

Automatic Classification of Ultrasound Screening Examinations of the Abdominal Aorta

Craig Morioka^{1,2,3} · Frank Meng^{2,3} · Ricky Taira^{2,3} · James Sayre^{2,3} · Peter Zimmerman^{1,3} · David Ishimitsu^{1,3} · Jimmy Huang^{1,3} · Luyao Shen³ · Suzie El-Saden^{1,2,3}

Published online: 11 July 2016
© Society for Imaging Informatics in Medicine 2016

Abstract Our work facilitates the identification of veterans who may be at risk for abdominal aortic aneurysms (AAA) based on the 2007 mandate to screen all veteran patients that meet the screening criteria. The main research objective is to automatically index three clinical conditions: pertinent negative AAA, pertinent positive AAA, and visually unacceptable image exams. We developed and evaluated a ConText-based algorithm with the GATE (General Architecture for Text Engineering) development system to automatically classify 1402 ultrasound radiology reports for AAA screening. Using the results from JAPE (Java Annotation Pattern Engine) transducer rules, we developed a feature vector to classify the radiology reports with a decision table classifier. We found that ConText performed optimally on precision and recall for pertinent negative (0.99 (0.98–0.99), 0.99 (0.99–1.00)) and pertinent positive AAA detection (0.98 (0.95–1.00), 0.97 (0.92–1.00)), and respectably for determination of non-diagnostic image studies (0.85 (0.77–0.91), 0.96 (0.91–0.99)). In addition, our algorithm can determine the AAA size measurements for further characterization of abnormality. We developed and evaluated a regular expression based algorithm using GATE for determining the three contextual conditions: pertinent negative, pertinent positive, and non-diagnostic from radiology reports obtained for evaluating the presence or absence of abdominal aortic aneurysm. ConText performed very well at identifying the contextual features. Our study also discovered

contextual trigger terms to detect sub-standard ultrasound image quality. Limitations of performance included unknown dictionary terms, complex sentences, and vague findings that were difficult to classify and properly code.

Keywords Classification · Coding · Natural language processing · Abdominal aortic aneurysm

Introduction

An abdominal aortic aneurysm (AAA) is the expansion of the infra-diaphragmatic aorta to a diameter of 3.0 centimeter (cm) or larger. AAA can be clinically silent but can rapidly expand and rupture becoming suddenly life-threatening if not surgically treated. The risk factors for AAA are age (≥ 65 years), smoking (≥ 100 cigarettes in a person's lifetime), gender (male), and previous family history [1–3]. Thus, the population at greatest risk for AAA and in whom surgical intervention will have greatest impact is male smokers, age 65–76 (after the age of 76 benefit of surgery drops off). This population constitutes the majority of the 4500 annual deaths in the USA from ruptured AAA with an additional 1400 deaths resulting from 45,000 repair procedures done to prevent rupture [4]. Within the USA, 4–8 % of older men and 0.5–1.5 % of older women have an AAA [2].

Several studies have shown that screening programs for AAA in male smokers greater than 65 years significantly reduce AAA-related mortality [1, 5]. Flemming et al. have shown that AAA screening program may reduce AAA-related mortality by 43 % in men age 65–75 years [6]. Ultrasonography of the abdomen is the preferred imaging modality for screening. Ultrasound (US) imaging is relatively inexpensive, accurate, and reliable for detecting AAA, with 95 % sensitivity and near 100 % specificity [7]. For these

✉ Craig Morioka
Craig.Morioka@va.gov

¹ Department of Radiology, VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA

² Medical Imaging Informatics, UCLA, Los Angeles, CA, USA

³ Department of Radiology, UCLA, Los Angeles, CA, USA

reasons, on January 1, 2007, Medicare began to offer US screening for AAA to all 65-year-old adult male enrollees with a smoking history of 100 or more cigarettes. In a 1997 study on the prevalence of AAA in a Department of Veterans Affairs (VA) male population between 50 and 79 years of age, 4.6 % had an AAA of 3.0 cm or larger, 1.4 % had AAA of 4.0 cm or larger, and 0.3 % had AAA of 5.5 cm or larger [8]. Therefore, in 2007, the Under Secretary for Health, Dr. Michael Kussman, implemented AAA screening throughout all Veteran Health Administration (VHA) facilities with the following recommended guidelines:

1. Men between the ages of 65 and 75 who have ever smoked are eligible for a one time screening exam preferably by ultrasound.
2. At a minimum, the screening study should examine the aorta from the renal arteries to the aortic bifurcation and report the greatest outside diameter measured perpendicular to the axis of the vessel. While ultrasound is the preferred modality, computer tomography (CT) and magnetic resonance imaging (MRI) can provide the required information and are acceptable alternatives. These patients may be studied by low-dose non-contrast CT.
3. Men with normal findings on screening (aortic diameter of less than 3.0 cm) do not need repeat imaging, as there is negligible health benefit in re-screening those with normal results. Men with aortic aneurysms of 3.0 to 3.9 cm need to be followed every 2–3 years, and those with aneurysms of 4.0 to 5.4 cm need to be followed every 6 months with repeat studies to monitor for growth of the aneurysms. If the aneurysm reaches 5.5 cm or larger, referral for surgical intervention needs to be considered.

In a 2005 survey of the VA enrollee population, 71.2 % of the population reported that they smoked at least 100 cigarettes during their lifetime [4]. A query to the VA analytics database shows that there were 1.5 million veterans between the age of 65 and 75 in 2013. If we utilize the 71.2 % smoking rate, then approximately one million veterans utilizing VHA services meet screening criteria. If all qualifying patients were to be referred for ultrasound in the first year, the imaging wait times would become unacceptably long for *all* ultrasound patients. Therefore, facilities need to be judicious in working down the backlog of patients to be screened. Radiologists at the VA Greater Los Angeles Healthcare System (VAGLA) are encouraged to report aortic diameter measurement on any imaging study that includes the abdominal aorta (CT, MR, PETCT) regardless of indication. Also, the radiologists are to utilize specific diagnostic codes on these reports in order to turn off the provider alerts once the screening exam is completed and to further alert the primary provider if the study is non-diagnostic or positive so that further action may be taken to complete the screening process. The automatic detection of

such data within the preexistent radiology reports would reduce demand for resources. Furthermore, the ability to mine current reports and correctly identify those patients in whom AAA has already been diagnosed, could allow for better tracking and follow-up. To that end, we developed and evaluated a ConText-based algorithm with the GATE development system to automatically classify 1402 AAA screening ultrasound free text reports for results of presence or absence of AAA.

The specifics of this particular mandate were used to generate the rules for classifying AAAs. Previously, natural language processing (NLP) algorithms have been used to extract concepts from free-text clinical reports to perform biosurveillance, clinical decision support, quality assurance, and summarization [9, 10]. In clinical documents, it may be important to determine if contextual features are negated, acute, chronic, or hypothetical. The identification of contextual features is not usually found in the lexical representation but in the context surrounding the clinical condition of interest. In our research, we utilize ConText, developed by Chapman et al. to identify contextual features to determine the clinical state of the patient [11].

Materials and Methods

Study Setting

This retrospective observational study was reviewed and approved by the VA Greater Los Angeles Research Institutional Review Board with a waiver of informed consent. The study population included ultrasound imaged patients in the Imaging Department from January 1, 2009 through Dec 31, 2013. We are a 740 bed urban teaching hospital with over 1,300,000 outpatient visits per year. The gender mix of our patient population is primarily adult male 91.6 % and adult female 9.4 %. The reports were dictated by five different fellowship trained body imagers. The average number of years' experience as an attending abdominal imager is 18.2 ± 10.4 years. Dictation was performed with speech recognition software integrated with our PACS system. Spell checking is done electronically but final approval on all changes must be approved by the radiologist.

The main objective of our research is to appropriately determine the correct VA national diagnostic code for AAA from radiology reports. The three steps required to complete the objective involve detection of the pertinent negative and positive AAA findings within the impression or conclusion section of an ultrasound report, (2) extract the measurement quantifying the maximal diameter from the report narrative, and (3) find the trigger words that define sub-optimal imaging studies. These three steps were implemented using the ConText algorithm. Table 1 shows the three AAA codes used to classify patients.

