NIHR Reference No. NIHR136014 (DAP76)

Version 1.1: 1 November 2023

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

External Assessment Group	CRD and CHE Technology Assessment Group, University of York, Heslington, York, YO10 5DD
Project Leads	Mark Simmonds Senior Research Fellow Centre for Reviews and Dissemination, University of York 01904 321091 mark.simmonds@york.ac.uk Robert Hodgson Senior Research Fellow Centre for Reviews and Dissemination, University of York 01904 321069 rob.hodgson@york.ac.uk

Note on the text

All commercial-in-confidence (CIC) data have been highlighted in blue and underlined, all depersonalised data (DPD) are highlighted in pink and underlined.

PROSPERO registration number: TBC

Source of funding

This project has been commissioned by the NIHR Evidence Synthesis Programme on behalf of the National Institute for Health and Care Excellence as project number NIHR136014.

PLAIN ENGLISH SUMMARY

Skin cancers are some of the most common types of cancer. Over 16,000 cases of melanoma (the deadliest form of skin cancer) and more than 210,000 cases of non-melanoma skin cancer are diagnosed every year in the UK. Skin cancer is normally diagnosed when a person notices a suspect mole or lesion on their skin. On seeing a general practitioner (GP), if the GP suspects the lesion might be cancerous the patient is referred to see a skin cancer specialist (a dermatologist) to make a diagnosis. As skin cancer is so common, this places a very high burden on dermatology clinics, which may lead to long waiting times for a diagnosis.

Artificial Intelligence (AI) may be helpful in the diagnosis of skin cancer. An AI system could potentially identify lesions that are or are not cancerous using a high-quality photograph of the lesion. This could be done by the AI system alone, or in combination with a dermatologist looking at the photograph (this is called teledermatology). People judged not to have cancer could then be quickly discharged, while people whose lesion may be cancerous will proceed to see a specialist in person. AI systems could therefore potentially speed up the diagnostic process and reduce the burden on the health service.

This project will investigate whether two such AI technologies: DERM (Skin Analytics) and MoleAnalyzer Pro (FotoFinder) could be useful. To do this, medical databases will be searched to identify all clinical studies of the technologies. These studies will be examined, alongside evidence supplied by the creators of the technologies, to investigate whether the technologies can accurately identify skin cancer cases, and whether their use might improve the diagnosis process for patients, and so improve their health outcomes. The project will also explore how the value of these technologies could be demonstrated in a future full cost-effectiveness model, and consider what evidence will be required before the technologies could be approved for widespread use on the NHS.

Table of Contents

Table of Tables

Table of Figures

1 DECISION PROBLEM

1.1 Purpose of the decision to be made

The purpose of this assessment is to investigate the use of Artificial Intelligence (AI) technologies for the analysis of skin lesions suspicious of cancer following a referral on the urgent suspected skin cancer pathway. The assessment will consider the use of two technologies: Deep Ensemble for Recognition of Malignancy (DERM) (Skin Analytics) and MoleAnalyzer Pro (FotoFinder systems). The assessment will consider existing evidence and identify potential evidence gaps on whether these technologies have the potential to be clinically useful and cost-effective to the NHS.

1.2 Interventions

This assessment will evaluate whether two AI technologies represent an effective and reliable means of triaging cancer from benign skin lesions, alongside current clinical practice.

1.2.1 Deep Ensemble for Recognition of Malignancy (DERM) (Skin Analytics)

DERM (Skin Analytics) is a UKCA class IIa AI-based skin lesion analysis technology intended for screening, triage and assessment of suspicious skin lesions. It is indicated for use on dermoscopic images of skin lesions where skin cancer is suspected in patients aged 18 years or over.

DERM uses AI-based algorithms to provide a suspected diagnosis of a given lesion and where applicable, a referral recommendation (for example, discharge and give safety netting advice or urgent referral for suspected cancer). DERM can classify lesions as: melanoma, squamous cell carcinoma, basal cell carcinoma, intra-epidermal carcinoma, actinic keratosis, atypical nevus or benign lesions (this includes benign vascular lesion, seborrheic keratosis, dermatofibroma, solar lentigo and melanocytic benign nevus). If a lesion exhibits features of more than one lesion type, DERM uses a risk hierarchy to return the more severe suspected diagnosis. The algorithm was trained on both historical (retrospectively) and prospectively collected images from populations in the UK, US and Italy. DERM uses a fixed algorithm and does not update itself automatically.

The technology has been deployed in the NHS since April 2020, including as a triage tool following a primary care referral. Over 51,000 patients have been assessed following a GP referral on the urgent suspected skin cancer pathway, to identify patients with benign lesions who can be discharged from the pathway without requiring specialist input from secondary care. People with suspicious lesions after DERM assessment have then been referred to a teledermatology review by a secondary care specialist.

1.2.2 MoleAnalyzer Pro (FotoFinder systems)

MoleAnalyzer pro (FotoFinder Systems) is a class IIa CE marked AI-based technology intended to be used by a medical professional for non-invasive visual documentation of skin lesions and aims to help the recognition of melanoma lesions. The technology is not intended to be used to confirm a clinical diagnosis of melanoma, and can be used for any age group. The target population is people with skin lesions, moles or multiple nevus syndrome. Lesions can be between 2 mm to 20 mm and should be on intact skin without additional psoriasis, eczema, acute sunburn or on hair-covered parts of the body.

MoleAnalyzer pro is used with the FotoFinder Universe software platform. The system requires a dermoscopic image for the AI score analysis. The software can only be used with the FotoFinder dermatoscopes: Dermlite Handyscope (this is compatible with any smartphone or tablet) and with Medicam 1000.

