

RESEARCH ARTICLE

Open Access



Early recognition of multiple sclerosis using natural language processing of the electronic health record

Herbert S. Chase^{*}, Lindsey R. Mitrani, Gabriel G. Lu and Dominick J. Fulgieri

Abstract

Background: Diagnostic accuracy might be improved by algorithms that searched patients' clinical notes in the electronic health record (EHR) for signs and symptoms of diseases such as multiple sclerosis (MS). The focus this study was to determine if patients with MS could be identified from their clinical notes prior to the initial recognition by their healthcare providers.

Methods: An *MS-enriched cohort* of patients with well-established MS ($n = 165$) and controls ($n = 545$), was generated from the adult outpatient clinic. A *random sample cohort* was generated from randomly selected patients ($n = 2289$) from the same adult outpatient clinic, some of whom had MS ($n = 16$). Patients' notes were extracted from the data warehouse and signs and symptoms mapped to UMLS terms using MedLEE. Approximately 1000 MS-related terms occurred significantly more frequently in MS patients' notes than controls'. Synonymous terms were manually clustered into 50 buckets and used as classification features. Patients were classified as MS or not using Naive Bayes classification.

Results: Classification of patients known to have MS using notes of the *MS-enriched cohort* entered after the initial ICD9[MS] code yielded an ROC AUC, sensitivity, and specificity of 0.90 [0.87-0.93], 0.75[0.66-0.82], and 0.91 [0.87-0.93], respectively. Similar classification accuracy was achieved using the notes from the *random sample cohort*. Classification of patients *not yet known* to have MS using notes of the *MS-enriched cohort* entered *before* the initial ICD9[MS] documentation identified 40% [23–59%] as having MS. Manual review of the EHR of 45 patients of the *random sample cohort* classified as having MS but lacking an ICD9[MS] code identified four who might have unrecognized MS.

Conclusions: Diagnostic accuracy might be improved by mining patients' clinical notes for signs and symptoms of specific diseases using NLP. Using this approach, we identified patients with MS early in the course of their disease which could potentially shorten the time to diagnosis. This approach could also be applied to other diseases often missed by primary care providers such as cancer. Whether implementing computerized diagnostic support ultimately shortens the time from earliest symptoms to formal recognition of the disease remains to be seen.

Keywords: Early Diagnosis [E01.390], Diagnostic errors [E01.354], Diagnosis, Computer-assisted [E01.158], Electronic health records [E05.318.308.940.968.249.500], Natural language processing [L01.224.050.375.580], multiple sclerosis [C10.114.375.500]

Background

Accurate and timely medical diagnosis is the sine qua non for optimal medical care. Unfortunately, diagnostic error, which includes delayed, incorrect or missed diagnosis, is common and accounts for a significant proportion of medical error [1, 2]. Studies looking at a variety of sources such

as malpractice claims, autopsy studies and manual review of patients' notes have revealed that misdiagnoses of disease as a common problem that leads to harm including death. Frequent errors include failure to order or interpret a test, follow-up on finding an abnormal test result or complete a proper history or physical exam [3, 4]. In the outpatient setting, the overall missed diagnosis rate has been estimated to be 5% [5, 6] and has been reported to be as high as 20% for various cancers [7] and chronic kidney disease [8].

* Correspondence: hc15@cumc.columbia.edu
Department of Biomedical Informatics, Columbia University Medical Center, PH-20, 622 West 168th street, New York, NY 10032, USA

Primary care providers, who are usually first to see patients with medical complaints, are well aware that a root cause of diagnostic error is inadequate knowledge and failure to consider a diagnosis [9–13]. They cannot be expected to have mastered the signs and symptoms of the literally thousands of conditions, many belonging to a rarified domain of specialists. With the sustained exponential growth of medical knowledge, accurate diagnosis of patients' conditions seems increasingly unachievable without assistance from computerized diagnostic support [14].

There are several diagnostic tools such as Isabel® (Isabel Healthcare Inc., USA), Watson® (IBM, Inc. USA), and DxPlain® (Massachusetts General Hospital, Laboratory of Computer Science, Boston, MS) that return a differential diagnosis based on signs, symptoms and demographic information entered by the provider. Although these tools have been shown to improve diagnosis [15–18] they are underutilized for a variety of reasons including the necessity of actively entering data and lack of integration into the EHR [19, 20].

A diagnostic tool that raises awareness of a specific disease could be integrated into the EHR. Illnesses characterized by abnormalities in structured data, easily accessible to a computer, such as laboratory values or vital signs, could be directly diagnosed by mining the patient's data. Some examples include identification of patients with chronic kidney disease [21, 22], acute kidney injury [23, 24], anemia [25] and sepsis [26].

