AJNR

Analysis of Stroke Detection during the COVID-19 Pandemic Using Natural Language Processing of Radiology Reports

M.D. Li, M. Lang, F. Deng, K. Chang, K. Buch, S. Rincon, W.A. Mehan, T.M. Leslie-Mazwi and J. Kalpathy-Cramer

AJNR Am J Neuroradiol 2021, 42 (3) 429-434
doi: https://doi.org/10.3174/ajnr.A6961
http://www.ajnr.org/content/42/3/429

Analysis of Stroke Detection during the COVID-19 Pandemic Using Natural Language Processing of Radiology Reports

M.D. Li, M. Lang, F. Deng, K. Chang, K. Buch, S. Rincon, W.A. Mehan, T.M. Leslie-Mazwi, and J. Kalpathy-Cramer

ABSTRACT

BACKGROUND AND PURPOSE: The coronavirus disease 2019 (COVID-19) pandemic has led to decreases in neuroimaging volume. Our aim was to quantify the change in acute or subacute ischemic strokes detected on CT or MR imaging during the pandemic using natural language processing of radiology reports.

MATERIALS AND METHODS: We retrospectively analyzed 32,555 radiology reports from brain CTs and MRIs from a comprehensive stroke center, performed from March 1 to April 30 each year from 2017 to 2020, involving 20,414 unique patients. To detect acute or subacute ischemic stroke in free-text reports, we trained a random forest natural language processing classifier using 1987 randomly sampled radiology reports with manual annotation. Natural language processing classifier generalizability was evaluated using 1974 imaging reports from an external dataset.

RESULTS: The natural language processing classifier achieved a 5-fold cross-validation classification accuracy of 0.97 and an F1 score of 0.74, with a slight underestimation (−5%) of actual numbers of acute or subacute ischemic strokes in cross-validation. Importantly, cross-validation performance stratified by year was similar. Applying the classifier to the complete study cohort, we found an estimated 24% decrease in patients with acute or subacute ischemic strokes reported on CT or MR imaging from March to April 2020 compared with the average from those months in 2017–2019. Among patients with stroke-related order indications, the estimated proportion who underwent neuroimaging with acute or subacute ischemic stroke detection significantly increased from 16% during 2017–2019 to 21% in 2020 ($P = .01$). The natural language processing classifier performed worse on external data.

CONCLUSIONS: Acute or subacute ischemic stroke cases detected by neuroimaging decreased during the COVID-19 pandemic, though a higher proportion of studies ordered for stroke were positive for acute or subacute ischemic strokes. Natural language processing approaches can help automatically track acute or subacute ischemic stroke numbers for epidemiologic studies, though local classifier training is important due to radiologist reporting style differences.

ABBREVIATIONS: ASIS = acute or subacute ischemic stroke; COVID-19 = coronavirus disease 2019; NLP = natural language processing

There is much concern regarding the impact of the coronavirus disease 2019 (COVID-19) pandemic on the quality of stroke care, including issues with hospital capacity, clinical resource re-allocation, and the safety of patients and clinicians.[1,2] Previous reports have shown that there have been substantial decreases in stroke neuroimaging volume during the pandemic.[3,4] In addition, acute ischemic infarcts have been found on neuroimaging studies in many hospitalized patients with COVID-19, though the causal relationship is unclear.[5,6] Studies like these and other epidemiologic analyses usually rely on the creation of manually curated databases, in which identification of cases can be time-consuming and difficult to update in real-time. One way to facilitate such research is to use natural language processing (NLP), which has shown utility for automated analysis of radiology report data.[7] NLP algorithms have been developed previously for the classification of neuroradiology reports for the presence of ischemic stroke findings and acute ischemic stroke subtypes.[8,9] Thus, NLP has the potential to facilitate COVID-19 research.

Received June 20, 2020; accepted after revision October 26.

From the Departments of Radiology (M.D.L., M.L., F.D., K.C., K.B., S.R., W.A.M., J.K.-C.) and Neurology and Neurosurgery (T.M.L.-M.), Massachusetts General Hospital, Mass General Brigham, Harvard Medical School, Boston, Massachusetts.

This research was supported by a training grant from the National Institute of Biomedical Imaging and Bioengineering the National Institutes of Health under award No. 5T32EB1680 and the National Cancer Institute of the National Institutes of Health under Award No. F30CA239407 to K. Chang.

