

Natural language processing of clinical notes for identification of critical limb ischemia

Naveed Afzal^a, Vishnu Priya Mallipeddi^b, Sunghwan Sohn^a, Hongfang Liu^a, Rajeev Chaudhry^c, Christopher G. Scott^a, Iftikhar J. Kullo^b, Adelaide M. Arruda-Olson^{b,*}

^a Department of Health Sciences Research, Mayo Clinic and Mayo Foundation, Rochester, MN, United States

^b Department of Cardiovascular Diseases, Mayo Clinic and Mayo Foundation, Rochester, MN, United States

^c Division of Primary Care Medicine, Knowledge Delivery Center and Center for Innovation, Mayo Clinic and Mayo Foundation, Rochester, MN, United States

ARTICLE INFO

Keywords:

Natural language processing
Electronic health records
Peripheral artery disease
Critical limb ischemia
Subphenotyping

ABSTRACT

Background: Critical limb ischemia (CLI) is a complication of advanced peripheral artery disease (PAD) with diagnosis based on the presence of clinical signs and symptoms. However, automated identification of cases from electronic health records (EHRs) is challenging due to absence of a single definitive International Classification of Diseases (ICD-9 or ICD-10) code for CLI.

Methods and results: In this study, we extend a previously validated natural language processing (NLP) algorithm for PAD identification to develop and validate a subphenotyping NLP algorithm (CLI-NLP) for identification of CLI cases from clinical notes. We compared performance of the CLI-NLP algorithm with CLI-related ICD-9 billing codes. The gold standard for validation was human abstraction of clinical notes from EHRs. Compared to billing codes the CLI-NLP algorithm had higher positive predictive value (PPV) (CLI-NLP 96%, billing codes 67%, $p < 0.001$), specificity (CLI-NLP 98%, billing codes 74%, $p < 0.001$) and F1-score (CLI-NLP 90%, billing codes 76%, $p < 0.001$). The sensitivity of these two methods was similar (CLI-NLP 84%; billing codes 88%; $p < 0.12$).

Conclusions: The CLI-NLP algorithm for identification of CLI from narrative clinical notes in an EHR had excellent PPV and has potential for translation to patient care as it will enable automated identification of CLI cases for quality projects, clinical decision support tools and support a learning healthcare system.

1. Introduction

Lower extremity peripheral artery disease (PAD) affects millions of people worldwide [1]. Advanced cases of PAD may manifest as critical limb ischemia (CLI) which is associated with considerable morbidity, mortality and high risk of major cardiovascular events [2]. Within one year of CLI diagnosis, 30% of patients undergo limb amputation while 25% die [3–5]. Despite the availability of state of the art revascularization procedures recommended by practice guidelines for treatment of CLI, high proportions of CLI patients undergo amputation without vascular evaluation in the previous year [6]. Due to population ageing and high prevalence of diabetes which are risk factors for CLI, it has been estimated that the consequent number of CLI patients is likely to increase in both developing and developed countries [7,8]. Moreover, CLI has been associated with significant health care resource utilization. The estimate of aggregate annual US national costs associated with CLI hospitalizations was approximately \$4.2 billion in 2013–2014 while the

30-day readmission rates for CLI contributed to over \$624 million in healthcare costs [9].

Electronic health records (EHRs) have been widely heralded for potential to improve the quality of patient care and as a source for rapid automated identification of patients for research studies [10]. However, the electronic ascertainment of CLI from EHRs has proved challenging due to absence of a single definitive ICD-9 or ICD-10 code. For this reason, prior studies have developed and validated billing code algorithms for ascertainment of CLI cases [11] using combinations of ICD-9 codes. Sensitivity of these billing code algorithms has varied by practice setting [11]. Importantly, the clinical diagnosis of CLI is based on the presence of signs and symptoms as recorded in clinical narrative [12] while billing codes are used primarily for administrative purposes. Billing codes are used to mine structured information while natural language processing (NLP) is used to extract meaningful information from unstructured data. ICD billing codes are used primarily for administrative transactions and reimbursements. Additionally, ICD codes are used for diverse secondary

* Corresponding author at: Cardiovascular Diseases, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, United States.

E-mail address: ArrudaOlson.Adelaide@mayo.edu (A.M. Arruda-Olson).

Table 1
CLI related keywords for CLI-NLP algorithm.

Diagnostic Keywords	Location Keywords
ischemia; ischemic ulcer; ischemic ulcers; ischemic wound; ischemic wounds; ischemic pressure wound; ischemic pressure wounds; gangrene; neuropathic ischemic wound; neuropathic ischemic wounds	limb; limbs; lower extremity; lower extremities; right lower extremity; left lower extremity; right lower extremities; left lower extremities; rle; lle; leg; legs; foot; feet; toe; toes; ankle; aorto bi-iliac; aorto bi-femoral; aorto iliac; aorta femoral; sfa; plantar; heel

purposes including epidemiology studies, cohort identification and for health services research [13]. Prior studies comparing administrative and clinical approaches indicated that administrative data may be less accurate for identification of certain patient characteristics [14–16]. Notably, the use of billing code algorithms for identification of other phenotypes from EHRs has also had disappointing accuracy, positive predictive value and/or sensitivity [17–19].