Many diseases, however, are not characterized by abnormalities manifested in routine laboratory tests. Consider neurological diseases, such as dementia, multiple sclerosis, myasthenia gravis or Parkinson's disease, or cancer, such as pancreatic or ovarian, illnesses characterized in their earliest stages by subtle signs and symptoms [27–30]. Implementing a diagnostic assistant to achieve early identification of patients with these illnesses is far more challenging because the computer would have to “read” the unstructured narrative portion of the medical record to identify the signs and symptoms characteristic of the particular disease using natural language processing (NLP). Based on the presence or absence of these signs and symptoms, an algorithm could provide an estimate of the likelihood that the patient had the illness in question.

The goal of this study was to test the hypothesis that patients with diseases characterized by subtle signs and symptoms could be identified early in the course of their illness using NLP of the narrative portion of the clinical notes in the EHR. We chose multiple sclerosis (MS) to explore this hypothesis because a formal diagnosis of MS is commonly missed [31] or delayed [32, 33] placing the patient at risk for irreversible complications. Early diagnosis is the key to treatment success [34, 35].

Methods

Study population

The patients whose notes were used for this IRB-approved study attend the Associates in Internal Medicine (AIM) clinic at Columbia University Medical Center (CUMC). The clinic is staffed by 150 medical residents and attending faculty who care for approximately 40,000 patients. Two patient cohorts were established for this study: an *MS-enriched cohort* consisted of patients with well-established MS and randomly selected controls (see Table 1). We used the presence (or absence) one or more ICD9 codes for MS (ICD9[MS]), “340”, as the gold standard for the presence (or absence) of MS. The *MS-enriched cohort* was used to identify predictive attributes, develop a classification model and determine achievability of early recognition. A *Random sample cohort* consisted of randomly selected patients from the AIM clinic, some of whom had MS. This cohort was used to identify patients who might have unrecognized MS. Only those MS and controls patients with clinical notes spanning at least two consecutive years in their record were included in the study.

Extraction and mapping of MS-related signs and symptoms from clinical notes

Classification models were developed using attributes (features) consisting of the well-known signs and symptoms of MS (terms), extracted from patients' free-text clinical notes using MedLEE, a natural language processing tool that recognizes terms and maps them to the United Medical Language System (UMLS) concept unique identifiers (CUIs) [36]. The patients' EHR notes were sourced from the clinical data warehouse of CUMC of the New York Presbyterian (NYP) system.

We gathered two collections (sets) of notes from the *MS-enriched cohort* based on the timing of the notes in relation to the date of the first entry of an ICD9[MS], (see Fig. 1). The *post-ICD9[MS]* set were notes written about the patients with MS two or more years *after* the initial entry of an ICD9 code for MS (ICD9[MS]) into the EHR, when the diagnosis was clearly established. The *pre-ICD9[MS]* set were notes on these same patients written *prior* to the initial entry of an ICD9[MS] code. Notes of the control patients were entered up to two years prior to the most recent note in the EHR. The

Table 1 Two cohorts used for MS classification study

Cohort	Purpose	Unique Patients	
MS-enriched	Identify predictive attributes, develop a classification model, and determine the achievability of early recognition	MS	165
		Controls	545
Random sample	Identify patients with unrecognized MS	MS	16
		Controls	2273

