

REAL-WORLD EVALUATION OF ARTIFICIAL INTELLIGENCE TOOLS IN RADIOLOGY

FELLOW — John L Chen

SUPERVISOR — Tilak Das

COLLABORATORS — Samia Nesar, Mateus A Esmeraldo, Fulvio Zaccagna, Abhishekh Ashok

INSTITUTION — Cambridge University Hospitals NHS Foundation Trust

INTRODUCTION

- There has been rapid development of artificial intelligence (AI) tools within (neuro)radiology, many of which are now commercially available
- It is critical that these are clinically evaluated, to ensure their safe and effective implementation, but there is a current paucity of real-world evaluation data

CLINICAL PROBLEM

- Multiple sclerosis (MS) is the most common disabling neurological disorder affecting young adults and is estimated to affect ~150,000 people in the UK (~3 million worldwide)
- Patients with MS are usually assessed with serial brain MRI studies to evaluate for any change in burden of demyelinating lesions, which can signify active versus inactive disease
- MS brain lesions can be numerous, and their accurate detection and characterisation can be challenging and time-consuming

ARTIFICIAL INTELLIGENCE SOLUTION

- Pixyl.Neuro.MS is a MDR Class IIa tool that has recently been deployed at our institution, aiming to enhance MS lesion detection and characterisation, and improve radiological reporting efficiency
- It can rapidly analyse MRI brain volumetric fluid-attenuated T2-weighted sequences to identify white matter lesions and classify these as stable, enlarging, or new, compared to previous studies

PROGRESS & MILESTONES

- Performed 3-month retrospective review of Pixyl.Neuro.MS at lesion level and case level
- Ground truth: Pixyl.Neuro.MS/consultant reports if concordant; researcher review if not
- 43 case-pairs (age 17–72, mean 42; male : female = 1 : 1.39; mean interval 16.5 months)
- Pixyl.Neuro.MS identified total of 2697 stable lesions across all patients and 75 new lesions across 18 patients (41.9%), of which 46 (61.3%) lesions were correctly identified
- Remaining 29 (38.7%), across 10 patients (23.3%), misidentified, which in 3 patients (7.0%), would have resulted in erroneous classification as active disease
- Pixyl.Neuro.MS failed to identify 1 new brainstem lesion in each of 2 patients (4.7%), which in 1 (2.3%) would have resulted in erroneous classification as inactive disease
- Pixyl.Neuro.MS identified 1 new lesion in a patient, not reported by neuroradiologist
- Promising tool for identifying new lesions (signifying active disease) or for confirming no new lesions (signifying inactive disease), with good performance at the case level (90.7% accuracy) and could identify new lesions not identified by a neuroradiologist
- However, did occasionally produce false positive and false negative results at the case level, and careful neuroradiologist scrutiny is still necessary
- Results presented at the *NHS & RCR Global AI Conference 2025 (London)*

Pixyl.Neuro.MS Longitudinal report

John Doe | M | 42yo | female | MS | 1877
 Pixl Date: May 22, 2024, New Pixl Date: Jan 26, 2024

Quality Control: ☒ Pass ☐ Observations

Disease Activity		T2 FLAIR lesions		T1 lesions with contrast	
New	Stable	Enlarging	Stable	Enlarging	Stable
0	0	0	0	0	0

Lesion Load

Volume (ml)	Change (ml)	Lesion count *
20.64	0.95	< 1
0.15	0.43	< 1
1.44	1.07	< 1
0.12	0.06	< 1
34.38	15.63	< 9 **

* New lesions identified in the 2024 dataset only. ** Lesions identified in the 2024 dataset only.

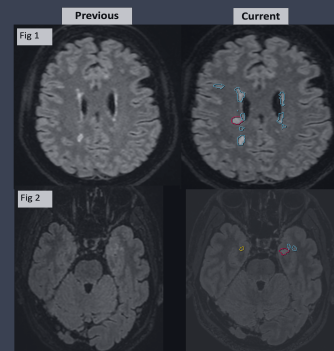


Figure 1: Pixyl.Neuro.MS correctly identifying a new lesion. Lesions classified by colour outline: new (red); stable (blue); enlarging (yellow).

Figure 2: Pixyl.Neuro.MS incorrectly identifying a new lesion.

Confusion Matrix for Pixyl.Neuro.MS Case-Level Prediction

Actual Label	Predicted Label	
	Active Disease	Inactive Disease
Active Disease	True Positives 15	False Negatives 1
Inactive Disease	False Positives 3	True Negatives 24

	Pixyl.Neuro.MS
Case-Level	
Sensitivity	93.4%
Specificity	88.9%
Positive Predictive	83.3%
Negative	96.0%
Accuracy	90.7%

NEXT STEPS

- Prospective study of Pixyl.Neuro.MS accuracy, comparing Pixyl.Neuro.MS alone vs. consultant neuroradiologist alone vs. both together, at both lesion level and case level
- Pre-implementation and post-implementation surveys of (neuro)radiologists (of various grades), regarding the perceived utility, benefits, and drawbacks of Pixyl.Neuro.MS
- Investigating impact on reporting efficiency, comparing reporting time and report turnaround time with and without Pixyl.Neuro.MS support

OTHER WORK

- Evaluation of multiple AI tools for detection of acute ischaemic stroke on non-enhanced CT and large vessel occlusion on CT angiography (paucity of real-world comparative evidence and understanding of reasons underlying differing performance)
- Evaluation of further AI tools within department (e.g. for multiple findings on CT head; pulmonary nodules on CT thorax)
- Development of general framework for evaluation of any clinical AI tool, informed by published guidance and encompassing multiple domains (strategic, implementation, technical, clinical, system, economic, reporting, monitoring)

REFERENCES

- Pixyl.Neuro.MS. <https://pixyl.ai/multiple-sclerosis/>
- Mendelsohn, Z. et al. Commercial volumetric MRI reporting tools in multiple sclerosis: a systematic review of the evidence. 2023. *Neuroradiology*, 65(1), 5-24.
- Roca, P. et al. Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. 2020. *Diagnostic and Interventional Imaging*, 101(12), 795-802.
- Giovannoni, G. et al. "Is it time to target no evident disease activity (NEDA) in multiple sclerosis?." *Multiple Sclerosis and Related Disorders* 4.4 (2015): 329-333.

ACKNOWLEDGEMENTS

- NHS Fellowship in Clinical AI
- NHS Health Education England — East of England