CM&R *Rapid Release*. Published online ahead of print May 25, 2012 as doi:10.3121/cmr.2012.1047

Original Research

Towards Automatic Diabetes Case Detection and ABCS Protocol Compliance Assessment

Ninad K. Mishra MD, MS†, Roderick Y. Son, PhD‡, and James J. Arnzen‡

†Centers for Disease Control and Prevention, Atlanta GA
‡Northrop Grumman, Atlanta GA

Word Count (Text): 4943 text; 286 abstract; 6 tables; 4 figures

Corresponding Author:
Ninad K. Mishra, MD, MS
Centers for Disease Control and Prevention
1600 Clifton Rd, Mail Stop E76
Atlanta, GA 30333
Phone: (404) 498-6289
Fax: (404) 498-6620
Email: nmishra@cdc.gov

Received: August 29, 2011
Revised: January 29, 2012
Accepted: February 22, 2012

doi:10.3121/cmr.2012.1047

Funding Source: CDC Funded Research

Copyright 2012 by Marshfield Clinic.

Abstract

Objective: According to the American Diabetes Association, the implementation of the standards of care for diabetes has been suboptimal in most clinical settings. Diabetes is a disease which had a total estimated cost of $174 billion in 2007 for an estimated diabetes-affected population of 17.5 million in the United States. With the advent of electronic medical records (EMR), tools to analyze data residing in the EMR for healthcare surveillance can help reduce the burdens experienced today. This study was primarily designed to evaluate the efficacy of employing clinical natural language processing to analyze discharge summaries for evidence indicating a presence of diabetes, as well as to assess diabetes protocol compliance and high risk factors.

Methods: Three sets of algorithms were developed to analyze discharge summaries for: 1) identification of diabetes, 2) protocol compliance, and 3) identification of high risk factors. The algorithms utilize a common natural language processing framework, which extracts relevant discourse evidence from the medical text. Evidence utilized in one or more of the algorithms includes assertion of the disease and associated findings in medical text, as well as numerical clinical measurements and prescribed medications.

Results: The diabetes classifier was successful at classifying reports for the presence and absence of Diabetes. Evaluated against 444 discharge summaries, the classifier's performance included macro and micro F-scores of 0. 9698 and 0. 9865, respectively. Furthermore, the protocol compliance and high risk factor classifiers showed promising results, with most F-measures exceeding 0.9.

Conclusions: The presented approach accurately identified diabetes in medical discharge summary and showed promise with regards to assessment of protocol compliance and high risk factors. Utilizing free-text analytic techniques on medical text can complement clinical-public health decision support by identifying cases and high risk factors.

Keywords: Natural Language Processing, Diabetes Mellitus, Public Health Informatics

Introduction

In 2010, it is estimated that 18.8 million people in the United States live with the diagnosis of diabetes.[1] The total estimated cost of diabetes in 2007 is $174 billion, including $116 billion in excess medical expenditures and $58 billion in reduced national productivity.[2] The burden of diabetes is shared by all sectors of the society – higher insurance premiums paid by employees and employers, reduced earnings due to disease and loss of productivity and most importantly reduced overall quality of life for people with diabetes.

Disease surveillance can be significantly overcome by the difficulty in identifying the target population. [3] Even though manual chart review remains the gold standard for the identification of individuals with diabetes, it can be supported by employing natural language processing (NLP) tools to reduce the burden of labor intensive manual processes. [4 5] This preliminary screening can be accomplished by analyzing physician's notes, medication lists, problem lists and discharge summaries by NLP tools as patient records are being increasingly stored in digital format. [6 7] In addition, it is also feasible to estimate adherence to quality of care protocols such as ABCS (*i.e.,* **A**1C - glycosylated hemoglobin, **B**lood Pressure, **C**holesterol, **S**moking) indicators for comparative effectiveness and population health measurement purposes.

In diabetes mellitus, higher amounts of glycosylated hemoglobin, indicates poor control of blood glucose in the past and have been associated with cardiovascular disease, nephropathy, and retinopathy. Monitoring the HbA1c in patients with type-1 diabetes improves treatment and could also provide potential benefit for the population level quality-of-care assessment.

According to the American Diabetes Association (ADA), the implementation of the standards of care for diabetes has been suboptimal in most clinical settings. A recent study maintains that

only about 57.1% of adults with diagnosed diabetes achieved an A1C of < 7%, only 45.5% had optimum blood pressure measurements, and about 46.5% had a total cholesterol < 200 mg/dl. Most importantly, only 12.2% of people with diabetes achieved all treatment goals of ABCS. [8 9]

"A major contributor to suboptimal care is a delivery system that too often is fragmented, lacks clinical information capabilities, often duplicates services and is poorly designed for the delivery of chronic care." [10]

While a simple solution to these problems is not readily available and beyond the scope of this study, having the patient data in digital format (whether structured or unstructured) provides an opportunity to screen medical records for mentions of diabetes and related medications. With the provisions of richer data sources, it may also be possible to screen medical digital records for adherence to protocols related to quality of care such as ABCS. Electronic Health Records (EHRs) have the potential to: 1) advance the quality of healthcare by providing timely access to medical information, 2) ensure that the patients receive guideline-protocol recommended care and 3) offer decision support for clinical and administrative healthcare providers. Developing tools to analyze data residing in the EHR for healthcare surveillance can help reduce the burdens experienced today. This potential has not been realized largely because the medical information is currently stored in an unstructured, free-text format and currently existing structured coding is often driven by financial imperatives.

Various natural language processing technologies have been applied to the clinical domain[11]. Table I lists related works and lists the distinctive characteristics of each work including: 1) disease domain applied, 2) data source that the classification was being performed against, 3) the classification approach employed (i.e., statistical vs. rule-based), 4) if it supports ABCS protocol compliance detection, and 5) if it supports high risk factors detection. Some of these have used

ConText [12 13] and NegEx [14] to perform concept extraction of medical terminology and apply a set of medically relevant descriptive qualifiers, such as negation, to the task. These two regular expression based concept extraction approaches have been applied as an initial processing step to various tasks including identifying the presence or absence of medical findings in support of clinical trials screening [7], obesity screening [15 16], smoking status[17], medication identification [18] and pathology report analysis [19]. This work is an extension of the authors' approach[15] employed in the i2b2 Obesity Challenge and extends the methodology to a different dataset and to novel application domains described below.

Reviewing the related i2b2 Obesity Challenge approaches, of the top five teams in the textual judgment task in the 2008 i2b2 Obesity Challenge, all used rule-based approaches. At a high level, all of these approaches examined the discharge summaries for the occurrence of clinical terms associated with each morbidity and tried to establish the context in which each term was mentioned. The context could be temporal (e.g. “history of _”), an indicator of the experiencer (e.g. the patient or a family member), or an indicator that the person did or did not have the relevant morbidity or if it was questionable that the patient had the morbidity. Once the term mentions and their contexts had been identified, rules were applied to integrate the results and establish a judgment for each document. Some of the major differences in the approaches would be due to text pre-processing, content of the dictionaries, the variation in the algorithms to identify positive, negative, and uncertain term mentions, the approach to ranking or ignoring certain sections of the discharge summaries, and the rules applied to integrate the results and establish a judgment for each document

The focus of the current work is to build on our obesity challenge work described above and extend these methodologies to the new domains of compliance assessment and risk factor

identification in addition to the classification task required by the i2b2 task. This work seeks to facilitate prescreening through analysis of free-text discharge summaries by applying ConText to check for linguistic evidence indicating a presence of diabetes, ABCS protocol assessment and high risk indicators. This work also explores the potential benefits of medication discourse to support the classification process.

The primary contributions of this work includes the novel application of clinical NLP to support ABCS protocol assessment and high risk detection, the development of concept dictionaries for ConText and heuristics diabetes classification rules to support the above classifiers in a limited data and resource environment.

Methods

We used a rule-based clinical NLP approach to classify discharge summaries to: 1) identify reports indicating diabetes, 2) identify reports demonstrating ABCS protocol compliance, and 3) identify reports with high risk factors such as high BP or cholesterol values (LDL≥ 100).

