

RESEARCH ARTICLE

Open Access



# Assessing stroke severity using electronic health record data: a machine learning approach

Emily Kogan<sup>1\*</sup> , Kathryn Twyman<sup>1</sup>, Jesse Heap<sup>1</sup>, Dejan Milentijevic<sup>2</sup>, Jennifer H. Lin<sup>2</sup> and Mark Alberts<sup>3</sup>

## Abstract

**Background:** Stroke severity is an important predictor of patient outcomes and is commonly measured with the National Institutes of Health Stroke Scale (NIHSS) scores. Because these scores are often recorded as free text in physician reports, structured real-world evidence databases seldom include the severity. The aim of this study was to use machine learning models to impute NIHSS scores for all patients with newly diagnosed stroke from multi-institution electronic health record (EHR) data.

**Methods:** NIHSS scores available in the Optum® de-identified Integrated Claims-Clinical dataset were extracted from physician notes by applying natural language processing (NLP) methods. The cohort analyzed in the study consists of the 7149 patients with an inpatient or emergency room diagnosis of ischemic stroke, hemorrhagic stroke, or transient ischemic attack and a corresponding NLP-extracted NIHSS score. A subset of these patients ( $n = 1033$ , 14%) were held out for independent validation of model performance and the remaining patients ( $n = 6116$ , 86%) were used for training the model. Several machine learning models were evaluated, and parameters optimized using cross-validation on the training set. The model with optimal performance, a random forest model, was ultimately evaluated on the holdout set.

**Results:** Leveraging machine learning we identified the main factors in electronic health record data for assessing stroke severity, including death within the same month as stroke occurrence, length of hospital stay following stroke occurrence, aphagia/dysphagia diagnosis, hemiplegia diagnosis, and whether a patient was discharged to home or self-care. Comparing the imputed NIHSS scores to the NLP-extracted NIHSS scores on the holdout data set yielded an  $R^2$  (coefficient of determination) of 0.57, an  $R$  (Pearson correlation coefficient) of 0.76, and a root-mean-squared error of 4.5.

**Conclusions:** Machine learning models built on EHR data can be used to determine proxies for stroke severity. This enables severity to be incorporated in studies of stroke patient outcomes using administrative and EHR databases.

**Keywords:** Database, Outcomes research, Real-world evidence

## Background

Stroke is the fifth leading cause of death in the US and a primary focus for improving patient outcomes and healthcare quality [1, 2]. The National Institutes of Health Stroke Scale (NIHSS) is a widely accepted, clinically-validated measurement of stroke severity. The NIHSS score serves as an important guide for clinicians

to effectively offer guidance about prognosis and disability associated with acute stroke [1, 3–5].

The NIHSS score is defined as the sum of 15 individually evaluated elements, and ranges from 0 to 42. Stroke severity may be categorized as follows: no stroke symptoms, 0; minor stroke, 1–4; moderate stroke, 5–15; moderate to severe stroke, 16–20; and severe stroke, 21–42 [6, 7]. NIHSS scores are not part of structured data in electronic health records (EHR); rather, stroke severity is recorded as free text in physician notes. The lack of a formal stroke severity assessment in large EHR databases

\* Correspondence: [ekogan@its.jnj.com](mailto:ekogan@its.jnj.com)

<sup>1</sup>Janssen Research & Development, LLC, Raritan, NJ, USA

Full list of author information is available at the end of the article



is a limitation of real-world evidence patient outcome studies related to stroke [1, 8]. Therefore, this type of machine learning approach may be useful for quantifying stroke severity given the limited availability of clinically assessed NIHSS scores in real-world evidence databases [9], aiding such practices as payer modeling for case mix risk adjustment or assessing quality outcomes. Using billing codes from administrative claims data of the single-payer, compulsory enrollment healthcare program in Taiwan, Sung and colleagues developed several models to derive a stroke severity index and validate its performance against the NIHSS [9–11]. Developing a machine learning model for stroke severity based on claims or EHR data from the United States presents unique challenges due to the fact that the US healthcare system is a multi-payer and provider system financed and delivered through a combination of private and public resources.

The objective of this study was to retroactively impute NIHSS scores for all patients with newly diagnosed stroke in a multi-institution EHR database by leveraging machine learning techniques. Imputed NIHSS scores will enable large-scale real-world observational studies to incorporate a measure of stroke severity in research studies of disease burden in these patients.