Please address correspondence to Matthew D. Li, MD, Department of Radiology, Massachusetts General Hospital, Mass General Brigham, Harvard Medical School, 55 Fruit St, Boston, MA 02114; e-mail: mdli@mgh.harvard.edu

Indicates open access to non-subscribers at www.ajnr.org

Indicates article with online supplemental data.

http://dx.doi.org/10.3174/ajnr.A6961

In this study, we developed an NLP machine learning model that classifies radiology reports for the presence or absence of acute or subacute ischemic stroke (ASIS), as opposed to chronic stroke. We used this model to quantify the change in ASIS detected on all CT or MR imaging studies performed at a large comprehensive stroke center during the COVID-19 pandemic in the United States. We also evaluated NLP model generalizability and different training strategies using a sample of radiology reports from a second stroke center.

MATERIALS AND METHODS

This retrospective study was exempted with waiver of informed consent by the institutional review board of Mass General Brigham (Boston, Massachusetts), the integrated health system that includes both Massachusetts General Hospital and Brigham and Women's Hospital.

Radiology Report Extraction

We used a custom hospital-based radiology report search tool to extract head CT and brain MR imaging study reports performed at Massachusetts General Hospital (hospital 1) and its affiliated imaging centers (a comprehensive stroke center) from March 1 to April 30 in each year from 2017 to 2020. At this hospital, head CT and brain MR imaging studies are routinely performed for patients with stroke. Head CTs included noncontrast and contrast-enhanced head CT and CT angiography studies. Brain MRIs included noncontrast and contrast-enhanced brain MRIs and MR angiography studies. After we removed outside imaging studies also stored in the data base, there were 15,627 head CT and 17,151 brain MR imaging reports (a total of 32,778 studies). Of these studies, 15,590 head CT and 16,965 brain MR imaging reports had study "Impressions", which restricted the analysis to 98.9% and 99.8% of the dataset, respectively. These studies formed the aggregate study cohort, which included a total of 32,555 brain MR imaging and head CT reports on 20,414 unique patients.

Of the original 32,778 study reports extracted, 1000 head CT and 1000 brain MR imaging studies were randomly sampled for manual annotation to serve as training and testing data for an NLP machine learning model. Of these studies, 1987 contained study Impressions (99.4%). The studies without study Impressions were predominantly pediatric brain MR imaging studies that involved a different structure for reporting.

Using a commercial radiology information system (Primordial/Nuance Communications), we also extracted an additional dataset of radiology reports from Brigham and Women's Hospital (hospital 2) and its affiliated imaging centers (also a comprehensive stroke center). We analyzed the overlap in radiologists and trainees involved in the dictation of these reports between hospitals 1 and 2. The first 500 consecutive head CTs and the first 500 brain MRIs performed in both April 2019 and April 2020 were obtained (a total of 1000 head CTs and 1000 brain MR imaging study reports), with the same inclusion criteria for noncontrast and contrast-enhanced studies, as well as angiographic studies. All of these reports had study Impressions. After removing duplicate study entries in this dataset (26, 1.3%), 1974 head CT and brain MR imaging reports remained for further analysis.

NLP Training Dataset Annotation

For NLP model training, the 1987 study reports sampled from hospital 1 and the 1974 study reports available from hospital 2 were manually annotated, each by a diagnostic radiology resident (F.D. for CT and M.L. for MR imaging from hospital 1 and M.D.L. for CT and MR imaging from hospital 2). The annotators classified each report for the presence of ASIS using the study "Impression." This finding could be explicitly or implicitly stated in the report, and reports that stated or suggested chronicity of an infarct were not considered to have this finding. For example, "old" or "chronic" infarct suggests chronicity, though more ambiguous terms like "age-indeterminate" or "unclear timeframe" were sometimes found. Reports with ambiguous terms were not considered to have ASIS, unless an expression of newness was conveyed in the report (eg, "new age-indeterminate infarct").

