

Evaluation of Deep Ensemble for the Recognition of Malignancy (DERM)

Overview

In England, 60% of dermatology referrals are made through the urgent suspected skin cancer pathway, but only 6% are converted to a confirmed case of skin cancer. This inefficiency could be addressed by safely triaging some referrals out of the suspected skin cancer pathway.

DERM is a UKCA class IIa AI technology, developed by Skin Analytics. This evaluation was commissioned by the Department of Health and Social Care (DHSC) as part of the Phase IV of the AI Awards in Health and Social Care. Skin Analytics have implemented DERM in six sites as part of the AI Awards, four of which were included in the evaluation, aiding the detection of skin cancers through triage of referrals. The sites and their associated patient numbers under relevant care models (blue for secondary care, green for community hub) were:

Chelsea and Westminster NHS Foundation Trust (CW)

4,010

University Hospitals Bristol and Weston NHS Foundation Trust (UHBW)

2,868

Ashford and St Peter's Hospitals NHS Foundation Trust (ASPH)

2,234

Birmingham PCNs / University Hospitals Birmingham NHS Foundation Trust (UHB)

537

Method



Quantitative insights

Performance analysis of DERM, subgroup analysis (health inequalities), and comparative analysis. There was a focus on safety, accuracy, effectiveness, and sustainability of DERM in deployment settings.



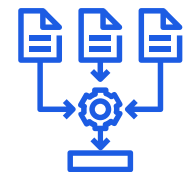
Qualitative insights

A mixed-method approach using a patient survey and interviews with patients and staff. There was a focus on safety, effectiveness, acceptability, and sustainability.



Machine learning principles review

Assessing whether best practice standards in machine learning were followed in developing the DERM algorithm. A focus on effectiveness, safety, accuracy, and sustainability.



Health economic modelling

Three models were developed: a cost-utility analysis (CUA), a cost-benefit analysis (CBA), and a budget impact model (BIM). There was a focus on value, effectiveness, and sustainability.

Quantitative study

In secondary care, DERM assessed 5,678 lesions out of 4,159 referred, yielding high pathway sensitivities for melanoma across all sites (higher than 90%), achieving its target rates and showing that DERM was effective in channelling high risk lesions to the appropriate management outcome.

In the community hub, DERM was correctly assigning pathways for lesions based on risk, with high pathway sensitivities, but with a wider confidence interval than seen in secondary care due to the smaller sample size.

Sensitivity
secondary care:

97.4%

Secondary care
conversion rate:

12.6%

Community hub
conversion rate:

18.2%

Community hub
discharge rate:

60.1%

Second read impact

36%

overturn rate



41%

overturned cases discharged
following trust review

The second read, involving the review (by a dermatologist consultant) of cases classified as low risk by DERM, is risk averse in its decision making. It overturned 36% (n = 754) of potential DERM discharges, of which 41% (n = 307) were discharged by trust dermatologists.

Across the evaluation, seven cancers were identified by the second read (out of 754 overturned discharge cases). The majority of these were low risk basal cell carcinomas (BCC).

There was no evidence of incorrect discharge by the second read, through patients returning within six months.

Subgroup analysis

While exploratory, results show that pathway sensitivities were similar between subgroups such as age and Fitzpatrick skin type. There was no indication of performance varying with respect to the likely distribution of cases according to deprivation.

Qualitative study

Patient survey results

Findings suggest that patient perceptions of AI-enabled teledermatology services were largely positive.

85%

rated AI-enabled teledermatology services as good or very good, suggesting a high level of acceptability

13%

were uncomfortable about the AI-enabled teledermatology being used to help determine their diagnosis

67%

acknowledged the value of using AI to speed up getting an appointment rather than waiting to be seen by a doctor

Interview results

The importance of appropriate referrals, high quality pictures, and a supported trained workforce were noted as key factors for a successful deployment.

Staff reported that teledermatology services using DERM had a 'transformational' effect on capacity. It was unclear whether this could be attributed specifically to DERM rather than teledermatology alone.

Both staff and patients were reassured by the use of the second read, some staff would not be in favour of removing it at this stage.



Machine learning principles review

The review reported DERM performing binary classification per lesion and achieved strong results, from data handling, image lesion distribution, sensitivity and perturbation analyses, and corner case classification.

Algorithm development was deemed reusable and reproducible, which is encouraging when considering implementation in future sites.

The existing teledermatology infrastructure was highlighted as an enabler to DERM deployment.

There was a need to ensure that cultural factors are appropriately considered.