Table 1 Veteran Affairs AAA screening national codes

Code no.	Representation	Description	Alert
1200	AAA not present	The maximum width of the infrarenal aorta is less than 3 cm	No
1201	AAA present	The maximum width of the infrarenal aorta is at least 3 cm or greater	Yes
1202	Does not satisfy screening for AAA	Exam is not technically adequate for AAA screening	No

Evaluation (Gold Standard)

Three board certified radiologists, with 10+ years of post-graduate experience, manually tagged each report from 2012 to 2013 for pertinent positive, pertinent negative, and unacceptable image quality. The AC1 statistic, which is similar to a generalized kappa, allows one to assess inter-rater reliability of multiple readers. For readers 1 and 2, AC1 = 0.9969, 95 % CI (0.9933–1.000), and 2-sided $p < 0.001$. Readers 1 and 3, AC1 = 0.9682, 95 % CI (0.9568–0.9796), and 2-sided $p < 0.001$. Finally, readers 2 and 3, AC1 = 0.9854, 95 % CI (0.9789–0.9919), and 2-sided $p < 0.001$. Our results indicate excellent agreement between independent manual tagging of the AAA reports by our three radiologists. The significant p value indicates that this strong agreement did not happen by chance. The final gold standard was decided by two additional radiologists who analyzed all the manually tagged results from the three radiologists and decided ground truth for those reports where disagreement occurred. The gold standard committee also reviewed other imaging exams (CT and/or MR) of the abdomen to determine ground truth when such concurrent studies were available.

Processing VA Patient Records

To obtain a set of AAA patient records, we queried the VA DHCP/VISTA/CPRS electronic medical record (EMR) system for all ultrasound (US) AAA screening examinations from 2009 to 2013 performed at VA Greater Los Angeles Healthcare System. The records in free-text format were then processed using the GATE (General Architecture for Text Engineering natural language processing system [12–14]. GATE is an open source text analytics engine for natural language processing developed at the University of Sheffield and includes an information extraction module called ANNIE (A Nearly-New Information Extraction System) which is composed of a tokenizer, gazetteer, sentence splitter, part of speech tagger, named entities transducer, and a co-reference tagger (Fig. 1) [12]. The text of each report was decomposed into sentences and each sentence in turn was parsed into tokens. Tokens were obtained by simply splitting the text of each sentence by using spaces as a delimiter. Terms of interest were then identified within each sentence and were mapped to their

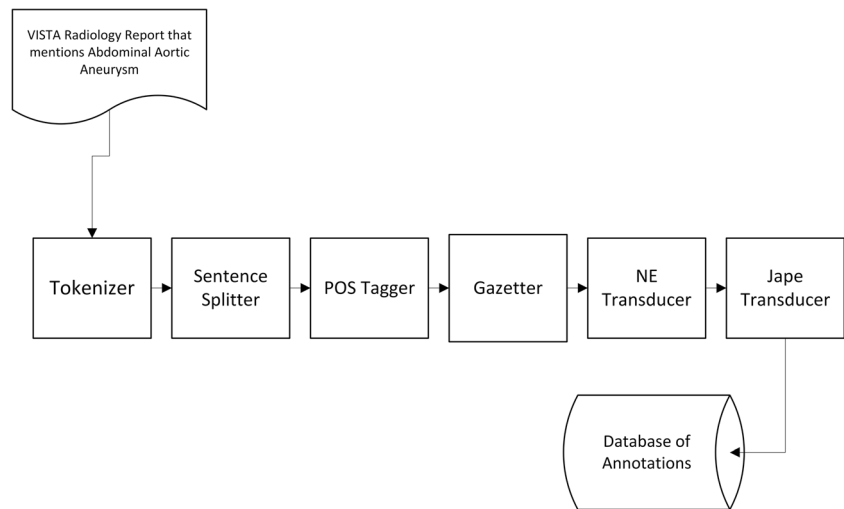
appropriate semantic representations. The semantic representation provides the meaning of the term of interest. The term of interest can be a negation, i.e., “no abdominal aortic aneurysm found,” positive affirmation, i.e., “3 cm abdominal aortic aneurysm found,” or ambiguous finding, i.e., “abdominal aortic aneurysm not visualized by poor quality of ultrasound exam.” In our particular case, the terms of interest consisted of phrase variations of AAA. The following is a sample list: abdominal aortic aneurysm, aortic aneurysm, and aneurysm of the abdominal aorta, ectasia of abdominal aorta, dilation of the abdominal aorta, etc. We were also interested in the size associated with each mention of AAA, and thus numeric terms were also identified as terms of interest. Numeric terms were extracted using a regular expression pattern. The advantage of a rule-based information extraction methodology includes the following: declarative statements, easy to incorporate domain knowledge, easy to comprehend and maintain, and easy to trace and fix errors. The disadvantages of rule-based text processing systems are the heuristic methodology, and tedious manual creation of new rules [15].