Classification Algorithms

Diabetes Classifier

The diabetes classifier, given a discharge summary, utilizes discourse evidence, both direct utterances of diabetes and diabetes medication utterances to identify reports associated with diabetes. In this study, discharge summaries are classified into one of three possible *judgment categories* (also referred to as *indication classes*) with respect to diabetes:

- **POSITIVE** - Positive indications of diabetes (*e.g.,* diabetes medication listed in discharge report; explicit statement of potential/existing diabetes condition)
- **NEGATIVE** - Negative indications of diabetes (*e.g.,* explicit statements supporting the absence of diabetes)
- **UNKNOWN** - Indeterminate indications of diabetes (*e.g.,* no mention of diabetes or related medications supporting or refuting the presence of diabetes)

Figure 1 illustrates the rules utilized to perform the classification. As shown in the diagram, explicit discourse by the physician indicating the presence or absence of diabetes takes precedence. Diabetes medication discourse can also be employed to affirm diabetes, but discourse mentioning non-use of such medications cannot exclude diabetes (e.g., non-insulin-

dependent diabetes mellitus). The $f_{status}(d, type)$ notation in Figure 1 is discussed later in this paper.

Discharge summaries that are deemed "POSITIVE" can be further analyzed to assess adherence to the ABCS protocol, as well as identify potential high risk factors.

ABCS Protocol Compliance Classifier

The ABCS protocol compliance classifier analyzes discharge summaries with a positive mention of diabetes to further assess compliance to ABCS protocol as follows:

- **SATISFIED** – Assessment for the risk factor has been performed (e.g., blood pressure was checked).
- **UNSATISFIED** – No mention of test, examination or risk factors (*e.g.,* physician makes no mention of blood pressure)

Figure 2 illustrates the rules utilized for 3 of the 4 ABCS protocol compliance assessment categories. The readers are referred to the i2b2 Smoking Competition [20] for smoking assessment. As demonstrated in the assessments, discourse evidence such as direct mentioning of the examination, abnormal indications, as well as measurements encountered in the discourse can be used to determine compliance.

High Risk Factors Classifier

In contrast to the compliance classifier described above, the high risk factor classifier looks at values and assertions to determine the presence or absence of a high risk factors such as LDL>110.

High risk factor assessment categories, based on ABCS indicators, are classified into one of two classes:

- **INDICATED** – High risk factor is present (e.g., hypercholesterolemia).
- **NOT INDICATED** – High risk factor is not present (*e.g.,* no mention of abnormal cholesterol or LDL measurement exceeding normal range).

Figure 3 illustrates the rules utilized for 2 of the 4 risk factor assessment categories. We used evidence from the clinical measurements mentioned in the text as well as clinical mention in the form of an assertion (e.g., "Patient X has a history of hypertension.") In addition to the smoking risk factor, which is again deferred to the i2b2 Smoking Competition results, the A1C risk factor is deferred for future investigation, due to paucity of data used in this study.

Classifier Framework

To perform each of the classification tasks described above, a clinical NLP framework using ConText has been employed for each of the classifiers. As illustrated in Figure 4, the analysis performed on a discharge document is broken down into multiple stages of: 1) Concept Extraction, 2) Concept Summarization, 3) Measurement Extraction (used only for the ABCS classifiers) and 4) Classification. In the Concept Extraction stage, an unstructured free-text document is transformed into a set of concept instances (*e.g.,* "diabetes") that are encountered in the documents with various attributes associated with the concepts. The concept instances are summarized based on the collective extraction results to characterize the likely presence or absence of the concept (*e.g.,* has indications of diabetes). The resulting output from Concept Extraction is a structured representation of the discourse within the document. Additional measurement extraction is performed to support the ABCS protocol compliance and high risk

factors classifiers through the extraction of relevant values, such as LDL and blood pressure measurements. Based on the structured document, a rule-based classification is performed, and the appropriate judgment class from above is assigned. The following subsections describe the details of each of these stages.

Concept Extraction

The initial step of classifying a discharge summary entails extracting *concepts of interest* at the utterance level from within the summary document. Employed in this initial step is ConText, a clinical natural language processing tool developed by Chapman, et al. [12 13], which utilizes a regular expression approach to extract concepts of interest. ConText extracts concepts from a document such that each concept is associated with its experiencer (*e.g.,* patient, other), affirm/negate state (*i.e.,* has or does not have indications of a condition/concept), historical/currency context and the sentence from which the concept was found. Examples of these concepts include “diabetes”, “glucose test”, “hypercholesterolemia”, and “hypertension”. These concepts are included in one or more dictionaries that are employed by ConText.

For this study, we defined seven different dictionaries employed by ConText, as shown in Table II. The first five dictionaries were used to identify affirmed or negated concepts within a document for their respective concept classes. Only concepts associated with the patient as the experiencer were considered. Through the compilation of the affirmative and negative concept mentions, 5 sets of affirmative/negative counts for each document were derived. The remaining two dictionaries were employed to identify candidate sentences that showed indications of blood pressure or LDL measurements associated with the patient, consistent with ABCS protocol compliance and high risk factor assessments.

The customized dictionaries were developed by domain experts consisting of a medical informatician and two medical doctors, including the principal author. The creation of the terminology lists included a review of the discourse in the "development" data set portion of the i2b2 data set used in this study, and supplemented by the domain experts.

Given a sentence in a discharge summary and a concept of interest from a dictionary, ConText checks for the existence of the concept within the sentence. If the concept is found, ConText determines whether or not the utterance of the concept refers to the existence (*i.e.,* an affirmative mention) or absence (*i.e.,* a negated mention) of that concept. Also, ConText determines the experiencer to whom the concept refers, such as the patient or other (*e.g.,* family member). In this study, we were only interested in concepts that were attributed directly to the patient. Thus, family history, although embedded within a discharge summary, was not used as a contributing factor towards assessing indications of diabetes for the patient. Using ConText, findings associated with the patient and others can be differentiated, and reduce the chance of false inclusion or exclusion of a patient status based on non-patient findings.

Upon processing a discharge summary d with the concepts of interest from each of the dictionaries listed in Table II, a set of these structured representations of concept instances $c \in d$ results. These concept instances (*e.g.,* utterance of "dm") can then be employed for determining the existence or absence of the concept class (*e.g.,* ""Diabetes - General") from the patient's discharge summary.

To better illustrate the utilization of ConText, consider the following example and the concept "diabetes":

His hemoglobin A1C levels actually came back within normal limits; however, the wide excursions of his sugars were consistent with possible dm.

The resulting characterization from ConText would be:

- *c.class*= `gen` [Diabetes - General]
- *c.term*= dm
- *c.status*= Affirmed
- *c.experiencer*=Patient
- *c.history*= Recent

where *c* is the concept utterance instance associated with the term "dm" encountered in the sentence.

Concept Summarization

After employing ConText to identify and characterize all utterances pertaining to the concepts of interest, a summarization step is employed to determine the status of the concept (*i.e.,* affirmed, negated or indeterminate) within the discharge summary. Within a discharge report *d*, all utterances pertaining to the same concept class *c.class* and associated with the patient are subsequently summarized based on the number of affirmed and negated utterances, as determined during ConText concept extraction. A function $f_{status}(d, type)$ is defined such that:

$$f_{status}(d,type) = \begin{cases} AFFIMED, & if\ aff\ d,type\ > neg(d,type) \\ NEGATED, & if\ aff\ d,type\ < neg(d,type) \\ INDETERMINATE, & otherwise \end{cases} \quad (1)$$

where *aff*(*d*,*type*)=|{ *c*∈*d*|*c.class*=*type*,*c.status*=AFFIRMED }| is the number of affirmed instances of the concept class *type* and *neg*(*d*,*type*)=|{ *c*∈*d*|*c.class*=*type*,*c.status*=NEGATED }| is the number of negated instances of the concept class *type*. The summarization is applied to all concept classes in Table II.

For each of the concept classes in this study, all concept term utterances associated with the class are treated as corresponding to a singleton entity instance. For example, diabetes (as characterized in the "Diabetes - General" dictionary) is treated as a singleton entity that exists or doesn't exist (or is otherwise indeterminate) for a given patient. Although special cases, such as gestational diabetes, can in fact be treated as separate instances over time, for this study, such a distinction is not made. Similarly, for the scope of this study, an exam was either performed or not performed; whether an A1C exam was performed multiple times or performed just a single time is not differentiated in this study. Concept classes (*e.g.,* tumors) that can pertain to multiple-instance entities (*e.g.,* multiple physical findings of tumor mentioned within a patient image report), where each entity is uniquely identified, are not addressed. Such tracking of multiple entity instances would require more extensive NLP analysis utilizing techniques, such as coreference resolution [21-24], which is outside the scope of this work.