Dermatologists required for a national scale-up of DERM with the second read should be considered, given the national workforce shortages.

Cost-benefit analysis

The cost-benefit analysis modelled the impact of teledermatology with and without DERM on skin cancer referrals and compared each against a non-digital face-to-face pathway. All instances highlighted savings for the NHS; DERM saved more dermatologist time than traditional teledermatology.



(*) Based on 4,320 suspected skin cancer referrals per year



Teledermatology

Potential implementation:
£2.5 returned for every £1 spent. Could increase to £3.7 in more mature sites.



DERM in secondary care

Actual implementation:
£1.7 returned for every £1 spent. Could increase to £2.0 if the second read is removed.



DERM in primary care

Actual implementation:
£1.1 returned for every £1 spent. Could decrease to £1.0 if the second read is removed.

Cost-utility analysis

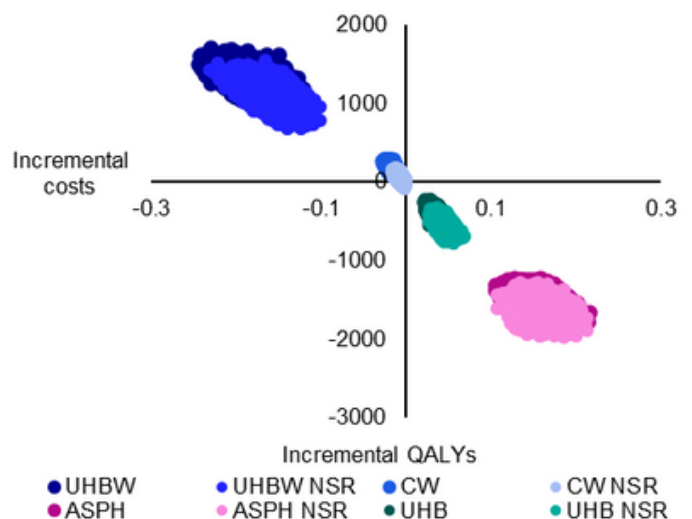
Results for each site were generated separately. They included the pathway with the second read, and with no second read (NSR).

The CUA showed varied outputs between sites, with UHBW and CW showing a higher cost and worse health effect, and UHB and ASPH showing lower cost and better health effect.

Removal of the second read resulted in slightly better cost and health effects at all sites.

The study remained broadly inconclusive due to inconsistencies in the data, overshadowing the relatively small quality of life impact of DERM that could be quantified.

Probabilistic sensitivity analysis showing variation in outputs



Recommendations



Collection and obtention of more complete site-level baseline data to improve robustness and conclusiveness of results.



Repeating the subset analyses with a larger dataset, and obtaining data on the performance of current pathways to treat patients with darker skin tones.



Future rollout studies to evaluate in detail how the technology is applied in different settings, to identify what the most effective model would be.



Further analysis regarding adherence, focusing on those patients who declined teledermatology.

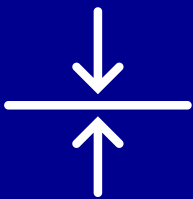


Investigation of the implementational and operational factors enabling DERM deployment, including infrastructure, workforce, and usability.



Further qualitative work regarding removal of the second read to understand perceptions of patients and staff in relation to clinical confidence and decision-making.

Limitations



Baseline data

Direct comparative data at a granular level could not be obtained, making inferences about the effect of DERM over and above existing pathways difficult to answer.



Data quality

Reliability of the aggregated comparative data skewed the performance of traditional teledermatology models towards best-in-class sites, likely inflating the return on investment.



Data availability

There was limited data available for some measures for example, impact of missed cancers, and quality of life data for outcomes such as anxiety.



Observable cases

Relatively few cases/lesions were observed in the limited deployment period at UHB, meaning the assessment of the DERM community hub model has been limited.

Conclusion

Findings suggest DERM performed accurately and safely across secondary care and community hub models, though the latter requires a larger sample size for validation. Unnecessary referrals were reduced and consultant time released due to DERM implementation.

The qualitative study found most participants suggested that the DERM service was very good or good. Staff interviews highlighted the platform's user-friendliness, efficiency, and the ability to discharge patients with benign lesions at the triage step as the main benefits of DERM. Both staff and patients were reassured by the use of the second read and some staff suggested they would not be in favour of removing it at this stage.

The CBA revealed cost savings with DERM over face-to-face pathways in secondary care, with modest savings in the community hub model. Findings from the CUA showed varied and mostly inconclusive results.