NLP applied to clinical narrative may overcome the limitations of billing code algorithms for identification of CLI by recognition of text which describes signs and symptoms used to establish a diagnosis. Indeed, previous studies have demonstrated that NLP methods outperform billing code algorithms for phenotype identification from narrative clinical notes from the EHR. Specifically, NLP-based phenotyping algorithms have been used for automated case identification for a variety of diseases including inflammatory bowel disease, multiple sclerosis, rheumatoid arthritis, asthma and pancreatic cancer [20–22]. In the present study, we developed an NLP-based algorithm for ascertainment of CLI from narrative clinical notes of a community-based PAD cohort and we compared the performance of the NLP algorithm with CLI-related billing codes. Both methods were compared to human abstraction as the gold standard. We tested the hypothesis that an NLP algorithm applied to narrative clinical notes will have superior performance compared to CLI-related billing codes for identification of CLI.

2. Methods

2.1. Study setting and population

The study was conducted at Mayo Clinic, Rochester MN and used the resources of the Rochester Epidemiology Project (REP) to assemble a community-based PAD cohort from Olmsted County [23,24]. The REP is an integrated health information system that links medical records of all residents of Olmsted County [23]. In the PAD inception cohort, all patients were diagnosed with PAD by an ankle-brachial index (ABI) test performed in the Mayo Clinic noninvasive vascular laboratory using standardized protocols [1]. The institutional review board approved the study and informed consent was attained for all subjects.

2.2. Study design

We retrieved all clinical notes through June 2015 of patients participating in this study from the Mayo clinical data warehouse. We applied the previously validated knowledge-driven NLP algorithm (PAD-NLP algorithm) to the dataset to automatically ascertain PAD status [25]. The PAD-NLP algorithm automatically ascertained cases from clinical notes using PAD-related keywords and a set of rules for classification of a PAD patient. The PAD-NLP algorithm consisted of two main components: text-processing and patient classification. The text-processing component analyzed the text of each clinical note by breaking down sentences into words using MedTagger [26], an open source clinical NLP system, and identified PAD-related concepts which were mapped to specific categories that were later used for patient classification. NLP technology was used for automated extraction and encoding of clinical information from narrative clinical notes. MedTagger is a knowledge driven clinical NLP system which enables sentence detection, word tokenization, section identification, contextual

information and concept identification. After sentence detection, word tokens are identified using space between two words. As recommended by the HL-7 CDA standard [27], clinical notes are divided into sections, (e.g. “impression, report and plan” and “diagnosis”) and MedTagger recognizes these note sections. Additionally, MedTagger identifies contextual information from clinical notes including: assertion, temporality and experienter. We used MedTagger to identify assertion, temporality and experienter of CLI-related keywords from clinical notes.

The steps for the NLP algorithm were: first, a list of PAD-related terms was identified by cardiovascular experts from narrative clinical notes. Second, these terms were mapped to corresponding concepts and their synonyms in the unified medical language system (UMLS) Metathesaurus which were also added to this list. Third, this list was further expanded during the interactive refinement of PAD-NLP algorithm [25] when additional synonyms and other lexical variants were identified in the clinical notes and added. Fourth, the PAD-NLP algorithm produced output on two levels: document and patient levels. At the document level, each clinical note was processed to find PAD-related keywords and if found produced output in the form of PAD-related keywords along with ± 2 sentences from clinical notes. Fifth, as CLI keywords were included in the list of keywords used by the PAD-NLP algorithm (Table 1) these concepts were also identified. Their relevant category along with certainty (positive, negative or possible), temporality (current, historical) and its experienter (patient or someone else) were also extracted during the text processing phase.

The CLI-related keywords were identified by cardiovascular experts and included in the list of keywords used in the PAD-NLP algorithm. A subphenotyping algorithm for identification of CLI cases was developed and used the document level (i.e. note level) output of the PAD-NLP algorithm and narrowed the focus of the algorithm to identify the subset of PAD cases with CLI. Hence, the CLI-NLP algorithm was derived from the PAD-NLP algorithm (Fig. 1). The performance of the billing codes for identification of CLI was compared with results obtained by the CLI-NLP algorithm. Both methods were then compared to human chart abstraction as the gold standard for validation (Fig. 1).

2.3. Rules for the CLI-NLP algorithm

The CLI-NLP algorithm identified keywords from the document level output. Keywords were categorized as diagnostic and location (Table 1). The list of diagnostic keywords included both the isolated mechanisms for wound or combined mechanisms (wounds which occurred as a consequence of a combination of co-existing mechanisms, e.g., ischemic/pressure wound or neuropathic ischemic wound – see Table 1). Importantly, “ischemia” (and the lexicon variations for this term) had to be considered as one of these mechanisms. In the absence of these criteria patients were classified as not having CLI (controls).

