 0.614 External Derm - 0.732	Derm 1 - 52.9 Derm 2 - 47 Derm 3 - 76.5 Derm4 - 82.4 Resident - 47 External Derm - 70.6	Derm 1- 69.7 Derm 2 - 71.2 Derm 3 - 48.5 Derm4 - 31.8 Resident - 75.8 External Derm - 76.8	Derm 1 - 66.4 Derm 2 - 66.4 Derm 3 - 54 Derm4 - 41.9 Resident - 70 External Derm - 74.72	0.702	40	89.6	64.7	75.76	73.56
Diaz-Ramon, J. L. et al, 2023[93]	diagnostic accuracy study	196	Diagnostic module - CNN	NR	histopathology								detect - 96.82	detect - 75.41	

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Esteva et al, 2017[57]	diagnostic accuracy study	NR	Inception v3	Dermatologist	histopathology					Melanoma (130 images) Auc 0.94 Melanoma (111dermoscopic images) AUC 0.91	-	-	-	-	-
Felmingham, C et al, 2023[94]	cohort study	214	CNN	Teledermatologists / registrar and dermatologists	consultant dermatologists' management decision and histopathology	Teledermatologist - 0.807 resident - 0.847 resident and CNN - 0.879	Teledermatologist - 89.5 (84.9-93.1)	Teledermatologist- 71.9 (67.7-75.9)	NR	0.837	NR	NR	95.8 (92.4-98.0)	71.5 (67.3-75.5)	NR
Fink, C. et al, 2020[95]	diagnostic accuracy study	72	CNN; based on GoogleNet Inception_v4	11 dermatologists with different levels of experience	expert consensus and unremarkable follow up of at least 2 years (when not excised) and histopathology		90.6 (84.1-94.7)	71.0 (62.6-78.1)					97.1 (82.7-99.6)	78.8 (62.8-89.1)	

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Giavina-Bianchi, M.; et al, 2021[76]	diagnostic accuracy study	NR	clinical model - DenseNet	Primary care physician (PCP)	labelled as malignant or benign by two dermatologists (nb biopsy confirmation of the majority of the images was not available)	-	-	-	-	0.935	0.5795	0.9727	0.8947	0.8525	0.8603
Giavina-Bianchi, M.; et al, 2021(76)	diagnostic accuracy study	NR	Dermoscopy model - EfficientNetB6	Primary care physician (PCP)	labelled as malignant or benign by two dermatologists	-	-	-	-	0.96	0.6457	0.9771	0.9056	0.89	0.8928
Giulini, M. et al, 2024(96)	diagnostic reader study	100	CNN	physician without CNN	histopathology		56.31	69.28		0.93 (CNN)			67.88	73.72	

Table 3 - Study Outcomes															
Study		Participant characteristics	Intervention	Comparator		Outcomes									
Author	Study design	Number of subjects	Machine Learning	Comparator	Gold standard - histopathology	ROC	Sensitivity %(95% CI)	Specificity %(95% CI)	Accuracy %(95% CI)	ROAUC	PPV	NPV	Sensitivity % (95% CI)	Specificity % (95% CI)	Accuracy %(95% CI)
Goessinger, E. V. et al, 2024a[97]	cohort	66	CNN-1: Dermoscopy explainable intelligence (DEXI) score, Canfield Scientific, CNN-2: MoleAnalyzer-Pro, FotoFinder ATBM® Systems GmbH,	CNN-2	histopathology or expert consensus review by 3 dermatologists with 1 year follow up of all available lesions										
Goessinger, E. V. et al , 2024b[98]	cohort	205 patients / 8 dermatologists	CNN-1: Dermoscopy explainable intelligence (DEXI) score, Canfield Scientific, CNN-2: MoleAnalyzer-Pro, FotoFinder ATBM® Systems GmbH,	-	-										

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Haenssle, H. A.; et al, 2018 [47]	diagnostic accuracy study	300	Google's inception v4 CNN additionally trained with more than 100,000 labelled digital images	58 dermatologists (readers) from 17 countries	histopathology. Non-excised lesions validated by follow-up examinations.	0.79	86.6	71.3	level - 1 reader study (image only)	0.86	-	-	86.6	82.5	-
Haenssle, H. A.; et al, 2018[47]	diagnostic accuracy study	300	Google's inception v4 CNN additionally trained with more than 100,000 labelled digital images	58 dermatologists (readers) from 17 countries	histopathology. Non-excised lesions validated by follow-up examinations.	0.82	88.9	75.7	level - 2 reader study (additional clinical information)	0.86	-	-	88.9	82.5	-
Haenssle, H. A.; et al, 2020[48]	diagnostic accuracy study	100	modified version of Google's inception v4 CNN	dermatologists	histopathology. Non-excised lesions validated by follow-up examinations over at least 2 years.		level-I 89.0 [87.4-90.6] / level-II 94.1 [93.1-95.1]	level-I 80.7% [78.8-82.6] level-II 80.4% [78.4-82.4]	level-I 84.0 level-II 85.9 [84.7-87.1]	0.918 [0.866 - 0.970]	-	-	95.0 [83.5-98.6]	76.7 [64.6-85.6]	84% [75.6-89.9]

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Haenssle, H. A.; et al, 2021[99]	diagnostic accuracy study	100	modified version of Google's inception v4 CNN trained with dermoscopic images	64 dermatologists after reviewing full case information	NR - all test images were provided from external institutions.	-	level-I 77.1 [74-80.2] level II 84.2% [82.2% – 86.2%]	Level-I 69.5% [66.3-72.7] level-II 69.4% [66.0%– 72.8%]	77.1	0.929 [0.880 – 0.978] 0.926 [0.912 - 0.940] in 'ISIC2018 data set', 0.939 [0.923 - 0.954] in 'MSK-1 data set', 0.945 [0.930 - 0.961] in 'Prospective data set' 0.814 [0.760 - 0.868] in	-	-	96.2% [87.0%– 98.9%]	68.8% [54.7%– 80.1%]	83

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
										'Australian data set'					
Hartman, R. I. et al, 2024[100]	cohort study	311	DermaSensor	Dermatologist	histopathology	0.747				0.758	16.0% (95% CI, 11.6% to 21.7%)	98.1% (95% CI, 91.8% to 99.6%)	95.5% (95% CI, 84.5% to 98.8%, 42 of 44 melanomas)	32.5% (95% CI, 27.2% to 38.3%)	

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Hekler, A.et al, 2024[101]	prospective, cohort study	617	CNN with ConveNeXT	Single View, MV-Artificial,	Histopathology	Single-View (0.905; 95% CI, 0.879-0.929; P <0.001) and Multiview-artificial approach (0.929; 95% CI: 0.908-0.948; P <0.001)				(0.930; 95% CI, 0.909-0.951)					
Huang, H. W. et al , 2021[102]	diagnostic accuracy study	1222	CNN, DenseNet and EfficientNet	Dermatologist	histopathology					0.74	75	96	37.50%	99.2	95.3