The general ConText algorithm in pseudo-code format:

1. Find trigger term in sentence,
2. If term is a pseudo-trigger term,
 - Find next trigger term,
3. Determine scope of trigger term,
4. If termination term within scope,
 - Terminate scope before termination term,
5. Assign appropriate contextual feature to corresponding clinical concept within scope. The contextual feature can be negation term, size measurement, or non-visual term applied to clinical concept.

ConText, which is an expansion of NegEx, relies on three types of terms: trigger terms, pseudo-trigger terms, and termination terms. In the case of negation, the trigger terms could be “no” or “not” along with a clinical concept (i.e., AAA). A clinical concept that falls within the scope of the trigger term will result in a negated clinical concept. Pseudo-trigger terms

Fig. 1 Data pipeline through GATE developer



contain negation trigger terms but are false positive for negation of a clinical concept. For example, “no increase in AAA diameter measurement,” “no increase” is a pseudo-trigger term that does not negate the clinical concept AAA. A termination term such as “but” can terminate the scope of the negation before the end of the window, as in “The AAA exists, but is not measurable due to the for-shortened view.” In our experiment, the window between the trigger term and clinical concept was up to five terms.

These are the steps performed in the training phase:

1. Input AAA reports into GATE as GATE document.
2. Add GATE documents into GATE corpus
3. Process corpus with ANNIE utilizing custom Java Annotation Patterns Engine (JAPE) transducers [16, 17]. JAPE transducers allow finite state processing of annotations with regular expressions.
4. Generate feature vector based on rules that were captured during step 3.
5. Respectively code reports based on feature vector utilizing decision table classifier.

A feature vector consisting of JAPE transducer rules was generated for each finding that mentioned AAA. The vector consisted of five rules: negative assertion of AAA, positive assertion of AAA, measurement of abdominal aorta, mild ectasia of AAA, and measure value of abdominal aorta ≥ 3.0 cm. The training phase was an iterative process to improve the precision and accuracy of detecting the non-diagnostic phrases, and pertinent positive and negative findings. This phase involved processing 468 ultrasound reports from the year 2009 to 2011 using GATE’s built-in JAPE. Using JAPE, regular expressions rules were written to discover pertinent negative and positive findings within the reports. The training set also was used to identify measurement patterns found in the free-text reports. AAA diameter sizes are

usually expressed as one-, two-, or three-dimension (anteroposterior (AP), transverse, and length) and described as (e.g., 2.9 cm, measuring 3.5×4.1 cm, $3.9 \times 4.8 \times 5.5$ cm). Finally, we discovered trigger words for sub-standard image exam. The training phase helped identify NLP findings under which concept matching (proper identification of AAA pertinent positive, pertinent negative, and non-visual study concepts) of desired term failed. In addition, we were able to determine contextual phrases that identified non-diagnostic image studies. All 468 reports were ultrasound screening procedures for AAA, and the analysis of this particular type of radiology report focused on the impression section. The second phase involves testing the final versions of the custom JAPE transducers on the test dataset. A separate set of 1402 ultrasound AAA screening reports from 2012 to 2013 were analyzed using GATE. Two of the authors identified all of the pertinent positive and negative AAA findings in the reports and the other two authors were involved in the failure analysis of the GATE output.

