NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Research plan

Artificial intelligence (AI) technologies for assessing and triaging skin lesions within the urgent suspected skin cancer pathway

15 August 2024

1 Purpose of this document

NICE recommends that more research is needed on using 2 artificial intelligence (AI) technologies (Deep Ensemble for Recognition of Malignancy [DERM] and Moleanalyzer pro) in teledermatology services to assess and triage skin lesions within the urgent suspected skin cancer pathway, before they can be used in the NHS. More evidence is needed to assess the risk and benefits of the technologies in routine care.

This plan outlines the evidence gaps and what real-world data needs to be collected for a NICE review of the technologies again in the future. It is not a study protocol but suggests an approach to collecting the information needed to address the evidence gaps.

The companies are responsible for ensuring that data collection and analysis takes place. An approach to research is through formation of a consortium, bringing analytical partners and implementation sites together with companies.

When suitable evidence has been collected, the companies should submit the evidence to NICE in a form that can be used for decision making. NICE will review all the evidence and assess whether the technologies can be routinely adopted in the NHS

August 2024

2   Evidence gaps

This section describes the evidence gaps, why they need to be addressed and their relative importance for future committee decision making.

The committee will not be able to make a positive recommendation without the essential evidence gaps (see section 2.1) being addressed. This will help the committee to make a recommendation by ensuring it has a better understanding of the patient or healthcare system benefits of the technologies.

2.1   Essential evidence for future committee decision making

How accurate AI technologies used in teledermatology services are at detecting cancer and non-cancer skin lesions compared with established teledermatology services alone

Collecting data on the accuracy of AI technologies in detecting cancer and non-cancer skin lesions compared with teledermatology alone is essential for determining whether AI can provide assessments suitable for routine clinical settings. The data will help to determine whether the AI technologies can effectively discharge non-urgent cases from the suspected skin cancer pathway while maintaining diagnostic accuracy for detecting high risk lesions. This data will help evaluate the potential of AI technologies to enhance the diagnosis of cancer skin lesions and optimise clinical resources by reducing the burden on dermatology services. The data will also help to assess whether AI technologies increase staff capacity and benefit people with non-cancer dermatological conditions, compared with teledermatology alone. Additionally, this can help inform whether such technologies can be used autonomously and if this is reliable and safe.

How accurate AI technologies are at detecting non-cancer and cancer skin lesions in people with black or brown skin

Collecting information about the accuracy of the technologies in detecting skin lesions on different skin colour is vital for assessing potential biases in performance and for ensuring equitable healthcare. Skin tone should ideally be measured using skin spectrophotometry.

Research plan – Artificial intelligence (AI) technologies for assessing and triaging skin lesions within the urgent suspected skin cancer pathway   Page 2 of 10

August 2024

The effect of using AI technologies in teledermatology services on the number of urgent suspected cancer referrals and for face-to-face dermatology appointments compared with a well-established teledermatology service alone

Collecting data on the number of face-to-face dermatology referrals generated by AI technology compared with established teledermatology services alone is crucial for understanding their impact on healthcare workflows. This information will help determine if AI can reduce unnecessary referrals, saving dermatologist time, and thereby optimising resources and improving timely access to care for people.

3 Approach to research

3.1 Evidence gaps and ongoing studies

Table 1 summarises the evidence gaps and ongoing studies that might address them. Information about evidence status is derived from the external assessment group's report; evidence not meeting the scope and inclusion criteria is not included. The table shows the evidence available to the committee when the guidance was published.

Table 1 Evidence gaps and ongoing studies

Evidence gap	Deep Ensemble for Recognition of Malignancy (DERM)	Moleanalyzer pro
How accurate AI technologies used in teledermatology services are at detecting cancer and non-cancer skin lesions compared with established teledermatology services alone	Limited evidence	Limited evidence
How accurate AI technologies are at detecting non-cancer and cancer skin lesions in people with black or brown skin	Limited evidence	Limited evidence
The effect of using AI technologies in teledermatology services on the number of urgent suspected cancer referrals and for face-to-face dermatology appointments compared with a well-established teledermatology service alone	Limited evidence	No evidence

No ongoing studies were identified in the external assessment group's report.

3.2   Data sources

NICE's real-world evidence framework provides detailed guidance on assessing the suitability of a real-world data source to answer a specific research question.