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Jahn, A.S. et al, 2022[78]	cohort study	114	CNN	skin lesions diagnosed by dermatologist, 2D automated total body mapping (ATBM) FotoFinder ATBM with MoleAnalyzer Pro, and 3D TBP Vectra WB360 with DexI.	histopathology, or for benign lesions with no histology dermatologists evaluation plus AI scores of two independent medical devices FotoFinder ATBM and Vectra WB360)					app vs dermatologist, FotoFinder and Vectra - 0.621 app vs histopathology - 0.717			app vs dermatologist, FotoFinder and Vectra - 41% app vs histopathology - 83%	app vs dermatologist, FotoFinder and Vectra - 83% app vs histopathology - 60%	
Jaklitsch, E. et al, 2023[75]	diagnostic reader study	57 primary care physicians	CNN	PCP without EES	dermoscopic image	0.62	67% (958/1425; 95% CI, 62%-72%)	53% (761/1425; 95% CI, 49%-57%)		0.68			88% (1261/1425; 95% CI, 84%-92%) with the device (*P*<.0001)	40% (577/1425; 95% CI, 37%-44%)	
Kommoss, K. S. et al, 2023[103]	diagnostic accuracy study	100	CNN - Google's Inception_v4	Dermatologist without CNN	histopathology or unremarkable clinical follow up for at least 2 years		1 dermoscopic image: 77.1%	1 dermoscopic image: 69.5%	1 dermoscopic image: 73.4%				96.2% [95% CI 87.0–98.9%],	68.8% [54.7–81.3%],	83.0% [74.5–89.8%]

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
							[74.0–80.1%], full case info: 84.2% [82.2–86.2%, p < 0.001]	[66.3–72.7%], full case info 69.4% [66.0–72.8%, p = 0.943],	[72.3–74.6%] full case info: 77.1% [76.1–78.1%, p < 0.001].						
Kränke, T. et al, 2023[104]	cohort study	238	2 CNNs Analyse (n - 238 patients and 1171 lesions) Detect (n - 92 patients and 552 lesions)	skin lesions diagnosed by 2 dermatologists.	Histology, if available or if benign 2 dermatologists were reference standard and digital follow up provided in 3-6 months.	-	-	-	-	NR	-	-	Analyse 95.35 [CI 93.45-97.25] Detect 96.4% [CI 93.94-98.85]	Analyse 90.32 [CI 88.1-92.54] Detect 94.85 [CI92.46-97.23]	NR
Liutkus, J. et al, 2023[105]	diagnostic accuracy study	NR	YOLO NNM - BottleNeckCSP CNN	skilled dermatologists and beginners	histologically confirmed diagnosis		Melanoma skilled: 0.98 (0.92–1.00) beginners: 0.83 (0.77-0.87)	Melanoma: skilled: 0.84 (0.51–0.96) beginners: 0.85 (0.77–0.90)		0.93 (0.88-0.98)			0.88 (0.71–0.96)	0.87 (0.76–0.94)	

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
MacLellan, A. N. et al, 2021[106]	cohort study	184	CNN NR however tradenames provided	MelaFind FotoFinder Tuebinger Mole Analyzer FotoFinder Moleanalyzer Pro Verisante Aura Dermatologist Teledermoscopic diagnosis	excision of all lesions and reviewed independently by 2 dermatopathologists		Dermatologist 96.6 (91.91 - 101.31) Teledermoscopic 89.8 (79.6-96.2)	32.2 (18.4-46.0) 66.0 (57.8-73.5)	-	-	-	-	MelaFind 82.5 (72.6-92.4) FotoFinder Tubeinger 83.1 (72.6-93.6) FotoFinder MoleAnalyzer 88.1 (79.4-96.9) Verisante Aura 21.4 (10.7-32.2)	52.4 (44.2-60.6) 75.2 (67.27-83.1) 78.8 (71.5-86.2) 86.2 (80.2-92.1)	-
Maier, T. et al, 2015[107]	cohort study	195	CNN	Clinical and dermoscopic diagnosis was independently documented by two dermatologists.	histopathology		88% (95%-CI: 0.69–0.98)	97% (95%-CI: 0.92–0.99)	0.95 (0.90-0.98)		49% (95%-CI: 0.32–0.65)	93% (95%-CI: 0.87–0.97)	73% (95%-CI: 0.52–0.88)	83% (95%-CI: 0.75–0.89)	81% (95%-CI: 0.74–0.87)
Manolakos, D. et al, 2024[108]	multicentre cohort study	394	CNN, Resnet 18	dermatologist	histopathology	0.785	96.45 %	56.10%		0.773	57.54	89.58	97.04%	26.22% (P<0.0001)	

Table 3 - Study Outcomes															
Study		Participant characteristics	Intervention	Comparator		Outcomes									
Author	Study design	Number of subjects	Machine Learning	Comparator	Gold standard - histopathology	ROC	Sensitivity %(95% CI)	Specificity %(95% CI)	Accuracy %(95% CI)	ROAUC	PPV	NPV	Sensitivity % (95% CI)	Specificity % (95% CI)	Accuracy %(95% CI)
Marchetti, M. A. et al, 2018[49]	diagnostic reader study	NR	CNN, Greedy fusion	8 international dermatologists	histopathology or for benign images reviewed by more than 2 dermatologists to confirm the benign nature.	0.71	82	59	-	0.86	-	-	82	76	-
Marchetti, M. A. et al, 2020[109]	diagnostic reader study	NR	CNN;	Performance of 8 international dermatologists and 9 USA resident physicians.	NR	dermatologist = 0.74 resident = 0.66	dermatologist 76 (71.5-80.1) Resident 56.0 (51.3-60.6),	Dermatologist 72.6 (69.4-75.7) Resident 76.3 (73.4-79.1)		0.87	-	-	76	85	-
Marchetti, M. A. et al, 2023[110]	cohort study	435	All Data are Ext (ADAE) open source, non-commercial AI algorithm Based on EfficientNet and ResNet architecture	dermatologist	histopathology	physician without AI - 0.780 (p=0.007) vs AI physician after AI - 0.82 (p=0.042 vs physician without AI)				0.857			96.8% (95% CI: 91.1–98.9%)	37.4% (95% CI: 33.3–41.7%)	

Table 3 - Study Outcomes

Study		Participant characteristics	Intervention	Comparator		Outcomes									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Marchetti, M. A. et al, 2023[111]	cohort study	35	VECTRA DermaGraphix research software	Dermatologist	histopathology					0.94 (95% CI: 0.92–0.96)					
Marsden, H. et al, 2024[83]	cohort trial, cost impact assessment	622	CNN	Clinical and dermoscopic image reviewed remotely with referral letter and patient reported medical history by consultant dermatologist (teledermatology service)	histopathology		97.0, 88.7–99.5	71.9, 68.4–75.1	PPV: 24.2, 19.3–29.9 NPV:99.6, 98.5–99.9 NNB (N, 95% CI)4.2 (3.3–5.5)	NNB:3 (2.4–3.7)	33.5, 26.8–40.9	99.0, 97.7–99.6	91.0, 80.9–96.3	83.2, 80.3–85.9	