The reports were processed using a Java-based driver for executing the GATE system in batch mode. The output of GATE is a set of annotations that provide syntactic or semantic labels to segments of text. More specifically, each annotation consists of a unique identifier, a document identifier, start and end indices, an annotation type, and a set of features. The document identifier and start and end indices uniquely specify the lexical location of each annotation. Annotation types are names given to the labels assigned to text segments, and features are other metadata associated with each annotation type. For instance, the “Token” annotation represents a token (e.g., a word) and token features may consist of the part-of-speech of the token, the token’s word root, or whether the token was capitalized. All annotations were stored in a backend relational database to facilitate retrieval and provide organizational structure. To maximize flexibility and reduce processing overhead, features were stored as Javascript Object Notation

(JSON) strings because different annotation types will be associated with different sets of features.

Diameter Extraction

AAA diameter sizes are usually expressed as one-, two-, or three-dimension (anteroposterior (AP), transverse, and length) and described as (e.g., 2.9 cm, measuring 3.5 × 4.1 cm, 3.9 × 4.8 × 5.5 cm). Although there maybe one or more dimensions describing the aneurysm size. Only the maximum size of either AP or transverse dimension is selected for the final diameter measurement.

$$\text{Diameter size accuracy} = \text{number of correct sizes} / \text{number of TP AAA cases} \tag{4}$$

We also calculated the 95 % confidence intervals (CI) for the outcome measures.

Results

Of the 1402 patients included in this study, the mean patient age was 66.5 ± 3.2 and 99 % were men, representative for a VA study population. We determined pertinent positives and negatives for AAA's using NegEx and Context algorithms implemented within the GATE development environment as a JAPE (Java Annotation Patterns Engine) transducer. Results are listed in Table 2. Approximately 89.1 % of the radiology reports were pertinent negative for (excluded) AAA (1249/1402). 4.4 % of the reports were pertinent positive for (diagnosed presence of) AAA (62/1402). Finally, the inadequately or non-visualized aorta (91/1402) was reported in approximately 6.5 % of the total number of cases. Table 3 shows the result of automatically processing the radiology reports for pertinent negative AAA—class 1, pertinent positive AAA—class 2, and poor/non-diagnostic image quality—class 3. Table 4 lists the trigger words that indicated sub-standard image quality. Comparing the manual diagnostic codes entered by the radiologist on completion of their report, we found that 6.06 % or (85/1402) of reports had no diagnostic code entered.

Table 2 Outcome measures for NLP processing on test set of 1402 ultrasound AAA screening reports—decision table classifier

Precision (95 % CI)	Recall (95 % CI)	F-measure	Class
0.99 (0.98–0.99)	0.99 (0.99–1.00)	0.99	1
0.98 (0.95–1.00)	0.97 (0.92–1.00)	0.97	2
0.85 (0.77–0.91)	0.96 (0.91–0.99)	0.90	3

Outcome Measures

For each contextual feature (i.e., proper meaning of AAA depending on the other terms surrounding the concept AAA), we compared ContText's value to a value assigned by the gold standard. We calculated recall, precision, and F-measure or harmonic mean of precision and recall using the following formulas:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \tag{1}$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \tag{2}$$

$$F = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \tag{3}$$

Three of the 85 (3.5 %) reports with no diagnostic code were in fact positive for AAA. The remainder of the (82/1402) reports should have been coded for no AAA. Additionally, in 4.35 % of the cases (61/1402 reports), we found a discrepancy between the diagnostic code that the radiologist entered and the one obtained through our AAA classifier. The accuracy of the diameter measurement was 96 % (76/79) using our ConText driven approach.

Discussion

Narrative free-text is the primary communication method between clinicians in the medical domain. NLP offers the possibility to convert free text data into structured representations. There have been many efforts to apply NLP technologies to clinical text. The Medical Language Extraction and Encoding System (MedLEE) was one of the first rule-based NLP systems developed by Friedman et al. in 1994 at Columbia-Presbyterian Medical Center [18]. MedLEE was initially applied to radiology for the occurrences of four diseases (neoplasm, congestive heart failure, acute bacterial pneumonia, and chronic obstructive pulmonary disease) Natural Language Processed from chest X-ray reports. In 2006, Pakhomov et al. used NLP to assign diagnostic codes to