NLP Machine Learning Model Training and Testing

We trained a random forest machine learning model that takes the radiology report free-text Impression as input and classifies the report for the presence or absence of an ASIS. To train a machine learning model to automatically parse the radiology report text, we used the *re* (Version 2.2.1), *sklearn* (Version 0.20.3), and *nltk* (Version 3.4) packages in Python (Version 3.7.1). Before model training, we used regular expressions to extract sentences with words containing the stems "infarct" or "ischem" from each study Impression. This step helped to focus the algorithm on sentences containing content relevant to the classification task. The words in the extracted sentences were stemmed using the snowball.EnglishStemmer from the *nltk* package. The extracted and stemmed sentences were then represented as vectors using bag-of-words vectorization with N-grams ($n = 2–3$; minimum term frequency, 1%), an approach previously used for radiology report natural language processing.[10] Negation was dealt with using the *nltk* mark_negation function, which appends a "_NEG" suffix between words that appear between a negation term and a punctuation mark. These vector representations of the radiology report Impression served as inputs to the random forest NLP classifier.

The random forest NLP classifier was trained using default hyperparameters in *sklearn*, Version 0.20.3, including 100 trees in the forest using the Gini Impurity for measuring the quality of the data split. Using the manually annotated datasets from hospitals 1 and 2, we evaluated 2 training strategies. First, we trained a classifier using the hospital 1 annotated dataset and tested performance using 5-fold cross-validation, stratified on outcome (ASIS), given the class imbalance. We also tested this classifier on the external hospital 2 annotated dataset. Second, we trained a classifier using the combined hospital 1 and 2 annotated datasets and tested performance using 5-fold cross-validation, also stratified on outcome, but only including hospital 1 reports in the cross-validation to assess performance on hospital 1 data specifically. We repeated this testing using only hospital 2 reports in the cross-validation to assess the performance on the hospital 2 data specifically. When the combined hospital 1 and 2 datasets were used for classifier training, the N-gram minimum term frequency was halved to 0.5%; empirically, the number of N-gram terms

was then similar between this classifier and the classifier trained on hospital 1 data only. The Python code for training these random forest classifiers is available at github.com/QTIM-Lab/asis_nlp.

The metrics used to assess model performance included accuracy, precision, recall, and F1 score. Performance was evaluated for CT and MR imaging reports combined, CT reports alone, and MR imaging reports alone, with 5-fold cross-validation when appropriate.

NLP-Based Epidemiologic Analysis

On the basis of the results of the NLP model testing, an NLP classifier was then applied to the complete cohort of 32,555 brain MR imaging and head CT reports from hospital 1 to estimate the number of patients with ASIS. Patients with at least 1 neuroimaging study (CT or MR imaging) with an ASIS during the time period in question were considered to have had an infarct. Demographic information associated with these patients was extracted along with the radiology report text.

Statistics

Statistical testing was performed using the *scipy* Version 1.1.0 package in Python. The Pearson χ^2 test of independence and 1-way ANOVA were used when appropriate. Statistical significance was determined a priori to be $P < .05$. Performance metrics were reported as the bootstrap median estimate with 95% confidence intervals.[11]

RESULTS

Manually Annotated Radiology Report Dataset Characteristics

Among the randomly sampled 1987 neuroimaging reports from hospital 1 used for NLP model development, 67 head CT and 68 brain MR imaging reports were manually classified as positive for ASIS (positive in 129 patients from 1904 total unique patients). Among the 1974 neuroimaging reports from hospital 2, 84 head CT and 91 brain MR imaging reports were manually classified as positive for ASIS (positive in 101 patients from 1514 total unique patients). The remainder of studies were negative for ASIS. In the hospital 1 annotated report dataset, 126 unique radiologists and trainees (residents and fellows) were involved in the dictation of these reports. In the hospital 2 annotated report dataset, there were 94 unique radiologists and trainees involved. There was an overlap of 3 radiologists and trainees between these 2 datasets due to radiologists/trainees moving between institutions. The hospital 1 and hospital 2 reports were all free-text without a standardized structure. The manual annotators who read the report Impressions found that they differed stylistically between the hospitals.

NLP Model Performance

Random forest NLP classifier testing performance is summarized in the Online Supplemental Data. The stratified 5-fold cross-validation performance of the NLP classifier trained on the hospital 1 annotated dataset showed an average accuracy of 0.97 (95% CI, 0.96–0.97) and an F1 score of 0.74 (95% CI, 0.72–0.76). When this NLP classifier was tested on the hospital 2 annotated dataset, the performance was lower, with an accuracy of 0.95 (95% CI, 0.94–0.96) and an F1 score of 0.66 (95% CI, 0.59–0.72). In both tests, when the performance results for CT and MR imaging were separately analyzed, we found that the model performed better for MR imaging reports compared with CT reports.