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Marsden, H. et al, 2023[77]	cross-sectional study	544	NR	Dermatologist (standard of Care)	histopathology	Biopsied (n=396) Melanoma 90.2% (80.2–100%)	81.2% (53.7–95.0%)	98.9% (97.6–99.6%)		Biopsied iPhone 11 - 91.8% (82.9–100%) iPhone 6S - 97.5% (94.8–100%) Samsung 10 - 94.4% (89.2–99.6%)	iPhone 6S - 8.1% (4.7–13.2%) iPhone 11 - 8.7% (5.0–14.4%) Samsung - 7.6% (4.6–12.3%)	iPhone 6S - 100% (98.8–100%) iPhone 11 - 99.8% (98.4–100%) Samsung - 100% (98.7–100%)	iPhone 6S 100% (74.7–100%) iPhone 11 - 93.3% (66.0–99.7%) Samsung - 100% (75.9–100%)	iPhone 6S - 69.6% (65.6–73.4%) iPhone 11 - 73.6% (69.6–77.1%) Samsung - 65.5% (61.4–69.4%)	
Menzies, S. W. et al., 2023[112]	prospective, cross-sectional study	diagnostic - 124 patients management - 66 patients	CNN - efficientNet-B3 or ResNet-10	dermatologist, experienced and novice	histopathology	NR	Melanoma Specialist 61·8% (47·7–74·6) Novice Dr 41·8% (28·7–55·9)	Specialist 85·5% (77·8–91·3) Novice Dr 72·6% (63·6–80·5)	Specialists 73·8% (51·0–96·6 Novice: 35·5% (2·8–67·7)	7 Class AI - 85.8% (79.4-92.2%) ISIC AI - 76.3% (77.5-89.6%)			7 class AI 50·9% (37·1–64·6) ISIC AI 16·4% (7·8–28·8)	7 class AI 94·0% (88·1–97·6) ISIC AI 98·3% (94·0–99·8)	7 class AI 65·9% (95% CI 26·9–100) ISIC AI 52·2% (5·4–98·9)
Miller, I. J.; et al, 2023[113]	cross-sectional study	423	CNN	-	histopathology					0.54			53.33	54.44	54.17

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Ngoo, A et al, 2018[42]	prospective cohort study	30	SkinVision iOS; SkinVision android; SpotMole direct; SpotMole indirect; Dr Mole android	dermoscopic diagnosis by 2 dermatologists (suspicious / benign / image not adequate)	dermatologist diagnosis using dermatoscope (not compared to histopathology result)	-	-	-	-	Dr Mole - Roc 0.52	SkinVision IOS- 77.8 SkinVision Android - 76.5% Spot mole direct - 85.7 Spot mole indirect - 81.8% Dr Mole - 100%	SkinVision IOS - 27.3 SkinVision Android - 23.1 Spot mole direct - 33.3 Spot mole indirect - 31.4 Dr Mole - 31.3	SkinVision IOS - 56.8% (40.8–72.7) SkinVision Android - 72.2% (57.6–86.9) Spot mole direct - 42.9 (27.9–57.8) Spot mole indirect - 42.9% (27.9–57.8) Dr Mole - 21.4% (9-33.8)	SkinVision IOS - 50.0 (21.7-78.3) SkinVision Android 27.3 (1-53.6) Spot mole direct 80.0 (59.8-100) Spot mole indirect 73.3 (51-95.7) Dr Mole 100 (100-100)	Concordance of app with dr decision (SkinVision IOS - 55.1 SkinVision Android - 61.7 Spot mole direct - 52.6 Spot mole indirect - 50.9 Dr Mole - 42.1
Oloruntoba, A. I.et al., 2022(114)	diagnostic accuracy study	Test set 1 n-519. NR test set 2 & 3 (image data set)	ResNet - 50 / ResNet-18	4 Teledermatologists	only some of the datasets images were biopsy-proven, remaining were labelled by dermatologists.		82.9% (80.8% -85%)	79.2% (78.5%-79.9%)	-	Danish test CNN-S0.759 CNN-S2 0.861 CNN-NS 0.831	-	-	CNN-S- 74.7 CNN-S2 - 71.5 CNN-NS - 56.3	CNN-S 72%; CNN-S2 65.2%; CNN-NS 46.7%	-

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Phillips, M et al, 2019[115]	prospective, cross-sectional study	514	DNN	Dermatologists	Histopathology	All lesions: 0.909	100	69.9		All lesions iPhone - 0.959 Galaxy - 0.938 DSLR - 0.918	All lesions iPhone - 17.9 Galaxy - 13.4 DSLR - 9.48	All lesions iPhone - 100 Galaxy - 100 DSLR - 100	All lesions iPhone - 100 Galaxy - 100 DSLR -100	All lesions iPhone - 64.8 Galaxy - 51.2 DSLR - 30.0	
Phillips, M et al, 2020[116]	diagnostic accuracy study	NR (image data set)	DNN	Compared to the diagnostic accuracy of doctors determined by a meta-analysis of studies.	Histopathology	PCP 0.83 (0.79-0.86) Dermatologists 0.91 (0.88-0.93)	PCP 79.9% Dermatologists 87.5%	PCP: 70.9% Dermatologists: 81.4%		0.93 (95% CI: 0.92-0.94)			85 (83.2-86.7)	85.3 (84.4-86.3)	-
Pyun, S. H. et al., 2022[117]	cohort study	353	DNN	histopathology	histopathology; benign confirmed by either histopathology or visually by physicians.	-	-	-	-	0.96	-	-	melanoma 92.9% (92.0-97.2)		-

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Gomez Rossi, J. et al, 2022[82]	Markov-model based cost effectiveness analysis; based on a diagnostic reader study (Brinker 2019)	NR	CNN ResNet50	dermatologist using a dermatoscope	histopathology										
Sangers, T et al, 2022[118]	Prospective cross-sectional study	372	CNN	Dermatologist	histopathology or for benign lesions monitor for 3 months					-	61.3	90.9	86.9 (82.3-90.7)	70.4 (66.2-74.3)	76.2 (73.0-79.1)
Thissen,M et al, 2017[119]	Prospective cross-sectional study	256	CNN	histopathology	histopathology	-	-	-	-	-	melanoma detection (before recalibration) :0.33 (0.17-0.53) melanoma detection (after recalibration using LMU database) 0.47 (0.33-0.62)	melanoma detection (before recalibration) :0.67 (0.56–0.77) melanoma detection (after recalibration using LMU database) 0.96 (0.90–0.99)	malignant or premalignant: 80% (95% CI 0.62–0.90) melanoma detection (before recalibration) :25% (0.13-0.43) melanoma detection (after recalibration using LMU database) 0.884 (0.68–0.96)	malignant or premalignant: 78.08% (95% CI 0.66–0.86) melanoma detection (before recalibration) :0.75 (0.63–0.84) melanoma detection (after recalibration using LMU database) 0.788 (0.70–0.85)	-