The rule for CLI cases was:

- One diagnostic keyword + one location keyword from Table 1 within two sentences anchored by a diagnostic keyword in the same note.

For controls (not having CLI), we used the following rule:

- If not satisfied, the rule for CLI described above.

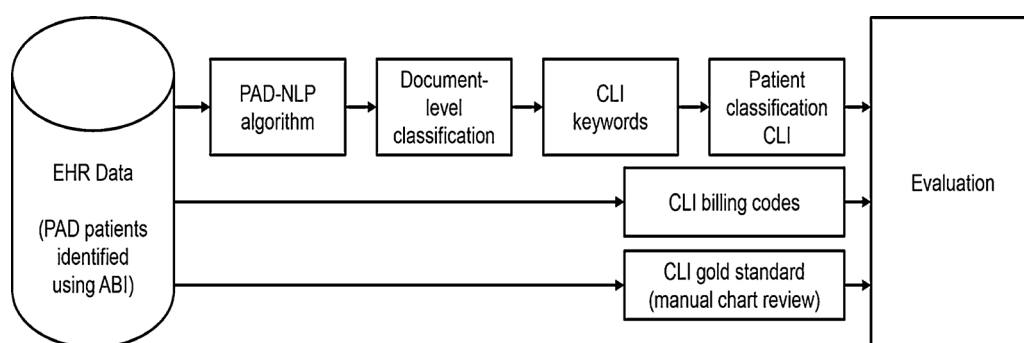


Fig. 1. Study design.

2.4. CLI ICD-9 diagnostic billing codes

For this study, we used the billing codes listed in Table 2 for identification of CLI. These billing codes were previously used by other studies for identification of CLI patients [9,11,28].

2.5. Gold standard: chart abstraction

One physician abstracted the information from the EHRs of each patient in the study dataset. After reviewing the written manual for abstraction and completion of specific training, the abstractor (a physician) reviewed the medical record of each patient to obtain the information necessary for classification according to the pre-determined criteria from the manual for abstraction which is summarized in Table 3. The abstracted information was then manually entered and stored in an electronic dataset. The physician abstracted CLI relevant information using keywords listed in Table 1 from each PAD patient clinical notes. A board certified cardiologist verified the abstracted information. The cardiologist conducted independent chart abstraction to clarify questions about patient classification raised by the primary abstractor while following predefined criteria documented in the manual for abstraction. A consensus decision was then reached by the two abstractors. In case of disagreement cases were reviewed by another cardiologist and a consensus opinion was used for classification. Table 3 shows the clinical criteria used for CLI diagnosis by manual chart abstraction. The inter-annotator agreement between the two annotators (physician and board certified cardiologist) was 95%. The clinical criteria for CLI diagnosis were based upon current clinical practice guidelines [29].

Table 2
ICD-9 billing codes for identification of CLI.

Code	Description
440.22	Atherosclerosis of the extremities with rest pain
440.23	Atherosclerosis of native arteries of the extremities with ulceration
440.24	Atherosclerosis of native arteries of the extremities with gangrene

Table 3
Gold standard diagnostic criteria for chart abstraction by humans.

One of the following:
a)
Ischemic rest pain
b) Wound or ulcer
c) Gangrene
PLUS – all listed below:
a)
Location: 1 or both lower extremities
b) Duration of symptom: chronic (≥ 2 weeks)
c) Objectively proven arterial occlusive disease

2.6. Statistical analysis

We compared both the CLI-related ICD-9 codes and CLI-NLP algorithm for identifying CLI with the gold-standard manual abstraction to calculate sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and F1-score of each method. We calculated 95% confidence intervals (CI) for each aforementioned measure. These measures were calculated as follows:

$$PPV = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

$$NPV = \frac{\text{true negatives}}{\text{true negatives} + \text{false negatives}}$$

$$F1 \text{ score} = 2 \times \frac{PPV \times \text{Sensitivity}}{PPV + \text{Sensitivity}}$$

Estimates of sensitivity, specificity, and F1-score were compared between algorithms using the McNemar test. Generalized score statistics were used to compare PPV and NPV. Analysis was performed in SAS v 9.4 (SAS Institute, Cary, NC), and significance was set using a two-sided p value of < 0.05 .

3. Results

The PAD cohort dataset contained 792 PAD cases. There were 295 CLI cases (37%) and 497 (63%) controls (without CLI). The dataset contained 270,336 clinical notes and on average each note contained 395 word tokens. The average age of patients in the dataset was 71 years. The dataset was comprised of 44% women and 90% of patients were white. Table 4 shows classification performance of the CLI-NLP algorithm in terms of a confusion matrix. The CLI-NLP algorithm had superior PPV (96%), compared to CLI-related billing code (67%); for identification of CLI ($p < 0.001$). Sensitivity of the two approaches was similar (CLI-NLP 84%; billing codes 88%; $p < 0.12$). However, the F1-score was superior for CLI-NLP (90%) compared to billing codes (76%) ($p < 0.001$). The CLI-NLP algorithm had higher specificity

Table 4
Classification performance of CLI-NLP algorithm.