We also trained a random forest NLP classifier using the combined annotated reports from hospitals 1 and 2. In the stratified 5-fold cross-validation performance with testing of only hospital 1 data in the validation folds, the average accuracy was 0.96 (95% CI, 0.96-0.96) and the average F1 score was 0.74 (95% CI, 0.72–0.76). This performance on hospital 1 data was similar compared with the NLP classifier trained using only hospital 1 data. In the stratified 5-fold cross-validation performance with testing of only hospital 2 data in the validation folds, the average accuracy was 0.96 (95% CI, 0.96-0.97) and the average F1 score was 0.79 (95% CI, 0.77–0.80). This performance on hospital 2 data was substantially improved compared with the NLP classifier trained using only hospital 1 data. Because the performance on hospital 1 data was similar between the NLP classifier trained on hospital 1 reports and the NLP classifier trained on hospitals 1 and 2 reports, we used the former classifier for further analysis of the complete hospital 1 dataset.

For the NLP classifier trained on hospital 1 reports, in the 5 cross-validation folds for the combined CT and MR imaging analysis, there was an average of 19.4 (95% CI, 18.6–20.2) true-positive, 6.2 (95% CI, 5.6–6.8) false-positive, 7.6 (95% CI, 6.8–8.4) false-negative, and 364.2 (95% CI, 18.6–20.2) true-negative classifications. In misclassified cases, the reports typically contained uncertainty regarding the chronicity of the infarct (eg, age-indeterminate or not otherwise specified in the study Impression). For each of the 5 cross-validation folds, there was an average of 25.6 positive results predicted (95% CI, 24.6–26.8), compared with 27.0 actual positive results in each validation fold, due to the stratification on outcome. The NLP predicted that the number of cases slightly underestimated the actual number of studies positive for ASIS in the validation folds (average difference, −1.4; 95% CI, 0.2–2.4; expressed as percentages, −5.1%; 95% CI, 0–8.8%).

To ensure that variations in reporting styles within the hospital 1 reports did not systematically differ by year (because our epidemiologic analysis would compare reports from different years), we performed leave-one-year-out cross-validation on the hospital 1 dataset, in which NLP classifiers were trained on data from all years except the year of the excluded validation set (eg, trained on reports from 2018, 2019, and 2020, and then tested on reports from 2020). We found that there was no substantial difference in model performance in each of those validation folds (with overlap of 95% CI), which shows that the model performed similarly across time periods at hospital 1 (Online Supplemental Data). The F1 score was 0.72 in 2020 versus between 0.68 and 0.73 from 2017 to 2019.

While the NLP model systematically slightly underestimated ASIS case numbers, because the model performed similarly from year-to-year, we used this random forest classifier to estimate changes in the numbers of ASISs detected in the complete hospital 1 study cohort of 32,555 head CT and brain MR imaging reports.

Table 1: Natural language processing-based analysis of all radiology reports from hospital 1

Time Period	Total No. of Neuroimaging Studies Performed (CT/MRI)	Estimated No. of Neuroimaging Studies with ASIS Detected (% of Total)	Total No. of Patients Undergoing Neuroimaging	Estimated No. of Patients with ASIS Detected (% of Total)	No. of Neuroimaging Studies per Patient
March-April 2020 (COVID-19 pandemic)	5709	428 (7.4%)	3977	231 (5.8%)	1.4
March-April 2017–2019[a]	8949	541 (6.0%)	6312	304 (4.8%)	1.4
2019	9403	586 (6.2%)	6770	313 (4.6%)	1.4
2018	9086	523 (5.8%)	6417	302 (4.7%)	1.4
2017	8357	515 (6.2%)	5750	297 (5.1%)	1.5

[a] Three-year average.

