IDENTIFICATION OF ABNORMAL MRI HEAD SCANS USING AI - A REAL WORLD FEASIBILITY STUDY

NHS Digital Academy

NHS England

Evaluating the feasibility, performance, usability, and secondary advantages of utilizing an AI tool for identifying abnormal MRI brain scans.

Dr Nikunj Davda

INTRODUCTION

The problem our group is looking to solve is the long and ever-growing waiting list of MRI Head scans waiting to be reported. Delays in reporting abnormal scans leads to delays in diagnosis, leading to poorer patient outcomes.

A possible solution is using AI software, that our group has developed, called **MIDI (MR Imaging deep learning identification)**. MIDI is a deep learning tool that labels MRI Head scans as either being normal or abnormal and could therefore accordingly triage abnormal scans to the top of the queue of scans waiting to be reported.

THE AI TOOL

The MIDI tool was developed using a neuroradiology report classifier trained by applying a large language model to MRI head scan reports, enabling classification of those reports (and their associated scans) as normal or abnormal. These labeled scans were then used to train a convolutional neural network for detecting abnormalities in new scans (see Figure 1).

The tool generates a document with:

- an axial image slice identified as most abnormal;
- an abnormality heat-map;
- a graph displaying abnormality levels for each slice;
- and an overall abnormality score for the study (Figure 2).

Based on the abnormality score threshold, scans can then be categorised as normal or abnormal.

A simulation study demonstrated that using the tool can cut the time from image acquisition to radiologist report generation by half.

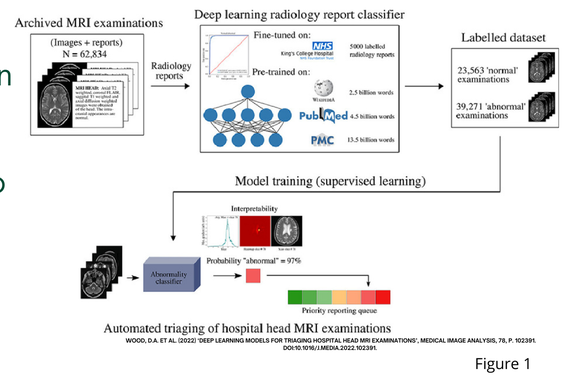

Figure 1

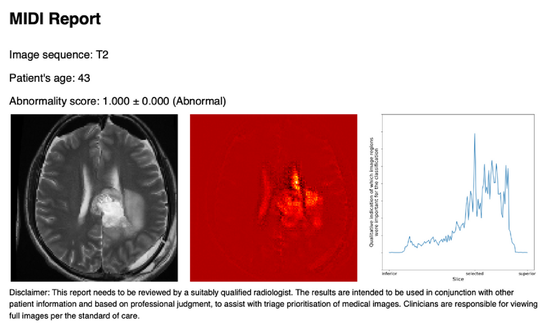

Figure 2

METHODOLOGY

1. For Aim 1, I worked with the PACS provider and the company behind AIDE to set up the pipeline and workflow shown in a figure 3 to test the feasibility of integrating the tool into a real clinical workflow. Feasibility will be confirmed by sending test scans from each site to AIDE manually and ensuring that the AI output PDF (showd in figure 2), is received back on PACS.

2. For Aim 2, more than 300 scans from each site (labeled by a radiologist as normal or abnormal) will be sent from PACS to AIDE. The scans will be processed by MIDI to determine an abnormality score, generate an output PDF, and send it back to PACS. The abnormality score will be used to label the scan (normal/abnormal) based on a threshold. The accuracy of the AI tool will be assessed by comparing the AI label to the human label.

3. For aim 3, to evaluate whether the AI tool reduced the time taken to read and report scans, I enlisted the help of 10 consultant neuro-radiologists. They were asked to report on 20 scans, each scan being reported twice: once with the AI information and once without it. Each radiologist received a list of 20 scans (similar to that on the right in figure 4), with 10 scans containing an AI label and accompanied by the AI output document (figure 2). The order of the scans was arranged so that the 10 AI scans presented the abnormal scans at the top of the list, simulating the tool's future function of prioritising abnormal scans. The other 10 scans had no AI information. Radiologists were instructed to report the scans as they normally would in a clinical setting, following the same procedures and conditions. After a two-month break, the same set of scans would be re-reported, but this time the AI-accompanied and non-AI-accompanied scans would be swapped. The radiolgists would use a proforma to note the time they started and finished reporting a scan. At the end of the study, the reporting times with and without AI can therefore be compared.