## Methods

### Database

The Optum© de-identified Integrated Claims-Clinical dataset combines adjudicated claims data with EHR data. The EHR database is derived from more than 50 healthcare provider organizations in the United States, including more than 700 hospitals and 7000 clinics that provide care to more than 91 million patients in the United States. Optum©'s Integrated Claims-EHR dataset is statistically de-identified under the Expert Determination method consistent with Health Insurance Portability and Accountability Act (HIPAA) and managed according to Optum© and customer data use agreements [12, 13]. The Integrated dataset links both claims and clinical data for approximately 14 million matched individuals, and is generated by Optum© using a proprietary algorithm that incorporates both salting and cryptographic hashing. The EHR information contained in the integrated dataset includes medications prescribed and administered, lab results, vital signs, body measurements, diagnoses, procedures, and information derived from physician notes using Natural Language Processing (NLP).

The NIHSS scores are a part of the information derived from the physician notes [14], and these scores were used as an outcome variable when training and evaluating model performance. Because some invalid values were originally extracted from the physician notes

(e.g. values which are not integers within NIHSS range), rigorous pre-processing was applied to exclude as many invalid NIHSS scores as possible. This exclusion criterion was defined in collaboration with Optum and evaluation was completed on the remaining extracted NIHSS values to ensure accuracy. When a patient had multiple NIHSS scores during their inpatient stay following stroke, the maximum score was used to capture the overall severity of the stroke. This study incorporated EHR data from January 2007 through September 2016.

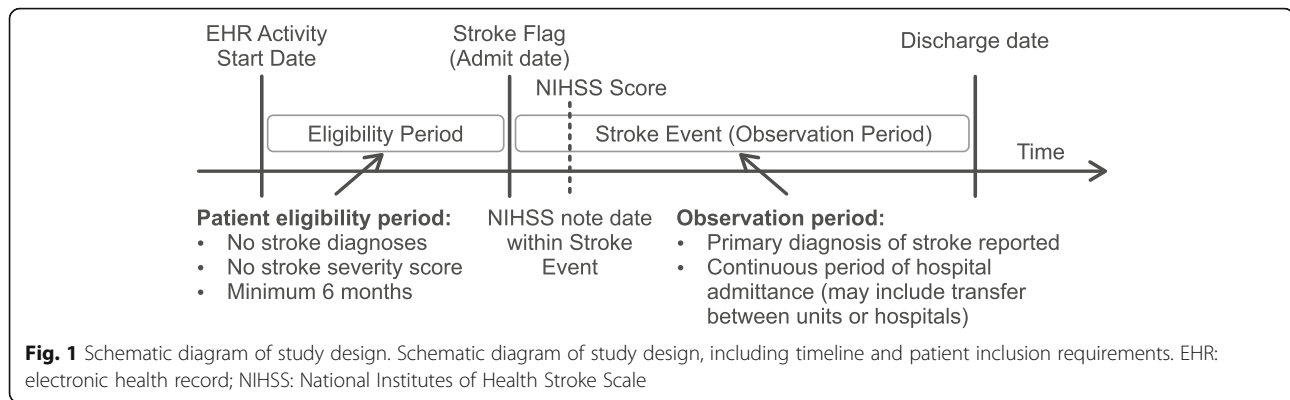
### Study population

Patients were included in the study if they had a primary diagnosis of stroke (hemorrhagic [the International Classification of Diseases (ICD)-9: 431; ICD-10 I61.XX], ischemic [ICD-9: 433.XX-434.XX, 436; ICD-10: I63.XXX], or transient ischemic attack (TIA) [ICD-9: 435.X; ICD-10: G45.9]) in an inpatient or emergency room setting, which was defined as the stroke event (Fig. 1). Additionally, patients were required to have a real NIHSS score (extracted from physician notes) during the stroke event. Patients were also required to have been in the database for at least 6 months prior to the stroke diagnosis.

From the Optum© de-identified Integrated dataset, 7149 patients were identified who met these inclusion and exclusion criteria. Prior to developing the machine learning model, patients were randomly grouped into a training set ( $n = 6116$ , 86%) and hold-out test set ( $n = 1033$ , 14.4%). Model development, parameter tuning, and feature selection were completed using the training dataset.