Table 3 Event classification table for 1402 ultrasound AAA screening reports

Test (classified as)	Response		
	1	2	3
1	1233	0	16
2	2	60	0
3	3	1	87

Table 4 Trigger words indicating sub-standard image quality

Trigger words
May not be completely visualized
Incompletely visualized on this study
Not well imaged
Only partially visualized
Not demonstrated on the images provided
Not well delineated on this study
Not well visualized
Not seen with sonography
Non-visualization
Incomplete visualization
Unable to visualize
Not visualized
Non-diagnostic study, exam, examination

patient encounters [19]. They used an example based and naive Bayes classifier to classify free-text to HICDA (Hospital International Classification of Disease Adaption) diagnostic codes. In 2006, HITex (Health Information Text Extraction) was developed by Goryachev et al. at Brigham and Women’s Hospital and Harvard Medical School [20, 21]. The system utilizes the Gate framework and was used to extract principal diagnosis, co-morbidity, and smoking status for asthma research from clinical notes. Topaz a rule-based approach to indexing concepts was developed by Harkema et al. is also built on the GATE platform and implements ConText as a GATE module that extracts 19 quality measures from free-text colonoscopy and pathology reports [22]. Similar AAA research was done by Sohn et al., they used NLP to find potential AAA reports based on keywords, and extract aneurysm size measurements from 650 radiology reports [23]. The NLP system was based on components from MedTagger, developed by the Mayo Clinic, and processed radiology reports utilizing Apache Unstructured Information Management Architecture (UIMA). The new contributions from our research include the following: identification of non-diagnostic quality image studies, the keywords used to identify poor image quality studies, larger document corpus for training and testing than previous studies, and quality control of manually diagnostic coded findings with NLP.

Limitations and Future Work

Some of ConText’s errors can be resolved by adding additional trigger and termination terms to the dictionary. There was a misspelled trigger term, i.e., “In the visible portion of the aorta there is no aneurysm but non- visualization of the complete abdominal aorta” (extra space after -) that was not detected by our algorithm which lead to a false negative for sub-standard image quality. This type of error was seen in Table 3, the test

case was 3 but the classifier response was 1 since there was no mention of an abdominal aneurysm and the non-visualization contextual term was not detected. Another example of the same type of errors: “Ectasia of the abdominal aorta, mid portion of the aorta not seen.” This report indicates possible portion of the aorta containing aneurysm, but after the comma the sentence acknowledges poor visualization of the middle portion of the abdominal aorta. Since a measurement is not given, the “ectasia” was not significant but the non-visualization of a portion of the aorta led the classifier to select response 3, but the majority of the abdominal aorta had no aneurysm greater than 3 cm. In this case, the report is insufficient in the description of the visual evidence. We also encountered a word pair form of aneurysm not seen in our training set “aneurysmal/dilation.” For example: “Extensive calcified atherosclerotic plaque throughout the abdominal aorta, without evidence of aneurysmal dilatation.” Another typical example: “Mild aneurysmal dilation of the infernal abdominal aorta.” Table 3 shows test 2 but the response is 1 in this case since our algorithm failed to identify an abdominal aortic aneurysm. The most common error for the non-visualization of the abdominal aorta was an additional finding that indicated negation of the AAA for the portion of the abdominal aorta that was visible. Another difficult case was the compound sentence, “mild ectasia of the abdominal aorta and the iliac artery was not visualized.” Our algorithm mistakenly identified the “not visualized” portion of the sentence with the abdominal aorta. This again is depicted as test 1 but response 3 in Table 3. Currently, our study represents only one VA institution, but generalization of this work to additional VA hospitals is possible by discovering each individual institutions particularly unique phrasing of positive, negative, and visually impaired AAA exams.

Conclusion

We developed and evaluated a regular expression based algorithm using GATE for determining the status of three contextual features in US radiology reports. We found that ConText performed very well at identifying pertinent negative AAA and moderately well at positive AAAs, as well as determining non-diagnostic ultrasound studies. In addition, our algorithm can determine the AAA size measurements from the radiology report to determine severity of the disease. We now have the ability to check the radiologist’s diagnostic code for accuracy compared to the findings in the final report. Our study also discovered contextual trigger terms to detect sub-standard image studies for ultrasound. We found limitations in our methods that involved unknown dictionary terms, complex sentences that required more than correlation between multiple sentences, and vague or contradictory findings that were difficult to classify and properly code.