FotoFinder provides two options: online AI where the algorithm is updated continuously and offline AI in which the algorithm can be updated annually. This AI score is based on comparisons with images of malignant skin tumours such as: melanoma, basal cell carcinoma, lentigo maligna, squamous cell carcinoma, actinic keratosis, and many others. The score indicates how similar a lesion is to these comparison images, therefore it is only meant to provide a statistical estimate of the similarity to the malignant lesion images. A score between 0-0.0.2 indicates the lesion is inconspicuous, 0.21-0.49 indicates further clarification is necessary, and 0.50-1.0 indicates a conspicuous lesion which should be observed with great attention. MoleAnalyzer is already in use in some NHS centres.

1.3 Populations and relevant subgroups

The population of interest is people with skin lesions suspicious of cancer, who have been referred from primary care for further evaluation. The particular setting of interest is patients undergoing is teledermatology assessments, but all settings after primary care referral will be considered.

Subgroups relevant to this appraisal will be according to skin colour and type, and socioeconomic status.

1.4 Place of the intervention in the treatment pathway

In the UK, dermatology services receive about 1.2 million referrals a year and about 60% of these are suspected skin cancer pathway referrals, but only about 6% are converted to a confirmed case of skin cancer.[1] A significant proportion of people referred by GPs may not require face-to-face appointments in dermatology departments. The GIRFT report on dermatology highlighted that there are shortages in the workforce leading to delays in the diagnosis and treatment of skin cancer.[2] Furthermore, experts in

dermatology mentioned there is a low threshold for referral because GPs don't receive in-depth dermatology training and many do not have access to dermatoscopes, which are essential for confidently identifying both benign skin lesions and skin cancer.

1.4.1 Types of skin cancer

This assessment will cover all types of skin cancer. This includes three main types of skin cancer: melanoma, squamous cell carcinoma (SCC) and basal cell carcinoma (BCC), as well as other, rarer, forms of skin cancer.

1.4.1.1 Melanoma

A melanoma is a malignant tumour arising from melanocytes in the skin and is usually seen as a pigmented lesion on the skin. Melanoma is the fifth most common cancer in the UK, accounting for around 4% of all new cancer cases and more cancer deaths than all other skin cancers combined. On average, between 2016 and 2018, 16,744 new cases of melanoma were diagnosed each year in the UK (NICE, 2022).[3] The incidence of melanoma is projected to increase by 7% in the UK between 2014 to 2035.

Prognosis is highly dependent on the stage at diagnosis. For people with stage 1 melanoma (thickness is 2 mm or less, no sign that it has spread) the 5-year survival rate is almost 100%, compared with 30% for people with stage 4 melanoma (spread to distant lymph nodes or other parts of the body).

A weighted 7-point checklist is used to assess pigmented skin lesions and determine the need for referral. A pigmented lesion scoring of 3 or more on the weighted 7-point checklist is referred to the suspected cancer referral pathway.[4]

Weighted 7-point checklist:

- Major features of the lesions (scoring 2 points each):
    - change in size
    - irregular shape
    - irregular colour.
- Minor features of the lesion (scoring 1 point each):
    - largest diameter 7mm or more
    - inflammation
    - oozing
    - change in sensation.

1.4.1.2 Squamous cell carcinoma (SCC)

Squamous cell carcinoma (SCC) is the second most common type of non-melanoma skin cancer. It starts in the cells lining the top of the epidermis (outermost layer of the skin) and accounts for about 20 in every 100 skin cancers (NHS, 2020).[5] Approximately 25,000 squamous cell carcinomas of the skin are diagnosed each year (NICE, 2021).[6]

There is a small risk (up to 5%) of SCC spreading to other parts of the body, such as the lymph nodes (NHS, 2020). The risk of spread with SCCs is greater than with BCCs especially for people who are immunosuppressed. Death from squamous cell carcinoma is rare.

Actinic keratoses are dry, scaly patches of skin caused by damage from sun exposure. There is a small risk that the patches could develop into SCC if untreated (NHS, 2020).[5]

1.4.1.3 Basal cell carcinoma (BCC)

BCC is the most common form of skin cancer and accounts for about 75 in every 100 skin cancers. Approximately 75,000 BCC of the skin are diagnosed each year (NICE, 2021).[6]

BCC does not usually spread to other parts of the body, but if left untreated for a long time, they may get larger and grow deep into the skin and destroy skin, tissue and bone. In rare cases BCC can spread to other parts of the body and sometimes become life-threatening (NHS, 2020).[5] Death from BCC is exceptionally rare.

1.4.1.4 Other rare skin cancers

There are 45 other types of non-melanoma skin cancers. Merkel cell carcinoma is rarer and more aggressive than melanoma cancer. It is usually found in the head and neck region. Other types of rare non-melanoma skin cancers can be found in Appendix 1 of the NICE CSG8 guideline.