Table 3 - Study Outcomes														
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**								
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Thomas, L. et al, 2023[120]	prospective, cohort study	8571	CNN	teledermatology	histopathology					Number needed to biopsy, treat or refer DermvA (UHB) 14.9 (1985/133) [12.7–17.6] Derm-vA (WSFT) 8.8 (281/32) [6.4–12.2] Derm-vB (UHB) 9.3 (540/58) [7.3–11.9] Derm-vB (WSFT) 7.8 (140/18) [5.2–12.1]	DermvA (UHB) 6.7% (133/1985) [5.7–7.9%] Derm-vA (WSFT)11.4% (32/281) [8.2–15.6%] Derm-vB (UHB) 10.7% (58/540) [8.4–13.6%] Derm-vB (WSFT) 12.9% (18/140) [8.3–19.4%]	DermvA (UHB) 99.7% (2643/2650) [99.5–99.9%] Derm-vA (WSFT) 99.8% (427/428) [98.7–100%] Derm-vB (UHB) 100.0% (2045/2045) [99.8–100.0%] Derm-vB (WSFT) 100% (502/502) [99.2–100%]	DermvA (UHB) 95.0% (133/140) [90–97.6%] Derm-vA (WSFT)97.0% (32/33) [84.7–99.5%] Derm-vB (UHB) 100.0% (58/58) [93.8–100%] Derm-vB (WSFT) 100.0% (18/18) [82.4–100%]	DermvA (UHB) 58.8% (2643/4495) [57.4–60.2%] Derm-vA (WSFT) 63.2% (427/676) [59.5–66.7%] Derm-vB (UHB) 80.9% (2045/2527) [79.3–82.4%] Derm-vB (WSFT) 80.4% (502/624) [77.2–83.4%]	
Tschandl, Philipp et al, 2019[121]	diagnostic accuracy study	2072 non-pigmented lesions	cCNN Inception V3 / ResNet50	95 human raters, beginner, intermediate or expert raters	histopathology	0.695 (0.676-0.713)*mean AUC of all human raters	77.6 (74.7-80.5)	51.3%(48.4-54.3%)	-	0.742 (0.729-0.755)	-	-	80.5% (79.0%-82.1%)	53.5% (51.7%-55.3%)	-

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Udrea, A et al, 2020[122]	diagnostic accuracy study	5603	CNN	Dermatologist	histopathology or for benign images review by a dermatologist								92.8% (95% CI, 87.8–96.5%)	78.3% (95% CI, 77.24–79.34%)	-
Van Molle, P.et al, 2023[123]	diagnostic accuracy study	501 (test set)	stochastic neural network - ResNet50	Dermatologists	histopathology	0.7	68%	73%		0.69			50%	88%	
Wies, C.et al, 2024[124]	diagnostic accuracy study	Dresden 126 / Erlangen 81 / Naples 50	CNN - ResNet	H&E slides, H& E slides and IHC slides.	histopathology	H&E - 0.81 MelanA and H&E - 0.85				MelanA - 0.82					

Table 3 - Study Outcomes																
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**										
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Winkler, J. K.et al, 2023[79]	prospective, cross-sectional study, cost impact assessment	188	CNN	Dermatologists alone; dermatologists and CNN.	histopathology, or follow-up for non-excised lesions	Dermatologist alone: 0.90; Dermatologist and CNN: AUC 0.97	Dermatologists: 84.2% (95% CI, 69.6%-92.6%) Dermatologist and CNN: 100% (95% CI, 90.8%-100.0 %;	Dermatologists: 72.1% (95% CI, 65.3%-78.0%) Dermatologist and CNN:83.7% (95% CI, 77.8%-88.3%; P < .001 vs dermatologist alone	Dermatologists alone: 74.1 (68.1-79.4) Dermatologist and CNN: 86.4 (81.3-90.3)	CNN alone: AUC0.90			81.6% (95% CI, 66.6%-90.8%)	88.9% (95% CI, 83.7%-92.7%)	87.7 (82.8-91.4)

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Winkler, J. K.et al, 2024[45]	prospective, cross-sectional study	236	NR	Dermatologists	histopathology, or follow-up for non-excised lesions			1075 CRML (clinically relevant melanocytic lesions) Follow up - 334 CRML new or changed lesions: 37							999/1075 CRML (92.9% 95% CI 91.2-94.3%) Follow up 323 /334 CRML (96.7%, 95%-CI: 94.2%–98.2%) new or changed lesions:37

Table 3 - Study Outcomes															
Study		**Participant characteristics**	**Intervention**	**Comparator**		**Outcomes**									
Author	**Study design**	**Number of subjects**	**Machine Learning**	**Comparator**	**Gold standard - histopathology**	**ROC**	**Sensitivity %(95% CI)**	**Specificity %(95% CI)**	**Accuracy %(95% CI)**	**ROAUC**	**PPV**	**NPV**	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Accuracy %(95% CI)**
Winkler, Julia K.et al, 2021[125]	diagnostic accuracy study	30	CNN - GoogleNet Inception v4	Moleanalyzer Pro	histopathology		Binary CoHI:82.4 % [59.0 %–93.8 %] individual dermatologist: 77.7 % [75.3 %–80.2 %]	Binary CoHI:76.9 % [49.7 %–91.8 %] individual dermatologist:73.0 % [70.6 %–75.4 %]	Binary CoHI:80.0 % [62.7 %–90.5 %] individual dermatologist:75.7 % [73.8 %–77.5 %] Multiclass: CoHI:79.2 % (19/24 correct diagnoses) Individual dermatologist:64.6 % [61.6 %–67.6 %]	binary: 0.765 (0.595 - 0.935)			binary CNN 70.6 % [46.9 %–86.7 %	binary CNN 69.2 % [42.4 %–87.3 %]	binary CNN 70.0 % [52.1–83.3 %] Multiclass CNN 62.5 % (15/24 correct diagnoses).