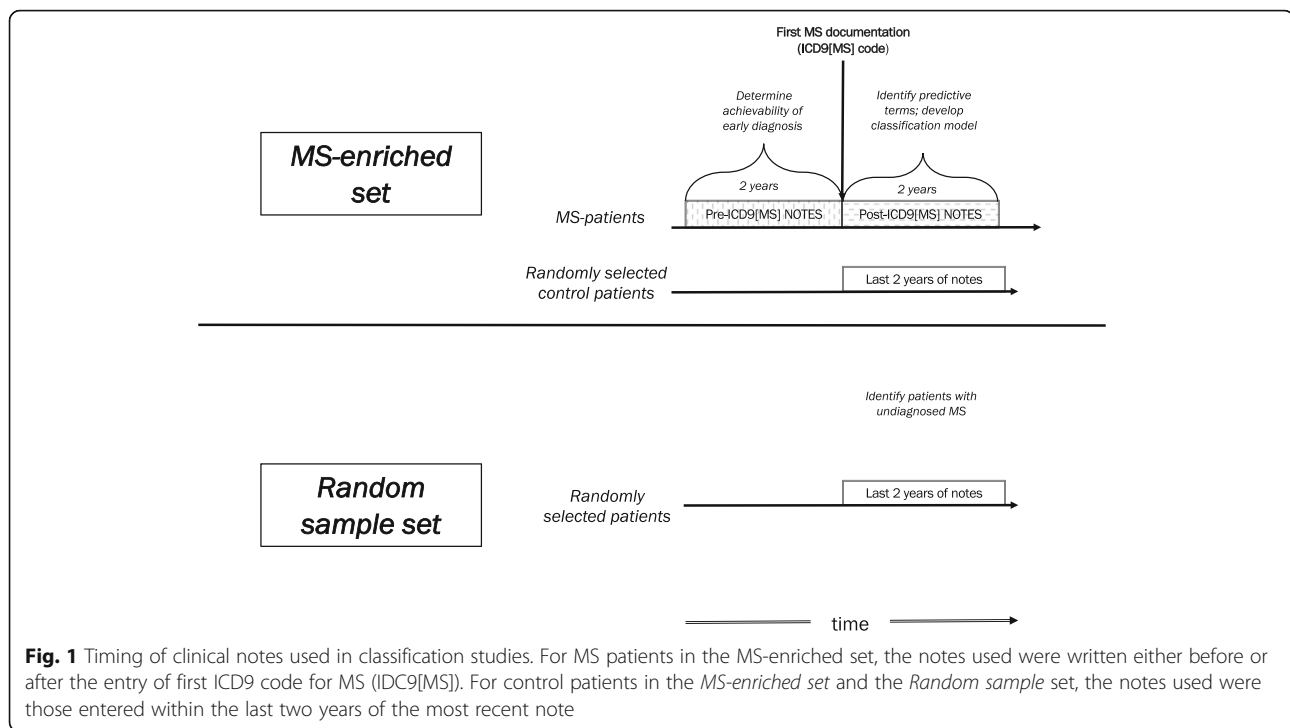


Fig. 1 Timing of clinical notes used in classification studies. For MS patients in the MS-enriched set, the notes used were written either before or after the entry of first ICD9 code for MS (ICD9[MS]). For control patients in the MS-enriched set and the Random sample set, the notes used were those entered within the last two years of the most recent note

Random sample set were notes entered up to two years before the last entry in the EHR in patients of the Random Sample cohort (see Fig. 1).

Classification

We trained a classifier to distinguish patients with known MS from those without MS, using the presence or absence of MS-related terms in patients' notes. We used the Naïve Bayes classification algorithm from the Weka suite to distinguish patients with MS from those without, having previously determined that this algorithm yielded the best results [10]. We assessed classification accuracy by the receiver operating characteristics area under the curve (ROC AUC), sensitivity and specificity and 95% confidence intervals.

Results

Characteristics of MS patients and control cohorts

The average age of MS patients in the MS-enriched cohort was 47.3 (SD 15.0), significantly lower than controls 54.1 (SD 16.6) ($p < 0.001$). Of the MS patients, 85.1% were female compared to 68.7% of the controls ($p < 0.001$). Both age and gender were subsequently used as attributes in classification.

Attribute (feature) selection

The first step in building an MS classifier was to identify terms in the patients' notes that communicated MS (such as "paresis" or "numbness") that should be more prevalent in patients with MS compared to controls.

From the *post-ICD9[MS]* set of notes of the MS-enriched cohort unique terms were extracted of which 1,057 were significantly more frequent in MS patients than in controls (Chi-squared analysis and an odds ratio at 5% significance level). Examples of terms with high odds ratios (CI) relating to MS symptoms include weakness in legs, 38.8 (2.3, 693.4), orbital pain 34.9 (2.0, 614.3) and paresthesia foot 38.8 (2.3, 693.4).

To amplify the signal and the potential value of pooling synonymous UMLS terms (CUIs) we manually aggregated individual synonyms into 66 individual buckets. For example, the 14 unique terms representing "paresis" (such as "facial paresis," "hemiparesis (right)," "hemiparesis (left)," and "limb paresis") were aggregated into the single bucket "paresis". The potential attribute strength of each of the buckets was measured by a Chi Squared comparison of the frequency of the terms in MS patients and controls. Of the original 66 buckets, 50 proved to have high attribute strength and were used in the classification modeling (see Table 2). Buckets were grouped to demonstrate the broad range of signs and symptoms of MS as well as the many related but specific categories of loss of function or disability. For example, there are nine different but related buckets of terms that reflect complications and involvement of the motor system.

For classification, if a patient had a single mention of any one of a particular bucket terms in any of their clinical notes they were scored a 1 for that bucket (and a 0 if not). Thus, for each of the MS and control patients, attribute columns were filled according to the presence or absence

Table 2 Buckets of MS-related UMLS terms used as attributes for MS classification. For each bucket, the frequency of terms in MS patients was compared to controls. The significance of the difference was measured by Chi squared analysis (in parenthesis). Buckets are grouped by common symptoms of loss of function