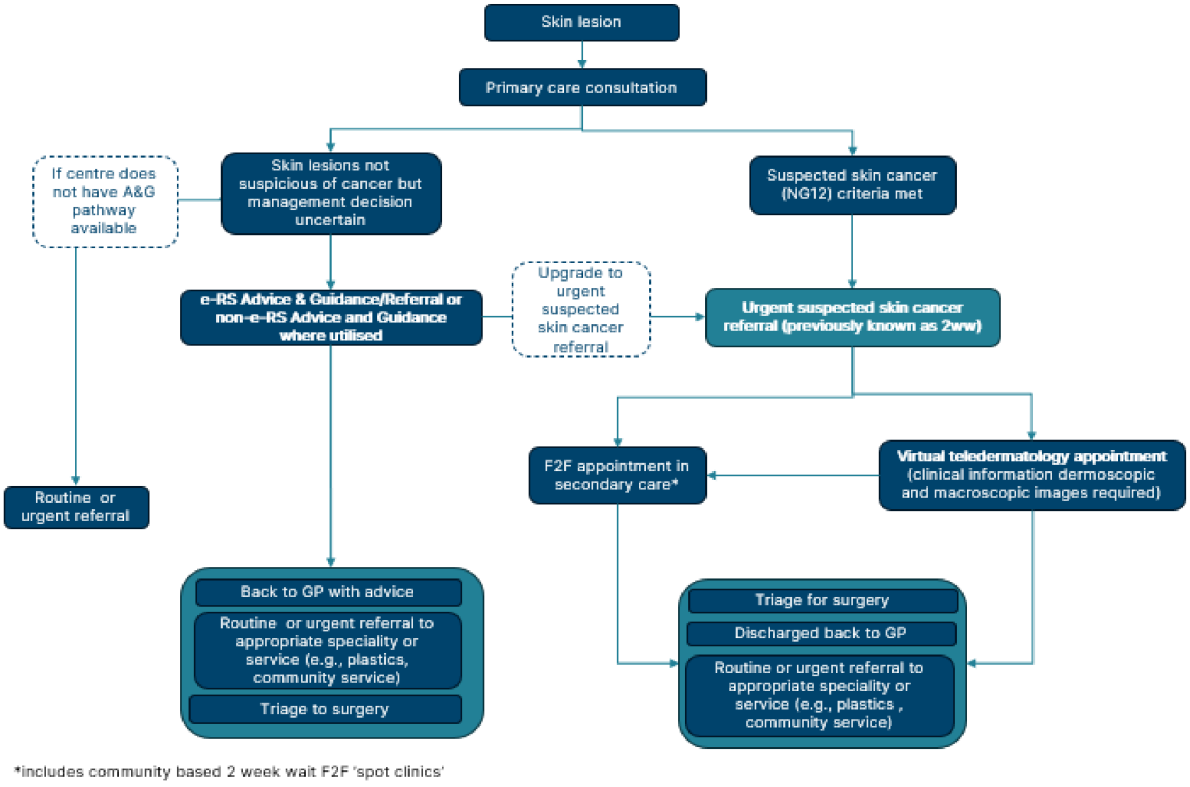
1.4.2 Current diagnostic pathway

The initial assessment of a person presenting with a skin condition occurs at the primary care level to determine the appropriate referral pathway. Traditionally, GPs directly referred everyone with suspicious skin lesions to secondary care through the urgent suspected skin cancer referral pathway where all referrals required people to attend a secondary care dermatology department for a face-to-face appointment with a consultant dermatologist. This pathway continues to exist where other clinical pathways are unsuitable or unavailable and is particularly well suited for people with multiple suspicious lesions, a history of skin cancer and other risk factors.

The NICE guidelines (NG12) on recognition and referral of skin cancer describes the use of the NHS e-RS advice and guidance (A&G) service as the main pathway in primary care for access to specialist

services, with the exception of the urgent suspected skin cancer pathway. However, there is variation in the pathway across the country.[7]

Figure 1 Current diagnostic pathway for suspect skin lesions (from NICE scope)



1.4.2.1 Urgent suspected skin cancer referral pathway

A person on the urgent suspected skin cancer referral pathway should receive a diagnosis or ruling out of cancer within 28 days of being referred urgently by their GP. For further details, see NHS England's webpage on faster diagnosis of cancer.[8] Section 1.7 of the NG12 guideline describes the criteria for an urgent referral for skin cancers (melanoma, SCC and BCC) to the urgent suspected skin cancer referral pathway.[7] These are summarised below.

Sections 1.7.1 to 1.7.3 of NICE guideline NG12 recommend that urgent referral using a suspected cancer pathway for melanoma should be arranged for people if:

- they have a suspicious pigmented lesion has a weighted 7-point checklist score of 3 or more,
- dermoscopy suggests melanoma,
- they have a pigmented or non-pigmented skin lesion that suggests nodular melanoma.

Additional criteria[9-11] also recommend urgent referral if:

- any new persistent skin lesion, especially if growing, pigmented, or vascular in appearance and the diagnosis is unclear,
- a new pigmented line in the nail (especially if there is associated damage to the nail), or a lesion growing under the nail,
- there is any doubt about the lesion, or there is a history of recent change,
- a biopsy has confirmed the diagnosis of malignant melanoma. Note: if a lesion is suspected to be melanoma, an urgent referral to a dermatologist or other suitable specialist with experience of melanoma diagnosis should be made, and excision in primary care should be avoided.
- a pigmented or non-pigmented skin lesion suggests nodular melanoma,
- any major features in the 7-point checklist, or any features of the ABCDE system.

Section 1.7.4 of NICE guidelines NG12 recommends a person is referred to an urgent suspected cancer pathway if they present with a skin lesion that raises the suspicion of squamous cell carcinoma. Section 1.7.5 to 1.7.6 of NICE guidelines NG12 recommend a routine referral for people if they have a skin lesion that raises the suspicion of a BCC. An urgent suspected cancer pathway referral should only be considered for a lesion that raises suspicion of BCC if there is a particular concern that a delay may have a significant impact, because of factors such as lesion site or size.

As shown in Figure 1, a referral to the urgent suspected skin cancer pathway results in either an urgent virtual teledermatology review, or an urgent face-to-face appointment in secondary care. If a primary care centre does not have a virtual teledermatology pathway available, the urgent face-to-face appointment pathway is used.

1.4.2.2 Teledermatology service

Teledermatology refers to the use of static digital images and relevant patient information to triage, diagnose, monitor or assess skin conditions remotely. If a person is referred through the urgent teledermatology referral pathway, clinical information along with a high-quality macroscopic image and dermoscopic images of the skin lesion are required. Images should be taken by a healthcare professional trained in medical photography. Images can be taken:

- in a GP surgery
- at a community diagnostic centre (CDC) close to a person's home
- at a teledermatology clinic based at a hospital.

Images are sent to be assessed by a consultant dermatologist using the teledermatology service and stored in the person's record. The person can be either be:

- booked directly for surgery
- discharged back to their GP
- referred for a routine or urgent referral to the appropriate speciality or service.

Virtual teledermatology cannot be used for lesions on difficult sites such as palms, soles, scalp and intimate areas, or for people with multiple lesions. Virtual teledermatology is not be used for children.