Table 4 Preference studies

Preference based studies		study population								
Author	Title	Risk	Number of subjects	Trade name	Image type	Intervention	Comparators	Perspective	Outcomes	Results
Jahn, A.S. et al 2022	Over-Detection of Melanoma-Suspect Lesions by a CE-Certified Smartphone App: Performance in Comparison to Dermatologists, 2D and 3D Convolutional Neural Networks in a Prospective Data Set of 1204 Pigmented Skin Lesions Involving Patients' Perception	High risk	114	SkinVision	Smartphone, 2D and 3D TBP	Smartphone app, 2d tbp, 3D TBP	Dermatologist	Patient and Dermatologist	Clinician and patient preference and trust of screening type	**Confidence in Dermatologists vs. Smartphone App** Most of the patients at high-risk for melanoma (55% (30/55)) and patients with melanoma (53% (31/59)) reported being very confident about a mole examination by a dermatologist Only (16% (9/55)) high-risk patients and (12% (7/59))melanoma patients felt very safe when being investigated by a smartphone app alone. **Trustworthiness of the Smartphone App** All participants primarily rated physician examination (100% (55/55) among high- risk patients, 100% (59/59) among melanoma patients) as trustworthy, and the majority did so for 2D TBP imaging (93% (51/55) resp. 88% (52/59)) and 3D TBP imaging (91% (50/55) resp. 90% (53/59)). The smartphone app was less frequently rated as trustworthy, with only 36% (20/55) among patients at high-risk for melanoma and 49% (29/59) among melanoma patients . The age revealed a significant correlation with the evaluation of smartphones' trustworthiness (p < 0.004), with older patients (>60 years old) having trusted the app three times more than younger patients (≤60 years old) Neither previous melanoma vs. high-risk criteria for melanoma nor sex significantly influenced the evaluation of the trustworthiness of the smartphone app. Most participants indicated that an examination by a physician reduced their fear of developing skin cancer, namely in 89% (49/55) among high-risk patients and 81% (48/59) among melanoma patients. Comparably, the 2D TBP imaging achieved the same effect in 78% (43/55) resp. 76% (45/59), and the 3D TBP device in 82% (45/55) resp. 75% (44/59). In contrast, the assessment with the smartphone app appeased fear of skin cancer in only 33% (18/55) of patients at high-risk for melanoma and in 32% (19/59) of patients with melanoma. **Patients' Subjective Assessment of the Accuracy of AI vs. Dermatologists** Patients expected reliable results with the highest accuracy by both the assessment by a physician (98% (54/55) among high-risk patients, 92% (54/59) among melanoma patients) and by the 2D TBP imaging (82% (45/55) resp. 86% (51/59)) as well as the 3D TBP device (89% (49/55) resp. 88% (52/59)). Only 16% (9/55) of high-risk patients and 31% (18/59) of melanoma patients expected reliable results from the smartphone app **Patient Preference for Skin Cancer Screening** Both cohorts favored a combination of dermatologist and 3D TBP risk assessment for the examination of pigmented skin lesions (64% (35/55) among patients at high-risk for melanoma, 51% (30/59) among melanoma patients), while neither preferred assessment by a smartphone app alone . The combination of dermatologist and smartphone app was favored by only 1.8% (1/55) of patients at high-risk for melanoma and 3.4% (2/59) of patients with melanoma Regarding patient preference for skin cancer screenings, almost all high-risk patients (98% (54/55)) and melanoma patients (95% (56/59)) indicated their belief that AI can improve a physician's diagnostic performance. Most patients (64% (35/55) among high-risk patients, 54% (32/59) among melanoma patients) would prefer that the physicians would always consider the result of AI in their diagnosis **Dermatologists' Perspective of Smartphone Apps for Melanoma Screening** Among the seven dermatologists, they stated in only 5.3% of the skin cancer screenings (6/114) that the smartphone app increased diagnostic confidence and in only 8.8% of the assessments (10/114) they trusted the app.
Winkler, J. K.et al 2023	Assessment of Diagnostic Performance of Dermatologists Cooperating With a Convolutional Neural Network in a Prospective Clinical Study: Human With Machine	high risk	188	Moleanalyzer Pro (FotoFinder Systems)	dermoscopic	Moleanalyzer Pro (market approved CNN)	Dermatologists alone; Dermatologists with help of CNN	N/A	clinician and patients attitdes toward CNN support system	For a majority of lesions dermatologists consented that CNN support was reassuring and/or helpful. dermatologists were optimistic about CNN support. Similarly, results of the study's questionnaire indicated that patients were open minded toward a CNN-based support system. However most patients still wished an interpretation of results by an expert clinician and rejected a full replacement of clinicians by neural networks.
Winkler, J. K.et al 2024	Performance of an automated total body mapping algorithm to detect melanocytic lesions of clinical relevance	high risk	236	ATMB master	2D Automated total body mapping (ATBM)	ATBM	Dermatologists	Patient	Patient perspectives	Written questionnaires on acceptance and confidence in the automated detection of lesions were returned from 233 of the patients. Most patients consented that ATBM examination gave them afeeling of increased safety (45.2% strongly agree, 47.9% agree). They considered the technology trustworthy (35.5% strongly agree, 57.5%agree) and agreed it might improve performance of dermatologists (44.4% strongly agree, 44.4% agree). Most patients disagreed to completely replace dermatologists' examinations (34.4% strongly disagree, 32.0% disagree) and demanded interpretation of results by an expert (54.8% strongly agree, 43.6% agree). The majority of patients accepted longer examination times for ATBM (26.6% strongly agree, 43.6% agree).

Reference list

1. Madan V, Lear JT, Szeimies R-M. Non-melanoma skin cancer. The Lancet. 2010;375(9715):673-85.

2. Curti BD, Vetto JT, Leachman SA. Cancer of the Skin. In: Longo D, Fauci A, Kasper D, Hauser S, Jameson JL, Loscalzo J, et al., editors. Harrison's Principles of Internal Medicine, 22nd Edition. New York, NY: McGraw Hill; 2026.

3. Cancer Council. Understanding Skin Cancer A guide for people with cancer, their families and friends. In: Cancer Council Australia, editor. 2024.

4. Australian Institute of Health and Welfare. Skin Cancer in Australia. Canberra: AIHW; 2016 2016. Contract No.: CAN 96.

5. Green AC, Olsen CM. Cutaneous squamous cell carcinoma: an epidemiological review. British Journal of Dermatology. 2017;177(2):373-81.

6. Australian Government DoH. AIHW 2024 Cancer Data in Australia Report. AIHW website; 2024.

7. Dika E, Scarfì F, Ferracin M, Broseghini E, Marcelli E, Bortolani B, et al. Basal Cell Carcinoma: A Comprehensive Review. International Journal of Molecular Sciences. 2020;21(15):5572.

8. Insight Economics. A report into Melanoma - A National Health Priority. Melanoma Institute Australia; 2022 2022.

9. Berwick M, Buller DB, Cust A, Gallagher R, Lee TK, Meyskens F, et al. Melanoma Epidemiology and Prevention. In: Kaufman HL, Mehnert JM, editors. Melanoma. Cham: Springer International Publishing; 2016. p. 17-49.

10. International Agency for Research on Cancer. Estimated age-standardized incidence rates (World) in 2020, Australia, both sexes all ages (excl NMSC) Cancer Today: GLOBOCAN; 2020 [Available from: https://gco.iarc.fr.

11. Ascierto PA, Del Vecchio M, Mandalá M, Gogas H, Arance AM, Dalle S, et al. Adjuvant nivolumab versus ipilimumab in resected stage IIIB–C and stage IV melanoma (CheckMate 238): 4-year results from a multicentre, double-blind, randomised, controlled, phase 3 trial. The lancet oncology. 2020;21(11):1465-77.

12. Long GV, Hauschild A, Santinami M, Atkinson V, Mandalà M, Chiarion-Sileni V, et al. Adjuvant Dabrafenib plus Trametinib in Stage III BRAF-Mutated Melanoma. The New England journal of medicine. 2017;377(19):1813-23.

13. Eggermont AMM, Blank CU, Mandala M, Long GV, Atkinson V, Dalle S, et al. Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma. The New England journal of medicine. 2018;378(19):1789-801.

14. Weber J, Mandala M, Del Vecchio M, Gogas HJ, Arance AM, Cowey CL, et al. Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma. The New England journal of medicine. 2017;377(19):1824-35.

15. Australian Institute of Health and Welfare. Data tables: CDIA 2023:Book 8 - Cancer incidence and survival by stage [updated 31Aug2023. Available from: https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/data?&page=2.

16. Eggermont AMM, Blank CU, Mandala M, Long GV, Atkinson VG, Dalle S, et al. Longer Follow-Up Confirms Recurrence-Free Survival Benefit of Adjuvant Pembrolizumab in High-Risk Stage III Melanoma: Updated Results From the EORTC 1325-MG/KEYNOTE-054 Trial. Journal of Clinical Oncology. 2020;38(33):3925-36.

17. Dummer R, Hauschild A, Santinami M, Atkinson V, Mandalà M, Kirkwood JM, et al. Five-Year Analysis of Adjuvant Dabrafenib plus Trametinib in Stage III Melanoma. The New England journal of medicine. 2020;383(12):1139-48.

18. Commonwealth of Australia. National targeted skin cancer screening health.gov.au; 2024 [Available from: https://www.health.gov.au/ministers/the-hon-mark-butler-mp/media/national-targeted-skin-cancer-screening-to-be-accelerated-with-103-million-investment.

19. Cust AE, Scolyer RA. Melanoma In Situ—Getting the Diagnosis and Prognosis Right. JAMA Dermatology. 2023;159(7):699-701.

