Post market surveillance of AIaMD: A medical algorithmic audit of a skin cancer assessment and triage tool

Dr Qasim Malik, Dr Aditya U Kale, Dr Sonam Vadera, Dr Jeffry Hogg, Ms Jaspret Gill, Prof Alastair K Denniston and Dr Xiaoxuan Liu

Introduction

The Medical Algorithmic Audit (MAA) is a safety monitoring framework for algorithmic error auditing and failure mode detection for artificial intelligence as medical devices, it is part of the post market surveillance safety auditing of Artificial Intelligence as medical devices (AIaMD).

This poster documents the method and results of undertaking a MAA. Here we describe the team's experience of working at national and local level to explore how the MAA could support a safer and more equitable approach to AI implementation in the national health service.

As part of the AI4VBH Fellowship our role was to assisst in the MAA including the failure models and effects analysis.

Methods

The MAA framework has been adopted by *University Hospitals Birmingham*, one of the Uk's largest health providers. The hospital has introduced the use of a deep learning algorithm for classification of skin cancers (DERM™), including assessment and triage.

MAA was applied to this deployment with the standard pathway and AI pathway mapped below in figure 1. All stages of the MAA included collaboration with the manufacturer, clinical specialists, informaticians and hospital's AI implementation team. The MAA included:

 1) Scoping of regulatory documentation for DERM,
2) Mapping of failure modes and effects analysis(FMEA)
3) Collection of artefacts including evidence from previous evaluations,
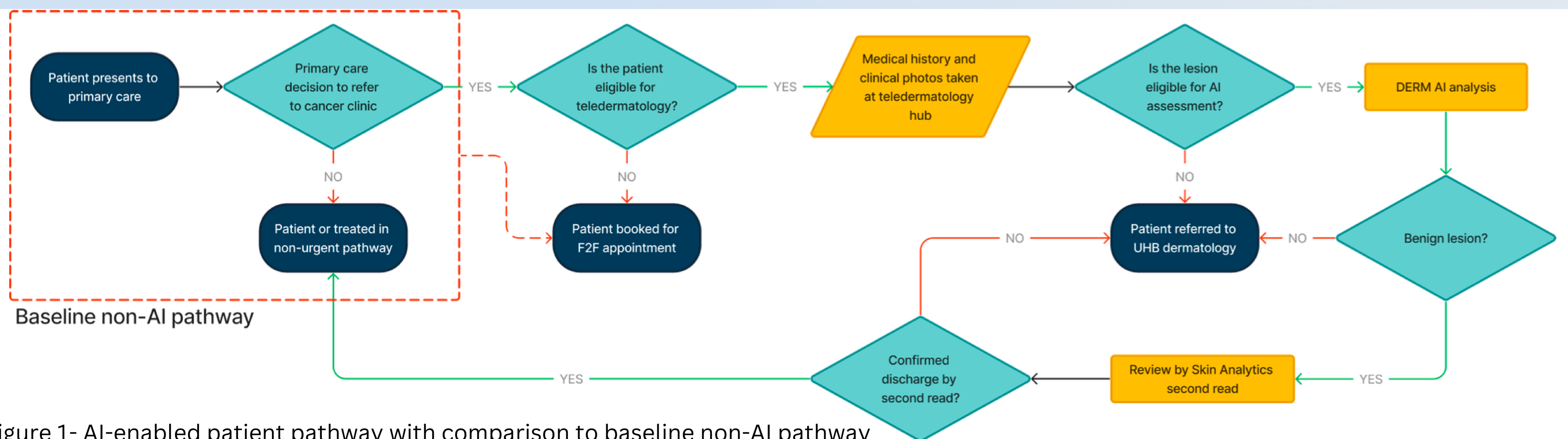4) Testing including exploratory error analysis, and 5) Reflection with key stakeholders.

Figure 1- AI-enabled patient pathway with comparison to baseline non-AI pathway

Results

Local Results
This audit examines **9783 cases**, evaluated through the skin cancer pathway, of which **7439** were assessed by AI. Algorithmic performance is shown in Table 1.

- 22 false negative results were identified using histology ground truth
- Error auditing identified that three lesions should have been excluded from AI evaluation, demonstrating that there may be some level of unintended misuse
- Furthermore 6 cancers could not have been diagnosed without clinical history highlighting the challenges associated with the use of isolated AI assisted diagnostic imaging
- The ground truth for the all lesion sensitivity is clinicial opinion and histology (** The overall specificity of DERM for benign lesions is 70.47% 95%CI [68.98, 71.93])

The MAA demonstrated that although the AI is operating above the performance targets agreed by key stakeholders. The maintenance of an error log to identify failure modes is essential.

Lesion type	Sensitivity	95% CI
Malignant lesions	97.26%	[95.88, 98.28]
Premalignant lesions	91.11%	[89.08, 92.86]
Malignant and premalignant lesions	93.97%	[92.74, 95.05]
All lesions*	81.28%*	[79.90, 82.60]

Table 1: Performance metrics for DERM (sensitivity and specificity**)

Discussion

The increasing deployment of AIaMD provides a need for continual monitoring and improvement of AI enabled clinical pathways. The MAA provides a framework for understanding key risks through a systems based approach, with prioritization of risk and auditing of errors. There is potential for the use of the MAA for assessing safety and cost efficacy and equity too.
 The findings from this MAA were presented locally and nationally to stakeholder groups with senior hospital leaders using this data to develop and plan local governance process for the surveillance of AIaMD.