Clinical advice indicates that only eight CDC are currently assessing skin lesions in the NHS.

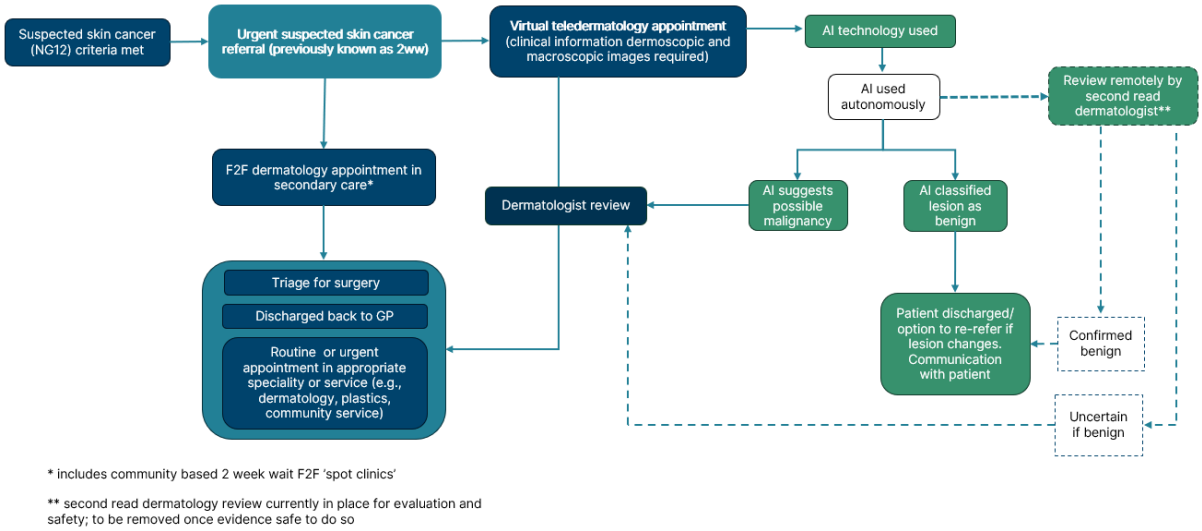
1.4.3 Potential positioning of AI technologies in the pathway

AI technologies to detect skin cancer could be used at various points in the diagnostic pathway:

1. By individuals concerned about suspect lesions, prior to consulting a GP.
2. As an adjunctive diagnostic in primary care settings (e.g. by a GP or nurse), to identify lesions that need referral
3. As an autonomous post referral assessment between primary and secondary care settings.
4. As an adjunctive diagnostic between primary and secondary care settings (e.g. teledermatology triage settings)
5. As an adjunctive diagnostic in a secondary care setting (e.g. by specialist dermatologists at face-to-face consultations)

This project will focus on settings 3 and 4, but will use evidence from other settings where it informs understanding. This aligns with where DERM is currently being used in a pilot to triage suspicious skin lesions after they have been referred by their GP on the urgent suspected skin cancer referral pathway. Figure 2 shows a possible pathway for AI use in post-referral that aligns with the current use of DERM. This post-referral assessment is used to identify those with benign lesions to be discharged from the urgent suspected skin cancer referral pathway. People identified with suspicious lesions (cases that contain at least one atypical, pre-malignant or malignant classification) from an AI assessment go on to a review by a specialist in secondary care.

Figure 2 Proposed positioning of AI technologies in post-referral setting (from NICE scope)



* includes community based 2 week wait F2F 'spot clinics'

** second read dermatology review currently in place for evaluation and safety; to be removed once evidence safe to do so

1.4.3.1 Adjunctive use of AI with dermatologist assessment

AI technologies could be used with a dermatologist review. After the AI assessment a dermatologist will review the results. This is done through virtual teledermatology with the aim of minimising false negative results (that is, cancerous lesions missed by the AI technology).

If the lesion is confirmed to be benign by the dermatologist, the patient is discharged from the pathway. The results are communicated to the patient and primary care referring clinician with safety net information to seek further medical advice if the lesion changes. If the dermatologist is uncertain about the diagnosis or if the AI suggests possible malignancy (whether the AI is used autonomously or not), the images are reviewed by a Trust dermatologist and triaged appropriately.

This "second read" dermatology review is currently in place for evaluation and safety, but the long-term plan is to remove this and for AI technologies to work autonomously, maximising efficient use of specialist dermatologists time (see below).

1.4.3.2 Autonomous use of AI

If AI technologies are used autonomously, a lesion classified as benign by the AI technology can be discharged without review by a dermatologist. The patient is discharged from the pathway and the results are communicated to the patient and primary care referring clinician with safety net information to seek further medical advice if the lesion changes. Lesions with suspected malignancy will be transferred to a dermatologist for teledermatology or face-to-face review.

1.4.4 Treatment of confirmed skin cancer

Treatment of skin cancer follows NICE guidance and British Association of Dermatologists guidelines.[11-13] In brief, early stage melanoma is usually treated by surgical excision; later stage

melanoma may also require lymph node resection, radiotherapy or chemotherapy. SCC and BCC are usually treated by surgical excision, but other treatments, including radiotherapy or chemotherapy may occasionally be used.

1.5 Relevant comparators

The comparator is clinical assessment and triage of suspicious lesions through the existing diagnostic pathway without use of AI. This can include assessment by specialist dermatologists either remotely or in person.

