Summative validation and usability testing of the MIDI tool – path to UKCA certification

Dr Michael Adofo Kwakye

Introduction – What is the clinical problem?

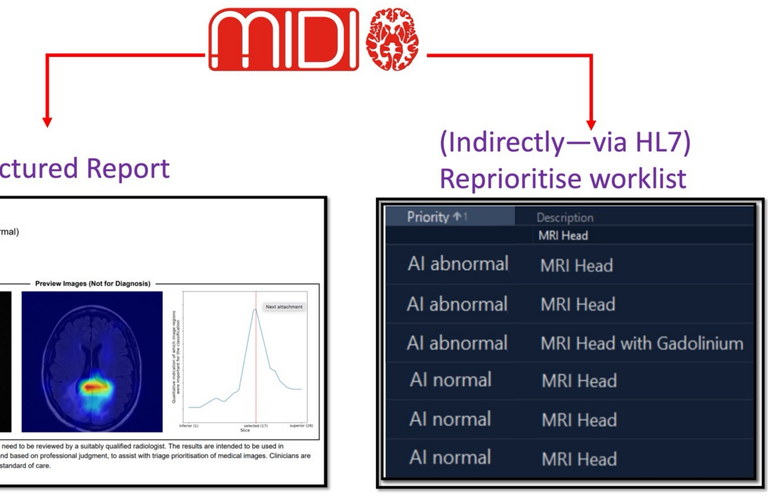
There is an increasing number of MRI Brain studies being performed across the NHS with a 9% year on year increase. Without a commensurate increase in reporting capacity, many of these studies do not meet their 28-day reporting target, and an increasing number are being outsourced, which costs NHS Trusts significantly, and produces reports of variable quality. The clinical problem is that **time critical abnormalities not identified expediently may lead to worse patient outcomes and increased healthcare costs.**

MIDI - The AI solution

Developed at King's College London, MIDI is a brain MRI abnormality detection AI tool. It classifies MRI head studies as either routine or abnormal according to a set of predefined actionable findings that could trigger downstream clinical intervention.

The tool outputs a structured PDF report comprising three elements: an abnormality score, an image/heatmap of the abnormality, and a graphical representation of its location within the MRI series (see image). Simultaneously, an HL7 message is sent back to the RIS that allows reprioritisation of the reporting worklist to reposition abnormal studies to the top for faster reporting.

MIDI has been trained on retrospective data using >70K head MRIs from King's Health Partners. (AUC 0.948). Prospective validation using dataset from external hospitals and a holdout internal dataset from King's College Hospital showed a small drop in accuracy (AUC 0.874).



Aims and Methods

As part of gaining UKCA Class 1 certification, MIDI underwent a summative validation and usability testing to assess technical and clinical interoperability, connectivity, and overall tool performance — including system design inputs and technical requirements defined in the Quality Management System. Testing was carried out in an environment as close to the intended use conditions as possible, with the operating point calibrated at 0.5.

A total of 100 human-labelled, previously unprocessed retrospective MRI head studies were identified and transferred to MIDI for processing. These included studies MIDI was not expected to process — e.g. paediatric, non-MRI, and low-resolution studies — as an operational check that MIDI triggered only on appropriate studies. Of the 100 studies, 75 were expected to be processed and 25 rejected. Logs from the vendor neutral platform (VNP), HL7 message logs, and structured PDF summary documents provided evidence for validation of system design inputs. Tool accuracy and reliability were analysed for all processed studies.

For usability testing, 15 radiologists watched a background video on MIDI before interacting with its output and providing feedback via a Likert-scale questionnaire.

Results

Of the 100 studies submitted, MIDI processed 80 — each generating a structured report and HL7 message — against an expected 75. MIDI correctly rejected 20 studies: those lacking T2/FLAIR sequences (n=5), CT brain studies (n=5), non-brain MRI studies (n=5), and low-resolution MRI brain studies (n=5). However, MIDI incorrectly processed paediatric studies sent to the VNP (n=5), failing this system input validation test. A bug fix was deployed in version 1.2.9, after which no paediatric studies were processed. Overall, mean MIDI processing time was 40 seconds for the 75 studies that were expected to be processed (95% CI 35.28 – 45.66).

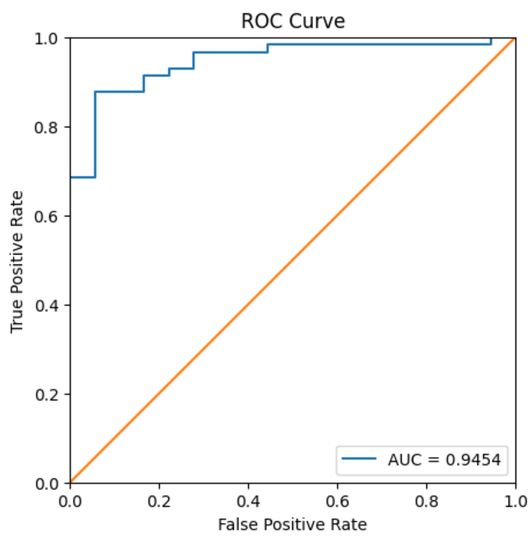
5/75 studies took longer than 90 seconds to process. The longest processing time was 153 seconds.