	BUCKET	SYMPTOM GROUP
1	Bladder dysfunction (115.8)	<i>Autonomic dysfunction</i>
2	Constipation (22.4)	
3	Memory (46.4)	<i>Cognition</i>
4	Cognitive (19.4)	
5	Ataxia (48.0)	<i>Coordination</i>
6	Balance (77.1)	
7	Cerebellar (25.0)	
8	Coordination (89.5)	
9	Dizziness (29.9)	<i>Dizziness and vertigo</i>
10	Vertigo (17.2)	
11	Diplopia (43.8)	<i>Eye and Vision</i>
12	Nystagmus (48.6)	
13	Optic neuritis (79.2)	
14	Orbital pain (30.0)	
15	Vision (88.9)	
16	Fatigue (30.7)	<i>Fatigue</i>
17	Weak (189.0)	
18	Headache (65.8)	<i>Headache</i>
19	Migraine (22.5)	
20	Depression (24.2)	<i>Mood</i>
21	Mood (36.5)	
22	Pain-musculoskeletal (12.1)	<i>Pain</i>
23	Pain-other (7.3)	
24	Atrophy (46.5)	<i>Motor</i>
25	Contracture (19.5)	
26	Dysphagia (23.4)	
27	Motor (59.6)	
28	Paresis (87.5)	
29	Reflex (50.6)	
30	Spastic (76.9)	
31	Speech (126.7)	
32	Stiffness (35.6)	
33	Burning (20.0)	<i>Sensory</i>
34	Lhermitte's sign (11.0)	
35	Neuritis (82.8)	
36	Numbness (93.2)	
37	Paresthesia (59.6)	
38	Sensory (67.9)	
39	Tingling (94.5)	

Table 2 Buckets of MS-related UMLS terms used as attributes for MS classification. For each bucket, the frequency of terms in MS patients was compared to controls. The significance of the difference was measured by Chi squared analysis (in parenthesis). Buckets are grouped by common symptoms of loss of function (Continued)

40	Epilepsy (13.3)	<i>Tremor, seizure</i>
41	Palsy (13.8)	
42	Seizures (33.9)	
43	Tremor (12.9)	
44	Fall (57.3)	<i>Walking (gait)</i>
45	Gait (79.2)	
46	Walk (45.8)	
47	Hearing loss (9.0)	<i>Miscellaneous</i>
48	Neurologic (43.4)	
49	Neuropathy (22.1)	
50	Sleep (12.6)	

of any one of the several terms within a bucket. If a patient had mention of a term on multiple dates or multiple terms in a bucket, he or she would still receive only a 1 for the bucket.

Classification of patients known to have MS

The first step to developing the diagnostic assistant was to determine if a classification approach using the extracted signs and symptoms of MS could be used to identify the MS patients already known to have the disease. We developed an MS classification model using the notes of patients with well-established MS (*post-ICD9[MS]* notes of the *MS-enriched set*) and controls. The classification model yielded an ROC AUC, sensitivity and specificity of 0.90 [0.87–0.93], 75% [66–82%], and 91% [87–93%] respectively.

Classification of the *Random sample set* enabled us to obtain a more accurate measurement of sensitivity and specificity of classification of patients known to have MS as well as the true prevalence of MS in our adult population which would allow calculation of positive and negative predictive values [37, 38]. Classification of the *Random sample set*, using the model described above, yielded an ROC AUC, sensitivity and specificity of 0.94 [0.93 - 0.95], 81% [54–95%], and 87% [86–89%], respectively. Based on the observed prevalence of MS in this AIM population of 0.8% (Table 1), and the sensitivity and specificity, the positive and negative predictive values were 3.8% (2.7–7.8%) and 0.15% (0.05–0.47%), respectively.

Classification of patients not known to have MS

We used the classification model, described above, to identify patients of the random sample who might have

unrecognized MS. These patients would be amongst those classified as having MS but lacking an ICD9[MS] code. Classification identified 295 such patients, out of a total cohort of 2289. We manually reviewed the notes of a random sample of 45 of these patients and identified four who could have unrecognized MS. For example, one 74-year-old woman complained of migraines, occipital neuralgia, had an abnormal EMG, hand numbness and bowel incontinence, all of which are important symptoms of MS. A second patient, a 75-year old woman complained of peripheral neuropathy, urinary incontinence and visual changes. The remainder of the reviewed patients had one or more medical conditions characterized by the signs and symptoms listed in Table 2, such as cerebrovascular accident, diabetic neuropathy, chronic migraine headaches, seizure disorders, brain cancer, metastatic cancer to the brain and traumatic brain injury. None of the four patients identified above as potentially having MS had any of these neurological conditions. However, inasmuch as none of the four patients suspected to have MS had been seen by a neurologist and were lacking additional follow-up, it was not possible to ascertain whether or not these patients actually had MS.

We next focused on the cohort of patients who were known to have MS (ICD9[MS]) and asked if the classifier would identify these patients as having MS using the notes entered *before* the initial entry of the ICD9[MS] code. Of the 165 patients with MS in the *MS-enriched cohort*, 30 had notes in the EHR up to two years *prior* to the first entry of an ICD9[MS] code. We trained a classifier using the *pre-ICD9[MS]* set of these 30 patients of the *MS-enriched cohort* (Fig. 1) and found that 40% of the patients, with documented MS, were identified by classification using notes entered before the initial ICD9[MS]. The ROC AUC, sensitivity and specificity were 0.71 [0.66 – 0.76], 40% [23–59%] and 97% [93–98%], respectively.