20. Whiteman DC, Olsen CM, MacGregor S, Law MH, Thompson B, Dusingize JC, et al. The effect of screening on melanoma incidence and biopsy rates*. British Journal of Dermatology. 2022;187(4):515-22.

21. Clark S, Rosendahl N, Rosendahl C. Melanoma overdiagnosis: What do we know and what do we do? Australian Journal for General Practitioners. 2024;53:612-6.

22. Curiel-Lewandrowski C, Chen SC, Swetter SM. Screening and prevention measures for melanoma: is there a survival advantage? Curr Oncol Rep. 2012;14(5):458-67.

23. Seidl-Philipp M, Nguyen VA. Cutaneous melanoma—The benefit of screening and preventive measures. memo - Magazine of European Medical Oncology. 2019;12(3):235-8.

24. Cancer Council Australia. Clinical practice guidelines for the diagnosis and management of melanoma 2021 Available from: https://files.magicapp.org/guideline/30c6978f-c8dd-4ef8-9aa2-941aa10235fe/published_guideline_8102-1_1.pdf. [Accessed 9May2024]

25. US Preventive Services Task Force. Screening for Skin Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2023;329(15):1290-5.

26. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. BMJ. 1994;309(6947):102.

27. Watts CG, Cust AE, Menzies SW, Mann GJ, Morton RL. Cost-Effectiveness of Skin Surveillance Through a Specialized Clinic for Patients at High Risk of Melanoma. J Clin Oncol. 2017;35(1):63-71.

28. Law CK, Cust AE, Smit AK, Trevena L, Fernandez-Penas P, Nieweg OE, et al. Long-term cost-effectiveness of a melanoma prevention program using genomic risk information compared with standard prevention advice in Australia. Genet Med. 2023;25(12):100970.

29. National Institute for Health and Care Excellence (NICE). Artificial Intelligence (AI) technologies for assessing and triaging skin lesions referred to the urgent suspected skin cancer pathway: early value assessment https://www.nice.org.uk/guidance/hte24: NICE; 2025.

30. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019;25(1):44-56.

31. TURING AM. Computing machinery and intelligence Mind. 1950;LIX(236):433-60.

32. Matheny ME, Whicher D, Thadaney Israni S. Artificial Intelligence in Health Care: A Report From the National Academy of Medicine. JAMA. 2020;323(6):509-10.

33. Stanford Centre for Health Education. Fundamentals of AI and Machine Learning for Healthcare Study guide: Stanford University,; 2023.

34. Badawy M, Ramadan N, Hefny HA. Healthcare predictive analytics using machine learning and deep learning techniques: a survey. Journal of Electrical Systems and Information Technology. 2023;10(1):40.

35. National Academy of Medicine. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. In: Whicher D AM, Israni ST, et al., , editor. The Learning Health System Series,. Washington (DC): National Academies Press (US),; 2023.

36. Du-Harpur X, Watt FM, Luscombe NM, Lynch MD. What is AI? Applications of artificial intelligence to dermatology. British Journal of Dermatology 2020;183(3):423-30.

37. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278-324.

38. Felmingham CM, Adler NR, Ge Z, Morton RL, Janda M, Mar VJ. The Importance of Incorporating Human Factors in the Design and Implementation of Artificial Intelligence for Skin Cancer Diagnosis in the Real World. American Journal of Clinical Dermatology. 2020;22(2):233-42.

39. Bakker GJ, Weischer S, Ortas JF, Heidelin J, Andresen V, Beutler M, et al. Intravital deep-tumor single-beam 3-photon, 4-photon, and harmonic microscopy. eLife. 2022;11.

40. de Hond AAH, Steyerberg EW, van Calster B. Interpreting area under the receiver operating characteristic curve. The Lancet Digital Health. 2022;4(12):e853-e5.

41. Australian Government DoH. SkinVision BV Medical Image management system application softward: Therapeutic Goods Administration; 2022 [cited 2022 16 June]. Available from: https://compliance.health.gov.au/artg/.

42. Ngoo A, Finnane A, McMeniman E, Tan JM, Janda M, Soyer HP. Efficacy of smartphone applications in high-risk pigmented lesions. Australasian Journal of Dermatology. 2018;59(3):e175-e82.

43. Chuchu N, Takwoingi Y, Dinnes J, Matin RN, Bassett O, Moreau JF, et al. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. Cochrane database of systematic reviews. 2018;2018(12):CD013192-CD.

44. Ackermann DM, Dieng M, Medcalf E, Jenkins MC, van Kemenade CH, Janda M, et al. Assessing the Potential for Patient-led Surveillance After Treatment of Localized Melanoma (MEL-SELF): A Pilot Randomized Clinical Trial. JAMA Dermatology. 2022;158(1):33-42.

45. Winkler JK, Kommoss KS, Toberer F, Enk A, Maul LV, Navarini AA, et al. Performance of an automated total body mapping algorithm to detect melanocytic lesions of clinical relevance. European Journal of Cancer. 2024;202.

46. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. European Journal of Cancer 2019;119:11-7.

47. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology. 2018;29(8):1836-42.

48. Haenssle HA, Fink C, Toberer F, Winkler J, Stolz W, Deinlein T, et al. Man against machine reloaded: performance of a market-approved convolutional neural network in

classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. Annals of Oncology. 2020;31(1):137-43.

49. Marchetti MA, Codella NCF, Dusza SW, Gutman DA, Helba B, Kalloo A, et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. Journal of the American Academy of Dermatology. 2018;78(2):270-7.e1.

50. Maron RC, Weichenthal M, Utikal JS, Hekler A, Berking C, Hauschild A, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. European Journal of Cancer 2019;119:57-65.

51. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115-8.

52. Australian Institute of Health and Welfare. Cancer Data in Australia, Web Report 2023 [9May24]. Available from: https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/overview-of-cancer-in-australia-2023.

53. International Agency for Research on Cancer. Melanoma of skin Fact Sheet: World Health Organization; 2022 [cited 2024 9 May 2024]. Available from: https://gco.iarc.fr.

54. Du-Harpur X, Watt FM, Luscombe NM, Lynch MD. What is AI? Applications of artificial intelligence to dermatology. British journal of dermatology (1951). 2020;183(3):423-30.

55. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. European journal of cancer (1990). 2019;119:11-7.

56. Maron RC, Weichenthal M, Utikal JS, Hekler A, Berking C, Hauschild A, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. European journal of cancer (1990). 2019;119:57-65.

57. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature (London). 2017;542(7639):115-8.

58. Hornung A, Steeb T, Wessely A, Brinker TJ, Breakell T, Erdmann M, et al. The value of total body photography for the early detection of melanoma: A systematic review. International journal of environmental research and public health. 2021;18(4):1-17.

59. Soyer P. Australian Centre of Excellence in Melanoma Imaging and Diagnosis Cohort Study ANZCTR2019 [Available from: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=378385&isReview=true.

60. Haggenmuller S, Maron RC, Hekler A, Utikal JS, Barata C, Barnhill RL, et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. European Journal of Cancer. 2021;156:202-16.

61. Jones OT, Matin RN, van der Schaar M, Prathivadi Bhayankaram K, Ranmuthu CKI, Islam MS, et al. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review. Lancet Digit Health. 2022;4(6):e466-e76.

62. Magalhaes C, Mendes J, Vardasca R. The role of AI classifiers in skin cancer images. Skin Res Technol. 2019;25(5):750-7.