Measurement Extraction

For blood pressure assessment and cholesterol assessment, an aggregate approach is employed, which incorporates both the concept extraction and summarization mechanisms described above, as well as employing blood pressure value extraction and LDL measurement extraction techniques, modeled after Turchin's regular expression blood pressure extraction patterns [25].

Both extraction techniques utilized regular expression patterns to identify the respective measurement values, which were then used as a part of the ABCS classification process.

Classification of Diabetes Indication

Once summarization for each of the concept classes has been performed, the discharge summary is classified into its judgment category by utilizing a heuristic set of rules focusing on the General Diabetes Discourse and Diabetes Medication concept dictionaries. The heuristics are based on the premise that a diabetes "POSITIVE" discharge summary would include discourse that explicitly states that the patient has diabetes, or may implicitly indicate having diabetes through medications associated with the patient that are known, or highly likely to be utilized for diabetes management.

The rule-set employed in this study for classifying reports of interest for diabetes is shown in Figure 1, where : $f_{status}(d,$ `gen`$)$ and $f_{status}(d,$ `med`$)$ are the summarized concepts for general diabetes terminology and diabetes medications, respectively.

Classification of ABCS Protocol Compliance & High Risk Factors Assessment

A similar strategy used in determining diabetes indication can also be applied to assess ABCS Protocol adherence for patients who have been classified as "POSITIVE" for diabetes. Utilizing the same concept extraction and summarization mechanisms employed in diabetes classification, information can be extracted from the report to determine ABCS adherence.

In the ABCS protocol compliance classifiers, simple concept extraction is employed to determine if A1C, blood pressure or cholesterol was mentioned, with the only restriction that the concepts are associated with the patient as the experiencer. For the abnormal findings, both concept extraction and summarization were employed to determine the presence of the respective

abnormalities: f_{status}*(d,* `hbp`*)=AFFIRMED* and f_{status}*(d,* `hc`*)=AFFIRMED*. Regular expressions were used to extract the blood pressure and LDL measurements from the discharge summaries, if present. For the high risk factor assessment classifier, the values of the measurements were examined to determine if they were outside the norm.

Because of the paucity of known diabetes discharge summary reports from the existing data set (i.e., only 56 positive diabetes cases), only preliminary results of the ABCS protocol assessment analysis are presented in this work. In this study, only ABC (i.e., Hb**A**1C, **B**lood Pressure and **C**holesterol) are evaluated. Strategies for smoking assessment are referred to the i2b2 Smoking Competition [20].

Data Set

To evaluate the technique for classifying discharge summaries with regards to diabetes, a de-identified set of free-text patient discharge summary reports from the First i2b2 Shared Task (i2b2 Smoking Challenge) [20] was utilized. The i2b2 data set consists of 889 de-identified discharge summaries [26]. The discharge summaries were tagged by the committee of domain experts, assessing whether or not a report indicated diabetes. The review process consisted of two reviewers independently classifying the reports for diabetes with the third reviewer being used when a disagreement was encountered. The manual tagging of the reports was used as the gold standard for the classification evaluation. For this study, the reports were divided into two sets: 445 reports were used as the development set and 444 reports were used as the test set to evaluate the technique. In the test set, there were 56 positive reports and 5 negative reports with the remaining 382 reports not having any indications supporting or refuting the existence of diabetes, as determined by the committee.

Classifier Evaluations

Three implementations of the diabetes classifier described above were employed in the evaluation of the efficacy of the presented methodology. Each variant employs the same concept extraction and concept summarization approach described above. However, each exercises a different variation of the classification algorithm illustrated in Figure 1.

- **Discourse Only**: Utilizes only general diabetes discourse indicators in classification process; exercises only the "Diabetes Indicated?" decision point. No mention of diabetes results in an "Unknown Indication of Diabetes".
- **Medication Only**: Utilizes only diabetes medication indicators in classification process; exercises only the "Diabetes Medication Indicated?" decision point.
- **Combined**: Incorporates both diabetes discourse and medication indicators; exercises complete algorithm.

Of the three implementations, the *Medication Only* classifier can only classify reports as "POSITIVE" or "UNKNOWN" for diabetes. The inability to classify to the "NEGATIVE" judgment category is due to the fact that an absence or explicit non-use of a diabetes medication does not preclude the existence of the disease (*e.g.,* non-insulin-treated type 2 diabetes). The *Discourse Only* and *Combined* solutions can classify to all three judgment categories.

The ABCS protocol compliance and high risk factor classifiers were also evaluated.

Evaluation Metrics

To evaluate each of the classifier's ability to correctly classify a document, the F-measure metric was used, which combines precision π and recall ρ into a single value. Both micro-averaged and macro-averaged F-measures were computed in this study [27].

The micro F-measure (F_μ) is a global computation of F-measure across all category decisions, and is computed using the equations below:

$$\pi = \frac{\sum_{i \in M} TP_i}{\sum_{i \in M} (TP_i + FP_i)} \quad (2)$$

$$\rho = \frac{\sum_{i \in M} TP_i}{\sum_{i \in M} (TP_i + FN_i)} \quad (3)$$

$$F_\mu = \frac{2\pi\rho}{\pi + \rho} \quad (4)$$

where π and ρ represent the aggregate precision and recall values.

The macro F-measure (F_M) is an average of each category's local F-measure, and is computed using the equations below:

$$\pi_i = \frac{TP_i}{TP_i + FP_i} \quad (5)$$

$$\rho_i = \frac{TP_i}{TP_i + FN_i} \quad (6)$$

$$F_i = \frac{2\pi_i\rho_i}{\pi_i + \rho_i} = \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (7)$$

$$F_M = \frac{\sum_{i \in M} F_i}{M} \quad (8)$$

where π_i, ρ_i and F_i represent the precision, recall and F-measure values for a given category *i* (*e.g.,* "POSITIVE", "NEGATIVE", or "UNKNOWN" for the diabetes classifier).

Both computations are a function of the number of true positive TP_i, false positive FP_i and false negative FN_i for each category in set *M*, the possible classification outcomes.

Results

The discharge summary classifier was developed based on a Java implementation of ConText.

Diabetes Classification Results

Based on the described methodology, three evaluations were performed, coinciding with the objectives targeted by this study: 1) determining indications of diabetes, and 2) examining the discriminating potential of the *Discourse Only*, *Medication Only* and *Combined* approaches. The classifier was executed against the 444 discharge summary reports for indications of diabetes utilizing the presented methodology.

The initial evaluation was performed on the three variations of the diabetes indication classifier (*i.e., Discourse Only*, *Medication Only* and *Combined*). The resulting precision, recall and F-measure results, as well as the aggregate F-measure results, for the three possible classification outcomes are shown in Table III.

A second analysis was performed where "NEGATIVE" and "UNKNOWN" judgment categories are combined into a single category, to compare *Medication Only* to *Discourse Only* and *Combined*, because *Medication Only* approach cannot definitively identify "NEGATIVE" cases.. Combining these judgment categories focuses the metric on high precision for positive cases, thus treating the remaining cases as "non-positive"; this statistical approach may be more appropriate for studies where falsely positive cases may more significantly bias a study than falsely negative cases. The results of the second analysis are shown in Table IV.

ABCS Protocol Compliance and High Risk Factor Assessment Results

To obtain a preliminary assessment of detecting ABCS protocol compliance, as well as risk assessment, two sets of analyses were performed using the above methodology. The ABCS

protocol compliance classifier was executed against the 56 known diabetes discharge summary reports (as determined by our gold standard) for ABC compliance utilizing the methodology described above. The resulting precision, recall and F-measure results, as well as the aggregate F-measure results, for the three possible classification outcomes are shown in Table V.

A further evaluation was performed to determine high risk factors, again using the presented methodology. Due to paucity of data, specifically the limited number of diabetes-related discharge summaries with abnormal A1C measurements in our data set, only high blood pressure and high cholesterol were evaluated. The associated results are shown in Table VI.

Discussion

Utilizing ConText for identifying diabetes "POSITIVE" discharge summaries shows promising results with macro and micro F-scores of 0.9500 and 0.9865 when employing a three-class classifier (*i.e.,* "POSITIVE", "NEGATIVE" and "UNKNOWN" indication classes), and macro and micro F-scores of 0.9698 and 0.9865 when combining "NEGATIVE" and "UNKNOWN", into a single category.

Both *Discourse Only* and *Combined* classifier results showed comparable F-score results with the *Combined* classifier having a slight improvements in F-score relative to *Discourse Only*, as shown in Table III. The improvement shown in the *Combined* classifier stemmed from the incorporation of medication evidence, resulting in two additional correctly classified "POSITIVE" discharge summary reports, without incurring appreciable addition of misclassified reports.