N = 792	Predicted Cases	Predicted Controls	Total
Actual Cases	TP = 247	FN = 48	295
Actual Controls	FP = 10	TN = 487	497
Total	257	535	

N = number of patients, TP = true positives, FN = false negatives, FP = false positives, TN = true negatives.

Table 5
CLI-NLP algorithm compared with CLI related billing codes for CLI identification.

	CLI-NLP (95% CI)	Billing codes (95% CI)	p-value CLI-NLP vs. Billing codes
Sensitivity (%)	84 (79, 88)	88 (84, 92)	0.12
Specificity (%)	98 (97, 99)	74 (70, 78)	< 0.001
PPV (%)	96 (94, 98)	67 (62, 71)	< 0.001
NPV (%)	91 (89, 93)	91 (88, 94)	0.87
F1-score (%)	90 (86, 93)	76 (71, 80)	< 0.001

CI = confidence interval, NPV = negative predictive value, PPV = positive predictive value.

compared to billing code (CLI-NLP 98%, billing codes 74%; $p < 0.001$) (Table 5).

During the interactive refinement of PAD-NLP algorithm, the clinical notes generated by selected specialties, note types and note sections which had best performance for identification of PAD cases were identified [25]. Accordingly, in this present study, the clinical notes from the following medical specialties were included: primary care, internal medicine (general internal medicine and family medicine), cardiology, vascular (vascular medicine and vascular surgery), urgent care and critical care. Fig. 2 shows the overall distribution of notes generated by providers practicing in these specialties. These notes were generated by care providers who evaluated patients during medical

encounters and included: staff physicians, residents, fellows, nurse practitioners and physician assistants. Similarly, the selected note types were relevant to PAD and CLI diagnosis [25]. These note types all documented a medical encounter, in the inpatient (hospitalized) or outpatient setting and included: consultations, subsequent visits, multisystem evaluation, specialty evaluation, supervisory notes and limited evaluation notes. Fig. 3 shows the distribution of these note types in the study cohort. The following relevant note sections were used: impression, report and plan, diagnosis, past medical and surgical history.

We analyzed reasons for false positives and false negatives of the CLI-NLP algorithm; Table 6 contains examples. False positives were due to the complex nature of natural language in the clinical notes when clinicians described possible future scenarios at follow-up. Another reason for false-positives was the absence of note section headers. In these records we found excluded sections embedded into other sections and missing section headers.

A reason for false negatives was the absence of location and/or diagnostic keywords within the \pm two sentence window (see Table 6). Another reason for false negatives was the absence of information in narrative clinical notes regarding a completed surgery to treat limb ischemia. In our dataset these surgical procedures dated to the transition period from paper records to electronic records when most clinical notes (e.g. progress notes) were in paper scanned into the EHR in PDF format. In addition, the surgical reports were in PDF format and consequently invisible to current NLP technology.

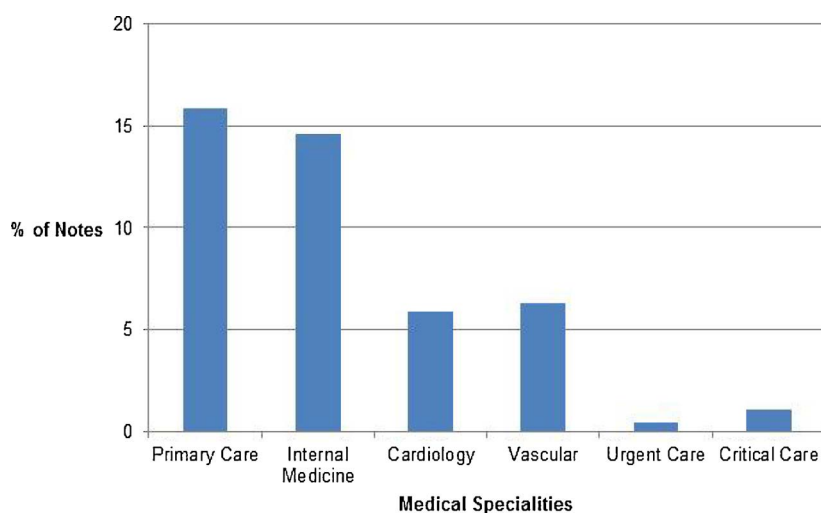


Fig. 2. Distribution of Clinical Notes by Medical Specialty.

