Natural language processing for stroke diagnostic imaging characterization

Brian Bursic, MEng, MD(c)[1], Andrew Cao, MD[1], Jinyue Feng, MSc, PhD(c)[2], Fahad Razak MD, MSc[1,3], Amol Verma, MD, MPhil[1,3], Moira Kapral, MD, MSc[1], Frank Rudzicz, MEng, PhD[2,4], Amy Y.X. Yu, MD, MSc[1]
Temerty Faculty of Medicine, University of Toronto[1]; Department of Computer Science, University of Toronto[2]; St. Michael's Hospital, Unity Health[3]; Faculty of Computer Science, Dalhousie University[4]

Background.

Despite significant advancements in therapeutics, stroke remains a leading cause of death and disability globally[1]. Large-scale studies evaluating stroke incidence and attributes are essential for disease surveillance, healthcare resource planning and patient care[2]. Neuroimaging reports contain important narrative data on clinically relevant stroke attributes, but harnessing this data on a macro level is limited by the laborious nature of manual chart review. Natural language processing (NLP) is a robust and flexible machine learning subdomain which has been used in healthcare settings to extract important information from unstructured radiology reports[3]. This study applies NLP methods to efficiently and accurately identify prespecified stroke attributes in unstructured neuroimaging reports of patients diagnosed with stroke. Leveraging the wealth of health data contained in these reports on a population level will allow for disease surveillance with increasing granularity and aid in future clinical decision making and research.

Methods.

Chart Selection: Full, free-text computed tomography angiography (CTA) neuroimaging reports were obtained of all patients admitted to a comprehensive stroke centre between 2017 and 2019.

Manual Chart Abstraction: Manual chart abstraction was completed on 1320 diagnostic imaging reports. Prespecified stroke attributes were tracked using a data collection form. The primary outcome was the presence of proximal large vessel occlusion defined as occlusion in the M1 segment of the middle cerebral artery (MCA) or A1 segment of the anterior cerebral artery (ACA). Secondary outcomes of interest included (1) Alberta stroke program early CT score (ASPECTS), (2) distal intracranial vessel occlusion, (3) presence of hemorrhage, (4) qualitative measure of collateral status, (5) various other intracranial vessel occlusions.

Model Training and Development: We trained machine learning classifiers to identify the stipulated primary and secondary outcomes. The dataset was largely made of rare, continuous outcomes of both binary and categorical types. Given the structured nature of radiology reports, we proposed to apply domain knowledge and develop sets of keywords associated with each outcome (for example, "M1" and "A1" for proximal large vessel occlusion). The keywords were applied as parts of regular expressions to anchor the most relevant sentences to each outcome. These predefined anchor sentences were then passed into a deep neural network with pretrained clinical word embeddings (ClinicalBERT). Token-level linear attention computed weighted representations of sentences were then fed into a linear layer with GELU activation, drop-out, and batchnorm. A final prediction layer used softmax before the model was trained using cross-entropy loss. This approach was compared against conventional NLP strategies including term frequency-inverse document frequency (TF-IDF) and bag-of-words (BoW).

Results.

Among the 1320 consecutive diagnostic imaging reports manually reviewed, chart abstractors identified large vessel occlusions in 161 (12.2%) reports, intracranial hemorrhage in 139 (10.5%) reports, and a reported ASPECTS score in 384 (29.1%) reports. Our deep neural network with ClinicalBERT currently achieves similar accuracy to standard approaches including TF-IDF and BoW for most studied outcomes (proximal intracranial vessel occlusion>0.97, distal intracranial vessel occlusion>0.89, presence of hemorrhage>0.95, collateral status>0.82, vertebral artery occlusion>0.97, basilar artery occlusion>0.98). Notably, the ASPECTS outcome achieved >0.98 accuracy using standard approaches.

Conclusion.

Our study and preliminary results demonstrate a robust NLP model for characterizing relevant clinical variables in the monitoring and treatment of acute ischemic stroke from free text radiology reports. Beyond fine tuning hyperparameters to achieve accuracy acceptable for the clinical setting (>0.95), next steps include integrating the individual predicted primary results into a model that simultaneously predicts all outcomes within a single CTA report and developing an end-to-end deep learning system that does not require definition of keywords. We will also validate the generalizability of our model on a held-out dataset from a separate regional stroke center. Our NLP methods have implications for stroke disease surveillance at a population level. The subsequent translation of this knowledge to the clinical setting will inform clinical risk prediction models and clinical decision making in acute stroke treatment.

1. Feigin VL, Stark BA, Johnson CO, Roth GA, Bisignano C, Abady GG, Abbasifard M, Abbasi-Kangevari M, Abd-Allah F, Abedi V, Abualhasan A. Global, regional, and national burden of stroke and its risk factors, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. The Lancet Neurology. 2021 Oct 1;20(10):795-820.
2. Ung D, Kim J, Thrift AG, Cadilhac DA, Andrew NE, Sundararajan V, Kapral MK, Reeves M, Kilkenny MF. Promising use of big data to increase the efficiency and comprehensiveness of stroke outcomes research. Stroke. 2019 May;50(5):1302-9.
3. Verma AA, Masoom H, Pou-Prom C, Shin S, Guerzhoy M, Fralick M, Mamdani M, Razak F. Developing and validating natural language processing algorithms for radiology reports compared to ICD-10 codes for identifying venous thromboembolism in hospitalized medical patients. Thrombosis research. 2022 Jan 1;209:51-8.