Netw Model Anal Health Inform Bioinforma (2015) 4:16
DOI 10.1007/s13721-015-0090-5

ORIGINAL ARTICLE

Mining clinical text for stroke prediction

Elham Sedghi[1] · Jens H. Weber[1] · Alex Thomo[1] · Maximilian Bibok[2] · Andrew M. W. Penn[2]

Received: 26 April 2015 / Revised: 26 June 2015 / Accepted: 30 June 2015 / Published online: 14 July 2015

Abstract One of the main problems in treating stroke patients is accurate and timely triage and assessment. Not all stroke events have direct severe consequences. Full strokes are often preceded by transient ischemic attacks (TIA) or mini strokes, which exhibit signs and symptoms similar to less concerning health events, e.g., migraines. In this paper, natural language techniques are presented to process a large collection of medical narrative descriptions extracting features that can be subsequently used for automatic classification using Data Mining algorithms. We reviewed 5658 cases and analyzed the chief complaint and history of the patient illness reported at stroke rapid assessment unit (SRAU) at Victoria General Hospital (VGH). Data were collected by neurologists and stroke nurses between years 2008 and 2013. Based on a clinician-supplied list of important sign and symptom terms, we translated narrative medical text into well-codified sentences achieving an impressive agreement with a human expert. Afterwards, Data Mining algorithms were applied on codified data and obtaining not only prediction models, but also important weights for the codified terms. An extensive experimental evaluation of several classifiers is provided based on past data to predict new cases. Notably, we achieved a sensitivity of about 84 % and specificity of 64 % using support vector machines (SVM). The top terms identified by data mining algorithms were responsible for most of the prediction quality; therefore, they can be used to build a questionnaire-like, online application that can be employed as a first-line screening in triage for detecting stroke/TIA or mimic and help triage decide for the next step of treatment or discharge the patient.

✉ Elham Sedghi
elham_Sedghi@yahoo.com; esedghi@uvic.ca

Jens H. Weber
jens@uvic.ca

Alex Thomo
thomo@uvic.ca

Maximilian Bibok
Maximilian.Bibok@viha.ca

Andrew M. W. Penn
Andrew.Penn@viha.ca

[1] Department of Computer Science, University of Victoria, Victoria, BC, Canada

[2] SpecTRA Research Project, Vancouver Island Health Authority, Victoria, BC, Canada

1 Introduction

Statistics Canada reported stroke as the third leading cause of death in Canada in 2012; 6 % of all deaths were due to stroke, with women being the major victims (Heart and Stroke foundation 2015). Timely detection of stroke can contribute significantly in preventing long-term patient disability and can have a great impact in public health, reducing care costs and preventing expensive and potentially harmful neuro-imaging tests.

The objective of this work is building an effective software system for fast detection of stroke/TIA or mimic at the triage stage via the analysis of past clinical reports. A TIA, or a mini-stroke, starts just like a stroke but then resolves leaving no noticeable symptoms or deficits (NIH 2015). For almost all TIAs, the symptoms go away within an hour and there is no way to tell whether it will be just a passing problem or persist and lead to death or disability (NIH 2015); therefore, all the signs and symptoms gathered by clinicians are valuable information for diagnosis.

Medical data are often represented in semi-structured or unstructured form, including textual narrative. Natural language processing (NLP) methods help to locate and extract information within clinical narrative text and are useful to transform unstructured texts to data in a machine interpretable format.

After preprocessing with NLP, data mining techniques are helpful in analyzing and interpreting data and can be used to create appropriate models for predicting disease based on signs and symptoms. Several studies showed data mining as a successful approach for extracting information from electronic health records (Warrer et al. 2012; Cerrito 2001; Glasgow and Kaboli 2010).

There were several challenges we needed to address in our work. Typically clinical narratives contain misspelled terms and incorrect grammar; also, most of these reports are full of abbreviations and clinical acronyms that are not found in dictionaries (Fiszman et al. 1999). One of the challenges in this work was data preprocessing and negation detection. More specifically, we identified medical problems in patient records by extracting predefined sign and symptom terms (or keywords) provided by stroke specialists. Afterwards, we reviewed different negation detection methods and adopted existing methods to fit our problem context. With negation detection rules, we determined whether each key sign or symptom is present, unmentioned, or declared absent (mentioned in a negated context) and generated structured (codified) data for the data mining process. After preprocessing, data mining algorithms were utilized to analyze the data and build models for predicting stroke or TIA in patients. We systematically evaluated different data mining algorithms and computed standard metrics, such as recall (sensitivity), specificity, precision, F measure, and ROC for each algorithm.