We sought to determine the temporal relationship between the dates of the first ICD9[MS] entry in the *pre-ICD9[MS]* set of the *MS-enriched cohort* and the first recognition of MS by a provider. We manually reviewed the notes of these patients from the time the initial ICD9 code was logged and back in time up to two years before. The dates of the first mention of the illness by the provider in a note and the dates of entry of an ICD9[MS] code were the same in 75% (58–93%) of these patients. Of the remaining patients, the dates of mention of MS (or the possibility of a demyelinating disease) in the notes was on average two months prior to the entry of the first ICD9[MS] code.

Discussion

Using NLP to identify patients with MS early in the course of the disease

The purpose of our study was to explore the possibility that patients with MS could be identified earlier in the

course of their illness using classification of the signs and symptoms in the clinical notes. Our results demonstrate that earlier identification is feasible. First, we found that 40% (23–59%) of the MS patients in the *MS-enriched set*, with clinical notes entered into the EHR up to two years prior to entry of the first ICD9[MS] code, were classified as having MS. The concordance between the date of entry of an ICD9[MS] code and first reporting of the illness by the provider in 75% (58–93%) of the reviewed patients suggests that classification identified patients who have not yet been recognized as having MS, not simply missing a timely entry of an ICD9[MS] code. Second, manual review of the patients in the *Random sample set*, those classified as has having MS but lacking an ICD9[MS] code identified four of 45 patients' whose clinical records suggested that they could have MS given that they had signs and symptoms of MS and did not have an obvious neurological condition that mimicked MS. These observations suggest the feasibility of building a diagnostic assistant to identify patients with MS early in the course of their disease perhaps before recognition by the primary provider.

Early diagnosis of a chronic illness through extraction and analysis of clinical terms has been explored previously in studies of celiac disease, well-known to elude diagnosis for years [39]. Two investigations achieved excellent results using different methods to classify patients with celiac disease based on the signs and symptoms present in the medical record [40, 41]. Neither study, however, determined if classification could identify patients before the definitive diagnosis of the disease. A recent study analyzed internet searching histories of patients with pancreatic cancer to identify typical searches and search terms *before* they knew that they had the illness [42]. They did not determine the sensitivity and specificity of using those search terms in identifying patients with pancreatic cancer.

Several groups have sought to identify patients with MS using the data in the EHR, most notably the eMERGE initiatives, seeking to identify patients with various conditions for genetic studies [43]. The eMERGE approach to identify patients with MS used billing codes as well as drugs used to treat MS and entry of “multiple sclerosis” in the notes (using NLP). Davis and coworkers used ICD-9 codes, text keywords, and medications to identify patients with MS and achieved excellent results [44]. Because our long-term goal was to identify patients who might have MS but had not yet been formally recognized as having the illness we specifically avoided using ICD9[MS] codes, drugs that treat MS, or the mention of MS in the notes as attributes for the classifier. Using only signs and symptoms of MS and none of the above-mentioned MS-specific attributes used in these other studies, our classifier

achieved a comparable level of accuracy in identifying patients already known to have MS.

Opportunity for EHR-based clinical decision support

Our results show that perhaps as many as 40% of MS patients could have been identified as having the illness from the signs and symptoms in their notes, well in advance of the initial recognition and assigning of an ICD9[MS] code. The prevalence of MS is sufficiently low, 0.8% (or 1/125 patients) that a busy primary care provider, who first see patients with MS [32], might not consider the illness. Integrating a classification-based clinical decision support tool has the potential to improve recognition of MS by providing a prompt to the provider early in the course of the clinical encounter [45–47].

Any MS diagnostic support tool embedded the EHR, however, would have to clearly communicate that the machine was providing a suggestion, not making a diagnosis. Given our observation that, at best, 4/45 randomly selected patients who were classified as MS but lacked an ICD9[MS] could have MS, the post-test (post-classification) probability would still only be in the neighborhood of 10%. Thus, 90% of the “flagged patients,” those classified as MS, would have some other neurologic condition. Strategies to improve the post-test probability could include removing patients with diseases known to have signs and symptoms similar to MS, such as cerebrovascular accident (CVA), diabetic neuropathy, or migraines. While removing these patients would increase specificity, it would also reduce the sensitivity of classification lessening its utility as a diagnostic prompt given that CVAs and migraines are common in patients with MS [48, 49]. Thus, a post-test probability of 10% may be the best that can be achieved using our approach. That said, a probability of 10%, more than 10 times higher than the prevalence of the disease in the general population, justifies the prompt “consider MS.”