In our error analysis of the *Medication Only* classifier, only one false positive was observed in the "POSITIVE" indication class. This false positive was caused by an observed diabetes

medication (*i.e.,* insulin), which was being employed for treatment of hyperkalemia. Although rare, the false positive error points out the potential for a diabetes medication observation to be misleading without consideration of other details in the discourse, such as the discourse of "hyperkalemia", a condition in which insulin is used sans diabetes. This high precision outcome resulted from the notion that reports including diabetes medications would have a high likelihood of indicating a patient's history of diabetes.

The F-scores for the *Medication Only* classifier were lower than either the *Discourse Only* and *Combined* classifiers. The absence of diabetes medication from the discharge summary does not strictly indicate an absence of diabetes as a disease condition, as hyperglycemia may also be managed by other interventions such as exercise and diet. Thus, medication evidence cannot be independently employed to classify discharge summaries for diabetes. However, combined with other discourse evidence, medication can supplement the classification process because of its high precision when diabetes medication discourse is encountered in a report.

Other contributors of misclassification that were encountered in the evaluation included an observation of diabetes discourse that was incorrectly associated to a patient when actually it was in reference to a family member's condition. In this case, the report would be incorrectly identified as positively indicating diabetes, when no other evidence supports a positive classification. Another error stemmed from a verbose list of principal diagnoses in one of the documents, which included a negation of an unrelated condition (*e.g.,* "without remission") along with the diagnosis of diabetes. Due to the structure of the discourse (*e.g.,* the list of diagnoses were uttered in a single sentence), the "without" negation was inadvertently associated with the both the "remission" and the "diabetes" utterances.

Although encumbered by limited data, ABCS protocol compliance and high risk factor classifiers show promising results using the concept extraction and summarization methodology with all but 2 F-measures above 0.9; the remaining two metrics were above 0.8.

One observation to note regarding the ABCS protocol compliance results is the apparent lack of A1C and cholesterol reporting for diabetes patients in contrast to blood pressure reporting. Although this observation might suggest a lack of ABCS protocol compliance, it may also suggest that evaluating ABCS protocol compliance should be performed across multiple reports over a range of time, as A1C and cholesterol tests may not be ordered during a given encounter, if they had been performed during a prior, recent visit.

Conclusion

In this study, we have demonstrated that it is possible to extract and classify concepts directly tied to diabetes from medical text with high precision and recall to support not only diabetes detection, but also protocol compliance and risk factor assessments. Natural language processing using ConText in conjunction with heuristic rules can help us to process a great number of medical records for preliminary screening at minimal cost and potentially augment better quality-of-care by tagging medical records for concepts of interest; such relatively simple to implement rule-based approaches can perform well, especially in circumstances where training data is limited, a potential shortcoming for statistical approaches.

One challenge in the current design of this study is distinguishing between different types of diabetes, for example, the differentiation between diabetes mellitus type 2 and gestational diabetes. In free-text discourse, understanding the true meaning of the utterance “diabetes” requires word sense disambiguation (WSD). Beyond WSD, disambiguation of the underlying

motivation for an event is also a challenge requiring further exploration to identify ways of resolving such ambiguities, as well understand the prevalence of potential false positive classification (e.g., misclassifying insulin usage as an indication of diabetes) resulting from these ambiguous concepts.

Although effective against discharge summaries, the true value of disease screening can only be realized when we use NLP based systems in conjunction with other sources of information such as procedure lists, problem lists, diagnosis codes etc. Most of the factual and objective pieces of information can be derived from rule based systems, but these systems lack the ability to be inferential which is a hallmark of a domain expert's judgment. Hence we recommend these tools be used as preliminary screening devices to assist the gold standard manual processes such as chart reading performed by people with required expertise.

The preliminary screening of textual medical records, like discharge summaries and clinical reports, facilitate medical and epidemiological study by providing statistically relevant data for analysis. The findings observed and reported for elements of a disease concept are of key importance in control and prevention of such diseases.[28] Although rule-based solutions, such as this work, do not necessarily require extensive training sets (like statistical methods); testing against a larger, diverse corpus of data can elucidate unusual presentations of diabetes, ABCS protocol compliance attributes and high risk factor indicators that are not accurately handled by the current classifiers. Future work includes testing against a broader set of data, as they become available.

Future work should also explore employing this approach on higher fidelity (*e.g.,* physician's notes) and more voluminous data sets to better assess and refine the current methodology.

Acknowledgment

The authors thank Wendy Chapman for her assistance with ConText. Deidentified clinical records used in this research were provided by the i2b2 National Center for Biomedical Computing funded by U54LM008748 and were originally prepared for the Shared Tasks for Challenges in NLP for Clinical Data organized by Dr. Özlem Uzuner, i2b2 and SUNY Authors also thank Danielle Kahn in assisting us with the literature search on use of NLP for disease screening and classification.

Reference

1. Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2011.
2. Dall T, Mann SE, Zhang Y, Martin J, Chen Y, Hogan P. Economic costs of diabetes in the US in 2007. *Diabetes Care* 2008;31(3):1-20.
3. Saydah SH, Geiss LS, Tierney E, Benjamin SM, Engelgau M, Brancati F. Review of the performance of methods to identify diabetes cases among vital statistics, administrative, and survey data. *Annals of epidemiology* 2004;14(7):507-16.
4. Turchin A, Kohane IS, Pendergrass ML. Identification of patients with diabetes from the text of physician notes in the electronic medical record. *Diabetes Care* 2005;28(7):1794-95.
5. Turchin A, Pendergrass ML, Kohane IS. DITTO - a tool for identification of patient cohorts from the text of physician notes in the electronic medical record. *AMIA Annual Symposium*. Washington, DC, 2005:744-8.
6. Solti I, Gennari JH, Payne T, Solti M, Tarczy-Hornoch P. Natural language processing of clinical trial announcements: exploratory-study of building an automated screening application. *AMIA Annual Symposium*. Washington, DC, 2008:1142.
7. Li L, Chase HS, Patel CO, Friedman C, Weng C. Comparing ICD9-encoded diagnoses and NLP-processed discharge summaries for clinical trials pre-screening: a case study. *AMIA Annual Symposium*. Washington, DC, 2008:404-08.
8. Cheung BMY, Ong KL, Cherny SS, Sham PC, Tso AWK, Lam KSL. Diabetes prevalence and therapeutic target achievement in the United States, 1999 to 2006. *American Journal of Medicine* 2009;122:443-53.
9. American Diabetes Association. Standards of Medical Care in Diabetes—2011. *Diabetes Care* 2011;34:S11-S61.
10. American Diabetes Association. Standards of Medical Care in Diabetes—2009. *Diabetes Care* 2009;32:S13-S61.
11. Cohen AM, Hersh WR. A survey of current work in biomedical text mining. *Briefings in Bioinformatics* 2005;6(1):57-71.
12. Chapman WW, Chu D, Dowling JN. ConText: An algorithm for identifying contextual features from clinical text. *BioNLP 2007: Biological, translational, and clinical language processing*. Prague, CZ, 2007:81-88.
13. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics* 2009;42(5):839-51.
14. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics* 2001;34(5):301-10.