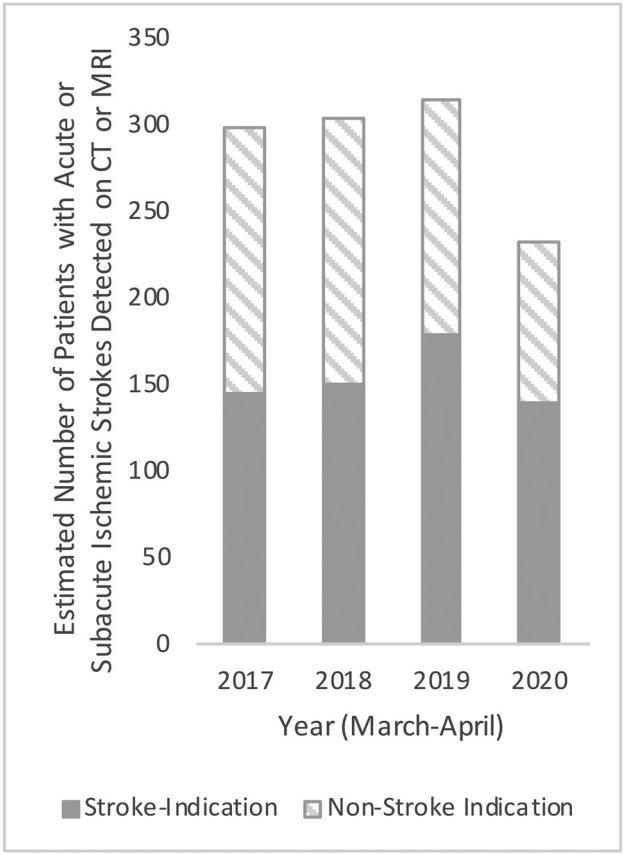

FIGURE. Estimated numbers of patients with acute or subacute ischemic strokes detected on CT or MR imaging in March and April from 2017 to 2020 at hospital 1.

ASIS during the COVID-19 Pandemic

Using this random forest NLP classifier, we estimated the number of neuroimaging studies performed and the number of patients with detected ASIS (Table 1). Patients with at least 1 neuroimaging study (CT or MR imaging) with an ASIS during the time period in question were considered to have had ASIS. There was an estimated 24% decrease in patients with ASIS reported on CT or MR imaging from March to April 2020 compared with the average of the same months from 2017 to 2019, after previous year-on-year growth from 2017 to 2019 (Figure). There was a concomitant decrease in the total number of neuroimaging studies performed and patients undergoing neuroimaging in March and April 2020 compared with 2019 (−39% and −41%, respectively).

In the complete cohort of 32,555 study reports, 32,358 of the reports (99.4%) included structured and/or unstructured text in the study indication field, entered at the time of order entry. Of those cases, we filtered for indications including "stroke," "neuro deficit," and "TIA," which resulted in 5204 study reports (Table 2). In these patients, we found an estimated 21% decrease in ASIS reported from March to April 2020 compared with March to April 2019 (Figure). In the subset of patients who underwent imaging with stroke-related indications, the estimated proportion of patients with ASIS detected increased from 16% during 2017–2019 to 21% in 2020 ($P = .01$) (Table 2). The estimated proportion of neuroimaging studies with ASIS detected increased from 20% during 2017–2019 and 24% in 2020 ($P = .01$).

The average age of patients with ASIS detected was 66 [SD, 17] years, and there was no significant difference in age among any of the years ($P = .9$). There was also no significant difference in the sex ratio of March to April 2020 compared with the March to April 2017–2019 time periods ($P = .8$). In aggregate, 56% of patients with ASIS were men. See Online Supplemental Data for data by year.

Neuroimaging Studies Performed per Patient

If the number of neuroimaging studies performed per patient differed between the prepandemic and pandemic time periods, the number of opportunities to detect ASIS in a patient could vary. However, this variance did not appear to be a confounding factor in our analysis because we found no significant difference in the number of neuroimaging studies performed per patient between the March and April 2020 time period and each of the March to April 2017, 2018, or 2019 time periods ($P > .2$).