Some data will be generated through the AI technologies themselves, such as the number of referrals that used the technology and the diagnostic outcomes predicted by the technology. This data can be integrated with other data collected.

The NHS England Secure Data Environment service could potentially support this research. This platform provides access to high standard NHS health and social care data that can be used for research and analysis. Local or regional data collections such as NHS England's sub-national secure data environments and databases like NHS England's National Cancer Registration and Analysis Service (NCRAS) already measure outcomes specified in the research plan. They could be used to collect data to address the evidence gaps. Secure data environments are data storage and

Research plan – Artificial intelligence (AI) technologies for assessing and triaging skin lesions within the urgent suspected skin cancer pathway   Page 4 of 10

August 2024

access platforms that bring together many sources of data, such as from primary and secondary care, to enable research and analysis. The sub-national secure data environments are designed to be agile and can be modified to suit the needs of new projects.

Datasets that are taken from general practice electronic health records with broad coverage, such as the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN) could be used to provide individual patient-level data. These could provide some useful information on referrals, diagnostic outcomes and patient characteristics.

The quality and coverage of real-world data collections are of key importance when used in research. Active monitoring and follow up through a central coordinating point is an effective and viable approach of ensuring good-quality data with broad coverage.

3.3 Evidence collection plan

Diagnostic accuracy study

A diagnostic accuracy study is used to assess the agreement between 2 or more methods. The study would assess the agreement between the diagnosis decision reached for each included case of suspected cancer by:

- AI technology alone (intervention)
- teledermatology unassisted by AI technology (comparator)
- a reference standard

There are several potential approaches that can be taken to collect the reference standard:

- Panel of experts: A consensus assessment by an expert panel, unassisted by AI technology, but ideally with access to clinical information that would be available at the time the AI is intended to be applied. This is the ideal approach for a comprehensive assessment of both the AI technology and teledermatology.

Research plan – Artificial intelligence (AI) technologies for assessing and triaging skin lesions within the urgent suspected skin cancer pathway   Page 5 of 10

August 2024

- Arbitration process: An arbitration process designed to resolve disagreements in diagnoses between the AI technology and teledermatology results. This method helps to determine the final diagnosis when there is discordance between the two approaches. For further details on implementing this method, please refer to the academic paper in this link. A limitation of this approach is that it does not collect specific information about sensitivity and specificity of the technology. Therefore, this information may need to be sourced through other methods. This is particularly important to help populate future economic models.

- Follow up: Monitoring of clinical progression to identify and assess any false negatives or false positives, ensuring that the accuracy of the initial diagnosis can be confirmed or corrected over time. This approach will suffer from differential verification bias and may require a considerable follow up period.

Representative image sets would be generated prospectively. These would then be processed by the AI technology and referred to usual teledermatology services. Cases that the AI technology was unable to analyse would be recorded. It is important to consider variation in skin colour as part of the study design, for example, ensuring a sufficient sample size to assess different skin colours (ideally, measured using skin spectrophotometry).

A comparison between the AI technology alone (intervention), the teledermatology unassisted by AI (comparator) and the reference standard could assess agreement between the diagnosis decision of each of the AI technologies, when used as intended. This comparison would also allow an assessment of the diagnostic accuracy of teledermatology services. Cases with disagreements in the diagnosis between each method, could be further explored to identify common characteristics, and reasons for disagreements could be considered.

For pragmatism, this study could be done as part of the 'Before' in the before after study. Care would need to be taken to control for confounders and blinding. This

Research plan – Artificial intelligence (AI) technologies for assessing and triaging skin lesions within the urgent suspected skin cancer pathway   Page 6 of 10

August 2024

could reduce the time to evidence generation and ensures the accuracy values are representative of the setting.

Real-world before-and-after implementation study

The results of the accuracy study should inform the population for a before-and-after implementation study, so that the AI technologies are implemented in populations in which they have been shown to be effective.

A before and after design allows for comparisons, particularly when there is considerable variation between services in the standards and mode of delivery of teledermatology across the NHS. It also facilitates an assessment of implementation costs, changes in referral rates, and the proportion of cases that are eligible for assessment by the AI technologies.

Before the AI technologies are implemented in a teledermatology service, data should be collected on the:

total number of referrals to that service
number of those referrals that resulted in a face-to-face appointment with a
   dermatologist
number of biopsies
number of referrals that resulted in a cancer lesion diagnosis.