15. Mishra NK, Cummo DM, Arnzen JJ, Bonander J. A Rule-based Approach for Identifying Obesity and Its Comorbidities in Medical Discharge Summaries. *Journal of the American Medical Informatics Association* 2009;16(4):576-79.
16. Ambert KH, Cohen AM. A system for classifying disease comorbidity status from medical discharge summaries using automated hotspot and negated concept detection. *Journal of the American Medical Informatics Association* 2009;16(4):590-95.
17. Savova GK, Ogren PV, Duffy PH, Buntrock JD, Chute CG. Mayo Clinic NLP system for patient smoking status identification. *Journal of the American Medical Informatics Association* 2008;15(1):25-28.
18. Hamon T, Grabar N. Linguistic approach for identification of medication names and related information in clinical narratives. *Journal of the American Medical Informatics Association* 2010;17(5):549-54
19. Nguyen AN, Lawley MJ, Hansen DP, Bowman RV, Clarke BE, Duhig EE, Colquist S. Symbolic rule-based classification of lung cancer stages from free-text pathology reports. *Journal of the American Medical Informatics Association* 2010;17(4):440-45.
20. Uzuner Ö. Second i2b2 workshop on natural language processing challenges for clinical records. *AMIA Annual Symposium*. Washington, DC, 2008:1252-3.
21. Lappin S, Leass HJ. An algorithm for pronominal anaphora resolution. *Computational Linguistics* 1994;20(4):535-61.
22. Ge N, Hale J, Charniak E. A statistical approach to anaphora resolution. *Sixth Workshop on Very Large Corpora*. Montreal, Quebec, CA, 1998:161-70.
23. Son RY, Taira RK, Kangarloo H. Inter-document coreference resolution of abnormal findings in radiology documents. *Medinfo 2004: Proceedings Of The 11th World Congress On Medical Informatics*. San Francisco, CA, 2004.
24. Son RY, Taira RK, Kangarloo H, Cardenas AF. Context-sensitive correlation of implicitly related data: an episode creation methodology. *IEEE Transactions on Information Technology in Biomedicine* 2008;12(5):549-60.
25. Turchin A, Kolatkar NS, Grant RW, Makhni EC, Pendergrass ML, Einbinder JS. Using Regular Expressions to Abstract Blood Pressure and Treatment Intensification Information from the Text of Physician Notes. *Journal of the American Medical Informatics Association* 2006;13(6):691-95.
26. Uzuner Ö, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. *Journal of the American Medical Informatics Association* 2007;14(5):550-63.
27. Yang Y. A study of thresholding strategies for text categorization. *The 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New Orleans, LA, 2001:137-45.
28. Solt I, Tikk D, Gál V, Kardkovács ZT. Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier. *Journal of the American Medical Informatics Association* 2009;16(4):580-84.
29. Wilke RA, Berg RL, Peissig P, Kitchner T, Sijercic B, McCarty CA, McCarty DJ. Use of an Electronic Medical Record for the Identification of Research Subjects with Diabetes Mellitus. *Clinical Medicine & Research* 2007;5(1):1-7.

30. Voorham J, Denig P. Computerized extraction of information on the quality of diabetes care from free text in electronic patient records of general practitioners. *Journal of the American Medical Informatics Association* 2007;14(3):349-54.

31. Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc* 2000;7(6):593-604.

32. Chapman W, Bridewell W, Hanbury P, Cooper G, Buchanan B. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics* 2001;34(5):301-10.

33. Chapman WW, Cooper GF, Hanbury P, Chapman BE, Harrison LH, Wagner MM. Creating a text classifier to detect radiology reports describing mediastinal findings associated with inhalational anthrax and other disorders. *J Am Med Inform Assoc* 2003;10(5):494-503.

34. Wilcox AB, Hripcsak G. The role of domain knowledge in automating medical text report classification. *Journal of the American Medical Informatics Association* 2003;10(4):330-38.

35. Chapman WW, Christensen LM, Wagner MM, Haug PJ, Ivanov O, Dowling JN, Olszewski RT. Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artif Intell Med* 2005;33(1):31-40.

36. Zeng Q, Goryachev S, Weiss S, Sordo M, Murphy S, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC medical informatics and decision making* 2006;6(1):30.

37. Friedlin J, McDonald CJ. Using a natural language processing system to extract and code family history data from admission reports. *AMIA Annual Symposium*. Washington, DC, 2006:925.

38. Clark C, Good K, Jezierny L, Macpherson M, Wilson B, Chajewska U. Identifying smokers with a medical extraction system. *Journal of the American Medical Informatics Association* 2008;15(1):36-39.

39. Childs LC, Enelow R, Simonsen L, Heintzelman NH, Kowalski KM, Taylor RJ. Description of a rule-based system for the i2b2 challenge in natural language processing for clinical data. *Journal of the American Medical Informatics Association* 2009;16(4):571-75.

40. Farkas R, Szarvas G, Hegedűs I, Almási A, Vincze V, Ormándi R, Busa-Fekete R. Semi-automated construction of decision rules to predict morbidities from clinical texts. *Journal of the American Medical Informatics Association* 2009;16(4):601.

41. Yang H, Spasic I, Keane JA, Nenadic G. A text mining approach to the prediction of disease status from clinical discharge summaries. *Journal of the American Medical Informatics Association* 2009;16(4):596-600.

42. Ware H, Mullett CJ, Jagannathan V. Natural language processing framework to assess clinical conditions. *Journal of the American Medical Informatics Association* 2009;16(4):585-89.

43. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, Dittus RS, Rosen AK, Elkin PL, Brown SH. Automated Identification of Postoperative Complications Within an

Electronic Medical Record Using Natural Language Processing. *JAMA: The Journal of the American Medical Association* 2011;306(8):848-55.

Tables

TITLE/PRINCIPAL INVESTIGATOR/AUTHOR	PUBLICATION YEAR	DESCRIPTION	APPLIED TO DISEASE CONDITION /DESCRIPTION	DATA SOURCE	CLASSIFICATION OF APPROACH	DETECT ABCS COMPLAINCE	DETECT HIGH RISK FACTOR
Towards Automatic Diabetes Case Detection and ABCS Protocol Compliance Assessment **Mishra, N. *(this work)***	**2012**	**Rule-based classifier incorporating the clinical natural language processing tool, ConText, to analyze discharge summaries for detecting diabetes, ABCS protocol compliance and high risk factors; extends upon published work originally applied to obesity.**	**Diabetes-screening**	**Medical discharge summaries**	**Rule-Based**	**yes**	**yes**
Identification of Patients With Diabetes From the Text of Physician Notes in the Electronic Medical Record Turchin, A.[4]	2005	DITTO-Diabetes Identification Through Textual element Occurrences Program written in Perl language that takes one or more text files containing patient notes as input	Diabetes-screening	Electronic medical record	Rule-Based	no	no
Use of an Electronic Medical Record for the Identification of Research	2007	Electronic case-finding algorithm that can be used for the identification of research subjects with DM in the PMRP DNA	Diabetes-screening	Electronic medical record	Rule-Based	no	no

Subjects with Diabetes Mellitus Wilke, R.[29]		biobank Combined medical record: including indices to all events and encounters with access all textual documentation such as office notes, operative reports and discharge summaries					
Computerized Extraction of Information on the Quality of Diabetes Care from Free Text in Electronic Patient Records ofGeneral Practitioners Voorham, J.[30]	2007	Evaluation of a computerized method for extracting numeric clinical measurements related to diabetes care from free text in electronic patient records converted numeric clinical information to structured data research and quality of care assessments for practices lacking structured data entry; 13 numeric clinical measurements considered relevant for evaluating the quality of diabetes care	Diabetes care	Electronic medical record	Rule-Based	no	yes
Automatic Detection of Acute Bacterial Pneumonia from Chest X-ray Reports Fiszman, M.[31]	2000	-Four physicians, three lay persons, a natural language processing system, and two keyword searches (designated AAKS and KS) detected the presence or absence of three pneumonia-related concepts and inferred the presence or absence of acute bacterial pneumonia from 292 chest x-ray reports.	pneumonia	chest x-ray reports	Statistical		

A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries Chapman, W.[32]	2002	NegEx – describes and tests a computationally simple algorithm that could be implemented quickly and easily to determine whether an indexed term is negated	Various diseases	Medical discharge summaries	Rule-Based		
Creating a Text Classifier to Detect Radiology Reports Describing Mediastinal Findings Associated with Inhalational Anthrax and Other Disorders Chapman, W.[33]	2003	Identify Patient Sets (IPS) system to create a key word classifier for detecting reports describing mediastinal findings consistent with anthrax and compared their performances on a test set of 79,032 chest radiograph reports	covert anthrax release	chest radiograph reports	Statistical		
The Role of Domain Knowledge in Automating Medical Text Report Classification Wilcox, A.[34]	2003	Inductively created classifiers for medical text reports using varying degrees and types of expert knowledge and different inductive learning algorithms. Six clinical conditions: congestive heart	Multiple Clinical Conditions	chest radiograph reports and discharge summaries	Both		

		failure, chronic obstructive pulmonary disease, acute bacterial pneumonia, neoplasm, pleural effusion, and pneumothorax					
Classifying free-text triage chief complaints into syndromic categories with natural language processing Chapman, W.[35]	2005	Natural language processing application for classifying chief complaints into syndromic categories for syndromic surveillance	Syndromic surveillance	Chief complaints	Statistical		
Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system Zeng, Q.[36]	2006	Health Information Text Extraction (HITEx) tool used to extract key findings for a research study on airways disease The principal diagnosis, co-morbidity and smoking status extracted by HITEx from a set of 150 discharge summaries were compared to an expert-generated gold standard.	asthma	discharge summaries	Rule-Based		

