A NATURAL LANGUAGE PROCESSING ALGORITHM TO IDENTIFY STROKE IN BRAIN IMAGING REPORTS ON A LARGE SCALE

W. Whiteley[1], C. Grover[2], B. Alex[2], C. Sudlow[1], G. Mair[1].
[1]University of Edinburgh, Centre for Clinical Brain Sciences, Edinburgh, United Kingdom.
[2]University of Edinburgh, School of Informatics, Edinburgh, United Kingdom.

Background

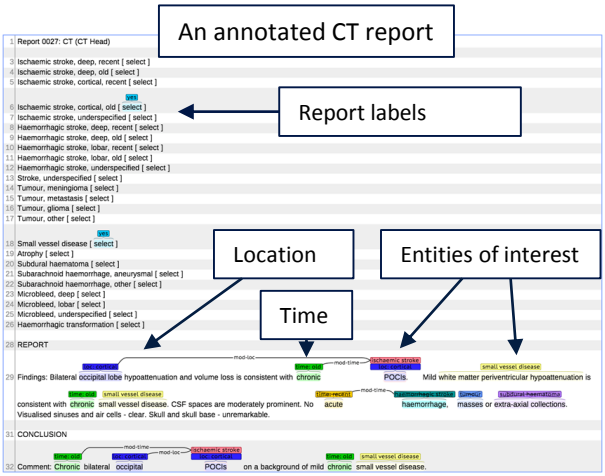
Large datasets of brain imaging (>1million) are collected in routine practice in Scotland. However, analysis of such data-sets is difficult.

We developed an automated system to read brain-imaging reports, to add validity to stroke diagnoses in electronic health records.

Methods

We obtained anonymised brain-imaging reports from the Edinburgh Stroke Study and NHS Tayside. Iteratively, we developed a rule-based natural language processing (NLP) system to identify stroke and its subtypes in radiologists' reports of brain CT and MR imaging.

We measured system positive predictive value (PPV) and sensitivity of the system by comparing system output with neurologist and neuro-radiologist ('expert') reading of unseen test data.



Results

We annotated 364 brain imaging reports to develop the NLP system. We used a different set of 266 annotated brain imaging reports to test the NLP system.

Label	PPV*	Sensitivity†
Ischaemic stroke		
Deep	98%	98%
Cortical	97%	96%
Recent	90%	100%
Old	98%	97%
Haemorrhagic stroke		
Deep	81%	100%
Cortical	86%	95%
Recent	78%	100%
Old	84%	85%

*PPV: positive predictive value, the % of labels identified by the system that agree with the expert reading;
† % of reports with an expert-identified labels that the NLP system also identifies

Conclusion

It is possible to automate the identification of ischaemic stroke age and location on brain imaging reports with NLP. The identification of haemorrhage is sensitive, though has a poorer positive predictive value

e: william.whiteley@ed.ac.uk
@edinburghstroke
m. +44-7811113610

THE UNIVERSITY of EDINBURGH

CCBS Centre for Clinical Brain Sciences