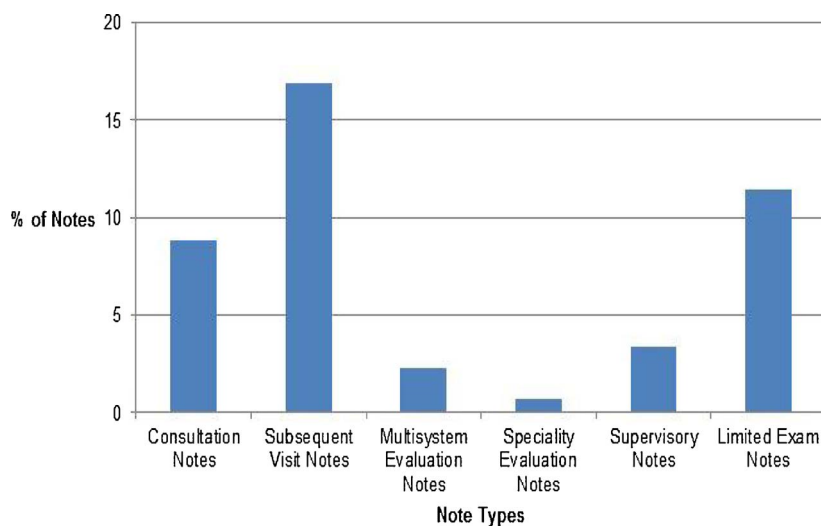


Fig. 3. Distribution of Note Types.

Table 6
Reasons for false positives and false negatives in CLI-NLP algorithm.

Category	Example
False Positives	
Complex nature of natural language	“the distal disease which would likely be found would be difficult to treat percutaneously and surgical bypass would be an option in the setting of critical limb ischemia, but not necessarily for limited claudication.” “should Mr. X develop disabling claudication or critical limb ischemia, MRA evaluation would be useful for revascularization options.”
Missing section header	Clinical notes are typically divided into sections or segments. Patient consent section was excluded however, in some false positives cases we found this excluded section embedded in another section.
False Negatives	
Absence of location/diagnostic keywords	No location keyword mentioned in two-sentence window: “... peripheral arterial disease with gangrene...”
Narrative clinical notes in PDF format due to transition from paper to electronic health records	CLI key words were not identified in clinical notes in PDF format (e.g. scanned progress notes).

4. Discussion

Clinical diagnosis of CLI is based on the presence of signs and symptoms which are recorded in narrative clinical notes. Billing codes under-performed for identification of CLI while NLP had excellent performance. This is because the NLP system enables automated data extraction from narrative clinical notes of the clinical characteristics used by clinicians to diagnose CLI, which are not expressed by billing codes. We previously demonstrated in a community-based dataset that billing codes have limited ability to identify PAD cases (PPV: 75% and sensitivity 68%) [30]. The PAD-NLP algorithm enhanced performance of the electronic approaches for automated PAD identification (PPV: 92.9%, sensitivity: 91.2%) [25]. In the present study, though sensitivity of billing codes for identification of CLI was comparable to the CLI-NLP algorithm, the PPV, F-1 scores and specificity were superior for the CLI-NLP algorithm.

In this study, we constructed and validated a novel NLP subphenotyping algorithm to identify CLI cases from narrative clinical notes. This CLI-NLP subphenotyping algorithm was derived from an existing and previously validated PAD-NLP algorithm [25]. A subphenotyping approach has been used previously to identify subtypes of asthma [31]. To the best of our knowledge, the present study is the first to use a NLP sub-algorithm for identification of patients with CLI from narrative clinical notes. Our study results clearly demonstrate higher performance of a subphenotyping NLP algorithm for ascertainment of CLI status from clinical notes. Importantly, the PAD-NLP algorithm identified PAD patients with a broad spectrum of clinically diagnosed PAD including CLI and other presentations of PAD as follows: claudication, abnormal ABI results, poorly compressible arteries, prior limb revascularization, and abnormal imaging results. Because of the under-diagnosis and under-treatment of CLI there is a clinical need for prompt and early identification of these patients with the most advanced form of PAD (CLI) using EHR notes.

Clinical notes in EHRs contain documentation of patient-provider encounters, as well as the assessment and plan for management which are not available in billing codes. The NLP technology enables automated data extraction from narrative clinical notes and has played a vital role in meaningful use of EHRs for clinical and translational research [32–36]. Moreover, the novel PAD-NLP subphenotyping algorithm described herein will eventually be linked to clinical decision support systems for prompt implementation of evidence-based management strategies at the point of care [29]. Therefore, automatic identification of CLI cases [5] may contribute to a learning health care system in cardiovascular care [37] thereby bringing innovation to healthcare delivery and responding to the 2017 Learning Healthcare System scientific statement from the American Heart Association [37].

Prior NLP studies in the clinical domain have used various types of EHR notes including radiology reports, dismissal summaries, and nursing notes to identify different phenotypes and syndromes [36]. In contrast, the study herein used narrative clinical notes which describe a

medical encounter. Prior studies have used dismissal summaries [38–40] as they contain a summary of events which occurred during the duration of a hospitalization and this type of note was one of the main reasons for false results during the interactive validation of the PAD-NLP algorithm consequently these note types were excluded. Additionally, during interactive validation the note sections which had the best performance for phenotype identification were identified and used for the NLP algorithm. In a prior study [41] the note section “chief complaint” was mined by NLP, however because this section included the reason for the visit, which often is ruled out during the encounter, this led to false results in our study and this section was excluded. Conversely the “impression, report and plan” had excellent performance for identification of PAD and CLI concepts and this section was used by the PAD-NLP and CLI-NLP algorithms.

