Review of Stroke-specific Natural Language Processing (NLP) and Machine Learning (ML) Applications with Unstructured Data

Tahani Daghistani
Ministry of National Guard Health Affairs
Riyadh, Saudi Arabia

ABSTRACT
The interest in Neutral Language Processing (NLP) and Machine Learning (ML) applications, in particular to stroke using unstructured data, has markedly increased in recent years. In this rapidly evolving context, it is necessary to learn and understand the novel approaches that complement and even exceed accumulated experience in the medical field. Various studies over the years conducted to demonstrate such applications with their ability to process large amounts of data, as one of the proposed approaches to support the neuroradiologists and improve the care of stroke. This study evaluates NLP/ML-based models with respect to (1) Application purpose, (2) the outcome considered, and (3) the approaches applied.

Keywords
Stroke, Neutral Language Processing (NLP), Machine Learning (ML), Unstrucured Data, Electronic Medical Records (EMR), Healthcare.

1. INTRODUCTION
In healthcare field, an accurate automated information extraction will be increasingly important as more medical researchers, organization systems, and academic institutions leverage “big data” from Electronic Medical Records (EMR). Unlike structured data, such as laboratory results, medical measures, or diagnoses codes, free text is challenging to analyze. However, clinicians frequently record essential observations, treatment recommendation, interpretations, and assessments that are otherwise unavailable from the rest of the medical record. In order to fully leverage the ability to access such data through the EMR, we must have validated methods or applications of accurate automated extraction of information through NLP [6, 11].

One particular medical application is diagnosing and treating of stroke. The Subsequent of demand for effective and efficient tools, and the complexity, variety and volume of stroke data make it a good target for NLP/ML applications. Indeed, the use of extracted relevant EMR data in combine with ML to achieve early diagnosis of conditions, such as stroke, has gained a lot of traction [2].

This study provides a comprehensive background of the performance of simple and complex stroke-specific Natural Language Processing (NLP) and Machine Learning (ML) approaches to detect, determine location, and severity of stroke from free text (unstrucured) reports.

2. NLP/ML-BASED APPLICATIONS
2.1 Applications for Data extraction
Stroke-specific NLP and ML and its application to the unstrucured medical data overcome the challenge of laborious manual reviews. It improves the finding of risk factor and provide information on the overall medical context. Several studies have recognized validated techniques in stroke care.

One of studies that applied NLP successfully, but not achieving a promising results by Elkins JS et al. aimed to expand the power of chart review by automate data acquisition tasks from of brain images reports using NLP. The accuracy of automated system was 84% (CI 83–85%) 95.7%, which is lower when compared to manual coding method that achieved 86% (CI 84–88%).The difference was small but statistically significant ($P = 0.026$) in term of accuracy [3]. Kim C. et al. used NLP-based Machine Learning (ML) classification for extracting stroke-related information from brain MRI reports. Features were extracted used to train ML algorithms. Single decision tree was the highest accuracy (98.0%) and F1-measure (93.2%) in classifying brain MRI reports and automate the identification of acute ischemic stroke (AIS) patients [8]. The prediction of schemic stroke (IS), subtype was also shown in the work of Garg R et al., where they extracted features from unstructured text-based. They applied natural language processing (NLP) on progress notes and neuroradiology reports in EHR. Then, several machine-learning methods used to identify TOAST subtype, which is the most widely used classification system for stroke, and compared results with manual classification performed by stroke neurologists. Machine-based classification performance increased from kappa of 0.25 and 0.57 using single data, either on progress notes or on neuroradiology reports, to 0.57 using combined data [4].