A crucial product of the prediction models we learned from codified data was a list of keywords weighted by their importance in the prediction quality (as captured by sensitivity, specificity, etc.). The top k keywords (typically less than 30) of the list were usually responsible for more than 95 % of the prediction quality.

In other words, considering only the top k keywords gave us models that performed almost as well as their counterparts built on the full set of keywords. Having the top k keywords allowed building of a questionnaire-like, online, application for triage staff to use. This was effective because the number of the top keywords was small. The backend part of the online application is a prediction model, which outputs mimic or stroke/TIA. Based on this output, the triage staff can tell whether the patient needs to be hospitalized or can be discharged. The workflow, from data preprocessing to stroke prediction, is shown in Fig. 1.

As shown in Fig. 1, the model (which represents the vector of features with specific weight for each feature) can be used in any type of application program (e.g., mobile, desktop, or online application) and clinicians can easily determine which feature is present, absent or unknown via a check-box driven form (e.g., Q1: Does patient have headache? yes, no, unknown) and once all the features are entered, the model can be called as a function and specifies whether the patient needs to be hospitalized (because of stroke/TIA) or can be discharged (mimic).[1]

Specifically, our contributions are:

- We analyzed unstructured text to provide a model for stroke prediction, thus providing a first inexpensive screening for stroke/TIA detection prior to confirmation using MRI or CT scan.
- We present a detailed account of processing narrative medical text describing stroke patient visits; we address important aspects of this processing, such as accurate negation detection, and codification in terms of a signs and symptoms list not biased towards diagnosis.
- We present a detailed study using supervised machine learning to predict stroke/TIA versus mimic based on visit descriptions and methodically compare several algorithms across a multitude of dimensions.
- We present a detailed evaluation of the supervised learning algorithms not only using cross-validation, but also employing a separate test set collected later in time than the set used for training.

The paper is structured as follows: Sect. 2 discusses related work, Sect. 3 outlines the proposed approach and explains in detail all the steps of text processing and implementation of the negation detection rules. Section 4 explains our experimental results and Sect. 5 concludes the paper and outlines future work directions.

2 Related work

A variety of medical language processing systems were developed to extract patient data from specific medical reports. Some prominent examples are SymText to identify pneumonia related concepts on chest X-ray (Fiszman et al. 1999), Regenstrief Extraction Tool (REX) to extract pancreatic cyst patient data from medical text files (Al-Haddad et al. 2010), and MedLee (Friedman et al. 1994) which was initially used to process radiological reports of the chest and then extended to other domains (e.g., mammography reports) (Friedman et al. 1996).

[1] Implementing the application is out of the scope of this project, but it is under progress.