DISCUSSION

In this study, we developed a random forest NLP algorithm for automated classification of ASIS in radiology report Impressions and applied this algorithm to reports during and before the COVID-19 pandemic. We found a substantial decrease in the number of patients with ASIS detected on all CT and MR imaging studies performed at a comprehensive stroke center during the pandemic in the United States. This decrease could be related to avoidance of the hospital due to fear of contracting COVID-19, as previously speculated.[12,13] Previous studies have shown a 39% decrease in neuroimaging studies performed primarily for stroke thrombectomy evaluation using commercial software in

Table 2: NLP-based analysis of radiology reports from hospital 1 containing "stroke," "neuro deficit," or "TIA" in the order indication

Time Period	Total No. of Neuroimaging Studies Performed (CT/MRI)	Estimated No. of Neuroimaging Studies with ASIS Detected (% of Total)	Total No. of Patients Undergoing Neuroimaging	Estimated No. of Patients with ASIS Detected (% of Total)
March-April 2020 (COVID-19 pandemic)	1046	254 (24.3%)	660	140 (21.2%)
March-April 2017–2019[a]	1386	274 (19.8%)	970	157 (16.2%)
2019	1720	314 (18.3%)	1283	178 (13.9%)
2018	1234	253 (20.5%)	820	149 (18.2%)
2017	1204	256 (21.3%)	808	144 (17.8%)

[a] Three-year average.

the United States and a 59.7% decrease in stroke code CT-specific cases in New York.[3,4] Our study differs because we sought to quantify the decrease in actual ASISs detected on such studies and the rate of detection. Among patients with stroke-related image -order indication, we found a significant increase in the proportion of neuroimaging studies positive for ASIS. This finding could suggest that during the COVID-19 pandemic, imaged patients had, on average, more severe or clear-cut stroke syndromes (with a higher pretest probability of stroke), implying that patients with mild or equivocal symptoms presented to the hospital less often than in previous years.

The NLP machine learning approach that we used in this study can also be applied to additional data relatively easily, which will allow us to continue to monitor the detection of ASIS on neuroimaging at our institution in the future. NLP algorithms have been used to analyze neuroradiology reports for stroke findings, specifically for the presence of any ischemic stroke findings or ischemic stroke subtypes.[8,9] The task in our model, however, is relatively challenging in that we sought to identify acute or subacute strokes specifically and deliberately excluded chronic infarcts. There is often uncertainty or ambiguity in radiology reports related to the timeframe for strokes, which can make this task challenging for the NLP algorithm. Thus, it is not surprising that our NLP model performed better on MR imaging reports compared with CT reports, given the superiority of MR imaging for characterizing the age of an infarct.

The NLP classifier trained on only hospital 1 reports showed lower performance when tested on radiology reports from an external site, which was likely due to systematic differences in linguistic reporting styles between the radiology departments in hospitals 1 and 2. While combining training data from hospitals 1 and 2 helped to create a more generalizable classifier with improved performance on hospital 2 data, the test performance of this classifier on data from hospital 1 was not substantially different from the classifier trained on only hospital 1 data. These findings highlight the importance of localized testing of NLP algorithms before clinical deployment. Nevertheless, a locally trained and deployed model can still be useful, as long as its specific use case and limitations are understood.[14]

Instead of using the radiology report NLP approach presented in our study, we could have used the International Classification of Diseases codes from hospital administrative and billing data. However, the International Classification of Diseases coding is known to have variable sensitivity and specificity for acute stroke in the literature[15] and may be particularly problematic for reliably differentiating stroke chronicity. Comparison of NLP analyses of radiology reports versus administrative data base International Classification of Diseases coding could be an avenue of future research.

There are important limitations to this study. First, we used an automated NLP approach for analysis, which systematically slightly underestimates the number of ASISs but may be scaled to analyze large numbers of reports. In the future, newer NLP technologies including deep learning–based algorithms may help improve the ability to perform studies like this one.[16] Second, the radiology report is an imperfect reference standard for assessment of ASIS, particularly for CT in which early infarcts may not be seen. In our epidemiologic analysis, patients with at least 1 neuroimaging study with ASIS during the time period of interest were counted as having ASIS. Thus, patients with early infarcts not reported on CT would still be counted as having ASIS if reported on the subsequent MR imaging, reducing the impact of false-negative CTs. However, false-positive head CTs would falsely elevate the count of ASIS. Third, identification of studies with stroke-related indications likely underestimates the total number of studies performed for suspicion of stroke because nonspecific indications like "altered mental status" were not included. This bias should be consistent across each year though; thus, it should not impact our comparison of the positive case rate between the time periods in question.

CONCLUSIONS

We developed an NLP machine learning model to characterize trends in stroke imaging at a comprehensive stroke center before and during the COVID-19 pandemic. The sequelae of decreased detection of strokes remains to be seen, but this algorithm and the shared code can help facilitate future research of these trends.