References

- Benson RA, Poole R, Murray S, Moxey P, Loftus IM. Screening results from a large United Kingdom abdominal aortic aneurysm screening center in the context of optimizing United Kingdom National Abdominal Aortic Aneurysm Screening Programme protocols. *J Vasc Surg.* 63(2):301–304, 2016.
- Pande RL, Beckman JA: Abdominal aortic aneurysm: populations at risk and how to screen. *J Vasc Interv Radiol* 19(6):52–58, 2008
- Lederle FA, Johnson GR, Wilson SE, et al: The aneurysm detection and management study screening program validation cohort and final results. *Arch Intern Med* 160(10):1425–1430, 2000
- Bondurant S, Wedge R: *Combating Tobacco use in Military and Veteran Populations.* Institute of Medicine of the National Academies. Washington, DC: The National Academies Press, 2009
- Ye Z, Bailey KR, Austin E, Kullo IJ: Family history of atherosclerotic vascular disease is associated with the presence of abdominal aortic aneurysm. *Vasc Med* 41(1): 41–46, 2016.
- Fleming C, Whitlock EP, Beil T, Lederle F: Screening for abdominal aortic aneurysm: a best-evidence systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* 142:203–211, 2005
- Lindholt JS, Vammen S, Juul S, Henneburg EW, Fasting H: The validity of ultrasonographic scanning as screening method for abdominal aortic aneurysm. *Eur J Vasc Endovasc Surg* 17:472–475, 1999
- Lederle FA, Johnson GR, Wilson SE, et al: for the Aneurysm Detection and Management (ADAM) Veterans Affairs Cooperative Study Group. “Prevalence and associations of abdominal aortic aneurysm detected through screening”. *Ann Intern Med* 126:441–449, 1997
- Demner-Fushman D, Chapman WW, McDonald CJ: What can natural language processing do for clinical decision support? *J Biomed Inform* 42(5):760–772, 2009
- Elkin PL, Froehling DA, Wahner-Roedler DL, Brown SH, Bailey KR: Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. *Ann Intern Med* 3(156):11–18, 2012
- Chapman WW, Chu D, Dowling JN: ConText: an algorithm for identifying contextual features from clinical text. *BioNLP*, 2007 pp 81–88
- Cunningham H, Maynard D, Bentchevz K, Tablan V, et al: *Text Processing with GATE (Version 6).* University of Sheffield Department of Computer Science. 15 April 2011
- Konchady M: *Building Search Applications: Lucene, Lingpipe and Gate.* Mustru Publishing, Seattle, 2008
- Bontcheva K, Tablan V, Maynard D, Cunningham H: Evolving GATE to meet New challenges in language engineering. *Nat Lang Eng* 10(3/4):349–373, 2004
- Chiticariu L, Yunyao L, Reiss FR: Rule-based information extraction is dead! Long live rule-based information extraction systems! *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp 827–832
- Thakker D, Oshman T, Lakin P: *JAPE Grammar Tutorial*, <http://gate.ac.uk/sale/thakker-jape-tutorial/GATE%20manual.pdf>
- Cunningham H, Maynard D, Tablan V: *JAPE: a Java Annotation Patterns Engine*, 2nd edition. Research Memorandum CS-00-10, Department of Computer Science, University of Sheffield, November 2000
- Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB: A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc* 1(2):161–174, 1994
- Pakhomov S, Buntrock J, Chute C: Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques. *JAMIA* 13(5):516–525, 2006
- Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R: Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decis Mak* 6:30, 2006
- Goryachev S, Sordo M, Zeng QT: A suite of natural language processing tools developed for the I2B2 project. *AMIA Annu Symp Proc*, 2006, p 931
- Harkema H, Dowling J, Thornblade T, Chapman W: Context: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *J Biomed Inform* 42(5):839–851, 2009
- Sohn S, Ye Z, Liu H, et al: Identifying abdominal aortic aneurysm cases and controls using Natural Language Processing of Radiology Reports. *AMIA Jt Summits Transl Sci Proc*; May 18, 2013, pp 249–253