Abstract citation ID: aakag023.1020

ABSTRACT NUMBER: ESOC2026A2214

COMBINING A LARGE LANGUAGE MODEL WITH TRADITIONAL SOFTWARE ENGINEERING TOOLS FOR AUTOMATED RES-Q REGISTRY VARIABLE EXTRACTION

Gertrūda Kaubrytė[1], Agnė Ulytė[2], Rytis Klimovič[3], Andrius Maslovas[3], Robert Mikulík[4], Dalius Jatuzis[5], Rytis Masiliunas[6]

[1]*Faculty of Medicine, Vilnius University, Vilnius, Lithuania*

[2]*Department of Quality Management, Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania*

[3]*Center of Informatics and Development, Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania*

[4]*Stroke Research Program, International Clinical Research Center, St Anne's University Hospital, Department of Neurology, Krajská nemocnice T. Bati, Health Management Institute, Brno, Czech Republic*

[5]*Clinic of Neurology and Neurosurgery, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University, Vilnius, Lithuania*

[6]*Department of Quality Management, Vilnius University Hospital Santaros Klinikos, Clinic of Neurology and Neurosurgery, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University, Vilnius, Lithuania*

Background and aims: Despite substantial advances in digital health data infrastructure, quality improvement efforts still rely on resource-intensive manual data collection. Using a combination of a large language model (LLM) and conventional programming

tools, we developed and evaluated the accuracy of a pilot algorithm that enables automated data extraction of the international Registry of Stroke Care Quality (RES-Q) variables.

Methods: We created a reference dataset containing 198 variables for 100 ischemic stroke cases, manually extracted from the electronic health records (overall 13,872 data points), taking an average of 20 minutes per patient. In automated extraction, 126 variables were extracted using conventional software engineering, and 72 using LLM. The accuracy of categorical variables was compared using Cohen's κ, numeric variables using Spearman's correlation, and free-text variables were assessed descriptively.

Results: The accuracy of the current version of the pilot algorithm was 89.2%, compared to the ground-truth dataset. Discrepancies included 457 (3.3%) missing and 1,042 (7.5%) incorrect values. 60 variables, including age, sex, prestroke modified Rankin scale score, index thrombectomy, mTICI score, puncture and reperfusion timestamps, showed Cohen's κ score of 1.0 or Spearman's coefficient of 1.0 (**Figure 1**). Overall, estimated programming time was 416 hours, and 19 algorithm iterations were performed using 5-30 patients' data in each iteration. 1.2% of values were found to be incorrectly entered manually into the reference dataset.

Conclusions: Automated tools can potentially improve stroke registry data collection efficiency by reducing data extraction time, with accuracy for some variables comparable to or better than trained staff. Further improvements in the algorithm are anticipated.

Conflict of interest: "All authors: nothing to disclose"

Figure 1 - belongs to Results