Another study by Ong et al. tested the combination of NLP and classification algorithms that can classify imaging findings in two steps. First, a tailored word-embedding approach developed to generate radiographic-specific word representations for vascular neurologic disorders. Text featurization (GloVe) used to convert text of both CT and MRIs into structured data with well performance (AUCs >0.9). Second, various classification models trained to identify three stroke tasks from radiology reports. RNN with GloVe algorithm was the best pair method that achieved accuracy of 0.92 for stroke presence, 0.89 for location and 0.93 for acuity [15]. The study by Kogan E et al. demonstrated the ability of random forest algorithm, built on patient treatment and demographic information, to determine National Institutes of Health Stroke Scale (NIHSS) scores. Real NIHSS recorded as free text and extracted from physician notes by applying NLP methods. The performance of model for imputing NIHSS scores compared with real NIHSS scores. The algorithm identified NIHSS scores with $R2$ (coefficient of determination), R (Pearson correlation coefficient) and root-mean squared error (RMSE) of 0.57, 0.76 and 4.5, respectively [9]. Yu et al.

applied NLP rule-based approach to automate the extraction of clinical data from neuroimaging reports. Extracted information used to identify the presence and location of vascular occlusions. The overall performance was 95.2%, 90.0%, 97.4% for accuracy, sensitivity, sensitivity, respectively. Moreover, NLP used to automate extracting and identifying information of distal or basilar occlusion and hemorrhage with high accuracy. While there were limitations in identifying other stroke-related attributes such as cerebral ischemia, ASPECTS, and collateral status [19]. Li MD et al. trained a machine-learning model using random forest algorithm and classified the report for the presence or absence of acute or subacute ischemic stroke (ASIS). The input of the model extracted from free-text in brain CTs and MRIs reports using NLP. The result showed a better performance on MR imaging reports compared with CT reports. Then, the model used to quantify the change in ASIS detected on reports and they found significant increasing in detection from 16% to 21% (P = 0.01). Last, applying the classifier on external data from a second stroke center to evaluate generalizability, NLP approach performed worse. To help creating a more generalizable model, they combined training data from both centers could, the test performance was not substantially different [10]. Gunter et al. analyzed 773 free-text radiology reports using rule-based NLP, CHARTextract. Automated data extraction offers an advantage over manual extraction and processed all reports within 5–10 s compared to 55 h for manual extraction with accuracy more than 90% [5].

2.2 Applications for Diagnosis

Even more, NLP and ML models were not only used to extract data of patients who already have the condition of interest, several studies have developed models to early predict or detect the conditions prior to their occurrence.

A system developed by Prevedello LM et al. for monitoring CT and MRI reports for specific content therefore categorize reports based on stroke possibility. This system, when executed on radiology reports, was able to use extracted key reporting component by Natural language processing (NLP) and create performance quality reports. The result of the categorization task was 58.1% of reports met system requirement for three components which are presence or absence of hemorrhage, mass, and acute infarction [16]. Mowery et al. extracted sections, structures, and expression types as an indicator for carotid stenosis using the Veteran Health Administration reports as part of a pilot study to support stroke prevention strategies. They assessed NLP ability to classify two types of free-text reports, radiology (RAD) and text integration utility (TIU) notes, to stenosis positive or stenosis negative. NLP outperformed with processing full RAD report with the highest sensitivity (88 %) rather than processing TIU notes [14]. Wheater, E. et al. used neuroradiology reports to develop rule-based NLP algorithm to automate diagnostic of more common cerebrovascular phenotypes. The model had very good performance with sensitivity >89% and specificity all > 97% to identify important brain imaging phenotypes such as ischaemic stroke, haemorrhagic stroke, brain tumours and cerebral small vessel disease and cerebral atrophy [18]. Alex B et al. described work of the development and evaluation of a system to identify phenotype information related to stroke from reports of brain imaging radiology. They used text mining techniques to identify entities, negation and relations between entities then label each report based on extracted indication. Automated reading and labeling system achieved >90% for predicting the document-level labels [1]. Study by Mayampurath A et al. utilized extracted text from paramedic reports using natural language to improve decisions about the most appropriate care to deliver, improve care pathways and linked services. The machine learning model outperformed both models of Cincinnati Prehospital Stroke Scale (CPSS) and the 3-Item Stroke Scale (3I-SS) in identifying risk associated with decisions about stroke (AUROC, 0.73) [12]. Zhao Y et al. focused on incident stroke ascertainment rather than stroke ascertainment. Sequence of events extracted from clinical notes using natural language processing including diagnosis codes, procedure codes, and clinical concepts. Several experiences were conducted using combinations of extracted features and machine learning algorithms. Random Forest model was the best-performing model identified stroke and achieved a positive predictive value of 86% and a negative predictive value of 96%. For stroke subtypes, heuristic rule based and achieved an accuracy of 83%. Further, they validated the model on general population sample and achieved 80% [20].