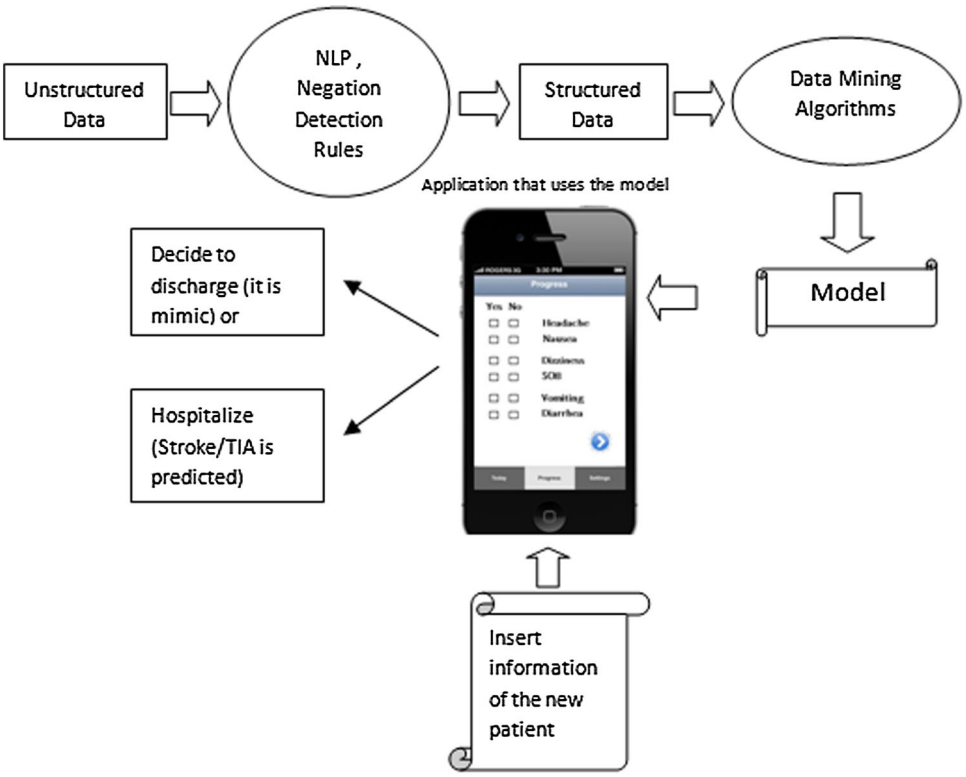

Fig. 1 The workflow from data processing to stroke prediction

There are also some studies that consider stroke prediction in particular by studying DNA and the number of single-gene disorders [cf. (Regnier 2012)]. Another study by Amini et. al. focused on prediction and control of stroke, however, it did not involve text mining but rather used a predefined list of factors to predict stroke (Amini et al. 2013).

As it has been shown by Elkins et al. (2000) and Hripcsak et al. (2002), NLP is useful for rapidly performing complex data acquisition and can be as accurate as expert human coders. While (Elkins et al. 2000) is a project in the stroke domain, it only deals with the neuroradiology reports of brain images, which are of a nature different from the medical description texts we consider in this paper.

3 Proposed approach

The study we report in this paper is part of a large-scale rapid stroke assessment project carried out at Victoria General Hospital (VGH) in British Columbia, Canada. The medical charts from 5658 patients collected between 2008 and 2013 at the SRAU were analyzed and the most important signs and symptoms were extracted from these data. The data were generated in SRAU and were unrelated to the Emergency Department (ED), even if the patient was referred to the unit from the ED. The fields were entered by keyboard and typed directly into the Stroke Guidance System (SGS) while consulting with a patient. Data elements included the patient medical history, the time when the stroke signs and symptoms first appeared, changes in symptoms over time, recent injuries, and information from the patients or bystanders, such as time of appearance of each problem or the last time the patient was without symptoms. Also, history of stroke, Transient Ischemic Attack (TIA), diabetes, hypertension and other clinical information were used in the analysis and the most important symptoms were extracted from these data.

A suite of regular expression scripts was developed to extract medical concepts and negative words. The extraction process is shown in Fig. 2. As mentioned, the work is classified into several phases and each phase is described in detail in the following.

3.1 Phase 1: data cleaning and splitting paragraphs into sentences

Patient medical history was stored in plain text in the system and records were distinguished from each other by the value of their primary key. In the first phase, the plain texts (paragraphs) were spell checked and a set of scripts were implemented to correct typographical errors and mistakes. The abbreviations were expanded to original

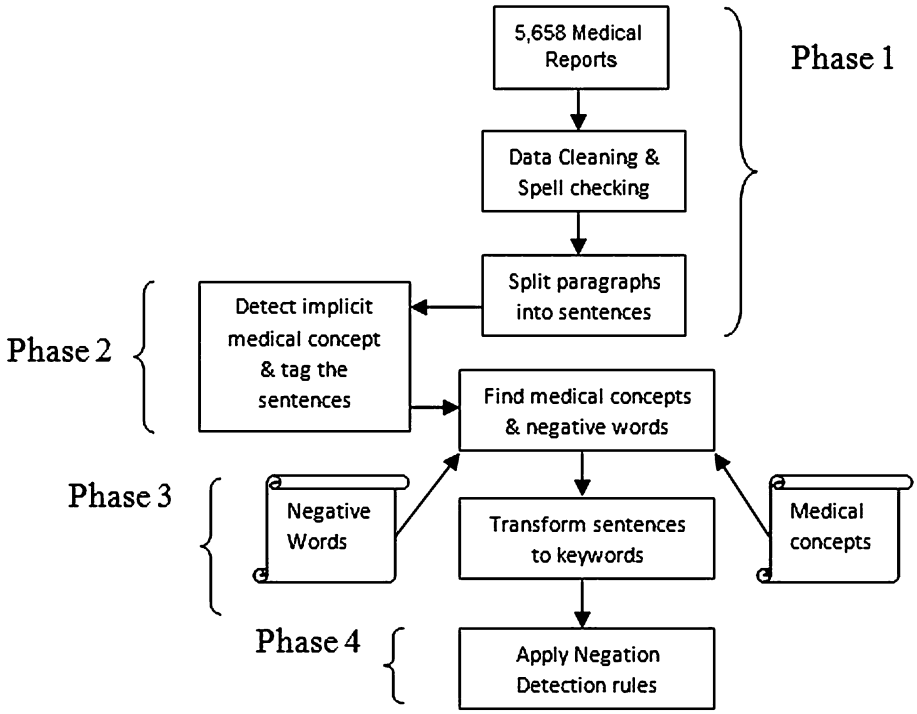

Fig. 2 The process of medical concept extraction

form (e.g., N/V was changed to nausea/vomiting, SZ was changed to seizure, etc.) and finally, a procedure was implemented to break paragraphs into sentences.