In the study herein, the CLI-NLP algorithm had better performance in terms of specificity, PPV and F1-score compared to CLI-related billing codes while CLI-related billing codes had sensitivity similar to the CLI-NLP algorithm. The excellent PPV, specificity and F1-score of the CLI-NLP algorithm could facilitate timely and automated identification of patients with CLI using narrative clinical notes from the EHR. In this study we used CLI-related diagnostic ICD-9 codes for comparison (Table 2) and did not limit the billing code algorithms to CLI-related procedural codes.

Coexisting factors contributed to the superior performance of the CLI-NLP algorithm. This algorithm extracted relevant clinical information from selected clinical notes, generated by selected departments, note types and note sections [25]. These were all previously identified during the interactive refinement of the PAD-NLP algorithm [25]. Another contributor was the collaboration between clinicians and computer scientists with complementary skill sets who worked together during all stages of development and validation of both NLP algorithms.

Using NLP approach meaningful information may be extracted from clinical notes in EHRs, which contain unstructured narrative text. NLP methodology consolidates information from clinical notes into coherent structure, which enables automated identification of cases that satisfy keywords and rules of an NLP algorithm [42]. The NLP-based approach in our study is independent of billing codes and relied on a specific set of keywords that are independent of EHR vendor and could be replicated in other EHR systems. We leveraged comprehensive data in the form of clinical notes from both inpatient and outpatient settings. A limitation of this study is that we used data for a PAD cohort from a single medical center and future studies should apply and validate this algorithm to other institutions to make the findings generalizable.

The narrative notes used for this study were created by clinicians for documentation of medical encounters from routine medical practice. The architecture of these notes follows the HL7 standard for clinical document architecture [27]. Clinicians dictated or typed their notes in the EHRs. The dictated notes were transcribed to the EHR. The investigators did not influence the pattern of documentation and only used the relevant medical documentation available in the EHR. An

important factor which may have accounted for the superior performance of the CLI-NLP algorithm was the contribution of clinicians, who are users of the EHRs for patient care activities. A recent literature review of clinical information extraction applications has shown that for effective and efficient NLP systems in the clinical domain there is a strong need for a close collaboration between NLP experts and clinicians [36]. Our study addressed this issue as computer scientists and clinicians collaborated in all stages of the development of the CLI-NLP algorithm. Clinicians contributed to identification of relevant medical terms, insights about clinician thought processes, as well as descriptions of signs and symptoms used to establish the diagnosis of CLI. Additionally, clinicians are very familiar with the note types, note sections and medical specialties used in EHR notes. Therefore, clinicians contributed to the validation and selection of relevant EHR notes which were used by the NLP algorithm. In summary, clinicians participated in all stages of creation and validation of the NLP algorithms which were used to extract information from narrative clinical notes created by clinicians.

To assure portability to other medical centers, the investigators are currently collaborating with two other academic medical centers in the US which participate in the electronic medical records and genomics (eMerge) network along with Mayo Clinic. In this collaborative study the PAD-NLP algorithm (from which the CLI-NLP algorithm has been derived) has been applied and sites are in the process of conducting validation studies in these centers.

5. Conclusions

The CLI-NLP algorithm for identification of CLI had excellent PPV with potential for translation to patient care for case identification and eventual linkage to NLP-based clinical decision support tools. The CLI-NLP algorithm for automatic identification of CLI cases from clinical notes may enhance CLI research and eventually lead to improved quality of care of CLI patients.

Authors contributions

Conception and design: NA, VPM, SS, CS, HL, IK, AA-O
 Analysis and interpretation: NA, VPM, SS, CS, HL, IK, AA-O
 Data collection: NA, VPM, SS, CS, AA-O
 Writing the article: NA, VPM, SS, CS, RC, HL, IK, AA-O
 Critical revision of the article: SS, RC, HL, IK, AA-O
 Final approval of the article: NA, VPM, SS, CS, RC, HL, IK, AA-O
 Statistical analysis: NA, CS, HL, IK, AA-O
 Obtained funding: AA-O, IK, HL
 Overall responsibility: AA-O, HL

Conflict of interest

The author(s) declare(s) that there is no conflict of interest.

Funding sources

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (award K01HL124045) and the NHGRI eMERGE (Electronic Records and Genomics) Network grants HG04599 and HG006379. This study was made possible using the resources of the Rochester Epidemiology Project supported by the National Institute on Aging of the National Institutes of Health (award R01AG034676) and the NLP framework established through the NIGMS award R01GM102282. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Summary points

- Critical limb ischemia (CLI) is a complication of advanced peripheral artery disease (PAD) with diagnosis based on the presence of clinical signs and symptoms. Automated identification of cases is challenging due to absence of a single definitive International Classification of Diseases (ICD-9 or ICD-10) code for CLI.
- We developed a natural language processing (NLP)-based algorithm for ascertainment of CLI from narrative clinical notes of a community-based PAD cohort and compared the performance of the NLP algorithm with CLI-related billing codes. Both methods were compared to human abstraction as the gold standard.
- The CLI-NLP algorithm for identification of CLI had excellent positive predictive value with potential for translation to patient care for case identification and NLP-based clinical decision support tools. The CLI-NLP algorithm for automatic identification of CLI cases from clinical notes may enhance CLI research and eventually lead to improved quality of care of CLI patients.