2.3 Advance Applications

Further studies used NLP and ML to solve other problems such as Complication, Intervention/decision support and mortality.

Intervention/decision support: Advance study by Sung et al. evaluated and treated patients with acute ischemic stroke (AIS) using natural language processing (NLP) techniques. NLP algorithm used to facilitate the decision-making process and provide timely information that help clinicians determine eligibility for intravenous thrombolysis (IVT) in patients. The classification accuracy of the algorithm was high and errors reduced in assessing eligibility criteria [17].

Complication: Miller et al. used NLP to automate the detection of important clinical features in patients with acute ischemic stroke from unstructured radiologic reports. They employed Rule-Based System (RBS) to identify the stroke complications including continuous midline shift (MLS) and intraventricular hemorrhage (IVH). RBS achieved sensitivities > 90% and specificities > 99% across different datasets [13].

Mortality: Huang, R., et al. developed multi-level model to predict short-term mortality (six months) ofhemorrhagic and ischemic stroke. They used bioassay structured data and unstructured data from radiology report for the development of an ensemble model by combining machine learning and deep learning. For haemorrhagic patients, model reached AUROC of 0.89, recall of 0.78, and precision of 0.52. For ischaemic patients the model reached AUROC of 0.88, recall of 0.80 and precision of 0.34. The prediction models can serve as risk assessment tool for early identification of patients at high-risk stroke mortality [7].

3. LIMITATIONS

State-of-the-art NLP and ML models struggle with a variety of limitations. The major limitation, most of models fail to perform well on new inputs. Either when trying to validate the model on larger sample or deploying the model in other systems. Which may reduce generalizability and ability to apply models to a wide range of real-world tasks and scenarios [8, 9, 14, 15, 20]. External validation of models, arising performance issues due to variation in reporting practices. various ways in which reports defines stroke factors regardless of modality [5, 13]. Also, lack of standard labeling of reports which reflects a professional interpretation of a patient's condition [13]. In this situation, the application of NLP rules may improve the standardization of data collection in the organization [19]. Additional challenge in extracting risk factors associated with stroke can be due to the usage of variable lexical expressions, spelling errors, ambiguous abbreviations and telegraphic constructions within unstrucured data [14]. Therefore, the results showed lower performance when models tested on reports from an external site or reports within same site [5, 10, 15]. Also, variation in accuracy in identifying types and subtypes of stroke. Extra point could affect the performance of the models which is the proportions of class subjected to predict in the training dataset, the proportion may vary significantly depending on the features of each hospital [8, 20]. There is a challenge in translating some scenarios into rules using NLP. Thereafter, rule-based algorithm will be able to capture only previously defined words in rules a sentence level. This will limit the extraction of stroke features if keywords scatter across different sentences, although reports often include more qualitative interpretation [5, 13, 19].

Furthermore, evaluating the real-world implementation of stroke applications have reported limitations in the performance. As with all studies based on real-world medical data, accurate assessment and distinct documentation of some variables are not avaliable. Other information which could be critical for model performance such as medical history may be hard to reliably collect. As instance, mixing the first observed stroke occurrence in the data with possibly later stroke diagnosis [9, 12]. More, there is dependence of subject matter experts to provide relevant clinical concepts. Advance feature engineering approaches could be explored to automate the identification of those relevant concepts. In addition, these NLP/ML applications may aid more practices such quantifying stroke severity which is limited available in EMR databases [9, 20].