### Feature engineering and feature selection

Relevant patient demographics (e.g. age, gender) and billing codes related to procedures, diagnoses, prescriptions/medications, hospital visit information, and comorbidities were collected to form the initial set of 8023 potential features. All features were created during the inpatient hospitalization following stroke occurrence (Fig. 1), except the Charlson Comorbidity Index, which was estimated based on data prior to the stroke [15]. Diagnoses codes were from the ninth and tenth revisions of the ICD-9 and ICD-10 [16]. In order to create features for machine learning models that are agnostic to coding version an equivalency mapping provided by The Centers for Medicare & Medicaid Services (CMS) was leveraged. As the granularities of diagnoses codes are different in ICD9 and ICD10 revisions this mapping includes many-to-many relationships. By starting with all diagnosis codes within the stroke events for the patient cohort, and recursively incorporating any diagnosis codes that are equivalent according to the CMS mapping, disjoint diagnosis code groups were created. Binary features were then formed by checking each patient for



the presence or absence of any diagnosis within a given ‘diagnosis code group’ during the patients’ stroke event.

Additionally, simple presence/absence features were created for procedures coded with the Current Procedural Terminology (CPT4), Healthcare Common Procedure Coding System – HCPCS procedure codes, Bergenson-Eggers Type of Service codes (BETOS), patient discharge status, diagnosis-related group assigned to the inpatient stay, drug class of prescriptions written, drug class of medications administered, and routes by which medications were administered. Counts of procedures (e.g. CPT4) and BETOS code groups within a patient’s stroke event were also included in the initial feature set. Other features included patient’s age at the time of stroke, gender, and length of hospital stay following stroke occurrence.

During the feature selection process, features with near zero variance or with high correlation ( $> 0.9$ ) to another feature were removed. In the latter case, only the feature more highly correlated with the response variable was retained. A response-balanced subset of the training cohort was created for this step in the process, by randomly selecting an equal number of patients from each of 5 stroke severity categories [6] ( $n = 183$  per category,  $n = 915$  total); this step was necessary such that the feature engineering process was not affected by the skewed distribution of stroke severity categories (i.e., more patients in less severe categories than in more severe categories, Fig. 2). After initial feature selection, the remaining 619 features were used for the subsequent modeling step.

#### Machine learning model development

The imputed NIHSS scores were ultimately compared to the real NIHSS scores (extracted from physician notes) in the hold-out test dataset to assess model performance using the coefficient of determination ( $R^2$ ), the Pearson correlation coefficient ( $R$ ), and root-mean-squared error (RMSE). During our initial model development, performance was compared across a set of models developed by several machine learning approaches including

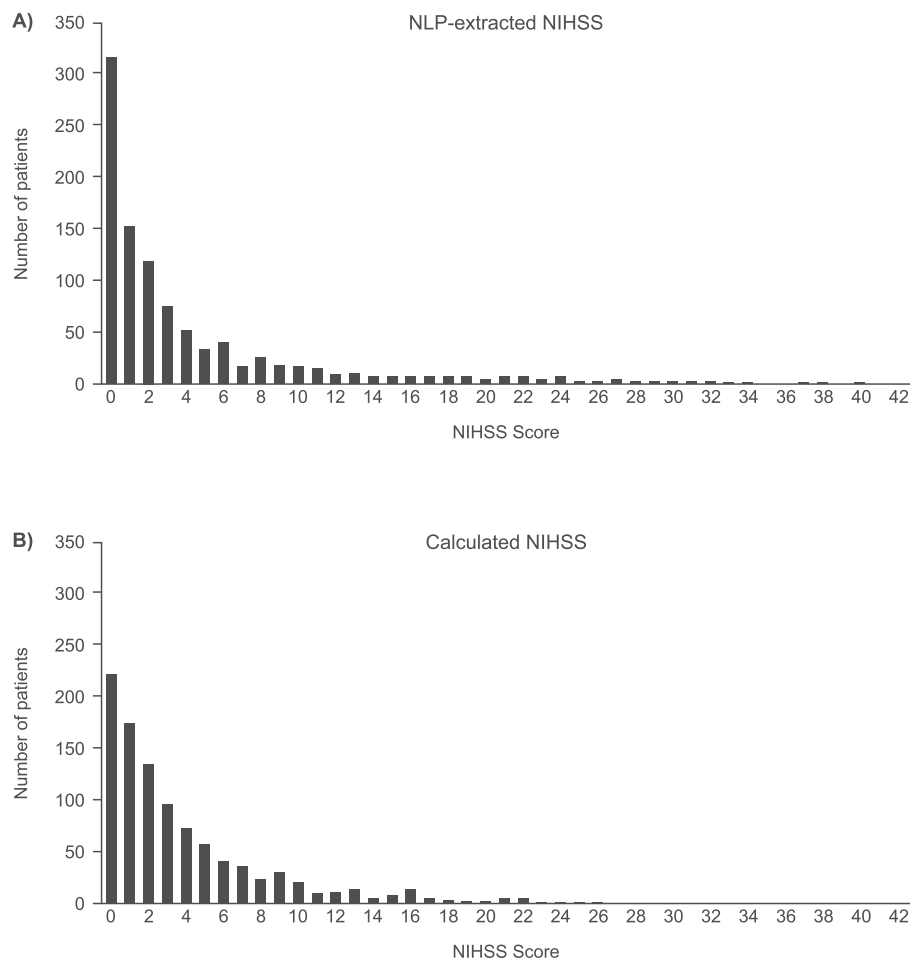
a random forest model, gradient boosting model, neural network, and linear regression. The random forest model, which is a meta estimator used to fit several classifying decision trees on various subsamples of the dataset, had the best performance. Model hyperparameters were optimized using a grid search and performance was evaluated using three-fold cross validation within the training data. (Additional file 1: Table S1). Recursive feature elimination performed on the training data reduced the 619 features further, with only the top 100 features included in the final model. The top 100 features were selected because only minor improvements in performance would have been gained for a substantial increase in model complexity with the inclusion of additional features (Fig. 3).