63. Popescu D, El-Khatib M, El-Khatib H, Ichim L. New Trends in Melanoma Detection Using Neural Networks: A Systematic Review. Sensors (Basel). 2022;22(2):10.

64. Zhang S, Wang Y, Zheng Q, Li J, Huang J, Long X. Artificial intelligence in melanoma: A systematic review. J. 2022;21(11):5993-6004.

65. PRISMA. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine. 2018;169(7):467-73.

66. Open Science Framework Available from: https://osf.io/.

67. TheEndNoteTeam. EndNote. Philidelphia,PA: Clarivate; 2022.

68. Veritas Health Innovation Ltd. Covidence 2022 Available from: www.covidence.org.

69. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated Reporting Guidance for Health Economic Evaluations. Value Health. 2022;25(1):3-9.

70. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020;2(10):e549-e60.

71. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. 2020;2(10):e537-e48.

72. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. JBI Evidence Implementation. 2015;13(3).

73. Tricco AL, E; Zarin, W and O'Brien, K;. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine. 2018;169(7):467-73.

74. DeepL. DeepL Translator Available from: https://www.deepl.com/en/translator.

75. Jaklitsch E, Thames T, de Campos Silva T, Coll P, Oliviero M, Ferris LK. Clinical Utility of an AI-powered, Handheld Elastic Scattering Spectroscopy Device on the Diagnosis and Management of Skin Cancer by Primary Care Physicians. Journal of Primary Care & Community Health. 2023;14:21501319231205979.

76. Giavina-Bianchi M, de Sousa RM, Paciello VZdA, Vitor WG, Okita AL, Prôa R, et al. Implementation of artificial intelligence algorithms for melanoma screening in a primary care setting. PLoS ONE. 2021;16:e0257006.

77. Marsden H, Morgan C, Austin S, DeGiovanni C, Venzi M, Kemos P, et al. Effectiveness of an image analyzing AI-based Digital Health Technology to identify Non-Melanoma Skin Cancer and other skin lesions: results of the DERM-003 study. Frontiers in Medicine. 2023;10.

78. Jahn AS, Navarini AA, Cerminara SE, Kostner L, Huber SM, Kunz M, et al. Over-Detection of Melanoma-Suspect Lesions by a CE-Certified Smartphone App: Performance in Comparison to Dermatologists, 2D and 3D Convolutional Neural Networks in a Prospective Data Set of 1204 Pigmented Skin Lesions Involving Patients' Perception. Cancers (Basel). 2022;14(15).

79. Winkler JK, Blum A, Kommoss K, Enk A, Toberer F, Rosenberger A, et al. Assessment of Diagnostic Performance of Dermatologists Cooperating With a Convolutional Neural Network in a Prospective Clinical Study: Human With Machine. JAMA Dermatology. 2023;159(6):621-7.

80. Smak Gregoor A, Sangers T, Bakker L, Hollestein L, Groot C, Nijsten T, et al. An artificial intelligence based app for skin cancer detection evaluated in a population based setting. npj Digital Medicine. 2023;6.

81. Lin A, Su B, Ning Y, Zhang L, He Y, editors. Convolutional Neural Networks in Medical Imaging: A Review. Advances in Swarm Intelligence; 2024; Singapore: Springer Nature Singapore.

82. Gomez Rossi J, Rojas-Perilla N, Krois J, Schwendicke F. Cost-effectiveness of Artificial Intelligence as a Decision-Support System Applied to the Detection and Grading of Melanoma, Dental Caries, and Diabetic Retinopathy. JAMA Network Open. 2022;5(3):e220269.

83. Marsden H, Kemos P, Venzi M, Noy M, Maheswaran S, Francis N, et al. Accuracy of an artificial intelligence as a medical device as part of a UK-based skin cancer teledermatology service. Frontiers in Medicine. 2024;11.

84. Agarwala S, Mata DA, Hafeez F. Accuracy of a convolutional neural network for dermatological diagnosis of tumours and skin lesions in a clinical setting. Clinical and Experimental Dermatology. 2021;46(7):1310-1.

85. Anderson JM, Tejani I, Jarmain T, Kellett L, Moy RL. Artificial Intelligence vs Medical Providers in the Dermoscopic Diagnosis of Melanoma. Cutis. 2023;111(5):254-8.

86. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. European Journal of Cancer. 2019;111:148-54.

87. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. European Journal of Cancer. 2019;113:47-54.

88. Brinker TJ, Schmitt M, Krieghoff-Henning EI, Barnhill R, Beltraminelli H, Braun SA, et al. Diagnostic performance of artificial intelligence for histologic melanoma recognition compared to 18 international expert pathologists. Journal of the American Academy of Dermatology. 2022;86(3):640-2.

89. Brodsky V, Levine L, Solans EP, Dola S, Chervony L, Polak S. Performance of Automated Classification of Diagnostic Entities in Dermatopathology Validated on Multisite Data Representing the Real-World Variability of Pathology Workload. Archives of Pathology & Laboratory Medicine. 2022;147(9):1093-8.

90. Cassalia F, Han SS, Cazzaniga S, Naldi L. Melanoma detection: Evaluating the classification performance of a deep convolutional neural network and dermatologist assessment via a mobile app in an Italian real-world setting. Journal of the European Academy of Dermatology and Venereology. 2024;38(9):e782-e4.

91. Cerminara SE, Cheng P, Kostner L, Huber S, Kunz M, Maul J-T, et al. Diagnostic performance of augmented intelligence with 2D and 3D total body photography and convolutional neural networks in a high-risk population for melanoma under real-world conditions: A new era of skin cancer screening? European Journal of Cancer. 2023;190:112954.

92. Crawford ME, Kamali K, Dorey RA, MacIntyre OC, Cleminson K, MacGillivary ML, et al. Using Artificial Intelligence as a Melanoma Screening Tool in Self-Referred Patients. Journal of Cutaneous Medicine and Surgery. 2024;28(1):37-43.

93. Diaz-Ramón JL, Gardeazabal J, Izu RM, Garrote E, Rasero J, Apraiz A, et al. Melanoma Clinical Decision Support System: An Artificial Intelligence-Based Tool to Diagnose and Predict Disease Outcome in Early-Stage Melanoma Patients. Cancers. 2023;15(7):2174.

94. Felmingham C, Pan Y, Kok Y, Kelly J, Gin D, Nguyen J, et al. Improving skin cancer management with ARTificial intelligence: A pre-post intervention trial of an artificial intelligence system used as a diagnostic aid for skin cancer management in a real-world specialist dermatology setting. Journal of the American Academy of Dermatology. 2023;88(5):1138-42.

95. Fink C, Blum A, Buhl T, Mitteldorf C, Hofmann-Wellenhof R, Deinlein T, et al. Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas. Journal of the European Academy of Dermatology and Venereology. 2020;34(6):1355-61.

96. Giulini M, Goldust M, Grabbe S, Ludwigs C, Seliger D, Karagaiah P, et al. Combining artificial intelligence and human expertise for more accurate dermoscopic melanoma diagnosis: A 2-session retrospective reader study. Journal of the American Academy of Dermatology. 2024;90(6):1266-8.

97. Goessinger EV, Cerminara SE, Mueller AM, Gottfrois P, Huber S, Amaral M, et al. Consistency of convolutional neural networks in dermoscopic melanoma recognition: A prospective real-world study about the pitfalls of augmented intelligence. Journal of the European Academy of Dermatology and Venereology. 2024;38(5):945-53.