<u>Performance Evaluation</u>

AUC was calculated for the 75 appropriately processed studies as 0.9454. A confusion matrix and ROC curve have been included (Table 1 and Figure 1 respectively) along with data on sensitivity, specificity, PPV and NPV measurements.

		MIDI predicted values	
		Positive	Negative
Ground Truth	Positive	55	2
	Negative	8	10

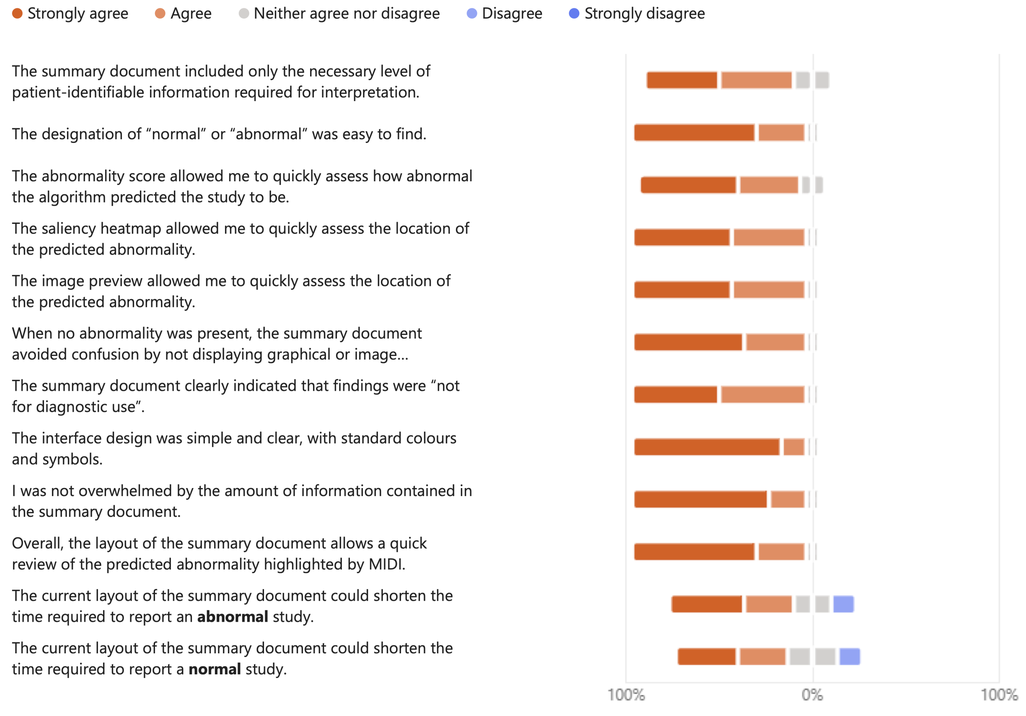
Sensitivity	0.965
Specificity	0.159
PPV	0.873
NPV	0.833



<u>MIDI output review - Usability</u>

A summary of user responses are included in Figure 2. No potentially unacceptable unforeseen difficulties were highlighted by clinicians after interacting with MIDI output, which is reassuring. In addition, clinicians understood and were satisfied with the information contained in the summary document.

5. **Task 4:** MIDI Structured Report Usability Questionnaire (https://tinyurl.com/MIDI-Reports)



Next Steps – TRIAGE Study

Having gained UKCA Class 1 certification, the next step is the TRIAGE study — a cluster randomised crossover trial across at least 12 sites, each randomised to an "MIDI on" period followed by washout and "MIDI off" (or vice versa). The primary aim is to evaluate MIDI's real-world clinical utility, specifically whether it reduces turnaround times for abnormal brain MRI scans. Seven secondary objectives examine cost-effectiveness, staff impact, safety, diagnostic accuracy, and patient outcomes in conditions including stroke and brain tumours.

	Run in Period		Phase I of Intervention: ON or OFF							Wash Out Period & Cross Over					Phase II of Intervention: ON or OFF							
	m1	m2	m1	m2	m3	m4	m5	m6	m7	m1	m2	m3	m4	m5	m1	m2	m3	m4	m5	m6	m7	
Hospital 1	X	X	Intervention: MIDI ON							Y*	X	X	X	X	Control: MIDI OFF							Y*
Hospital 2	X	X	Control: MIDI OFF							Y*	X	X	X	X	Intervention: MIDI ON							Y*

Fellowship in Clinical AI
NHS Digital Academy Programme