#### Results

Table 1 provides demographic data for patients at the time of stroke for each of the two cohorts (training and hold-out test set). Patients in the hold-out test set ( $n = 1033$ ) were only used once to assess the performance of the final optimized random forest model with 100 selected features. This ensured that performance metrics were not biased by over-fitting and that the model for predicting stroke severity scores is generalizable to data not previously included in model development.

The random forest model achieved an  $R^2$  of 0.57, an  $R$  of 0.76, and a RMSE of 4.5. Figure 4 presents imputed versus actual NIHSS scores. The median (interquartile range, IQR) NIHSS score in the hold-out test cohort was 2 (6) for both the real and imputed NIHSS scores. The distribution of the real NIHSS scores and imputed NIHSS scores for the hold-out test cohort are shown in Fig. 2.

A detailed list of the 100 features included in the final model is shown in Additional file 1: Table S2, ranked in order of relative importance. Top features included death within the same month as stroke occurrence, length of hospital stay following stroke occurrence, aphagia/dysphagia diagnosis, hemiplegia diagnosis, and whether a patient was discharged to home or self-care.

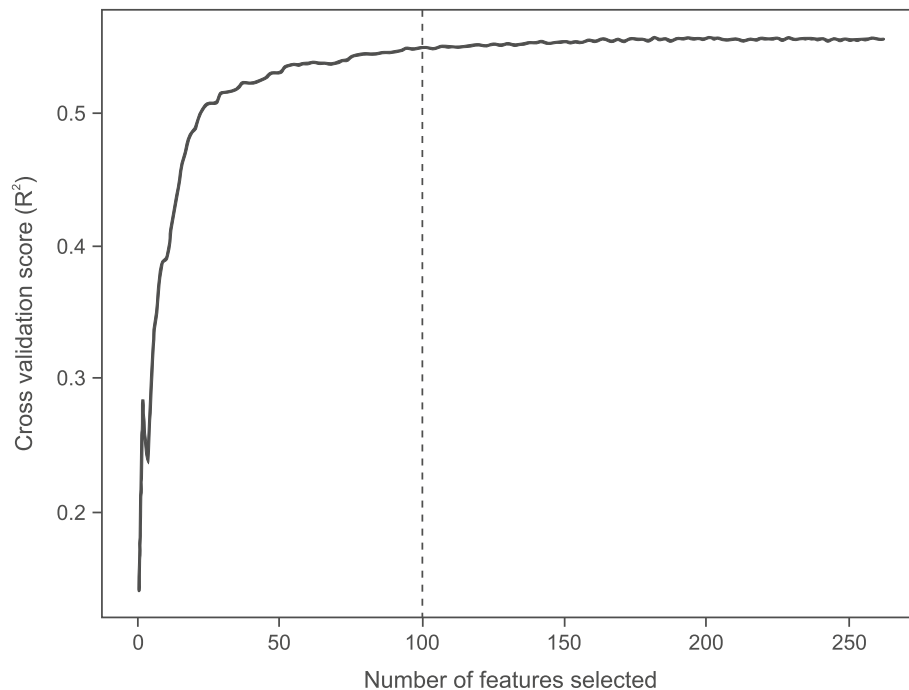


**Fig. 2** NIHSS Distribution. Distribution of (a) real NLP-extracted NIHSS scores and (b) model-imputed NIHSS scores for the hold-out test set ( $n = 1033$ ). Model-imputed scores were rounded down to nearest integer