Limitations of the study

There are several important limitations of our study. First, although there were a sufficient number of patients with MS and notes after entry of the ICD9[MS] code to train a classifier to identify patients with recognized MS, there was an insufficient number who had notes entered *before* the ICD9[MS] code to ascertain if classification could identify them accurately before being recognized as having the disease. Nor could we conclusively determine to what extent we could use the date of the initial ICD9[MS] code as the date of recognition of the disease by the provider. Thus, the classification sensitivity based on the pre-ICD9[MS] notes is likely to be lower than the observed 40%.” Clearly, a much larger set, probably derived from data from several institutions, will be necessary to validate our findings. Second, patients in the

random sample whom we suspect may have MS, had received no follow-up or referral to a neurologist (at the time of writing) to ascertain whether or not they had MS. We will have to follow these patients over the next several years before knowing for certain. Third, the classifier was built and tested on our local patient population which has a high proportion of Hispanic, largely female elderly patients. Classification accuracy could be different if applied to a largely white, younger population given the known racial differences in the prevalence of MS [50]. Fourth, we used MedLEE to parse the notes. Many institutions have home-grown term-extractors that would be used to replicate the classification approach described in this study. It is not known how well these other NLP systems would perform [51]. Last, the presence or absence of an ICD9[MS] code was used as the gold standard for the presence of MS in both patient sets. It is known that ICD coding for MS is inaccurate meaning that a proportion of patients with MS will lack an ICD9[MS] code while those who do not have MS receive an ICD9[MS] code [52]. Although this coding inaccuracy influences the accuracy of the classification, it is expected that the accuracy of the classification of patients with known MS would be less than the true accuracy. Thus, our estimates of specificity and sensitivity of classification of patients with known MS (ICD9[MS]) are likely an underestimate of the true values. We could have used the eMERGE criteria to identify patients with MS for the classification studies more accurately. However, given that our goal was to identify patients not known to have MS by their providers, we chose not to base the gold standard definition on the eMERGE criteria which are derived from patients with known MS.

Conclusions

Diagnostic error, a common cause of medical error, might be reduced with diagnostic assistants that recognize diseases by mining patients’ clinical notes for signs and symptoms using NLP. We built a classifier, which identified 40% of patients with MS disease before formal documentation of MS by providers suggesting that earlier recognition of the illness is possible. This approach could be applied to other diseases often missed by primary care providers such as cancer [45]. Whether implementing diagnostic assistants based on this classification approach ultimately shortens the time from earliest symptoms to formal recognition of the disease remains to be seen and will be the focus of future studies.

Abbreviations

AIM: Associates in internal medicine; CUIs: Concept unique identifiers; CUMC: Columbia University Medical Center; EHR: Electronic health record; ICD9[MS]: ICD9 codes for MS; MS: Multiple sclerosis; NLP: Natural language processing; NYP: New York Presbyterian; ROC AUC: Receiver operating characteristics area under the curve; UMLS: United medical language system

Acknowledgements

The authors thank Lyudmila Ena, who maintains the MedLEE database for the department, Carol Friedman, PhD, for her generous support and helpful suggestions in preparing the manuscript and Adler Perotte, MD and Hojjat Salmasian, MD, PhD, for their critical reviews of the manuscript.

Funding

U.S. Department of Health and Human Services; National Institutes of Health; National Institute on Aging: 5T35AG044303; U.S. Department of Health and Human Services; National Institutes of Health; National Library of Medicine: LM010016.

Availability of data and materials

Due to HIPPA restrictions, supporting data cannot be made openly available. For further information about the data and conditions for access please contact Herbert S. Chase directly via email: hc15@cumc.columbia.edu.

Authors' contributions

HSC extracted the data from the CUMC clinical data warehouse and MedLEE database, prepared the data for analysis, analyzed the classification results and prepared the manuscript. LRM analyzed the extracted data and identified the key attributes (signs and symptoms of MS) that had predictive value in identifying patients with MS, and prepared the manuscript. DJF and GGL performed classification experiments on the extracted data. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The protocol followed in this study was approved by the CUMC Institutional Review Board (IRB): CUMC IRB protocol "Medical Diagnostic Assistant," IRB-AAAI0688. Obtaining consent was waived given that it would not be feasible to obtain consent on the thousands of patients used in the study and that using the data posed no risk to the patients.