Disclosures: Ken Chang—*RELATED*: *Grant*: National Institutes of Health, *Comments*: Research reported in this publication was supported by a training grant from the National Institute of Biomedical Imaging and Bioengineering the National Institutes of Health under award No. 5T32EB1680 and the National Cancer Institute of the National Institutes of Health under Award No. F30CA239407 to K. Chang.* Karen Buch—*UNRELATED*: *Employment*: Massachusetts General Hospital. William A. Mehan, Jr—*UNRELATED*: *Consultancy*: Kura Oncology, *Comments*: independent image reviewer for head and neck cancer trial; *Expert Testimony*: CRICO and other medical insurance companies, *Comments*: expert opinion for medicolegal cases involving neuroimaging studies. Jayashree Kalpathy-Cramer—*UNRELATED*: *Grants/Grants Pending*: GE Healthcare, Genentech Patient Foundation*; *Travel/*

Accommodations/Meeting Expenses Unrelated to Activities Listed: IBM. *Money paid to the institution.

REFERENCES

1. Leira EC, Russman AN, Biller J, et al. **Preserving stroke care during the COVID-19 pandemic: potential issues and solutions.** *Neurology* 2020;95:124–33 CrossRef Medline
2. Markus HS, Brainin M. **COVID-19 and stroke: a global World Stroke Organization perspective.** *Int J Stroke* 2020;15:361–64 CrossRef Medline
3. Phillips CD, Shatzkes D, Moonis G, et al. **From the eye of the storm: multi-institutional practical perspectives on neuroradiology from the COVID-19 outbreak in New York City.** *AJNR Am J Neuroradiol* 2020;41:960–65 CrossRef Medline
4. Kansagra AP, Goyal MS, Hamilton S, et al. **Collateral effect of Covid-19 on stroke evaluation in the United States.** *N Engl J Med* 2020;383:400–01 CrossRef Medline
5. Mahammedi A, Saba L, Vagal A, et al. **Imaging in neurological disease of hospitalized COVID-19 patients: an Italian multicenter retrospective observational study.** *Radiology* 2020;297:E270–73 CrossRef Medline
6. Jain R, Young M, Dogra S, et al. **COVID-19 related neuroimaging findings: a signal of thromboembolic complications and a strong prognostic marker of poor patient outcome.** *J Neurol Sci* 2020;414:116923 CrossRef Medline
7. Pons E, Braun LM, Hunink MG, et al. **Natural language processing in radiology: a systematic review.** *Radiology* 2016;279:329–43 CrossRef Medline
8. Wheater E, Mair G, Sudlow C, et al. **A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records.** *BMC Med Inform Decis Mak* 2019;19:184 CrossRef Medline
9. Garg R, Oh E, Naidech A, et al. **Automating ischemic stroke subtype classification using machine learning and natural language processing.** *J Stroke Cerebrovasc Dis* 2019;28:2045–51 CrossRef Medline
10. Hassanpour S, Langlotz CP, Amrhein TJ, et al. **Performance of a machine learning classifier of knee MRI reports in two large academic radiology practices: a tool to estimate diagnostic yield.** *AJR Am J Roentgenol* 2017;208:750–73 CrossRef Medline
11. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall/CRC; 1994
12. Oxley TJ, Mocco J, Majidi S, et al. **Large-vessel stroke as a presenting feature of Covid-19 in the young.** *N Engl J Med* 2020;382:e60 CrossRef Medline
13. Zhao J, Rudd A, Liu R. **Challenges and potential solutions of stroke care during the coronavirus disease 2019 (COVID-19) outbreak.** *Stroke* 2020;51:1356–57 CrossRef Medline
14. Futoma J, Simons M, Panch T, et al. **The myth of generalisability in clinical research and machine learning in health care.** *Lancet Digit Heal* 2020;2:e489–92 CrossRef Medline
15. McCormick N, Bhole V, Lacaille D, et al. **Validity of diagnostic codes for acute stroke in administrative databases: a systematic review.** *PLoS One* 2015;10:e0135834 CrossRef Medline
16. Wolf T, Debut L, Sanh V, et al. **HuggingFace's Transformers: State-of-the-art Natural Language Processing.** Oct 9, 2019. https://arxiv.org/abs/1910.03771. Accessed June 15, 2020.