Using A Natural Language Processing System to Extract and Code Family History Data from Admission Reports Friedlin, J.[37]	2006	Extracts and code family history data from hospital admission notes	Family history of 12 diseases	Hospital admission notes	Rule-Based		
Identifying Smokers with a Medical Extraction System Clark, C.[38]	2007	Rule-based extraction engine with machine learning algorithms to identify and categorize references to patient smoking in clinical reports	Smoking-screening	Electronic medical record	Both		
Mayo Clinic NLP System for Patient Smoking status Identification Savova, G.[17]	2008	Identifying the smoking status of patients; system of reusable text analysis components built on the Unstructured Information Management Architecture and Weka; system has three layers of sentence classification	Smoking status	Electronic medical record	Rule-Based		
Comparing ICD9-Encoded Diagnoses and NLP-Processed Discharge Summaries for Clinical Trials Pre-Screening: A Case Study Li, L.[7]	2008	Strengths and weaknesses of using structured or textual clinical data for clinical trials eligibility pre-screening Medical Language Extraction and Encoding System (MedLEE)6 - processes narrative discharge summaries, MedLEE generates structured, computer-	Screening for clinical trials	Clinical Data Warehouse (CDW)	Rule-Based		

		recognizable, XML-coded concepts using semantic classes for problem, finding, body location, certainty, degree, region, etc.					
A Rule-based Approach for Identifying Obesity and Its Comorbidities in Medical Discharge Summaries Mishra, N.[15]	2009	Free-text classifier based on keyword identification, negation detection, and simple scoring rules.	Obesity and comorbiditie s-screening	Medical discharge summaries	Rule-Based		
A System for Classifying Disease Comorbidity Status from Medical Discharge Summaries Using Automated Hotspot and Negated Concept Detection Ambert, K.[16]	2009	A text mining system for classifying patient comorbidity status, based on the information contained in clinical reports No aspect of their system requires a-priori knowledge of a disease, expert medical knowledge, or manual examination of the patient records beyond the training labels themselves.	Obesity and comorbiditie s-screening	Medical discharge summaries	Statistical		
ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports	2009	Algorithm for determining whether clinical conditions mentioned in clinical reports are negated, hypothetical, historical, or experienced by someone other than the patient	Various diseases-	Electronic medical record- six report types-	Rule-Based		

Harkema, H.[13]							
Description of a Rule-based System for the i2b2 Challenge in Natural Language Processing for Clinical Data Childs, L.[39]	2009	Lockheed Martin's rule-based natural language processing (NLP) capability, ClinREAD; "read" a patient's clinical discharge summary and replicate the judgments of physicians in evaluating presence or absence of obesity and 15 comorbidities	Obesity and 15 comorbidities	clinical discharge summary	Rule-Based		
Semi-automated Construction of Decision Rules to Predict Morbidities from Clinical Texts Farkas, R.[40]	2009	Development of automatic systems that analyzed clinical discharge summary texts and addressed the following question: "Who's obese and what co-morbidities do they (definitely/most likely) have?"	Obesity and 15 comorbidities	clinical discharge summary	Both		
A Text Mining Approach to the Prediction of Disease Status from Clinical Discharge Summaries Yang, H.[41]	2009	Assembled a set of resources to lexically and semantically profile the diseases and their associated symptoms, treatments, etc. These features were explored in a hybrid text mining approach, which combined dictionary look-up, rule-based, and machine-learning methods.	Obesity and 15 comorbidities	clinical discharge summary	Both		
Semantic Classification	2009	Context-aware rule-based semantic	Obesity and	clinical discharge	Rule-Based		

of Diseases in Discharge Summaries Using a Context-aware Rule-based Classifier Solt, I.[28]		classification technique for use on clinical discharge summaries	15 comorbidities	summary			
Natural Language Processing Framework to Assess Clinical Conditions Ware, H.[42]	2009	Natural language processing (NLP) framework that could be used to extract clinical findings and diagnoses from dictated physician documentation; authors used a combination of concept detection, context validation, and the application of a variety of rules to conclude patient diagnoses.	Obesity and 15 comorbidities	clinical discharge summary	Both		
Symbolic rule-based classification of lung cancer stages from free-text pathology reports Nguyen, A.[19]	2010	Classify automatically lung tumorenodeemetastases (TNM) cancer stages from free-text pathology reports using symbolic rule-based classification.	Lung cancer-classification of stages	free-text pathology reports	Rule-Based		
Automated Identification of Postoperative Complications Within an Electronic Medical Record Using Natural Language Processing	2012	Classifier focused on the 6 postoperative complications (acute renal failure requiring dialysis, sepsis, deep vein thrombosis, pulmonary embolism, myocardial infarction, and pneumonia; used Multi-threaded Clinical Vocabulary	postoperative surgical complications	Electronic medical record- surgical inpatient admissions	Rule-Based		

Murff, H.[43]		Server natural language processor system					

Table I: Review of Related Works

Dictionary Name	Abbreviation	Associated Classification Task	Example Concept Term(s)	# of Terms
Diabetes - General	gen	Indication	diabetes, dm	21
Diabetes – Medication	med	Indication	insulin, lispro, glucophage	75
Diabetes – Examination	exam	**A**BCS (Protocol Only)	A1C	7
High Blood Pressure	hbp	A**B**CS (Protocol + Risk)	hypertensive, hypertension	3
High Cholesterol	hc	AB**C**S (Protocol + Risk)	hypercholesterolemia	3
Blood Pressure Value Assessment	bp	A**B**CS (Protocol + Risk)	bp, blood pressure	4
LDL Assessment	ldl	AB**C**S (Protocol + Risk)	ldl	5

Table II: Dictionaries Utilized with ConText for Concept Extraction

Indic. Class	Metrics	Discourse Only	Med. Only	Combined
+	π_+	0.9455	0.9783	0.9310
	$(TP_+ + FP_+)$	(52+3)	(45+1)	(54+4)
	ρ_+	0.9286	0.8036	0.9643
	$(TP_+ + FN_+)$	(52+4)	(45+11)	(54+2)
	F_+	0.9369	0.8824	0.9474
–	π_-	0.8333	N/A	0.8333
	$(TP_- + FP_-)$	(5+1)	(0+0)	(5+1)
	ρ_-	1.0000	0.000	1.000
	$(TP_- + FN_-)$	(5+0)	(0+5)	(5+0)
	F_-	0.9091	0.0000	0.9091
?	$\pi_?$	0.9922	0.9598	0.9974
	$(TP_? + FP_?)$	(380+3)	(382+16)	(379+1)
	$\rho_?$	0.9922	0.9974	0.9896
	$(TP_? + FN_?)$	(380+3)	(382+1)	(379+4)
	$F_?$	0.9922	0.9782	0.9934
	F_M	0.9461	0.6202	0.9500
	F_μ	0.9842	0.9617	0.9865

Table III: Precision[π], Recall[ρ], F-measure[F], aggregate macro F-measure[F_M] and aggregate micro F-measure [F_μ] for Diabetes Indication Classifiers Based on Three-Classes (Positive[+], Negative[–], Unknown[?])

Indic. Class	Metrics	Discourse Only	Med. Only	Combined
+	π_+	0.9455	0.9783	0.9310
	$(TP_+ + FP_+)$	(52+3)	(45+1)	(54+4)
	ρ_+	0.9286	0.8036	0.9643
	$(TP_+ + FN_+)$	(52+4)	(45+11)	(54+2)
	F_+	0.9369	0.8824	0.9474
!	$\pi_!$	0.9897	0.9724	0.9948
	$(TP_! + FP_!)$	(385+4)	(387+11)	(384+2)
	$\rho_!$	0.9923	0.9974	0.9897
	$(TP_! + FN_!)$	(385+3)	(387+1)	(384+4)
	$F_!$	0.9910	0.9847	0.9922
	F_M	0.9640	0.9335	0.9698
	F_μ	0.9842	0.9730	0.9865

Table IV: Precision[π], Recall[ρ], F-measure[F], aggregate macro F-measure[F_M] and aggregate micro F-measure [F_μ] for Diabetes Indication Classifiers Based on Two-Classes (Positive[+] and Combined Negative/Unknown[!])

