NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Artificial Intelligence technologies for assessing skin lesions for referral on the urgent suspected cancer pathway to detect benign lesions and reduce secondary care specialist appointments: early value assessment

Matthew Walton,[1] Alexis Llewellyn,[1] Eleonora Uphoff,[1] Joseph Lord,[1] Melissa Harden,[1] Robert Hodgson[1] and Mark Simmonds[1*]

[1]Centre for Reviews and Dissemination, University of York, Heslington, York, UK

[*]**Corresponding author** mark.simmonds@york.ac.uk

Published February 2026
DOI: 10.3310/GJMS0317

Scientific summary

Artificial Intelligence technologies for assessing skin lesions for referral on the urgent suspected cancer pathway to detect benign lesions and reduce secondary care specialist appointments: early value assessment

Health Technology Assessment 2026; Vol. 30: No. 10
DOI: 10.3310/GJMS0317

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Skin cancers are some of the most common types of cancer. Over 16,000 cases of melanoma, and more than 210,000 cases of non-melanoma skin cancer are diagnosed every year in the UK. In current practice, patients with suspicious skin lesions are referred to secondary care through the urgent suspected skin cancer referral pathway, where people attend a secondary care dermatology department for a face-to-face appointment with a consultant dermatologist. As benign skin lesions and skin cancer are so common, this places a very high burden on dermatology clinics, which may lead to a reduction in capacity to handle other skin conditions.

Artificial intelligence (AI) may be helpful in the diagnosis of skin cancer. An AI system could potentially identify which referred lesions are not cancerous using a high-quality photograph. An AI system could be used alone, or in combination with a dermatologist looking at the photograph. People judged not to have cancer could then be quickly discharged prior to secondary care consultation, while people whose lesion may be cancerous may be seen by a specialist in person. AI systems could therefore potentially speed up the diagnostic process and reduce the burden on the health service. AI systems are already used in the NHS in a research context, but there is a need to evaluate their clinical impact and value.

This project investigated whether two such AI technologies – Deep Ensemble for Recognition of Malignancy (DERM; Skin Analytics) and Moleanalyzer Pro (FotoFinder) –can produce clinically meaningful benefits for skin cancer diagnosis, and whether they have the potential to be cost-effective.

Objectives

The aim of the project was to investigate the clinical and cost-effectiveness of the two AI technologies, DERM and Moleanalyzer Pro, as decision aids to triage and diagnose suspicious skin lesions following a referral on the urgent suspected skin cancer pathway. To achieve this, the following objectives were proposed:

- To perform a rapid systematic review, narrative synthesis, and, where feasible, a meta-analysis, of the diagnostic accuracy, clinical impact and practical implementation of the included AI technologies.
- To perform a rapid systematic review of published cost-effectiveness studies of diagnostic strategies used to aid the diagnosis of skin cancer.
- To develop a conceptual model that will identify likely drivers of health benefit, harms and costs associated with implementing the included AI technologies in the NHS and identify areas for further research.

Methods

Data sources

MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials and the Association for Computing Machinery Digital Library were searched in November 2023. Clinical trial registries were searched. Unpublished material supplied by the included companies was also assessed.

Inclusion criteria

Any clinical study evaluating DERM or Moleanalyzer Pro in people with skin lesions suspicious of cancer, presenting in primary care, rapid diagnostic clinic, teledermatology or secondary care settings were eligible for inclusion. Included studies must report diagnostic accuracy, clinical outcomes, or evidence on implementation. The comparator was clinical judgement by dermatologists, but this did not need to be reported for a study to be eligible. The preferred reference standard for diagnosis was histology, but for unbiopsied lesions, clinical confirmation of non-malignancy was accepted.

The cost-effectiveness review included any economic evaluation including budget impact models, return on investment analysis, and other cost-only analyses of either DERM or Moleanalyzer Pro in the above population and setting. It was anticipated that no relevant studies would be identified for the named technologies; therefore, additional searches were also conducted to identify cost-effectiveness studies looking at any technology used to aid diagnosis of skin cancer in an NHS setting.

Data extraction

An initial scoping of studies was performed by extracting data on intervention, study location, size, setting, type of outcomes reported, and design and key quality indicators. Only studies with prospective recruitment of patients were taken forward for full data extraction and synthesis. For those studies, full data on the intervention, patient characteristics and all reported outcomes were extracted. Risk of bias was assessed using quality assessment of diagnostic accuracy studies-2 and quality assessment of diagnostic accuracy studies-comparative.

Identified economic evaluations were reviewed and discussed in detail, with the aim of informing the design and parameterisation of conceptual model. Material provided by submitting stakeholders pertaining to the value case for their product was also reviewed.