## Discussion

The results of this study demonstrate that machine learning algorithms built on patient treatment and demographic information can be used to determine proxies for NIHSS scores in a real-world evidence database. This is a novel advancement in the ability to quantify stroke severity in large population-based cohorts. This methodology also has the advantage of obtaining these values without the significant cost and time investment required for retrospective chart review studies and enables assignment of a proxy NIHSS value even if the assessment was not completed or documented at the time of stroke occurrence. Applying this model to patients with stroke in the Optum® de-identified Integrated dataset who did not have a real NIHSS score increased the number of patients with stroke with a measure of stroke severity by 6-fold, thus significantly increasing the potential cohort size for any follow-on real-world evidence studies.

Although other methods have been used to determine proxy measures of stroke severity using clinical features available in EHR and claims databases, these methods (such as linear regression) often assume there are linear relationships between the features and stroke severity, which is not always the case [11]. For example, the length of a hospital stay varies based on stroke severity: patients who suffer a moderately severe stroke tend to have a longer length of hospital stay compared to those with low severity strokes who recover more quickly. However, patients with severe stroke have high mortality rates within the first few days of admittance for their strokes and therefore tend to have shorter hospital stays compared to patients with moderately severe stroke. Machine learning algorithms can identify non-linear relationships between features and stroke severity and can incorporate the complex relationships and interactions between features such as the clinical diagnoses relevant to stroke outcomes, treatments including medications



**Fig. 3** Feature selection based on cross-validated model performance. Feature selection based on cross-validated model performance on the training data. Mean model performance was calculated on 3-fold cross-validated subsets for feature selection. Maximum cross-validated performance was achieved with the top 226 features incorporated into the model; however, minimal gains in performance were seen beyond 100 features

and procedures administered to patients at stroke onset, as well as the medical history of patients prior to stroke diagnosis. Machine learning methods are thus well suited to the task of assessing stroke severity based on clinically available information from EHRs.

The distribution of imputed NIHSS scores calculated using this machine learning model as consistent with observations from the population-based Greater Cincinnati/Northern Kentucky Stroke Study [17]. The Cincinnati study determined NIHSS scores from a retrospective chart abstraction of 2233 ischemic stroke cases identified during a 12-month period, with median (IQR) NIHSS values of 3 (6) [17]. This is similar to the results obtained using the machine learning model, where the median (IQR) was 2 (6), and both studies showed a skewed distribution toward less severe strokes. The slightly lower median in the machine learning study may be caused by the inclusion of all types of stroke, including TIAs which tend to be much less severe compared to ischemic stroke. These results contrast with randomized clinical trials in which the enrolled patient population is often selected to include patients with more severe stroke. For example, the Albumin in Acute Stroke study required patients to have an ischemic stroke and baseline NIHSS score of 6 or higher [18] and the Desmoteplase in Acute Ischemic Stroke3 study required patients to have an ischemic stroke and NIHSS score of 4–24 [19].

Documentation of NIHSS scores was evaluated as part of the Get With The Guidelines – Stroke [20]. Over the 10-year study period from 2003 to 2012, the documentation rate was 56.1%, with a median NIHSS score of 4 (IQR, 2–9) and mean of 6.7 (SD, 7.4) [20]. Characteristics associated with NIHSS documentation were those related to eligibility for thrombolysis (e.g. arrival by ambulance and within 3 h of symptom onset) [20]. A modest selection bias was observed reflecting the tendency of hospitals with lower documentation rates to selectively report higher NIHSS scores. The ability to impute NIHSS score with machine learning algorithms may eliminate incomplete documentation issues.