Tokenization is an important component of language processing and there is no widely accepted method to tokenize English texts, including biomedical texts (Barrett and Weber-Jahnke 2011). If the text contains well-formed sentences, then it may be possible to use existing software (e.g., Punkt) to segment text into sentences with few errors (Barrett and Weber-Jahnke 2011). In this study, regular expression commands were utilized to evaluate each period to determine whether it is sentence terminator or not. Periods followed by title prefix (e.g., Dr., Mr., Mrs., Miss.), periods used as a decimal point (e.g., 1.2 mg ASA), a row of three periods (...) that means "and so forth", and periods followed by date (e.g., September.21.2009) were all replaced by white space. Finally, a procedure was implemented to break the paragraphs into sentences using the remaining sentence separators (periods). We randomly selected 50 records consisting of 400 sentences and evaluated them manually to make sure the segmentation was done properly. The evaluation process is described in Phase 6.

3.2 Phase 2: detection of implicit medical concepts and tag sentences

In this project, signs and symptoms were target of the feature extraction process. The stroke terms (variables) used in this study were selected based on an exploratory data analysis of historical stroke data by attending neurologists and stroke clinicians at VGH. Overall, 126 signs and symptoms were defined as stroke signs and symptoms (terms) that were used as attributes to be extracted from the raw data set. Finding the top terms by data mining was the goal of our study to detect stroke/TIA or mimic and help triage decide for the next step of treatment or discharge the patient.

The negative words utilized in this project were partially borrowed from negation words/phrases used in NegEx (e.g., no, not, without, denies, etc.) (Wendy 2001) and the negative words in the Wiktionary website (Wiktionary 2013).

Once the list of negative words and medical concepts was compiled, a unique id was assigned to each term. For example, letter "n" was assigned to negative words (e.g., n1 = "no", n2 = "not", etc.), letter "k" to medical symptom terms (e.g., k1 = "HTC", k2 = "numbness", etc.), letter b to conjunctions meaning BUT (e.g., b1 = "but", b2 = "although", b3 = "however") and letter "s" to the laterality (e.g., s1 = left, s2 = right, etc.).

We implemented a program to find and tag phrases or sentences that explicitly referred to a predefined term. We defined a key for each term and the tagging application finds the appropriate terms or phrases for a given symptom and tags the sentence with the key assigned to that term. For example, diplopia means double vision and "k48" was assigned to this concept as a key or concept ID. Whenever the tagger application encounters the word diplopia or any phrase or keyword that means diplopia (e.g., "double

vision", "everything went double"), it tags that sentence with k48.

3.3 Phase 3: transforming sentences to sequences of negation signals and medical symptom terms

In this phase, the concepts were mapped to unique ids and each sentence was translated to a string containing n, k, b and s. For example, if s2 = right, k15 = hand, and k42 = numb, the following sentence:

"The patient had finished making dinner when her right hand went numb"
was translated to
"s2 k15 k42"
Another example is: "No leg weakness/numbness." It was translated to "n1 k26 k40 k42"

3.4 Phase 4: applying negation detection rules

In this phase, different negation detection methods were reviewed and the list of the negative words were borrowed from NegEx. NegEx identifies negation signals and negates all the medical terms within a window of five words of the negation signal (Kelleher and Mac Namee 2008). In the context of our study, we found that clinicians often negate more than five words; therefore, several rules were implemented to determine whether a concept is positive or negated. Also, compound sentences are usually composed of two or more independent clauses that are joined by conjunctions (e.g., "but") which alter the context of the second clause (Averbuch et al. 2004). The rules that we defined, assigned the appropriate value to each medical term and suitable data was prepared for applying data mining algorithms.

The medical concepts were extracted and inserted into a table of concept vectors. The concepts were divided into two categories: single concepts that included the body parts (e.g., face, eye, arm, leg, etc.) and the compound concepts that described the symptoms, signs (e.g., loss of consciousness, shortness of breath, etc.) and problems with specific parts of body (e.g., left eye droop, right arm numbness, etc.). −1 was assigned when the concept was negated, +1 when the concept was present, and 0 if the concept was absent or not mentioned in the sentence. As mentioned, several rules were defined to assign the correct value to the medical concepts. Rules were defined based on the order of k, n, and b in each sentence. The rules are described in the following.