Indic. Class	Metrics	A1C Mentioned	BP Mentioned	Cholesterol Mentioned
+	π_+	1.0000	0.9535	1.0000
	$(TP_+ + FP_+)$	(5+0)	(41+2)	(9+0)
	ρ_+	1.0000	0.9318	1.0000
	$(TP_+ + FN_+)$	(5+0)	(41+3)	(9+0)
	F_+	1.0000	0.9425	1.0000
-	π_-	1.0000	0.7692	1.0000
	$(TP_- + FP_-)$	(51+0)	(10+3)	(47+0)
	ρ_-	1.0000	0.8333	1.0000
	$(TP_- + FN_-)$	(51+0)	(10+2)	(47+0)
	F_-	1.0000	0.8000	1.0000
	F_M	1.0000	0.8713	1.0000
	F_μ	1.0000	0.9107	1.0000

Table V: Precision[π], Recall[ρ], F-measure[F] , aggregate macro F-measure[F_M] and aggregate micro F-measure [F_μ] for ABCS Protocol Compliance Assessment Classifiers Based on (Satisfied [+] and Not Satisfied [-])

Indic. Class	Metrics	BP Risk	Cholesterol Risk
+	π_+	0.9429	0.7500
	$(TP_+ + FP_+)$	(33+2)	(3+1)
	ρ_+	0.9167	0.6000
	$(TP_+ + FN_+)$	(33+3)	(3+2)
	F_+	0.9296	0.6667
-	π_-	0.8571	0.9615
	$(TP_- + FP_-)$	(18+3)	(50+2)
	ρ_-	0.9000	0.9804
	$(TP_- + FN_-)$	(18+2)	(50+1)
	F_-	0.8780	0.9709
	F_M	0.9038	0.8188
	F_μ	0.9107	0.9464

Table VI: Precision[π], Recall[ρ], F-measure[F] , aggregate macro F-measure[F_M] and aggregate micro F-measure [F_μ] for High Risk Factor Classifiers (Indicated [+] and Not Indicated [-])