1.6 Key outcomes to be addressed

Outcomes fall into four main areas:

- Diagnostic accuracy
- Implementation, resource use, and practicality
- Clinical impact and patient benefit.
- Costs

1.6.1 Diagnostic accuracy

Diagnostic outcomes will be:

- Diagnostic test accuracy (sensitivity and specificity, area under ROC curve)
    - Where available, separately for each type of skin cancer (melanoma, BCC, SCC, rare skin cancers)
- Proportion of cancers missed and detected
- Proportion of benign lesions missed and detected
- Proportion of referrals confirmed to be skin cancer (positive predictive value)

1.6.2 Implementation, resource use, and practicality

Key outcomes will relate to resource use and timing:

- Proportion of urgent cancer referrals:
    - needing a face-to-face hospital appointment with a specialist for review of lesion
    - converted to routine referral pathway
    - resulting in a diagnostic biopsy
    - booked for surgical procedure
    - discharged back to GP
- Time to:

    - diagnosis
    - discharge
    - face-to-face consultant appointment
    - treatment (surgery)
- Cancer stage at detection
- Ease of use/acceptability of AI software by healthcare professionals
- Number of people consenting to use the technology
- Test failure rates (with reasons, e.g. image capture issues)
- Proportion of suspicious skin lesions/patients excluded (with reasons, e.g. due to lesion location or scarring)

1.6.3 Clinical impact and patient benefit.

- Clinical morbidities
    - Including distant metastases and adverse outcomes of treatment
- Mortality
- Health-related quality of life
- Non-clinical benefits to patients
    - Reassurance that lesion is not cancerous.
    - Anxiety associated with waiting for a diagnosis.
    - Acceptability of AI technologies or processes

1.6.4 Costs

Costs will be considered from an NHS and Personal Social Services perspective. Costs for consideration will include:

- Cost of annual subscription for AI software
- Cost of training healthcare professionals to take images and to interpret AI software results
- Cost of consultant dermatologist face-to-face appointments
- Cost of staff time to upload images to AI software platforms and to interpret the results
- Costs related to missed cancers
- Costs of consultant dermatology triage team
- Costs of teledermatology
- Costs of new services required to support AI technologies (such as establishing new teledermatology services and setting up image capture)

1.7 Objectives

The aim of the project is to investigate the clinical and cost-effectiveness of AI technologies as decision aids to triage and diagnose suspicious skin lesions, specifically the two technologies (DERM and MoleAnalyzer Pro) described in Section 1.2.

To achieve this, the following objectives are proposed:

Clinical effectiveness

- To perform a rapid systematic review, and if feasible a meta-analysis, of the diagnostic accuracy of the included AI technologies
- To perform a rapid systematic review with a narrative synthesis of the clinical impact and practical implementation of the AI technologies
- Based on the results of the rapid review, to identify evidence gaps and formulate recommendations for future research.

Cost effectiveness

- To perform a rapid systematic review of published cost-effectiveness studies of alternative diagnostic strategies used to aid the diagnosis of skin cancer. This will focus on the included AI technologies but will also include alternative strategies if no evidence is identified for the included technologies.
- To develop a conceptual model that will identify likely drivers of health benefit, harms and cost associated with implementing the included AI technologies in the NHS.
- If evidence and time allows: to develop a budget impact model capturing the direct resource implications of implementing the included AI technologies in the NHS. This may additionally include threshold analysis to explore how health effects or indirect costs may impact cost-effectiveness.

2 METHODS FOR SYNTHESISING EVIDENCE OF CLINICAL EFFECTIVENESS

The systematic review will be conducted following the general principles recommended in CRD's guidance and reported in accordance with the PRISMA statement.[14, 15]

The review will be conducted as a rapid review, aiming to scope the relevant literature and synthesise studies of key relevance to the UK health setting.

2.1 Search strategy

Searches of the literature will be conducted to identify published and unpublished primary studies relating to the use of the proposed AI technologies (DERM and MoleAnalyzer Pro) for identifying skin cancer.

An Information Specialist will design a search strategy in Ovid MEDLINE in consultation with the research team. The strategy will consist of search terms for skin cancer (in line with those types of skin cancer specified in the NICE scope document) combined with terms for AI and dermoscopy. Terms for the two specific tests and their company names will also be included in the strategy. The searches will be limited to records from 2015 onwards, reflecting the recent development of these technologies.

Bibliographic databases have been prioritised for searching, based upon relevance to the topic area of the assessment: MEDLINE ALL (Ovid), Embase (Ovid), Cochrane Central Register of Controlled Trials (CENTRAL, Wiley) and the Association for Computing Machinery (ACM) Digital Library.

As well as the database listed above, the following registers will be searched to identify any additional ongoing or unpublished studies: ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform (WHO ICTRP).

A draft search strategy for Ovid MEDLINE is included in Appendix 1. This strategy will be adapted to run on all databases and resources mentioned above. Records from the search will be imported into EndNote 20 (Clarivate Analytics, US.) for deduplication.

Additionally, company websites will be searched to identify relevant publications and other materials relating to the technology. The companies will be contacted (via NICE) to provide details of all studies (completed or ongoing) that they have conducted.

2.2 Study selection

One of two reviewers will screen all titles and abstracts. Papers will be prioritised for screening according to keywords in title and abstract (such as names and manufacturers of technologies). Full papers of any records that may be relevant will be obtained where possible and independently screened by two reviewers according to the inclusion criteria listed below. Papers that examine AI technologies but where the technology used is unclear will be identified, but will not proceed to full text assessment. Any disagreements will be resolved through discussion and, where necessary, consultation with a third reviewer.

A two-phase scoping process to identify relevant studies will be used. At the first phase, all relevant studies, (according to the inclusion criteria in Section 2.3) will be identified. A scoping process will then be used to identify studies of highest quality and most relevance to the decision problem for full data extraction and synthesis (see Section 2.4).

2.3 Inclusion criteria

Population

People with skin lesions suspicious of cancer, presenting in primary care, rapid diagnostic clinic, teledermatology or secondary care settings. The applicability of populations and settings to the NICE scope will be assessed and accounted for.

Interventions

DERM (Skin Analytics) and MoleAnalyser Pro (FotoFinder systems) used either alone, or in combination with clinical judgment. All versions of the technologies will be considered, and their applicability to current NHS practice will be assessed and accounted for.