- Rule 1: the value of concept K is +1 if K is not preceded by a negation signal N or if there is no N in the sentence. In the following sentence: "he has headache", the word "headache" is a medical concept which is translated to "k11" and its value is +1 because it is not negated.
- Rule 2: if a negative word, N, comes first, it negates the subsequent medical concept. In the following sentence: "he has no headache", headache is negated by "no", so the sentence is translated to "n1 k11" where "no" negates concept "headache". Thus, for this sentence, the value of headache is −1.
- Rule 3: if there is a conjunction indicating an exception B, such as "but", "although", "however", and the order of words in the sentence is NKBK, the value of the last concept is +1. For example: "The patient reports no diplopia but feels his left eye a bit droopy" is translated to "n1 k48 b1 s1 k31 k41". Here, "diplopia" or k48 is negated and its value is −1, but the value of the concept "left-eye-droop" is +1.
- Rule 4: if an exception B is indicated in the sentence and the order of words is KBNK, then the value of the last concept is -1. In the following sentence, "The patient has headache, but no dizziness", the value of headache is +1, but dizziness is negated, so the value of concept dizziness is −1.

At the end of the process, the sentences were combined into paragraphs and the results were summed up for each record. A given concept might appear more than once in the same paragraph. Unlike Negex, if a concept was positive at least once, then all occurrences of it in that paragraph were considered to be positive.

3.5 Phase 5: data mining process

Waikato environment for knowledge analysis (WEKA) is a popular suite of machine learning software and it contains a collection of algorithms and visualization tools for data analysis and predictive modeling (University of Waikato, New Zealand 2014). WEKA was employed in this study to apply different data mining algorithms on the stroke dataset.

To enhance the data set for analysis, fourteen attributes were added from SGS which were provided by the unit staff after the patient was examined. These data carried other patient information, such as age, gender, blood pressure, ABCD score[2] (Wikipedia 2014), smoking status, diabetes status, and so on.

By the end of this process, we had a data set with 140 attributes (variables) and 5,658 records. Recall that each sign and symptom term was represented by an attribute with values −1 (for negated), 0 (for absent or not mentioned), +1 (for present).

[2] The ABCD score alone did not give us acceptable levels of sensitivity and specificity.

The class values we considered for classification were "stroke/TIA" and "mimic". In other words, we perform binary classification of "stroke/TIA" versus "mimic". Differentiating between full stroke and TIA was not important for the clinicians in the project. This is because at triage stage, once someone is identified to have potentially suffered a stroke or TIA, the distinction is not crucial in the further process, i.e., in both the stroke and TIA cases, the patient will be immediately admitted to the hospital and proper tests will be done.

3.6 Phase 6: evaluation

To evaluate the correctness of term mapping and the quality of negation detection rules, we randomly selected 50 records from the set of 5658 reports and a human expert manually segmented and translated the sentences into structured terms. The expert detected the negation manually and assigned an appropriate value to each medical term. The results gained from automated negation rules were compared against the results provided manually for the selected records. The expert used Kappa statistics and the level of agreement between the negation detection rules and human evaluation was 0.9.[3]

To better understand the quality of classification methods, the data were divided into two sets: a training set (3520 records) and a test set (2138 records). The training set contained the data from four years (2008–2011) and the test set contained the data from 2 years (2012 and 2013). In total, the number of cases who experienced stroke/TIA were 3275 and the number of negative cases (mimic) was 2383. Tenfold cross-validation was also run; the results obtained were similar to those on the aforementioned test set.

4 Experimental results

The results on the full set of 5658 records with tenfold cross-validation showed that a parameter-tuned SVM with RBFKernel provides the highest accuracy (75.5 %) followed by logistic regression (73.67 %), NaiveBayes (72.48 %) and J48-Decision Tree (70.91 %). Normalizing the value of some of the continuous attributes, such as age and blood pressure (systolic and diastolic), did not provide significant change in the results.

Accuracy, however, is not considered most important in medical studies. Therefore, in the following we focus on recall (sensitivity), specificity, precision, *F* measure and ROC area. The results shown here were obtained using the test set of 2138 records as described earlier.

To establish a baseline for the algorithms, we first considered raw text data. SVM, logistic regression, and neural network provided the highest recall (sensitivity), about 80 %, among the rest of the methods. The detailed results are presented in Table 1. (Cross-validation results can be found in the supplementary file.)

Next the analysis was performed on codified data obtained as described in detail in the previous section. Table 2 shows the results of different classification methods on codified data. Again, SVM and logistic regression provided the highest recall (sensitivity) of over 83 %. The results show that training classifiers on codified data significantly outperformed training them on raw data. The main added benefit of using codified data is of course the ranked list of terms we obtain as a side effect of the data mining algorithms. Having a ranked list of terms allows building a user-friendly software application for use in the triage phase.