Acknowledgements

We thank Jared Robb, LuAnne Koenig and Cynthia Regnier for data collection, Carin Smith for statistical analysis, and Rebecca M. Olson for secretarial support.

References

- [1] L.J. Kullo, T.W. Rooke, Peripheral artery disease, *N. Engl. J. Med.* 2016 (374) (2016) 861–871.
- [2] M.D. Gerhard-Herman, H.L. Gornik, C. Barrett, et al., 2016 AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: a report of the American college of cardiology/American heart association task force on clinical practice guidelines, *J. Am. Coll. Cardiol.* 69 (11) (2017) e71–e126.
- [3] L. Norgren, W.R. Hiatt, J.A. Dormandy, M.R. Nehler, K.A. Harris, F.G. Fowkes, Inter-society consensus for the management of peripheral arterial disease (TASC II), *J Vasc Surg.* (2007) S5–67.
- [4] A. Farber, R.T. Eberhardt, The current state of critical limb ischemia: a systematic review, *JAMA Surg.* 151 (11) (2016) 1070–1077.
- [5] M.H. Shishehbor, C.J. White, B.H. Gray, et al., Critical limb ischemia: an expert statement, *J. Am. Coll. Cardiol.* 68 (18) (2016) 2002–2015.
- [6] P.P. Goodney, L.L. Travis, B.K. Nallamothu, et al., Variation in the use of lower extremity vascular procedures for critical limb ischemia, *Circ. Cardiovasc. Qual Outcomes* 5 (1) (2012) 94–102.
- [7] P.R. Cavanagh, B.A. Lipsky, A.W. Bradbury, G. Botek, Treatment for diabetic foot ulcers, *Lancet* 366 (9498) (2005) 1725–1735.
- [8] P. Burns, S. Gough, A.W. Bradbury, Management of peripheral arterial disease in primary care, *BMJ* 326 (7389) (2003) 584–588.
- [9] D. Kolte, K.F. Kennedy, M.H. Shishehbor, et al., Thirty-day readmissions after endovascular or surgical therapy for critical limb ischemia: analysis of the 2013 to 2014 Nationwide Readmissions Databases, *Circulation* 136 (2) (2017) 167–176.
- [10] B.E. Himes, Y. Dai, I.S. Kohane, S.T. Weiss, M.F. Ramoni, Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records, *JAMIA* 16 (3) (2009) 371–379.
- [11] W. Bekwelem, L.G. Bengtson, N.C. Oldenburg, et al., Development of administrative data algorithms to identify patients with critical limb ischemia, *Vasc Med.* 19 (6) (2014) 483–490.
- [12] M.D. Gerhard-Herman, H.L. Gornik, C. Barrett, et al., AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: executive summary a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines, *J. Am. Coll. Cardiol.* 69 (11) (2017) 1465–1508.
- [13] K.J. O'Malley, K.F. Cook, M.D. Price, K.R. Wildes, J.F. Hurdle, C.M. Ashton, Measuring diagnoses: ICD code accuracy, *Health Serv. Res.* 5 (Pt 2) (2005) 1620–1639.
- [14] E.R. McPeck Hinz, L. Bastarache, J.C. Denny, A natural language processing algo-