The machine learning model described in this study was developed on a US-based data source and achieved similar performance to models developed from the more uniform, single payer, compulsory healthcare data from Taiwan [9–11]. The R was 0.76 for this US based study and ranged between 0.68 and 0.73 for different models developed in the Taiwan-based study. In the Taiwanese study, NIHSS scores were assessed on admission and recorded directly in the national stroke registry, patients were primarily managed by neurologists, and the features were based on medical billing codes rather than the diagnosis and procedure codes, as these are considered more accurate in Taiwanese health databases due to Taiwan's universal coverage for hospitalizations and

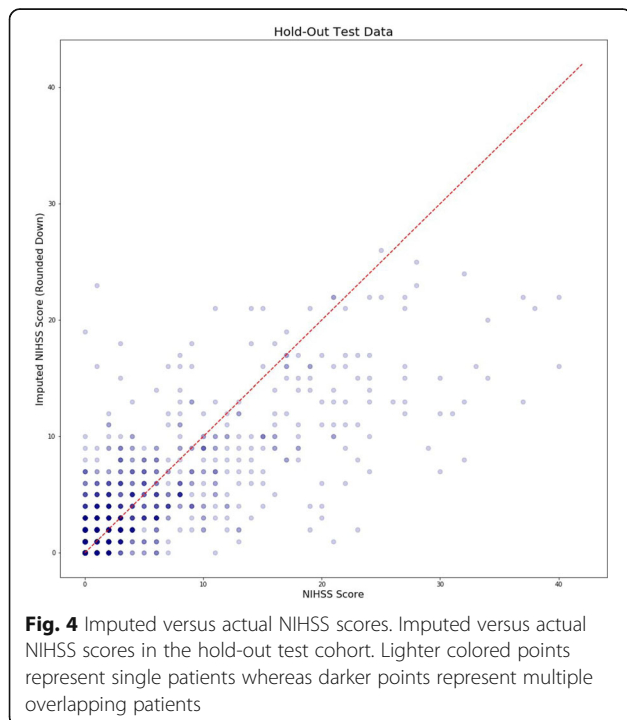
**Table 1** Patient Demographics and Characteristics

Characteristic	Training set (n = 6116)	Hold-out test set (n = 1033)	Overall population (n = 7149)
Demographics			
Age, mean (SD)	66 (14)	67 (14)	66 (14)
Female, n (%)	3196 (52)	568 (55)	3764 (53)
Region			
Northeast, n (%)	464 (8)	84 (8)	548 (8)
Midwest, n (%)	2388 (39)	389 (38)	2777 (39)
South, n (%)	2957 (48)	496 (48)	3453 (48)
West, n (%)	186 (3)	40 (4)	226 (3)
Other/Unknown, n (%)	121 (2)	24 (2)	145 (2)
EHR data			
NIHSS, median (IQR)	2 (6)	2 (6)	2 (6)
LOS, median (IQR)	3 (5)	2 (4)	3 (5)
Type of stroke <sup>a</sup>			
Ischemic, n (%)	4328 (70.8)	710 (68.7)	5038 (70.5)
Hemorrhagic, n (%)	605 (10.0)	113 (10.9)	718 (10.0)
TIA, n (%)	2235 (36.5)	384 (37.2)	2619 (36.6)
Charlson Comorbidity Index <sup>b</sup> , median (IQR)	1 (3)	1 (3)	1 (3)

SD standard deviation, EHR Electronic Health Record, NIHSS National Institutes of Health Stroke Scale, IQR interquartile range, LOS length of stay, TIA transient ischemic attack

<sup>a</sup>Based on ICD diagnosis codes during stroke event; more than one type may be coded per patient stroke event

<sup>b</sup>Calculated based on patients' diagnosis codes prior to stroke [15]



**Fig. 4** Imputed versus actual NIHSS scores. Imputed versus actual NIHSS scores in the hold-out test cohort. Lighter colored points represent single patients whereas darker points represent multiple overlapping patients

reimbursement system. In contrast, for this study using the US-based data source, real NIHSS scores were extracted using NLP from free text physician notes, attending physician specialty varied, and diagnostic and procedural coding can vary in the multi-provider US healthcare system. Using a more complex machine learning algorithm with significantly more features enabled the US-based model to achieve similar performance to the Taiwan-based model and demonstrates the ability of machine learning methods to handle the systematic differences from diverse EHR systems across various US providers.

This study is subject to several limitations worthy of consideration. Although the current EHR database has captured comprehensive information on diagnoses, administration of treatments and procedures during stroke occurrence, other information which could be critical for model performance including imaging of brain scans was not available. As with all studies based on real-world data, there is the potential for missing records. In addition, healthcare information in the database was not available until January 2007, which precluded the study from capturing information in patients who might have stroke-related diagnosis prior to the year 2007. As such, the first observed stroke occurrence in the data could be a mix of 1st and possibly later stroke diagnosis.