Besides recall (sensitivity) in Fig. 3a, specificity, precision, *F* measure and ROC area were also computed for both approaches and the results are shown in separate figures. In each figure, each classifier is represented by two bars, the first showing the performance achieved using raw data and the second showing the performance achieved using codified data.

Figure 3b presents the specificity gained from different methods using raw text and codified data. Specificity is the number of true negatives (TN) divided by the total number of negatives (N). A sensitivity level of 79 % or greater and a specificity level of 60 % or greater were deemed suitable by the clinicians in this project. SVM provided 84 % sensitivity and 63.3 % specificity, and logistic regression provided 83.2 % sensitivity and 63.7 % specificity using the codified data.

F measure and precision are depicted in Fig. 3c, d, respectively. SVM and logistic regression provided the highest *F* measure (78.1 and 77.8 %, respectively) using codified data. The differences in precision between the two approaches did not turn out to be significant. The values provided by different algorithms varied between 58.6 and 79.2 % for raw text classification and 64.4–76.4 % with codified approach.

Receiver operating characteristic (ROC) curves are an important outcome measure as they display the trade-off between sensitivity and specificity (Florkowski 2008).

The ROC curve results are shown for both approaches in Fig. 4.

Remark We would expect that working with codified data will be better than working with raw text. However, in reality the process of codifying data from raw text may also

[3] If the estimate is 0.8 or above, there is excellent agreement between the algorithm and the human assessment, the score between 0.6 and 0.8 is considered good agreement (Goryachev et al. 2006).

Table 1 Results using raw text data

	Logistic (%)	Naive bayes (%)	SVM (%)	Network neural (%)	IBK (%)	J48 (%)	Random forest (%)
Recall (sensitivity)	79.9	64.7	79.8	79.1	77.6	72.5	75.3
Specificity	57.2	80.0	61.9	64.8	35.3	52.5	54.5
Precision	68.7	79.2	71.2	72.6	58.6	64.3	66.1
F measure	73.9	71.2	75.2	75.7	66.8	68.1	70.4
ROC	76.3	78.4	70.8	79.0	58.0	62.1	70.7

Table 2 Results using codified data

	Logistic (%)	Naive bayes (%)	SVM (%)	Neural network (%)	IBK (%)	J48 (%)	Random forest (%)
Recall (sensitivity)	879	6192	8495	639p	6195	6896	8291
Specificity	1796	1896	1797	6.94	3593	3495	3391
Precision	6795	6492	6795	6194	1494	1198	1894
F measure	6698	6392	6892	6192	1p96	6.97	6494
ROC	8.92	6p9p	6796	8597	179	1693	6192

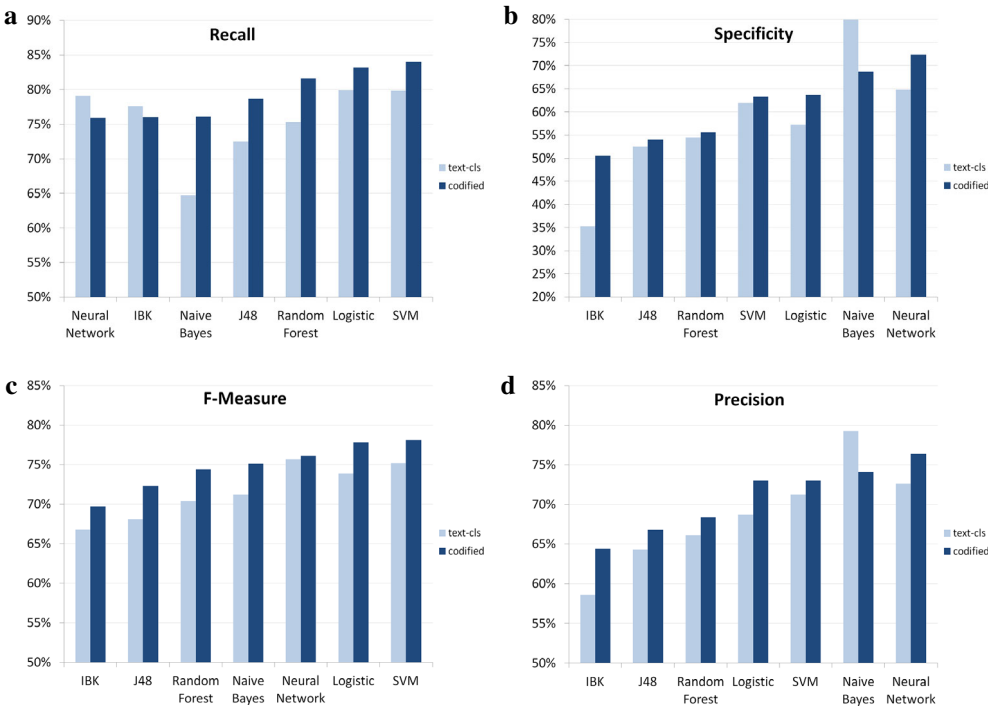

Fig. 3 **a** Recall for different methods. **b** Specificity for different methods. **c** *F* measure for different methods. **d** Precision for different methods