Comparators

Clinical judgement and triage of suspect skin lesions as part of the current diagnostic pathway, without AI use. This includes, but is not restricted to, urgent teledermatology services and urgent face-to-face secondary care appointments. Assessment of lesions without AI by primary care physicians may also be included, at lower priority, where it informs understanding. The applicability of comparators to the NICE scope will be assessed and accounted for. Studies without a comparator will also be eligible.

Reference standard

Histological confirmation or rejection of malignancy from a biopsy of the suspect lesion. For unbiopsied lesions, confirmation of non-malignancy by specialist dermatologists, or ground truth as established by panels of dermatologists, will be accepted.

Outcomes

See Section 1.6 for a full list of intended outcomes.

Study designs

All studies that included adult patients with skin lesions suspicious of cancer, of any design. Priority will be given to studies with prospective recruitment of participants over retrospective reviews. Proof-of-concept, simulations and algorithm training studies will be excluded.

2.4 Data extraction

A data extraction form will be developed and piloted. For the initial scoping process data on intervention, study location and size, setting, type of outcomes reported, design and key quality indicators (randomisation, whether studies are comparative, prospective vs. retrospective design etc.) will be extracted.

For studies selected for full data extraction and synthesis, full data on the intervention, patient characteristics and all reported outcomes will be extracted by one reviewer and independently checked by a second reviewer. Discrepancies will be resolved by discussion, with involvement of a third reviewer where necessary. Where feasible and appropriate, data will be electronically extracted from figures presented in publications.

Data from relevant studies with multiple publications will be extracted and reported as a single study. The most recent or most complete publication will be used in situations where we cannot exclude the possibility of overlapping populations. Where there is evidence that an AI technology has developed or changed over time, only the most recent and complete studies will be included. Studies reported as conference abstracts only will be excluded.

2.5 Quality assessment

At the scoping phase, all studies will be assessed for broad quality using the hierarchy presented in Table 1.

If feasible given the volume of identified evidence, diagnostic accuracy studies will be assessed using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) tool[16] and comparative diagnostic accuracy studies will be assessed using the QUADAS-C (Quality Assessment of Diagnostic Accuracy Studies-Comparative) tool, which include items on the quality and applicability of studies.[17] If a full assessment of the studies is not feasible, the principles of QUADAS assessment will be used to broadly examine the quality and applicability of the studies.

Studies not reporting diagnostic accuracy data will not be formally assessed for quality, but all studies will be examined for any general quality or risk of bias concerns, following the general principals of the ROBINS-I quality assessment tool.

2.6 Methods of analysis and synthesis

2.6.1 Scoping review

Initially a scoping process will be used to classify identified studies for relevance to the decision problem, based on study quality, setting, outcomes reported and relevance to the NHS and population in the NICE scope (people referred on the urgent suspected skin cancer pathway). The priority hierarchy for the quality of diagnostic accuracy and clinical evidence studies that will be used is presented in Table 1.

Table 1 Study priority hierarchy for scoping review

Priority level	Diagnostic accuracy	Clinical and implementation evidence
1 (highest)	Prognostic cohort comparative studies	RCTs
2	Prognostic cohorts of AI technology only	Non-randomised cohort studies
3 (lowest)	Retrospective and case-control studies	Retrospective and case-control studies. Patient or clinician surveys

For each included AI technology only, the studies at highest priority level for that technology will be taken forward for full data extraction and narrative synthesis. For example, if there are RCTs of a technology, non-randomised studies will not be considered further. Studies at lower priority levels may be taken forward if they are of particular relevance to the NHS and the population in scope or report outcomes not presented in higher-quality studies.

Studies conducted in teledermatology settings, or equivalent early diagnostic clinics, will be preferred for full data extraction and synthesis. However, given variation in diagnostic processes in different countries, other settings, including primary care and specialist dermatology clinics, in studies outside of the UK will be considered where no evidence in the preferred settings is available.

2.6.2 Narrative synthesis

For studies taken forward from the scoping phase for full synthesis, a narrative synthesis approach will be used following appropriate guidance.[18] The results of data extraction for each outcome will be presented in structured tables and figures as appropriate, with a text summary. Studies will be grouped by population and intervention characteristics. Tabulated results will then be compared across studies, interventions and outcomes to identify the broader evidence of effectiveness. Evidence will be

summarised for specified subgroups (skin colour, skin type and socioeconomic status) where available.

2.6.3 Meta-analysis

If sufficient data on diagnostic accuracy are available, sensitivity and specificity data will be pooled using bivariate meta-analytic techniques. Subgroup analyses will be performed for relevant subgroups (skin colour, skin type and socioeconomic status) where there are sufficient data. Heterogeneity will be investigated by examining data plots, ROC plots, considering the I^2 statistic, and if feasible, by performing separate meta-analyses in different subgroups of studies or participants.

3 METHODS FOR SYNTHESISING EVIDENCE OF COST EFFECTIVENESS

Relevant cost-effectiveness evidence on the use of AI technologies with Class IIa designation (DERM, MoleAnalyzer Pro) for early detection of benign skin lesions will be identified, appraised for quality and narratively summarised. The aim of the review will be to examine any existing decision-analytic models used to assess the cost-effectiveness of the named AI software options against any comparator(s), in order to inform parameterisation of a conceptual model, which can be used to identify key issues, evidence gaps, and areas of uncertainty to help direct future data collection and research.

3.1 Identifying and reviewing published cost-effectiveness studies

The results of the literature search carried out to identify studies relating to the use of AI software to detect skin cancer (Section 2) will be used to identify any relevant studies of the cost-effectiveness of the named technologies in people with suspicious skin lesions in any setting.

Study designs will include budget impact models, return on investment analysis, and other cost-only analyses, as well as full economic evaluations considering both costs and consequences (including cost-effectiveness, cost-utility and cost-benefit analyses). Formal synthesis will be limited to studies involving one or more of the named technologies in an NHS setting If no relevant studies are identified, cost-effectiveness studies which look more broadly at other technologies used to aid diagnosis of skin cancer in the NHS urgent cancer referral pathway (per NG12) will be reviewed. Cost-effectiveness studies of other technologies will be identified using additional pragmatic searches of the literature.