98. Goessinger EV, Niederfeilner J-C, Cerminara S, Maul J-T, Kostner L, Kunz M, et al. Patient and dermatologists' perspectives on augmented intelligence for melanoma screening: A prospective study. Journal of the European Academy of Dermatology and Venereology.n/a(n/a).

99. Haenssle HA, Winkler JK, Fink C, Toberer F, Enk A, Stolz W, et al. Skin lesions of face and scalp – Classification by a market-approved convolutional neural network in comparison with 64 dermatologists. European Journal of Cancer. 2021;144:192-9.

100. Hartman RI, Trepanowski N, Chang MS, Tepedino K, Gianacas C, McNiff JM, et al. Multicenter prospective blinded melanoma detection study with a handheld elastic scattering spectroscopy device. JAAD International. 2024;15:24-31.

101. Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron RC, et al. Superior skin cancer classification by the combination of human and artificial intelligence. European Journal of Cancer. 2019;120:114-21.

102. Huang H-W, Hsu BW-Y, Lee C-H, Tseng VS. Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers. The Journal of Dermatology. 2021;48(3):310-6.

103. Kommoss KS, Winkler JK, Mueller-Christmann C, Bardehle F, Toberer F, Stolz W, et al. Observational study investigating the level of support from a convolutional neural network in face and scalp lesions deemed diagnostically ‘unclear’ by dermatologists. European Journal of Cancer. 2023;185:53-60.

104. Kränke T, Tripolt-Droschl K, Röd L, Hofmann-Wellenhof R, Koppitz M, Tripolt M. New AI-algorithms on smartphones to detect skin cancer in a clinical setting—A validation study. PLOS ONE. 2023;18(2):e0280670.

105. Liutkus J, Kriukas A, Stragyte D, Mazeika E, Raudonis V, Galetzka W, et al. Accuracy of a Smartphone-Based Artificial Intelligence Application for Classification of Melanomas, Melanocytic Nevi, and Seborrheic Keratoses. Diagnostics. 2023;13(13):2139.

106. MacLellan AN, Price EL, Publicover-Brouwer P, Matheson K, Ly TY, Pasternak S, et al. The use of noninvasive imaging techniques in the diagnosis of melanoma: a prospective diagnostic accuracy study. Journal of the American Academy of Dermatology. 2021;85(2):353-9.

107. Maier T, Kulichova D, Schotten K, Astrid R, Ruzicka T, Berking C, et al. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. Journal of the European Academy of Dermatology and Venereology. 2015;29(4):663-7.

108. Manolakos D, Patrick G, Geisse JK, Rabinovitz H, Buchanan K, Hoang P, et al. Use of an elastic-scattering spectroscopy and artificial intelligence device in the assessment of lesions suggestive of skin cancer: A comparative effectiveness study. JAAD International. 2024;14:52-8.

109. Marchetti MA, Liopyris K, Dusza SW, Codella NCF, Gutman DA, Helba B, et al. Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: Results of the International Skin Imaging Collaboration 2017. Journal of the American Academy of Dermatology. 2020;82(3):622-7.

110. Marchetti MA, Cowen EA, Kurtansky NR, Weber J, Dauscher M, DeFazio J, et al. Prospective validation of dermoscopy-based open-source artificial intelligence for melanoma diagnosis (PROVE-AI study). npj Digital Medicine. 2023;6(1):127.

111. Marchetti MA, Nazir ZH, Nanda JK, Dusza SW, D'Alessandro BM, DeFazio J, et al. 3D Whole-body skin imaging for automated melanoma detection. Journal of the European Academy of Dermatology and Venereology. 2023;37(5):945-50.

112. Menzies SW, Sinz C, Menzies M, Lo SN, Yolland W, Lingohr J, et al. Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial. The Lancet Digital Health. 2023;5(10):e679-e91.

113. Miller IJ, Stapelberg M, Rosic N, Hudson J, Coxon P, Furness J, et al. Implementation of artificial intelligence for the detection of cutaneous melanoma within a primary care setting: prevalence and types of skin cancer in outdoor enthusiasts. PeerJ. 2023;11:e15737.

114. Oloruntoba AI, Vestergaard T, Nguyen TD, Yu Z, Sashindranath M, Betz-Stablein B, et al. Assessing the Generalizability of Deep Learning Models Trained on Standardized and Nonstandardized Images and Their Performance Against Teledermatologists: Retrospective Comparative Study. JMIR Dermatol. 2022;5(3):e35150.

115. Phillips M, Marsden H, Jaffe W, Matin RN, Wali GN, Greenhalgh J, et al. Assessment of Accuracy of an Artificial Intelligence Algorithm to Detect Melanoma in Images of Skin Lesions. JAMA Network Open. 2019;2(10):e1913436-e.

116. Phillips M, Greenhalgh J, Marsden H, Palamaras I. Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy. Dermatol Pract Concept. 2020;10(1):e2020011.

117. Pyun SH, Min W, Goo B, Seit S, Azzi A, Yu-Shun Wong D, et al. Real-time, invivo skin cancer triage by laser-induced plasma spectroscopy combined with a deep learning based diagnostic algorithm. Journal of the American Academy of Dermatology. 2023;89(1):99-105.

118. Sangers T, Reeder S, van der Vet S, Jhingoer S, Mooyaart A, Siegel DM, et al. Validation of a Market-Approved Artificial Intelligence Mobile Health App for Skin Cancer Screening: A Prospective Multicenter Diagnostic Accuracy Study. Dermatology. 2022;238(4):649-56.

119. Thissen M, Udrea A, Hacking M, von Braunmuehl T, Ruzicka T. mHealth App for Risk Assessment of Pigmented and Nonpigmented Skin Lesions-A Study on Sensitivity and Specificity in Detecting Malignancy. Telemed J E Health. 2017;23(12):948-54.

120. Thomas L, Hyde C, Mullarkey D, Greenhalgh J, Kalsi D, Ko J. Real-world post-deployment performance of a novel machine learning-based digital health technology for skin lesion assessment and suggestions for post-market surveillance. Frontiers in Medicine. 2023;10.

121. Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, Braun RP, et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. JAMA Dermatol. 2019;155(1):58-65.

122. Udrea A, Mitra GD, Costea D, Noels EC, Wakkee M, Siegel DM, et al. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. Journal of the European Academy of Dermatology and Venereology. 2020;34(3):648-55.

123. Van Molle P, Mylle S, Verbelen T, De Boom C, Vankeirsbilck B, Verhaeghe E, et al. Dermatologist versus artificial intelligence confidence in dermoscopy diagnosis: Complementary information that may affect decision-making. Experimental Dermatology. 2023;32(10):1744-51.

124. Wies C, Schneider L, Haggenmüller S, Bucher T-C, Hobelsberger S, Heppt MV, et al. Evaluating deep learning-based melanoma classification using immunohistochemistry and routine histology: A three center study. PLOS ONE. 2024;19(1):e0297146.

125. Winkler JK, Sies K, Fink C, Toberer F, Enk A, Abassi MS, et al. Collective human intelligence outperforms artificial intelligence in a skin lesion classification task. JDDG: Journal der Deutschen Dermatologischen Gesellschaft. 2021;19(8):1178-84.