Synthesis

A scoping process was used to classify identified studies for relevance to the decision problem, based on study quality, setting, outcomes reported and relevance to the NHS. For studies taken forward from the scoping phase for full synthesis, a narrative synthesis was performed. Results are presented in structured tables and figures as appropriate, with a text summary. Random-effects meta-analyses of sensitivity and specificity were performed to pool diagnostic accuracy estimates across studies.

Evidence related to cost-effectiveness studies was reviewed and synthesised narratively.

Modelling

The conceptual model described sought to provide an overview of the structure of a cost–utility model and key evidence required for the assessment of AI technologies for the identification of benign lesions among suspected cancer cases referred on the urgent referral pathway. The structure of the conceptual model was designed considering the strengths and limitations of previously published diagnostic models for skin cancer in an NHS setting, and evidence submitted by stakeholders.

Results

Diagnostic accuracy and clinical impact of DERM

Six studies of DERM were identified, of which four were considered for full synthesis. Those four studies were all conducted in the UK. All studies excluded a substantial proportion of participants from assessment, which may produce biased results.

Meta-analysis of diagnostic accuracy data supplied by the company suggested that DERM has a high sensitivity of 96.1% to detect any malignant lesion [95% confidence interval (CI) 95.4 to 96.8], at a specificity of 65.4% (95% CI 64.7 to 66.1). The diagnostic accuracy for detecting melanoma or squamous cell carcinoma specifically was similar. For the detection of benign lesions, the sensitivity was 71.5% (95% CI 70.7 to 72.3) for a specificity of 86.2% (95% CI 85.4 to 87.0). This appears to be comparable in diagnostic accuracy to that achieved by dermatologists without the use of DERM. The diagnostic accuracy of combining DERM with assessment by a dermatologist could not be assessed.

Data on the clinical impact of using DERM were limited, and mostly unpublished. Some trial data suggested that autonomous use of DERM would lead to approximately half of patients being referred to a dermatologist for further assessment, and half being discharged. However, around 1% of people would be discharged with malignant lesions [mostly basal cell carcinomas (BCCs)]. DERM could potentially be used as part of a teledermatology service. However, use of DERM may slow progress to diagnosis.

Patient and clinical opinions of DERM were generally favourable towards accepting its use as part of the diagnostic pathway. However, there was very substantial resistance, particularly among clinicians, to using DERM without any assessment of lesions by a dermatologist.

Diagnostic accuracy and clinical impact of Moleanalyzer Pro

Seventeen publications of Moleanalyzer Pro were identified, but these were mostly retrospective reviews, and two prospective studies were eligible for full data extraction. The applicability of the evidence for Moleanalyzer Pro to practice is limited, notably due to the lack of studies from the UK and the lack of data for non-melanocytic lesions.

When pooled, these studies found that Moleanalyzer Pro alone had a sensitivity of 84.4% (95% CI 73.9 to 91.0) and a specificity of 84.5% (95% CI 72.0 to 92.1) to detect melanoma. Moleanalyzer Pro had a lower sensitivity and higher specificity to detect melanoma when compared with face-to-face dermatologist and remote teledermatology. There was no evidence on the diagnostic accuracy of Moleanalyzer Pro to detect other skin cancers, and no evidence was found on its clinical impact.

Economic evidence

No economic studies relating to the named technologies were identified from searches of the literature. Broader searches for any technology used to aid diagnosis of skin cancer in an NHS setting identified three studies. Although relevant to this review, none related to the use of AI for the detection of skin cancer and considered populations which were not relevant to the decision problem. While all identified studies adopted similar model structures, the mechanisms by which diagnostic accuracy generated value (in terms of either cost savings or quality-adjusted life-year gain) differed across these models. For instance, diagnostic sensitivity had less value in some models with value instead generated by the avoidance of unnecessary referral and diagnostic procedures.

Economic evidence on the cost-effectiveness of DERM was submitted by Skin Analytics and NHS England. This evidence was preliminary and did not include an executable model. The most relevant analysis was a cost–utility model developed by the Exeter Test Group and Skin Analytics. The Evidence Assessment Group (EAG) considered the model structure largely appropriate to capture important direct cost and health implications of AI technologies for directing discharge in a post-referral setting. However, a lack of key comparative data meant the relative clinical and cost-effectiveness of alternative pathways was necessarily based on often optimistic assumptions. The model suggested DERM could be highly cost-effective in the NHS, but we note that results may be very sensitive to the use of alternative sources of diagnostic accuracy data. We also noted several issues which may mean that the main value drivers were not appropriately characterised. Namely, the model imposed disincentives for the correct diagnosis and treatment of BCC; structurally imposed assumed sensitivity benefits for any strategy incorporating a triage step; used costs associated with biopsy and treatment which were inconsistent with sources generally used in National Institute for Health and Care Excellence appraisals, and may overvalue specificity in terms of generating cost savings.