be a source of error. It is, therefore, to be expected that the observed improvements are not necessarily consistent across all measures. Nonetheless, we argue that still, working with codified data in this study, gives better performance overall even in those cases when some measures score better for raw text.

Specifically, let us focus, for example, on the Naive Bayes (NB) classifier, which shows the biggest positive difference in specificity for raw text compared to the codified approach. This is indeed true for the default classification threshold of 0.5. [Recall that Naive Bayes produces in fact a probability score, which is compared to a

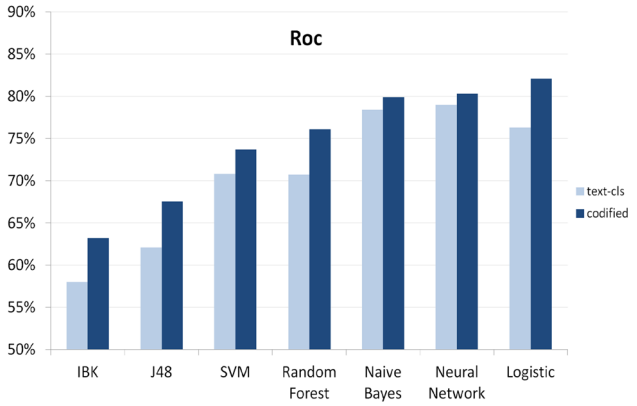

Fig. 4 ROC area plotted for different methods.

threshold to produce a binary classification.] However, NB with raw text scores quite poorly with respect to recall (sensitivity) compared to the codified approach.

For different classification thresholds, both the recall (sensitivity) and specificity will vary. We can better see the sensitivity/specificity behavior by building an ROC curve for each approach (raw text and codification). An ROC curve plots sensitivity versus (1-specificity) and each point in it corresponds to a different classification threshold. To not rely on visual inspection of different points in the ROC curves, we use the well-known measure of the area under the curve (AUC). The bigger the AUC, the better in general the classifier. With respect to AUC (see Fig. 4), NB performs better when using codification than raw text. In fact, we see that all the classifiers we consider perform better with respect to AUC when using codification.

A similar discussion can be made for the recall/precision combination. Naive Bayes scores better with respect to precision when using raw text. However, it is worst in terms of recall. Typically, we combine recall and precision into their harmonic mean, which is the *F* measure. In terms of the latter, NB does worst using raw text than codification.

Finally, we report here few of the most important concepts discovered by our data mining process. The important warning signs and symptoms to detect stroke/TIA were face droop, visual loss, diplopia, language disturbance, speech disturbance, swallow difficulty, drunk, drag, etc. In addition, some of the most important concepts to detect mimic were headache, seizure, migraine, anxiety, fatigue, amnesia, photophobia, tremor, stress, etc.

5 Conclusions

Natural language processing (NLP) methods were used to identify and extract information from medical charts of patients collected between 2008 and 2013 at SRAU of Victoria General Hospital (VGH). The unstructured texts narratives were transformed to codified data in a computable format and data mining methods were utilized to build models for stroke/TIA prediction. Our clinical NLP-based system consists of several components to extract and analyze data.

Various algorithms were utilized on codified data and compared against baselines on raw data. Evaluation metrics were computed for both approaches and showed that the codification approach outperformed the approach using raw data. The recall (sensitivity) provided by SVM and logistic regression showed that these two classifiers can provide reliable models to predict stroke/TIA based on patients' signs and symptoms instead of immediately using costly tests, such as MRI or CT scan.

The list of symptoms that play the most important role in stroke detection can be identified via the machine learning process. This list can be used in stroke assessment forms to help stroke nurses to decide for the next step of treatment in a more timely fashion.[4]

In this study, we used the history of patient illness and chief complaint information of the patients who experienced stroke/TIA or discharged with the conditions that mimic the symptoms of stroke to implement a model for stroke prediction in new patients with the same symptoms. In this analysis, we considered two class values "stroke/TIA" and "mimic", but future analysis will contain more possible values to determine different types of stroke (e.g., PACS, POCS, TACS, LACS, etc.) and different types of mimics such as migraine, TGA, BPV, etc.