The review will examine existing economic models in detail, with the aim of informing the design and parameterisation of conceptual model. This will include identifying important structural assumptions, highlighting key areas of uncertainty and defining the key parameter inputs necessary in order to develop a full decision analytic model. Any material provided by submitting stakeholders pertaining to the value case for their product will also be reviewed and is likely to contribute to the development of a budget impact model, and a conceptual cost-effectiveness model.

We will aim to answer the following decision questions on the basis of the identified published evidence, and material submitted by the developers of the included technologies:

1. What are the cost and resource use implications of the use of AI technologies following an urgent suspected skin cancer referral to identify benign skin lesions?

2. What would a health economic model to estimate the cost-effectiveness of AI technologies to identify benign skin lesions in this setting look like, and what are the key evidence requirements necessary to populate such a model?

3.2 Development of a conceptual cost-effectiveness model

The structure of a conceptual model for AI tools will be necessarily pragmatic and flexible in terms of the number of different diagnostic and care pathways included, and the two potential implementation models for AI technologies in a post-referral setting.

The conceptual model framework will seek to identify the key input parameters required to capture the impact of the modelled technologies. Key inputs necessary for the linkage of short-term diagnostic accuracy metrics with long-term outcomes will be identified. This will be extended to cover the indirect effects of modifying clinical decision-making, including the relationship between time to diagnosis and treatment outcomes, if appropriate.

The conceptual model will be developed in alignment with the NICE reference case.

3.3 Development of a budget impact assessment

If data and time allow, we will aim to assess whether the use of AI technologies for the analysis of suspicious skin lesions on the urgent suspected skin cancer pathway has the potential to be cost saving compared to triage via teledermatology services, or referral straight to face to face appointments with a dermatologist.

A budget impact model will be developed with reference to the scope of the present assessment, if data allows. Detailed national-level, and ideally trust-level service use data will be required to accurately model the resource and cost consequences of the technologies versus each of the comparator diagnostic pathways.

The budget impact model will consider how the costs of implementation and ongoing operation of AI software compare to the costs offset by avoidance of additional referrals generated by current diagnostic pathways with reference to the key elements depicted in Figure 1 of the final NICE scope. The model will also consider costs related to missed cancers, and the processing of skin lesions not eligible for assessment using the included AI technologies.

The budget impact will be based on the cost of software installation, unit cost per image taken and analysed, the number of individuals currently referred through the suspected cancer referral pathway, and the number of cases of the target conditions identified. Simple threshold analysis may also be

undertaken to use known data on offset costs to identify the maximum number of false negatives per patient screened for AI technologies to maintain a positive net monetary benefit.

4 HANDLING INFORMATION FROM THE COMPANIES

All data submitted by the company(s) will be considered if received by the EAG no later than the end of October 2023. Data supplied during November 2023 may be evaluated if time permits. Data arriving after the end of November 2023 will not be considered. If the data meet the inclusion criteria for the review they will be extracted and quality assessed in accordance with the procedures outlined in this protocol.

Any 'commercial in confidence' data provided by a company and specified as such will be highlighted in blue and underlined in the assessment report (followed by an indication of the relevant company name e.g. in brackets). Any 'academic in confidence' data provided by the manufacturers or study analysts will be highlighted in yellow and underlined in the assessment report (followed by an indication of the relevant company name e.g. in brackets). Confidential data will be stored securely and will only be accessible to members of the project team.

5 COMPETING INTERESTS OF AUTHORS

The authors have no competing interests to declare.

6 TIMETABLE/MILESTONES

Milestone	**Date to be completed**
Submission of final protocol	30/10/2023
Submission of progress report	16/11/2023
Submission of draft Diagnostic Assessment Report	8/12/2023
Submission of final Diagnostic Assessment Report	3/1/2024

7 REFERENCES

1. *The two-week wait skin cancer pathway: innovative approaches to support early diagnosis of skin cancer as part of the NHS COVID-19 recovery plan. 4 April 2022 Version 1*: NHS and British Association of Dermatologists; 2022.

2. Levell N. *Dermatology. GIRFT Programme National Specialty Report*: GIRFT and NHS; 2021.

3. *Clinical Knowledge Summary - Melanoma*. National Institute of Health and Care Excllence (NICE); 2022. URL: https://cks.nice.org.uk/topics/melanoma/ (accessed 26th October, 2023).

4. *Clinical Knowledge Summary - Melanoma. How should I assess a lesion?* National Institute of Health and Care Excellence (NICE); 2022. URL: https://cks.nice.org.uk/topics/melanoma/diagnosis/assessment/ (accessed 26th October, 2023).

5. *Non-melanoma skin cancer*. NHS. URL: https://www.nhs.uk/conditions/non-melanoma-skin-cancer/ (accessed 26th October, 2023).

6. *Clinical Knowledge Summary - Skin cancers - recognition and referral*. National Institute of Health and Care Excellence (NICE); 2021. URL: https://cks.nice.org.uk/topics/skin-cancers-recognition-referral/ (accessed 26th October, 2023).

7. National Institue for Health and Care Excellence. *Suspected cancer: recognition and referral (NG12)*. London: NICE; 2023.

8. *Faster diagnosis*. NHS England. URL: https://www.england.nhs.uk/cancer/faster-diagnosis/ (accessed 26th October, 2023).

9. Marsden JR, Newton-Bishop JA, Burrows L, Cook M, Corrie PG, Cox NH*, et al.* Revised UK guidelines for the management of cutaneous melanoma 2010. *J Plast Reconstr Aesthet Surg* 2010;**63**:1401-19. https://doi.org/https://doi.org/10.1016/j.bjps.2010.07.006

10. Bhave P, Wu Y, Fernandez-Peñas P. *Melanoma*. BMJ Best Practice; 2023. URL: https://bestpractice.bmj.com/topics/en-gb/268 (accessed 26th October, 2023).