Moreover, generalizability of the model to another database remains unclear, as the current model was trained and validated only within a single database. As such, future work is planned to validate the current model in a different EHR database.

## Conclusions

Applying this machine learning method to assess patient's stroke severity in real-world databases where NIHSS scores are not available enables large scale health-economics and long-term patient outcome studies to incorporate stroke severity. However, in any such endeavor, it would be important to ensure the removal of all features related to any outcomes being studied (and model performance reassessed after removal) to avoid artificially elevated associations. These enhanced studies can potentially accelerate the development of better clinical management and improve patient quality of care [3, 8]. This study represents a novel advanced analytics application to real-world data that could significantly impact drug development and patient outcomes.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12911-019-1010-x>.

**Additional file 1: Table S1.** Random Forest Hyperparameters. The parameters of the final model which were obtained through hyperparameter optimization are presented here. **Table S2.** List of Final Model Features. Here, features are ranked by the expected fraction of the samples they contribute to as a measure of feature importance.

## Abbreviations

BETOS: Bergenson-Eggers Type of Service codes; CAFR: Chinese Atrial Fibrillation Registry; CMS: Centers for Medicare & Medicaid Services; CPT4: Current Procedural Terminology; EHR: Electronic health records; HCPCS: Healthcare Common Procedure Coding System; HIPAA: Health Insurance Portability and Accountability Act; ICD: International Classification of Diseases; IQR: Interquartile Range; NIHSS: National Institutes of Health Stroke Scale; NLP: Natural Language Processing; R: Pearson Correlation Coefficient; R<sup>2</sup>: Coefficient of Determination; RMSE: Root-mean-squared error; TIA: Transient Ischemic Attack

## Acknowledgments

The authors wish to thank Erik Sjoeland for his valuable guidance and input in drafting the manuscript. Editorial assistance was provided by Matthew Michelson, PhD, of Evid Science (El Segundo, CA) and Ashley O'Dunne, PhD, of MedErgy (Yardley, PA).

## Authors' contributions

EK, KT, DM contributed to conceptualization of the study. EK and KT contributed to data curation and formal analysis and validation and visualization. JH contributed to project administration work. EK, KT, DM, JHL, MA contributed to investigation and methodology of the study. DM contributed to funding acquisition. EK, KT, DM, JH, JHL and MA contributed to writing, review, and editing of the work. All authors reviewed and approved the manuscript for submission.

## Funding

This research and preparation of this manuscript were supported by Janssen Scientific Affairs, LLC. The funder did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available because the data source is owned by a third-party (Optum© Integrated Claims-Clinical dataset). Janssen Pharmaceuticals has a license for analysis of data from this source. As such, the authors cannot provide the raw data themselves. Other researchers could access the data by purchase through Optum©.

## Ethics approval and consent to participate

No patient's personal identifying information is included in this manuscript. All study data were accessed with protocols compliant with US patient confidentiality requirements, including the Health Insurance Portability and Accountability Act of 1996 regulations. Because this study was non-interventional and used only statistically deidentified patient records, it was exempt from institutional review board review.

## Consent for publication

Not applicable.

## Competing interests

EK, JH and DM are employees of Janssen; KT and JHL were employees of Janssen during the study and manuscript development; and MA is an employee of Hartford HealthCare.

## Author details

<sup>1</sup>Janssen Research & Development, LLC, Raritan, NJ, USA. <sup>2</sup>Janssen Scientific Affairs, LLC, Titusville, NJ, USA. <sup>3</sup>Hartford HealthCare, Hartford, CT, USA.