With AI (these are made up examples)
accession number RJZ134343949. Abnormal
accession number RJZ144338299. Abnormal
accession number RJZ134343949. Abnormal
accession number RJZ144338299. Abnormal
accession number RJZ134343349. Abnormal
accession number RJZ144118299. Normal
accession number RJZ134343949. Normal
accession number RJZ148938299. Normal
accession number RJZ123443949. Normal
accession number RJZ144338343. Normal

Without AI (these are made up examples)
accession number RJZ134343949. No category
accession number RJZ144338299. No category
accession number RJZ134343949. No category
accession number RJZ144338299. No category
accession number RJZ134343349. No category
accession number RJZ144118299. No category
accession number RJZ134343949. No category
accession number RJZ148938299. No category
accession number RJZ123443949. No category
accession number RJZ144338343. No category

Figure 4

AIMS AND DEPLOYMENT

The **aims** of the study were to determine if the tool could:

1. Work as part of a real clinical pathway (i.e. whether it is feasible).
2. Perform accuracuately compared to a human labelled ground truth
3. Reduce the time taken to read and report a scan
4. Is helpful and easy to use.

I designed a clinical workflow that integrated the tool into a real clinical pathway safely in shadow mode to avoid any impact on real patient care. This integration involved deploying the tool at the three participating trusts: Guys and St Thomas' (GSTT); Kings College Hospital (KCH); and East Kent (EK), working with their PACS (picture archiving and communicating system) departments and their PACS providers. It also involved co-ordination with a third party company that developed AIDE (AI deployment engine) which is a platform onto which AI tools can be onboarded (like an app store) and which has terminals at each participating site, allowing the tool to be onboarded to AIDE centrally but still operate within trust firewalls. This workflow is summarised below (figure 3).

Study sites and aims:

Guy's and St Thomas' NHS Foundation Trust
King's College Hospital NHS Foundation Trust
East Kent Hospitals University NHS Foundation Trust

Aims:

1. Determine the technical feasibility of the tool
2. Assess performance
3. Assess reporting time - (human vs AI+human)
4. Obtain usability metrics

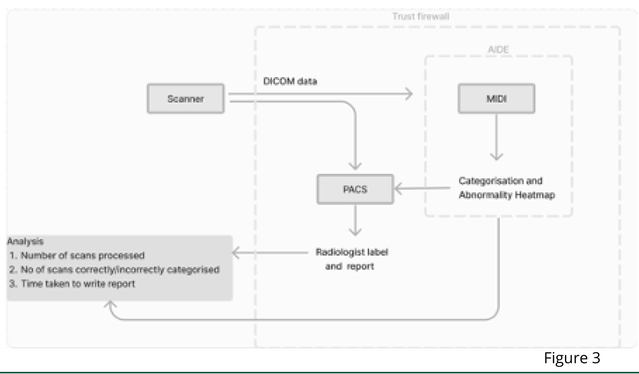

Figure 3

4. For Aim 4 (testing usbility), I designed a mixed methods usability study. To obtain structured data, I created a questionnaire which was designed and structured according to international usability testing standards (e.g. ISO 9241) and industry best practices to test efficacy, efficiency and satisfaction. I performed a literature search on AI usability studies in healthcare and extracted additional useful questions from similar usability studies. The questionnaire is accompanied with semi-structured interviews to gain further, more open-ended feedback. This usability sub-study runs on completion of the multi-reader study described in the previous paragraph.

PROGRESS AND NEXT STEPS

Progress update as of June 7, 2024 for each part of the study:

- **Aim 1**: Scans are now successfully being sent from PACS to AIDE and back to PACS with the desired AI document. However, more work is needed to display the AI output in a user-friendly way for radiologists on PACS before confirming feasibility. Progress is close to completion, with numerous technical interoperability issues addressed.
- **Aim 2**: Only a few scans have gone through the workflow and received an AI-labelled abnormality score. Performance evaluation will occur once all scans are processed.
- **Aim 3**: The multi-reader reporting study is ongoing, with initial results from some participating radiologists received for the first reporting round.
- **Aim 4**: A literature search on usability testing for healthcare AI tools was conducted, and a mixed methods usability study design, including a questionnaire and interviews, has been developed. This study will start after the completion of the multi-reader study.

CONTRUBUTORS

- **Study PI:** Dr Thomas Booth
- **Consultant neuro-radiologists:** Sina Kafiabadi, Berna Aygun, Abdalla Gehad, Emily Guilhem, Mark MacDonald

With thanks to the Fellowship in Clinical Artificial intelligence programme team. Contributions from other members of the host research team at King's College London, other AI fellows and programme managers, in addition to knowledge gained from monthly workshops was invaluable in successfully designing and running this study.