If teledermatology is already established in the service then the number of lesions that are not eligible for assessment by this service should also be recorded and the reasons why. The AI technology should then be implemented into the service and all implementation and training costs should be collected. After leaving a period of time to account for learning effects, the outcomes on referral rates, appointments, and biopsies should be collected again in a period after implementation. The number of lesions that are not eligible for assessment by the AI technology should also be collected in the after-implementation study, and the reason why.

In a phased approach, a comparison between the AI technology's diagnosis and a dermatologist's opinion, and ideally, the final clinical outcome, can help to predict the

likely impact of autonomous use of the AI technology before moving into, and testing fully autonomous use.

This study could be done at a single centre with an established teledermatology service or ideally, replicated across multiple centres. This could show how the AI technologies can be implemented across a range of services, representative of the variety in the NHS. Outcomes may reflect other changes that occur over time in the population, unrelated to the interventions. Additional robustness can be achieved by collecting data in a centre that has not implemented an AI technology but is as similar as possible (in terms of clinical practice and patient characteristics) to a service where an AI technology is being used or ideally, a stepped wedge design. This could help control for changes in referral rates over time that might have occurred anyway.

3.4   Data to be collected

The following information has been identified for collection:

Diagnostic accuracy study

- Classifications made using teledermatology unassisted by AI technology, and by AI technology alone, and by the reference standard.
- Information on lesions that are not eligible for assessment by teledermatology and not eligible for assessment by AI technology and the reasons.
- Whether or not the AI technology was able to process each image.
- Performance of the AI technologies and teledermatology compared with the reference standard, for example, diagnosis of any malignant lesions, melanoma, squamous cell carcinoma, basal cell carcinoma or other non-cancer lesions classified by the technology.
- Accuracy in people with black or brown skin (ideally measured using skin spectrophotometry).
- Cases of diagnostic disagreement and the likely reason for disagreement (given by reference standard).

Real-world before-and-after implementation study

- Patient information, for example age, sex and ethnicity.

- Number and proportion of suspected skin cancer cases that are not eligible for teledermatology before implementation of AI technologies and the reasons why.

- Total number of referrals through the urgent suspected skin cancer pathway (during the before-and-after implementation periods).

- Number and proportion of referrals that had appointments with a dermatologist (during the before-and-after implementation periods).

- Number and proportion of appointments with a dermatologist that resulted in a biopsy or diagnosis of a cancer lesion (for the before and after implementation periods).

- For the after-implementation period, the number and proportion of suspected cancer cases that are not eligible for assessment by AI technologies.

- The number and proportion of suspected cancer cases that are judged to be 'indeterminate' or cannot be processed by the AI technologies (technical failure and rejection rate).

Data collection should follow a predefined protocol and quality assurance processes should be put in place to ensure the integrity and consistency of data collection. See NICE's real-world evidence framework, which provides guidance on the planning, conduct, and reporting of real-world evidence studies.

Information about the technologies

Information about how the technologies were developed, the update version tested, and how the effect of future updates will be monitored should also be reported. See the NICE evidence standards framework for digital health technologies.

4 Implementation considerations

Companies should work with providers and central NHS England teams to begin the research. Planning for a prespecified period for the set-up of the technologies is

Research plan – Artificial intelligence (AI) technologies for assessing and triaging skin lesions within the urgent suspected skin cancer pathway　　　　　Page 9 of 10

August 2024

advised. During this period, training and implementation should be done before data collection is started, to account for learning effects. The following considerations around implementing the research process have been identified through working with system partners:

the technology companies should provide training for staff in using the AI-derived software.

services should be carefully selected to, when appropriate, maximise data collection for subgroups of interest.

- to assess the potential for automation bias after deployment, companies may want to track the rate of diagnostic disagreement over time.

Potential barriers to implementation include:

the availability of research funds for data collection, analysis and reporting

the availability of NHS funding to cover the costs of implementing the technologies in clinical practice

lack of expertise and staff to collect data

burden on clinical staff, including the need to have training ahead of implementation, data collection and follow up.

differences in practice between primary care settings across the NHS

- differences in skill and experience among staff when using the AI technologies.

ISBN: [to be added at publication]