11. National Institute of Health and Care Excellence. *Improving outcomes for people with skin tumours including melanoma. Cancer service guideline [CSG8]*. NICE; 2010. URL: https://www.nice.org.uk/guidance/csg8 (accessed 26th October, 2023).

12. National Institute of Health and Care Excellence. *Melanoma: assessment and management. NICE guideline [NG14]*. NICE; 2022. URL: https://www.nice.org.uk/guidance/ng14/chapter/Recommendations#assessing-melanoma (accessed 26th October, 2023).

13. Keohane SG, Botting J, Budny PG, Dolan OM, Fife K, Harwood CA*, et al.* British Association of Dermatologists guidelines for the management of people with cutaneous squamous cell carcinoma 2020. *Br J Dermatol* 2021;**184**:401-14. https://doi.org/10.1111/bjd.19621

14. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. York: CRD, University of York; 2009.

15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD*, et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* 2021;**10**:89. https://doi.org/10.1186/s13643-021-01626-4

16. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB*, et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529-36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

17. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF*, et al.* QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med* 2021;**174**:1592-9. https://doi.org/10.7326/m21-2234

18. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S*, et al.* Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ* 2020;**368**:l6890. https://doi.org/10.1136/bmj.l6890

APPENDICES

Appendix 1: Example MEDLINE search strategy

Database: Ovid MEDLINE(R) ALL <1946 to October 26, 2023>
Search Strategy:
1 exp Skin Neoplasms/ (144346)
2 melanoma/ or hutchinson's melanotic freckle/ or melanoma, amelanotic/ (99020)
3 exp Carcinoma, Basal Cell/ (19779)
4 Carcinoma, Squamous Cell/ (141618)
5 Bowen's Disease/ (2003)
6 Carcinoma, Merkel Cell/ (3171)
7 Carcinoma, Neuroendocrine/ (5883)
8 exp Nevus/ (17235)
9 (skin adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (51761)
10 (cutaneous adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (18507)
11 melanoma$.ti,ab. (138809)
12 (nonmelanoma$ or non-melanoma$ or NMSC).ti,ab. (7220)
13 (melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab. (75642)
14 ((melanotic or malignan$ or Hutchinson$) adj2 freckle$).ti,ab. (66)
15 (lentigo$ adj2 maligna$).ti,ab. (1363)
16 ((basal adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)) or basalioma$ or BCC).ti,ab. (30172)
17 ((squamous cell adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)) or SCC or cSCC).ti,ab. (137009)
18 (Bowen$ adj3 (disease$ or lesion$ or cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (2464)
19 (Merkel cell adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (4165)
20 ((intra-epiderm$ or intraepiderm$ or intra-derm$ or intraderm$) adj3 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (864)
21 ((neuroendocrine or neuro-endocrine) adj2 (cancer$ or carcinoma$ or tumour$ or tumor$ or neoplas$ or oncolog$ or adenoma$ or adenocarcinoma$ or epithel$ or malignan$ or premalignan$ or precancer$)).ti,ab. (28676)
22 ((skin or cutaneous or pigmented or nonpigmented) adj3 (lesion$ or nodul$ or macule$)).ti,ab. (59203)
23 (mole$1 or nevus or nevi or naevus or naevi).ti,ab. (43253)
24 ((acitinic or solar or senile) adj2 kerato$).ti,ab. (535)
25 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 (617130)
26 Artificial Intelligence/ (40789)
27 algorithms/ (306181)
28 exp Machine Learning/ (60979)
29 exp neural networks, computer/ (61553)
30 ((artificial$ or machine$ or computational$) adj2 intelligen$).ti,ab. (35619)
31 computer vision.ti,ab. (7402)
32 (AI or AIDHT or AIaMD).ti,ab. (47366)
33 (augment$ adj2 intelligen$).ti,ab. (209)
34 algorithm$.ti,ab. (366114)
35 deep learning.ti,ab. (46615)
36 machine learning.ti,ab. (85699)
37 ((supervised or unsupervised or semi-supervised) adj2 learning).ti,ab. (11772)
38 ((neural or convolutional) adj2 network$).ti,ab. (99895)
39 (CNN or CNNs or DCNN or DCNNs).ti,ab. (17959)
40 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 (726461)
41 25 and 40 (8426)
42 Dermoscopy/ (5905)
43 (dermoscop$ or dermascop$ or dermatoscop$).ti,ab. (7648)
44 ((skin or cutaneous or epidermis) adj3 (microscopy or microscopies)).ti,ab. (1060)
45 (epiluminescen$ adj3 (microscopy or microscopies)).ti,ab. (229)
46 (teledermoscop$ or teledermascop$ or teledermatoscop$).ti,ab. (150)
47 (videodermoscop$ or videodermascop$ or videodermatoscop$).ti,ab. (188)
48 (Dermlite Handyscope$ or "Medicam 1000").ti,ab. (4)
49 (teledermatolog$ or tele-dermatolog$).ti,ab. (1280)
50 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 (11806)

51 41 and 50 (986)
52 "Deep Ensemble for the Recognition of Malignancy".af. (1)
53 (DERM and (Algorithm$ or Artificial Intelligen$ or AI)).tw. (12)
54 "Melanoma Image Analysis Algorithm".af. (0)
55 (Skin Analytics$ or SkinAnalytics$).af. (6)
56 (Moleanalyzer$ or Mole analyzer$ or Moleanalyser$ or Mole analyser$ or FotoFinder$).af. (63)
57 52 or 53 or 54 or 55 or 56 (79)
58 51 or 57 (1041)
59 exp animals/ not humans.sh. (5163374)
60 58 not 59 (1036)
61 limit 60 to yr="2015 -Current" (719)