4. CONCLUSIONS

In conclusion, all reviewed studies applied the combination of NLP with ML to pinpoint patients at high risk of conditions related to stroke using EMR as data sources. Starting with structured data, followed by using NLP to leverage and synthesize stroke features from clinical notes and reports (unstrucured data) into structured form where the ML applied on data contained in these records. With stroke as our case study application, the current applications have allowed for vast opportunities in advancing stroke care and improve clinical outcomes by employing all parts of the patient journey to accurately predict outcome of the case.

NLP/ML models offer multiple perspectives in stroke research nevertheless future research will be essential for medical and scientific community to overcome the challenges and limitations. In fact, proposed models should also be considered as applications in which reveal the full potential of NLP and ML in the field and support clinical practice.

5. REFERENCES

[1] Houssein, E.H., Mohamed, R.E. and Ali, A.A., 2021. Machine learning techniques for biomedical natural language processing: a comprehensive review. IEEE Access, 9, pp.140628-140653.

[2] Malmasi, S., Hosomura, N., Chang, L.S., Brown, C.J., Skentzos, S. and Turchin, A., 2017. Extracting healthcare quality information from unstructured data. In AMIA Annual Symposium Proceedings (Vol. 2017, p. 1243). American Medical Informatics Association.

[3] De Rosario, H., Pitarch-Corresa, S., Pedrosa, I., Vidal-Pedrós, M., de Otto-López, B., García-Mieres, H. and Álvarez-Rodríguez, L., 2023. Applications of Natural Language Processing for the Management of Stroke Disorders: Scoping Review. JMIR Medical Informatics, 11, p.e48693.

[4] Elkins JS, Friedman C, Boden-Albala B, Sacco RL, Hripcsak G. Coding neuroradiology reports for the Northern Manhattan Stroke Study: a comparison of natural language processing and manual review. Comput Biomed Res. 2000 Feb;33(1):1-10. doi: 10.1006/cbmr.1999.1535. PMID: 10772780.

[5] Kim C, Zhu V, Obeid J, Lenert L. Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke. PLoS One. 2019 Feb 28;14(2):e0212778. doi: 10.1371/journal.pone.0212778. PMID: 30818342; PMCID: PMC6394972.

[6] Garg R, Oh E, Naidech A, Kording K, Prabhakaran S. Automating Ischemic Stroke Subtype Classification Using Machine Learning and Natural Language Processing. J Stroke Cerebrovasc Dis. 2019 Jul;28(7):2045-2051. doi: 10.1016/j.jstrokecerebrovasdis.2019.02.004. Epub 2019 May 15. PMID: 31103549.

[7] Ong, C.J., Orfanoudaki, A., Zhang, R., Caprasse, F.P.M., Hutch, M., Ma, L., Fard, D., Balogun, O., Miller, M.I., Minnig, M. and Saglam, H., 2020. Machine learning and natural language processing methods to identify ischemic stroke, acuity and location from radiology reports. PloS one, 15(6), p.e0234908.

[8] Kogan E, Twyman K, Heap J, Milentijevic D, Lin JH, Alberts M. Assessing stroke severity using electronic health record data: a machine learning approach. BMC Med Inform Decis Mak. 2020 Jan 8;20(1):8. doi: 10.1186/s12911-019-1010-x. PMID: 31914991; PMCID: PMC6950922.

[9] Yu, A.Y., Liu, Z.A., Pou-Prom, C., Lopes, K., Kapral, M.K., Aviv, R.I. and Mamdani, M., 2021. Automating Stroke Data extraction from free-text radiology reports using natural language processing: instrument validation study. JMIR Medical Informatics, 9(5), p.e24381..