No economic evidence related to Moleanalyzer Pro was identified.

Conceptual model

We developed a conceptual model aimed at providing an alternative to that presented in the Skin Analytics submission. While the proposed model retained the structure reported by Skin Analytics, the EAG propose an alternative structure for patients with BCC, aimed at better capturing the cost and health consequences of BCC, particularly with reference to disease recurrence.

We consider the current evidence inadequate to fully address the decision problem. Current evidence for both DERM and Moleanalyzer Pro is lacking with regard to the diagnostic accuracy of the whole diagnostic pathway (i.e. inclusive of subsequent steps). Availability of these data is essential to understanding the likelihood of missed cases which cannot be inferred from the partial data currently available. Similarly, comparable diagnostic accuracy data describing current service provision is lacking, particularly for the teledermatology pathway.

Conclusions

Impact on practice

The diagnostic accuracy of DERM suggests that it has potential for use within a post-primary care referral setting. This could be either alongside assessment by dermatologists or as an autonomous tool within the post-referral pathway within a subset of patients. However, the practical impact and clinical benefit of using DERM in a post-referral setting is currently unclear. In particular, the impact on referrals and secondary care appointments, the burden on clinicians and the subsequent clinical impact on patients are largely unclear. Although Moleanalyzer Pro shows promising accuracy for diagnosing melanoma, its evidence base is currently too limited to fully assess its clinical value.

Evidence on the diagnostic accuracy and clinical value of AI in people with darker skin tones or with lesions that are more difficult to assess (such as when versions are large, or obscured by scarring, tattooing or hair) was largely absent. Only a small number of people with darker skin tones were recruited to the included studies, and people with hard-to-assess lesions were often excluded. This raises concerns as to whether AI could be used in these people.

Current economic evidence supporting the cost-effectiveness of DERM is limited, and it is unclear whether the plausible advantages of DERM represent value for money relative to other strategies. Company-sponsored analyses suggested that DERM used autonomously and with a second read could be highly cost-effective compared to current 2-week wait diagnostic models. However, much of this value is generated through potentially optimistic assumptions around the diagnostic accuracy of comparators, and of the surrounding pathway (confidential information has been removed). Notably, the magnitude of uncaptured non-cash-releasing benefits remains unquantified.

There is currently no economic evidence supporting the use of Moleanalyzer Pro, but assuming a similar use case to DERM and appropriate data collection, the value of Moleanalyzer Pro could be assessed using the conceptual framework presented by the EAG.

Future research needs

The diagnostic accuracy of AI in a post-primary care referral pathway is uncertain and requires further evaluation. A lack of key comparative data on diagnostic accuracy means the relative clinical and cost-effectiveness of pathways incorporating AI technologies and teledermatology remains highly uncertain. Assessments of diagnostic accuracy of AI in people with darker skin tones or with hard-to-assess lesions are urgently needed.

Directly comparable evidence on the diagnostic accuracy of AI technologies and teledermatology in a post-referral setting compared with unassisted teledermatology is required to assess the potential value of AI technologies. This would require studies comparing AI with dermatologists' assessments, recruiting a representative population and case-mix, use of up-to-date versions of AI and dermoscopy, and with a robust independent reference standard for all patients.

A better understanding of the clinical benefits and resource implications associated with the implementation of AI technologies will also require further research to set up AI and teledermatology services in the NHS. Further research must also be undertaken to quantify the benefits to population health within skin cancer and other dermatological indications associated with any release of NHS consultant dermatologist resource, and understand the effects of these technologies on waiting times for final diagnosis.

This could potentially be achieved through continuations and extensions of existing ongoing pilot studies of DERM, but truly comparative evidence may also be required. Moleanalyzer Pro requires evaluation within a UK teledermatology setting.

The substantial resistance from both patients and clinicians to using AI without any human dermatological assessment means that if AI is to be used to direct discharge autonomously, more evidence is needed to demonstrate that it has clear benefits to patients, without sacrificing accuracy.

Study registration

This study is registered as PROSPERO CRD42023475705.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR136014) and is published in full in *Health Technology Assessment*; Vol. 30, No. 10. See the NIHR Funding and Awards website for further award information.

Health Technology Assessment

NICE TAR and DAR

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal

Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was commissioned and funded by the Evidence Synthesis Programme on behalf of NICE as award number NIHR136014. The protocol was agreed in October 2023. The draft manuscript began editorial review in March 2024 and was accepted for publication in April 2025. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).