Acknowledgments The authors would like to acknowledge Kristine Votova, Ph.D., the project manager for the SpecTRA Research Project and the Island Health clinical research team at the Stroke Rapid Assessment Unit for their support. Funding for the natural experiment in stroke care and the large-scale personalized medicine for mass spectrometry in rapid TIA triage comes from Canadian Institute of Health Research (2009–2012) and Genome Canada/BC (2013–2017).

References

Al-Haddad MA, Friedlin J, Kesterson J, Waters JA, Aguilar-Saavedra JR, Schmidt CM (2010) Natural language processing for the development of a clinical registry: a validation study in intraductal papillary mucinous neoplasms. HPB 12(10):688–695

Amini L, Azarpazhouh R, Farzadfar MT, Mousavi SA, Jazaieri F, Khorvash F, Norouzi R, Toghianfar N (2013) Prediction and control of stroke by data mining. Int J Prev Med 4(2):245

[4] Due to IP restrictions, we cannot provide here the list of important terms and the weights we derived for them. However, the interested readers can contact Dr. Andrew Penn on how to obtain this information. Also, the front-end application is not part of this study. Again, details on the front-end application can be obtained from Dr. Andrew Penn.

Averbuch M, Karson T, Ben-Ami B, Maimon O, Rokach L (2004) Context-sensitive medical information retrieval. In: Proceedings of the 11th World Congress on Medical Informatics (MEDINFO-2004), Citeseer. 1–8

Barrett N, Weber-Jahnke J (2011) Building a biomedical tokenizer using the token lattice design pattern and the adapted viterbi algorithm. BMC Bioinform 12(3):1

Cerrito P (2001) Application of data mining for examining polypharmacy and adverse effects in cardiology patients. Cardiovasc Toxicol 1(3):177–179

Elkins JS, Friedman C, Boden-Albala B, Sacco RL, Hripcsak G (2000) Coding neuroradiology reports for the northern manhattan stroke study: a comparison of natural language processing and manual review. Comput Biomed Res 33(1):1–10

Fiszman M, Chapman WW, Evans SR, Haug PJ (1999) Automatic identification of pneumonia related concepts on chest X-ray reports. In: Proceedings of the AMIA Symposium, American Medical Informatics Association. 67

Florkowski CM (2008) Sensitivity, specificity, receiver-operating characteristic (roc) curves and likelihood ratios: communicating the performance of diagnostic tests. Clin Biochem Rev 29(1):S83

Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB (1994) A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1(2):161–174

Friedman C, Shagina L, Socratous SA, Zeng X (1996) A web-based version of medlee: a medical language extraction and encoding system. In: Proceedings of the AMIA Annual Fall Symposium, American Medical Informatics Association. 938

Glasgow JM, Kaboli PJ (2010) Detecting adverse drug events through data mining. Am J Health Syst Pharm 67(4):317–320

Goryachev S, Sordo M, Zeng QT, Ngo L (2006) Implementation and evaluation of four different methods of negation detection. DSG, Boston

Heart and Stroke foundation (2015) Statistics. http://www.heartandstroke.com/site/c.ikIQLcMWJtE/b.3483991/k.34A8/Statistics.htm. Accessed Jan 2015

Hripcsak G, Austin JH, Alderson PO, Friedman C (2002) Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports 1. Radiology 224(1):157–163

Kelleher JD, Mac Namee B (2008) A review of negation in clinical texts: dit technical report: Soc-aig-001-08. http://www.comp.dit.ie/bmacnamee/papers/negationinclinicaltexts_article.pdf

NIH (2015) Stroke, hope through research. http://www.ninds.nih.gov/disorders/stroke/detail_stroke.htm. Accessed Jan 2015

Regnier M (2012) Focus on stroke: predicting and preventing stroke. http://blog.wellcome.ac.uk/2012/05/07/focus-on-stroke-predicting-and-preventing-stroke/. Accessed Jan 2015

University of Waikato, New Zealand (2014) Weka (machine learning). http://en.wikipedia.org/wiki/Weka(machine learning). Accessed Dec 2014

Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L (2012) Using text-mining techniques in electronic patient records to identify adrs from medicine use. Br J Clin Pharmacol 73(5):674–684

Wendy W (2001) Chapman, will bridewell, paul hanbury, gregory f. cooper, and bruce g. buchanan. 2001. a simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 34(5):301–310

Wiktionary (2013) Category:english words suffixed with -n't. http://en.wiktionary.org/. Accessed Dec 2014

Wikipedia (2014) Abcd score. http://en.wikipedia.org/wiki/ABCD_score. Accessed Dec 2014