Received: 30 October 2016 Accepted: 10 February 2017

Published online: 28 February 2017

References

- Committee on Diagnostic Error in Health Care BoHCS, Institute of Medicine, The National Academies of Sciences, Engineering, and Medicine. Improving diagnosis in health care. Washington (DC): National Academies Press (US); 2015.
- Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Safety*. 2013;22 Suppl 2:i21–7.
- Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, Studdert DM. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med*. 2006;145(7):488–W183.
- Callen JL, Westbrook JJ, Georgiou A, Li J. Failure to follow-up test results for ambulatory patients: a systematic review. *J Gen Intern Med*. 2012;27(10):1334–48.
- Singh H, Giardina T, Meyer AD, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med*. 2013; 173(6):418–25.
- Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Safety*. 2014;23(9):727–31.
- Feldman MJ, Hoffer EP, Barnett GO, Kim RJ, Famiglietti KT, Chueh H. Presence of key findings in the medical record prior to a documented high-risk diagnosis. *J Am Med Inform Assoc*. 2012;19(4):591–6.
- Chase HS, Radhakrishnan J, Shirazian S, Rao MK, Vawdrey DK. Under-documentation of chronic kidney disease in the electronic health record in outpatients. *J Am Med Inform Assoc*. 2010;17(5):588–94.
- Sarkar U, Bonacum D, Strull W, Spitzmueller C, Jin N, López A, Giardina TD, Meyer AND, Singh T. Challenges of making a diagnosis in the outpatient setting: a multi-site survey of primary care physicians. *BMJ Qual Safety*. 2012; 21(8):641–8.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explor Newsl*. 2009;11(1):10–8.
- Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med*. 2009;169(20):1881–7.
- Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med*. 2003;78:775–80.
- Friedman CP, Gatti GG, Franz TM, Murphy GC, Wolf FM, Heckerling PS, Fine PL, Miller TM, Elstein AS. Do physicians know when their diagnoses are correct? implications for decision support and error reduction. *J Gen Intern Med*. 2005;20:334–9.
- El-Kareh R, Hasan O, Schiff GD. Use of health information technology to reduce diagnostic errors. *BMJ Qual Safety*. 2013;22 Suppl 2:i40–51.
- Graber ML, Mathew A. Performance of a Web-based clinical diagnosis support system for internists. *J Gen Intern Med*. 2008;23(1):37–40.
- Svenstrup D, Jørgensen HL, Winther O. Rare disease diagnosis: a review of web search, social media and large-scale data-mining approaches. *Rare Dis*. 2015;3(1):e1083145.
- Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. *J Gen Intern Med*. 2012;27(2):213–9.
- Elkin PL, Liebow M, Bauer BA, Chaliki S, Wahner-Roedler D, Bundrick J, Lee M, Brown SH, Froehling D, Bailey K, et al. The introduction of a diagnostic decision support system (DXplain™) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging diagnostic related groups (DRGs). *Int J Med Inform*. 2010;79(11):772–7.
- Berner ES. What can be done to increase the use of diagnostic decision support systems? *Diagnosis*. 2014;1(1):119–23.
- Nurek M, Kostopoulou O, Delaney BC, Esmail A. Reducing diagnostic errors in primary care. A systematic meta-review of computerized diagnostic decision support systems by the LINNEAUS collaboration on patient safety in primary care. *Eur J Gen Pract*. 2015;21(sup1):8–13.
- Wang V, Maciejewski ML, Hammill BG, Hall RK, Van Scoyoc L, Garg AX, Jain AK, Patel UD. Recognition of CKD after the introduction of automated reporting of estimated GFR in the Veterans Health Administration. *Clin J Am Soc Nephrol*. 2014;9(1):29–36.
- Wang V, Hammill BG, Maciejewski ML, Hall RK, Scoyoc LV, Garg AX, Jain AK, Patel UD. Impact of automated reporting of estimated glomerular filtration rate in the Veterans Health Administration. *Med Care*. 2015;53(2):177–83.
- Flynn N, Dawnay A. A simple electronic alert for acute kidney injury. *Ann Clin Biochem*. 2015;52(2):206–12.
- Lachance P, Villeneuve P-M, Wilson FP, Selby NM, Featherstone R, Rewa O, Bagshaw SM. Impact of e-alert for detection of acute kidney injury on processes of care and outcomes: protocol for a systematic review and meta-analysis. *BMJ Open*. 2016;6(5):e011152.
- Matsumura Y, Yamaguchi T, Hasegawa H, Yoshihara K, Zhang Q, Mineno T, Takeda H. Alert system for inappropriate prescriptions relating to patients' clinical condition. *Methods Inf Med*. 2009;48(6):566–73.
- Calvert JS, Price DA, Chettipally UK, Barton CW, Feldman MD, Hoffman JL, Jay M, Das R. A computational approach to early sepsis detection. *Comput Biol Med*. 2016;74:69–73.
- Gartlehner, Gerald, Thaler, Kylie, Chapman, Andrea, Kaminski H, Angela, Berzaczky, Dominik, et al. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. *Cochrane Database of Syst Rev*. 2013;30(4): CD009632. doi:10.1002/14651858.CD009632.pub2.
- Gullo L, Tomassetti P, Migliori M, Casadei R, Marrano D. Do early symptoms of pancreatic cancer exist that can allow an earlier diagnosis? *Pancreas*. 2001;22(2):210–3.
- Risch HA, Yu H, Lu L, Kidd MS. Detectable symptomatology preceding the diagnosis of pancreatic cancer and absolute risk of pancreatic cancer diagnosis. *Am J Epidemiol*. 2015;182(1):26–34.
- Goff B. Symptoms associated with ovarian cancer. *Clin Obstet Gynecol*. 2014;55(1):36–42.
- Solomon AJ, Weinshenker BG. Misdiagnosis of multiple sclerosis: frequency, causes, effects, and prevention. *Curr Neurol Neurosci Rep*. 2013;13(12):403.
- Fernández O, Fernández V, Arbizu T, Izquierdo G, Bosca I, Arroyo R, García Merino JA, de Ramón E. Characteristics of multiple sclerosis at onset and delay of diagnosis and treatment in Spain (The novo study). *J Neurol*. 2010;257(9):1500–7.
- Kingwell E, Leung AL, Roger E, Duquette P, Rieckmann P, Tremlett H. Factors associated with delay to medical recognition in two Canadian multiple sclerosis cohorts. *J Neurol Sci*. 2010;292(1–2):57–62.
- Leary SM, Porter B, Thompson AJ. Multiple sclerosis: diagnosis and the management of acute relapses. *Postgrad Med J*. 2005;81(955):302–8.