[10] Li MD, Lang M, Deng F, Chang K, Buch K, Rincon S, Mehan WA, Leslie-Mazwi TM, Kalpathy-Cramer J. Analysis of Stroke Detection during the COVID-19 Pandemic Using Natural Language Processing of Radiology Reports. AJNR Am J Neuroradiol. 2021 Mar;42(3):429-434. doi: 10.3174/ajnr.A6961. Epub 2020 Dec 17. PMID: 33334851; PMCID: PMC7959438.

[11] Gunter, D., Puac-Polanco, P., Miguel, O., Thornhill, R.E., Yu, A.Y., Liu, Z.A., Mamdani, M., Pou-Prom, C. and Aviv, R.I., 2022. Rule-based natural language processing for automation of stroke data extraction: a validation study. Neuroradiology, 64(12), pp.2357-2362.

[12] Prevedello LM, Farkas C, Ip IK, Cohen AB, Mukundan S, Sodickson AD, Khorasani R. Large-scale automated assessment of radiologist adherence to the Physician Quality Reporting System for stroke. J Am Coll Radiol. 2012 Jun;9(6):414-20. doi: 10.1016/j.jacr.2012.01.014. PMID: 22632668.

[13] Mowery, D.L., Chapman, B.E., Conway, M., South, B.R., Madden, E., Keyhani, S. and Chapman, W.W., 2016. Extracting a stroke phenotype risk factor from Veteran Health Administration clinical reports: an information content analysis. Journal of Biomedical Semantics, 7(1), pp.1-12.

[14] Wheater, E., Mair, G., Sudlow, C., Alex, B., Grover, C. and Whiteley, W., 2019. A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records. BMC Medical Informatics and Decision Making, 19(1), pp.1-11.

[15] Alex B, Grover C, Tobin R, Sudlow C, Mair G, Whiteley W. Text mining brain imaging reports. J Biomed Semantics. 2019 Nov 12;10(Suppl 1):23. doi: 10.1186/s13326-019-0211-7. PMID: 31711539; PMCID: PMC6849161.

[16] Mayampurath A, Parnianpour Z, Richards CT, Meurer WJ, Lee J, Ankenman B, Perry O, Mendelson SJ, Holl JL, Prabhakaran S. Improving Prehospital Stroke Diagnosis Using Natural Language Processing of Paramedic Reports. Stroke. 2021 Aug;52(8):2676-2679. doi: 10.1161/STROKEAHA.120.033580. Epub 2021 Jun 24. PMID: 34162217; PMCID: PMC8514137.

[17] Zhao Y, Fu S, Bielinski SJ, Decker PA, Chamberlain AM, Roger VL, Liu H, Larson NB. Natural Language Processing and Machine Learning for Identifying Incident Stroke From Electronic Health Records: Algorithm Development and Validation. J Med Internet Res. 2021 Mar 8;23(3):e22951. doi: 10.2196/22951. PMID: 33683212; PMCID: PMC7985804.

[18] Sung, S.F., Chen, K., Wu, D.P., Hung, L.C., Su, Y.H. and Hu, Y.H., 2018. Applying natural language processing techniques to develop a task-specific EMR interface for timely stroke thrombolysis: a feasibility study. International journal of medical informatics, 112, pp.149-157.

[19] Miller, M.I., Orfanoudaki, A., Cronin, M., Saglam, H., So Yeon Kim, I., Balogun, O., Tzalidi, M., Vasilopoulos, K., Fanaropoulou, G., Fanaropoulou, N.M. and Kalin, J., 2022. Natural Language Processing of Radiology Reports to Detect Complications of Ischemic Stroke. Neurocritical Care, pp.1-12.

[20] Huang, R., Liu, J., Wan, T.K., Siriwanna, D., Woo, Y.M.P., Vodencarevic, A., Wong, C.W. and Chan, K.H.K., 2023. Stroke mortality prediction based on ensemble learning and the combination of structured and textual data. Computers in Biology and Medicine, 155, p.106176