Figure Legends

Figure 1 Diabetes Indication Classifier Rule

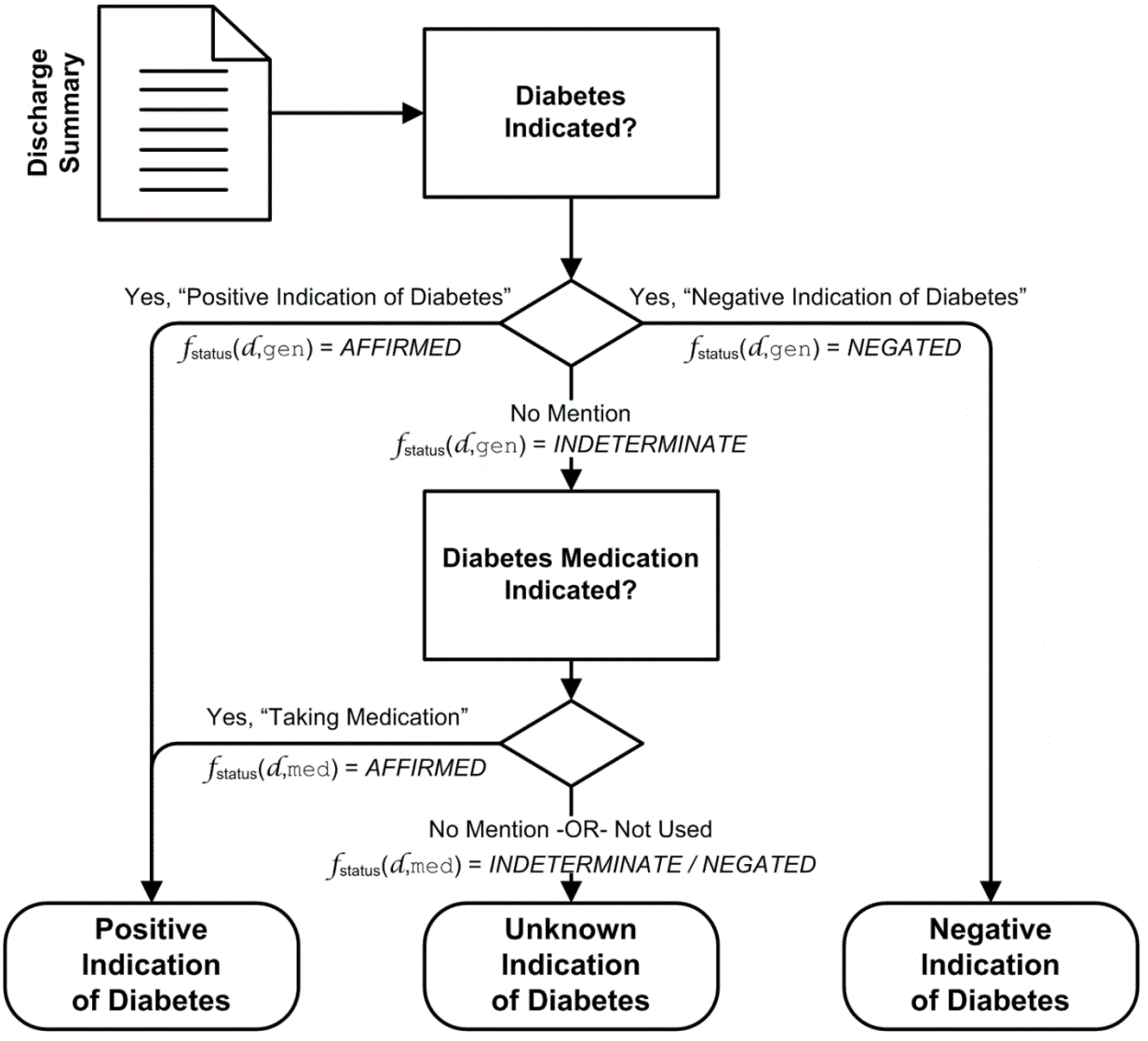

Figure 2 ABCS Protocol Compliance Assessment Strategy for Each Indicator

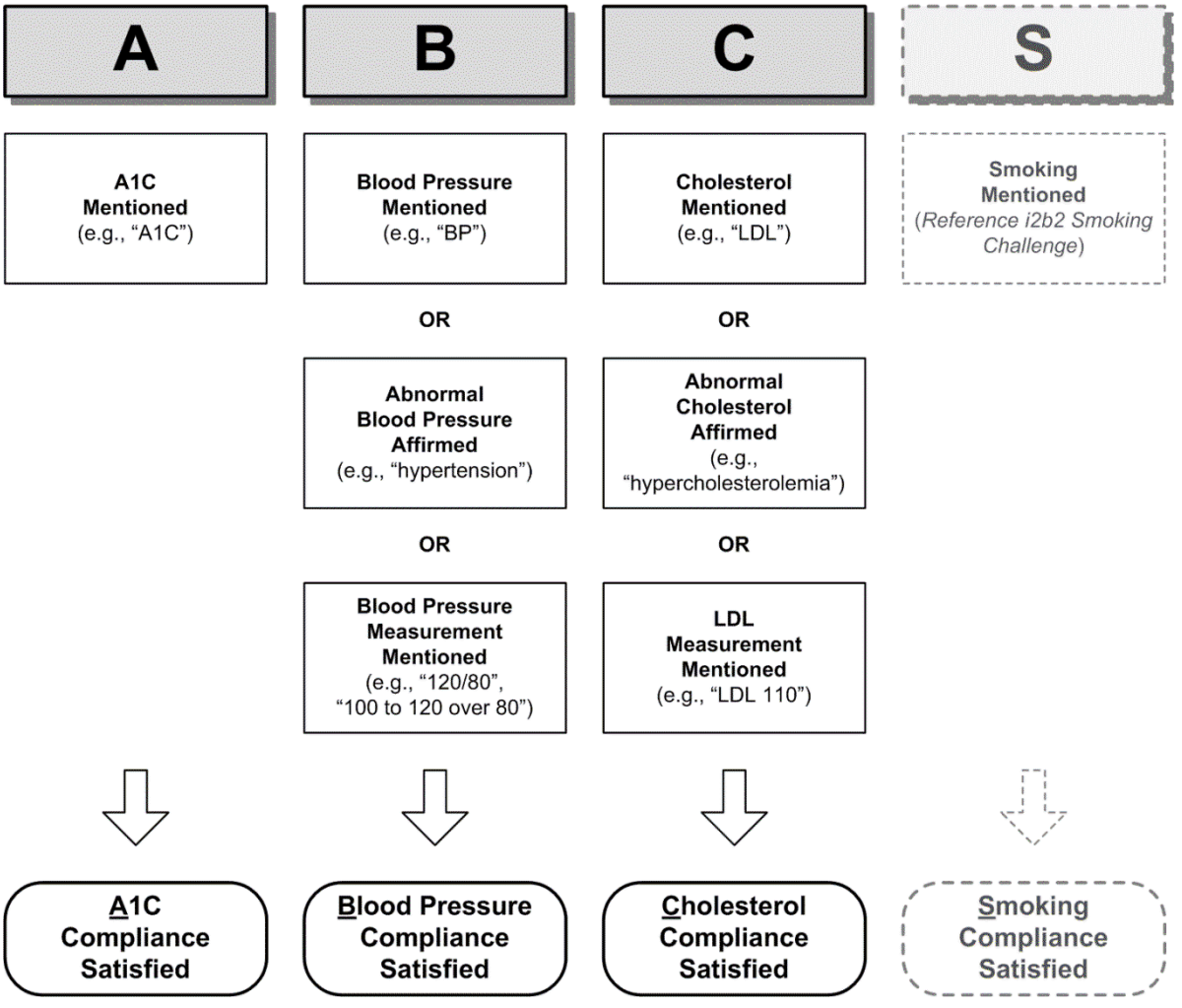

Figure 3 High Risk Factors Assessment Strategy for Each Indicator

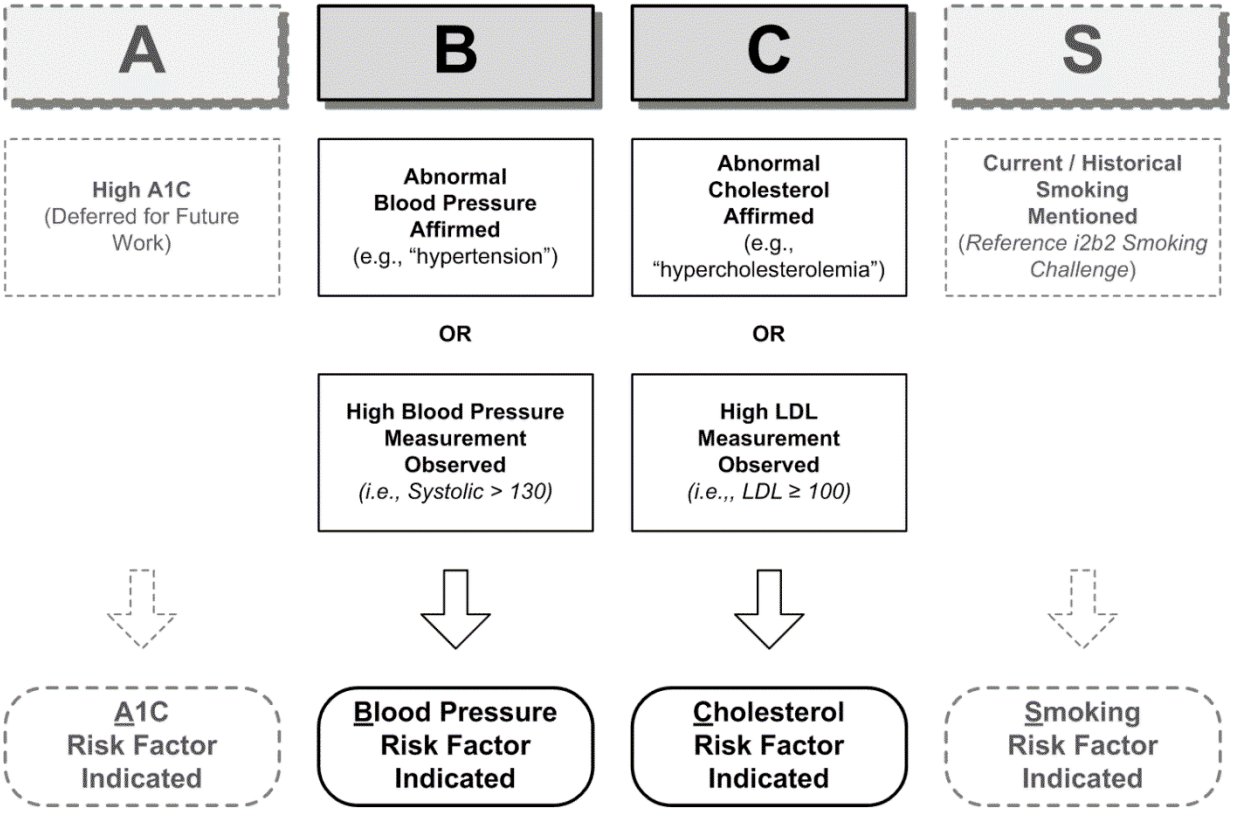
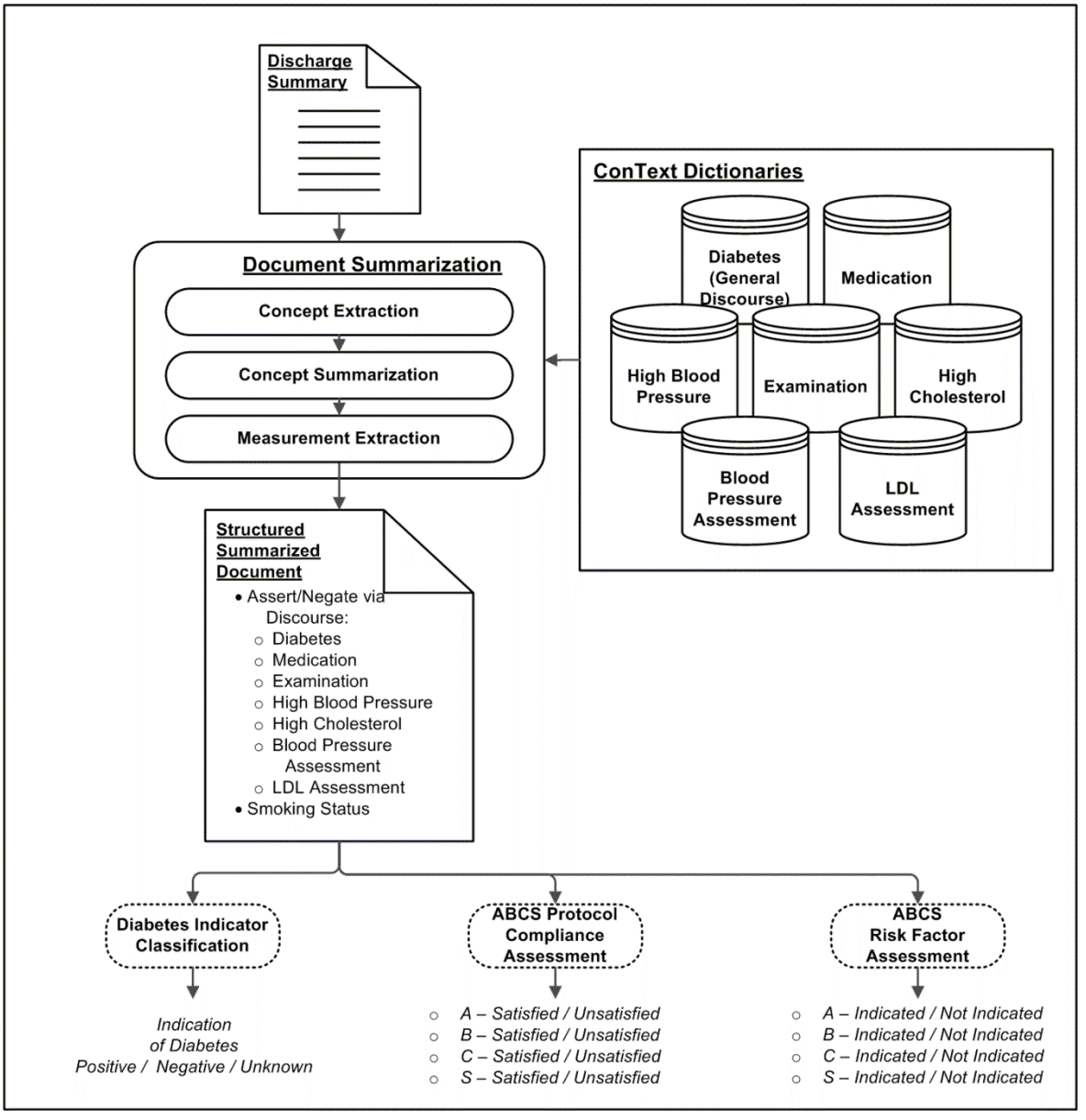
Figure 4 Overview of Discharge Summary Analysis for Diabetes