35. Kennedy P. Impact of delayed diagnosis and treatment in clinically isolated syndrome and multiple sclerosis. *J Neurosci Nurs*. 2013;45(6 Suppl 1):S3–S13.
36. Friedman C, Alderson PO, Austin JHM, Cimino JJ, Johnson SB. A general natural-language test processor for clinical radiology. *J Am Med Inform Assoc*. 1994;1(2):161–74.
37. Berkson J. Limitations of the application of fourfold able analysis to hospital data. *Biom Bull*. 1946;2(3):47–53.
38. Snoep JD, Morabia A, Hernández-Díaz S, Hernán MA, Vandembroucke JP. Commentary: a structural approach to Berkson's fallacy and a guide to a history of opinions about it. *Int J Epidemiol*. 2014;43(2):515–21.
39. Norström F, Lindholm L, Sandström O, Nordyke K, Ivarsson A. Delay to celiac disease diagnosis and its implications for health-related quality of life. *BMC Gastroenterol*. 2011;11(1):1–8.
40. Ludvigsson JF, Pathak J, Murphy S, Durski M, Kirsch PS, Chute CG, Ryu E, Murray JA. Use of computerized algorithm to identify individuals in need of testing for celiac disease. *J Am Med Inform Assoc*. 2013;20(e2):e306–10.
41. Tenório JM, Hummel AD, Cohrs FM, Sdepanian VL, Pisa IT, de Fátima MH. Artificial intelligence techniques applied to the development of a decision-support system for diagnosing celiac disease. *Int J Med Inform*. 2011;80(11):793–802.
42. Paparrizos J, White RW, Horvitz E. Screening for pancreatic Adenocarcinoma using signals from web search logs: feasibility study and results. *J Oncol Pract*. 2016;12:737–44.
43. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, Basford MA, Brown-Gentry K, Balsler JR, Masys DR, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet*. 2010;86(4):560–72.
44. Davis MF, Sriram S, Bush WS, Denny JC, Haines JL. Automated extraction of clinical traits of multiple sclerosis in electronic medical records. *J Am Med Inform Assoc*. 2013;20(e2):e334–40.
45. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care—a systematic review. *Fam Pract*. 2008;25(6):400–13.
46. Kostopoulou O, Lionis C, Angelaki A, Ayis S, Durbaba S, Delaney BC. Early diagnostic suggestions improve accuracy of family physicians: a randomized controlled trial in Greece. *Fam Pract*. 2015;32(3):323–8.
47. Kostopoulou O, Rosen A, Round T, Wright E, Douiri A, Delaney B. Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. *Br J Gen Pract*. 2015;65(630):e49–54.
48. Tseng CH, Huang WS, Lin CL, Chang YJ. Increased risk of ischaemic stroke among patients with multiple sclerosis. *Eur J Neurol*. 2015;22(3):500–6.
49. Sahai-Srivastava S, Wang SL, Ugurlu C, Amezcuca L. Headaches in multiple sclerosis: cross-sectional study of a multiethnic population. *Clin Neurol Neurosurg*. 2016;143:71–5.
50. Langer-Gould A, Brara SM, Beaber BE, Zhang JL. Incidence of multiple sclerosis in multiple racial and ethnic groups. *Neurology*. 2013;80(19):1734–9.
51. Stanfill MH, Williams M, Fenton SH, Jenders RA, Hersh WR. A systematic literature review of automated clinical coding and classification systems. *J Am Med Inform Assoc*. 2010;17(6):646–51.
52. St. Germaine-Smith C, Metcalfe A, Pringsheim T, Roberts JL, Beck CA, Hemmelgarn BR, McChesney J, Quan H, Jette N. Recommendations for optimal ICD codes to study neurologic conditions: a systematic review. *Neurology*. 2012;79(10):1049–55.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