- rhythm to define a venous thromboembolism phenotype, AMIA Annu. Symp. Proc. 2013 (2013) 975–983.
- [15] J.S. Tieder, M. Hall, K.A. Auger, et al., Accuracy of administrative billing codes to detect urinary tract infection hospitalizations, *Pediatrics* 128 (2) (2011) 323–330.
 - [16] C.I. Wi, S. Sohn, M. Ali, et al., Natural language processing for asthma ascertainment in different practice settings, *J. Allergy Clin. Immunol. Pract.* (2017).
 - [17] E. Birman-Deych, A.D. Waterman, Y. Yan, D.S. Nilasena, M.J. Radford, B.F. Gage, Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors, *Med. Care* 43 (5) (2005) 480–485.
 - [18] M. Schmiedeskamp, S. Harpe, R. Polk, M. Oinonen, A. Pakyz, Use of international classification of diseases, ninth revision clinical modification codes and medication use data to identify nosocomial *Clostridium difficile* infection, *Infect. Control Hosp. Epidemiol.* 30 (11) (2009) 1070–1076.
 - [19] E.F. Kern, M. Maney, D.R. Miller, et al., Failure of ICD-9-CM codes to identify patients with comorbid chronic kidney disease in diabetes, *Health Serv. Res.* 41 (2) (2006) 564–580.
 - [20] S.T. Wu, S. Sohn, K. Ravikumar, et al., Automated chart review for asthma cohort identification using natural language processing: an exploratory study, *Ann. Allergy Asthma Immunol.* 111 (5) (2013) 364–369.
 - [21] J. Friedlin, M. Overhage, M.A. Al-Haddad, et al., Comparing methods for identifying pancreatic cancer patients using electronic data sources, *AMIA Annu. Symp. Proc.* 2010 (2010) 237–241.
 - [22] K.P. Liao, T. Cai, G.K. Savova, et al., Development of phenotype algorithms using electronic medical records and incorporating natural language processing, *BMJ* 350 (2015) h1885.
 - [23] J.L. St Sauver, B.R. Grossardt, B.P. Yawn, L.J. Melton 3rd, W.A. Rocca, Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project, *Am. J. Epidemiol.* 173 (9) (2011) 1059–1068.
 - [24] J.L. St Sauver, B.R. Grossardt, B.P. Yawn, et al., Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system, *Int. J. Epidemiol.* 41 (6) (2012) 1614–1624.
 - [25] N. Afzal, S. Sohn, S. Abram, et al., Mining peripheral arterial disease cases from narrative clinical notes using natural language processing, *J. Vasc. Surg.* 65 (6) (2017) 1753–1761.
 - [26] H. Liu, S.J. Bielinski, S. Sohn, et al., An information extraction framework for cohort identification using electronic health records, *AMIA Jt. Summits Transl. Sci. Proc.* (2013) 149–153.
 - [27] R.H. Dolin, L. Alschuler, S. Boyer, et al., HL7 clinical document architecture, release 2, *J. Am. Med. Inform. Assoc.* 13 (1) (2006) 30–39.
 - [28] M.R. Nehler, S. Duval, L. Diao, et al., Epidemiology of peripheral arterial disease and critical limb ischemia in an insured national population, *J. Vasc. Surg.* 60 (3) (2014) 686–695 (e682).
 - [29] M.D. Gerhard-Herman, H.L. Gornik, C. Barrett, et al., AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: executive summary a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines, *Circulation* (2016) 2016.
 - [30] J. Fan, A.M. Arruda-Olson, C.L. Leibson, et al., Billing code algorithms to identify cases of peripheral artery disease from administrative data, *J. Am. Med. Inform. Assoc.* 20 (e2) (2013) e349–e354.
 - [31] M.C. Prosperi, S. Marinho, A. Simpson, A. Custovic, I.E. Buchan, Predicting phenotypes of asthma and eczema with machine learning, *BMC Med. Genom.* 7 (Suppl 1) (2014) S7.
 - [32] M.S. Safarova, H. Liu, I.J. Kullo, Rapid identification of familial hypercholesterolemia from electronic health records: the SEARCH study, *J. Clin. Lipidol.* 10 (5) (2016) 1230–1239.
 - [33] L. Zheng, Y. Wang, S. Hao, et al., Web-based real-time case finding for the population health management of patients with diabetes mellitus: a prospective validation of the natural language processing-based algorithm with statewide electronic medical records, *JMIR Med. Inform.* 4 (4) (2016) e37.
 - [34] H. Mo, W.K. Thompson, L.V. Rasmussen, et al., Desiderata for computable representations of electronic health records-driven phenotype algorithms, *J. Am. Med. Inform. Assoc.* 22 (6) (2015) 1220–1230.
 - [35] E. Ford, J. Carroll, H. Smith, et al., What evidence is there for a delay in diagnostic coding of RA in UK general practice records? An observational study of free text, *BMJ Open* 6 (6) (2016) e010393.
 - [36] Y. Wang, L. Wang, M. Rastegar-Mojarad, et al., Clinical information extraction applications: a literature review, *J. Biomed. Inform.* (2017).
 - [37] T.M. Maddox, N.M. Albert, W.B. Borden, et al., The learning healthcare system and cardiovascular care: a scientific statement from the American Heart Association, *Circulation* (2017).
 - [38] C. Friedman, C. Knirsch, L. Shagina, G. Hripcsak, Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries, Paper presented at: Proc AMIA Symp. (1999).
 - [39] M. Fiszman, W.W. Chapman, D. Aronsky, R.S. Evans, P.J. Haug, Automatic detection of acute bacterial pneumonia from chest X-ray reports, *J. Am. Med. Inform. Assoc.* 7 (6) (2000) 593–604.
 - [40] G. Hripcsak, J.H. Austin, P.O. Alderson, C. Friedman, Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports 1, *Radiology* 224 (1) (2002) 157–163.
 - [41] W.W. Chapman, L.M. Christensen, M.M. Wagner, et al., Classifying free-text triage chief complaints into syndromic categories with natural language processing, *Artif. Intell. Med.* 33 (1) (2005) 31–40.
 - [42] F. Liu, C. Weng, H. Yu, Natural language processing, electronic health records, and clinical research, *Clinical Research Informatics*, Springer, 2012, pp. 293–310.