Received: 2 August 2019 Accepted: 17 December 2019

Published online: 08 January 2020

## References

- Katzan IL, Spertus J, Bettger JP, Bravata DM, Reeves MJ, Smith EE, et al. Risk adjustment of ischemic stroke outcomes for comparing hospital performance: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2014;45(3):918–44.
- National Center for Health Statistics. Health, United States, 2016: with Chartbook on long-term trends in health. Hyattsville: National Center for Health Statistics; 2017.
- Rost NS, Bottle A, Lee JM, Randall M, Middleton S, Shaw L, et al. Stroke severity is a crucial predictor of outcome: an international prospective validation study. *J Am Heart Assoc*. 2016;5(1). <https://doi.org/10.1161/JAHA.115.002433>.
- Fonarow GC, Saver JL, Smith EE, Broderick JP, Kleindorfer DO, Sacco RL, et al. Relationship of national institutes of health stroke scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. *J Am Heart Assoc*. 2012;1(1):42–50.
- Phan TG, Clissold BB, Ma H, Ly JV, Srikanth V. Predicting disability after ischemic stroke based on comorbidity index and stroke severity-from the virtual international stroke trials archive-acute collaboration. *Front Neurol*. 2017;8:192.
- Hage AV. The NIH stroke scale: a window into neurological status. *Nurs Spectr*. 2011;24(15):44–9.
- Samuel OW, Fang P, Chen S, Geng Y, Li G. Activity recognition based on pattern recognition of myoelectric signals for rehabilitation. In: Khan SU, Zomaya AY, Abbas A, editors. *Handbook of large-scale distributed computing in smart healthcare*. Basel: Springer International Publishing AG; 2017. [https://doi.org/10.1007/978-3-319-58280-1\\_16](https://doi.org/10.1007/978-3-319-58280-1_16).
- Fonarow GC, Alberts MJ, Broderick JP, Jauch EC, Kleindorfer DO, Saver JL, et al. Stroke outcomes measures must be appropriately risk adjusted to ensure quality care of patients: a presidential advisory from the American Heart Association/American Stroke Association. *Stroke*. 2014;45(5):1589–601.
- Sung SF, Hsieh CY, Kao Yang YH, Lin HJ, Chen CH, Chen YW, et al. Developing a stroke severity index based on administrative data was feasible using data mining techniques. *J Clin Epidemiol*. 2015;68(11):1292–300.
- Sung SF, Hsieh CY, Lin HJ, Chen YW, Chen CH, Kao Yang YH, et al. Validity of a stroke severity index for administrative claims data research: a retrospective cohort study. *BMC Health Serv Res*. 2016;16(1):509.

11. Sung SF, Chen SC, Hsieh CY, Li CY, Lai EC, Hu YH. A comparison of stroke severity proxy measures for claims data research: a population-based cohort study. *Pharmacoepidemiol Drug Saf.* 2016;25(4):438–43.
12. 45 CFR 164.514(b)(1) Implementation specifications: Requirements for de-identification of protected health information. Federal Government of the United States. <https://www.govinfo.gov/content/pkg/CFR-2017-title45-vol1/pdf/CFR-2017-title45-vol1-sec164-514.pdf>. Accessed 15 Feb 2019.
13. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Information Insurance Portability and Accountability Act (HIPAA) Privacy Rule (Dated as September 4, 2012, as first released on November 26, 2012).
14. Nunes AP, Yang J, Radican L, Engel SS, Kurtyka K, Tunceli K, et al. Assessing occurrence of hypoglycemia and its severity from electronic health records of patients with type 2 diabetes mellitus. *Diabetes Res Clin Pract.* 2016;121:192–203.
15. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care.* 2005;43(11):1130–9.
16. CMS' ICD-9-CM to and from ICD-10-CM and ICD-10-PCS Crosswalk or General Equivalence Mappings. <http://www.nber.org/data/icd9-icd-10-cm-and-pcs-crosswalk-general-equivalence-mapping.html>. Accessed 12 Apr 2018.
17. Reeves M, Khoury J, Alwell K, Moomaw C, Flaherty M, Woo D, et al. Distribution of National Institutes of Health stroke scale in the Cincinnati/northern Kentucky stroke study. *Stroke.* 2013;44(11):3211–3.
18. Ginsberg MD, Palesch YY, Hill MD, Martin RH, Moy CS, Barsan WG, et al. High-dose albumin treatment for acute ischaemic stroke (ALIAS) part 2: a randomised, double-blind, phase 3, placebo-controlled trial. *Lancet Neurol.* 2013;12(11):1049–58.
19. Albers GW, von Kummer R, Truelsen T, Jensen J-KS, Ravn GM, Grønning BA, et al. Safety and efficacy of desmoteplase given 3–9 h after ischaemic stroke in patients with occlusion or high-grade stenosis in major cerebral arteries (DIAS-3): a double-blind, randomised, placebo-controlled phase 3 trial. *Lancet Neurol.* 2015;14(6):575–84.
20. Reeves MJ, Smith EE, Fonarow GC, Zhao X, Thompson M, Peterson ED, et al. Variation and trends in the documentation of National Institutes of Health stroke scale in GWTG-stroke hospitals. *Circ Cardiovasc Qual Outcomes.* 2015; 8(6 Suppl 3):S90–8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

