Automated Clinical Documentation Improvement

1st Vatsal Shah
Research Lab
ezDI Inc.
Louisville, USA
vatsal.s@ezdi.com

2nd Raxit Goswami
Research Lab
ezDI Inc.
Louisville, USA
raxit.g@ezdi.com

3rd Vivek Kumar
Research Lab
ezDI Inc.
Louisville, USA
vivek.k@ezdi.com

4th Dr. Binni Shah
Research Lab
ezDI Inc.
Louisville, USA
binni.s@ezdi.com

5th Helly Shah
Research Lab
ezDI Inc.
Louisville, USA
helly.s@ezdi.com

***Abstract*—Complete and accurate clinical documentation in the medical record has a direct impact on the assignment of codes, more accurate levels of reimbursement, and is critical to the higher quality of patient care. This paper describes the development of a system which can automatically flag the cases if there is an opportunity of improvement in patient clinical documents. Automated Clinical Documentation Improvement (CDI) leverages the natural language processing (NLP) and contextual understanding of health record structure with additional business rules logic, helping CDI specialists identify critical documentation information that may be missing from the medical record. This results in more specific coding opportunity and better understanding of the clinical complexity for accurate reimbursement. This system helped increase CDI specialists' productivity by efficiently filtering cases which need more attention from them.**

***Index Terms*—Clinical documentation improvement, CDI, Automated CDI, CA-CDI, Computer Assisted Coding, Healthcare Knowledge Representation, CAC, NLP, Clinical Decision Support.**

I. Introduction

Clinical documentation is maintained to monitor a patient's health and share the physician's actions and thoughts to other members of the care team. Hence it is critical to the patient, the physician, and the health-care organizations [1]. Clinical Documentation Improvement (CDI) is a process of analyzing the documents to find out such instances where more specific documentation may help. According to the Association of Clinical Documentation Improvement Specialists (ACDIS), the main objective of CDI is to review the medical record for increasing the accuracy, clarity, and specificity of provider documentation. It is used to initiate concurrent and retrospective reviews of health records for conflicting, incomplete, or non-specific provider documentation [2]. Automatic or Computer-assisted Clinical Documentation Improvement (CA-CDI) system can be developed for helping CDI specialists by automatically suggesting cases that need to be reviewed for improvement. This can help in better coding, better patient acuity scores and increased Case Mix Index (CMI).

Following is the flow of the paper. Section II describes the current CDI process and motivations for developing this system. Section III describes the approach taken to develop the system, Section IV analyses the results and performance of the system and we conclude in Section V.

II. Motivation And Opportunities

A. Background

AHIMA defines clinical documentation as any manual or electronic notation made by a clinical care provider. It's also noted that rapidly changing healthcare environment and the variety of uses and users of clinical documentation has led to the ever-increasing importance of CDI program implementation [3]. In October 2007 Centers for Medicare and Medicaid Services (CMS) implemented Medicare Severity Diagnosis Related Groups (MS-DRGs) for hospital inpatient prospective payment in order to better reflect the patient's severity of illness and expected risk of mortality. The principal diagnosis assigned to a patient case and other comorbid conditions determine these assessments. Later in October 2008, CMS made the presence of a Present on Admission (POA) indicator mandatory for all coded diagnoses in order to distinguish the conditions that are present when a patient is admitted from those that are acquired once in the hospital. Thus, the need for complete and accurate documentation has taken a more important role. Acute care hospitals have now become more dependent on clinical documentation in order to comply with the regulations regarding quality and reimbursement. The health record has become an essential legal document with requirements for non-modification and retention to define the work product for which physicians were paid and drive the education of medical students and trainees. The Meaningful Use program has made the medical documentation to include specific information which can be used for payer quality measures and health information sharing with patients, families, and caregivers [4]. Fig. 1 shows the major drivers for CDI.

Payers now require more documentation for pre authorizations and payments. A number of private and public entities have started requiring the patient documentation for tracking quality, public health initiatives, and research. As such, CDI programs are important to any facility that recognizes the necessity of complete and accurate patient documentation. They have become an essential part in health-care organizations for clinical data quality, compliance and revenue improvement.

B. Limitations of existing CDI programs

Most CDI programs implemented in healthcare organizations are manual. According to the results of the 2017 ACDIS CDI Week Survey, 48.42% of the respondents said that they

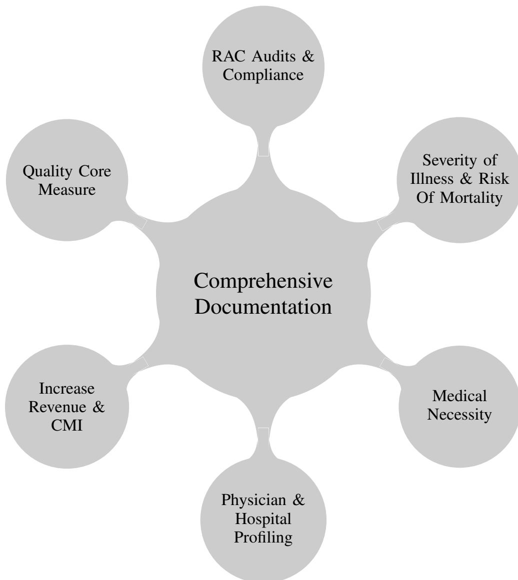

Fig. 1. Clinical Documentation Impact

dont use computer assisted technologies, CAC or NLP, (or are planning to use them in future) to support with record reviews [5]. They depend on the CDI specialists checking each and every document in the medical record of a single patient to identify the areas in which the document is not specific. This is a slow process and most of the time is wasted on checking the documents which may not require any improvement. Most often, CDI specialists only work on a small part of the patient population, based on the severity of the case or predefined service-lines only, which lead to reduced coverage. In many cases, the queries are identified and submitted after the patient has left the organization and the physician may not have fresh memory about those.

C. Need for automation

10th revision of International Classification of Diseases (ICD10) from WHO is now being used in most countries (with or without some modifications) which contains over 16000 codes. This codeset has greater specificity and requires clinical detail to be provided for better clinical decision-making, outcomes and research [6]. Moreover, the United States uses a modified version of this codeset, called ICD10-CM (Clinical Modification) which contains about 68,000 codes and a separate system for procedures, ICD10-PCS (Procedure Coding System) containing almost 76,000 codes [7]. To incorporate these changes, clinical documentation is undergoing significant changes, especifically in the areas of specificity of diagnoses, clarification of anatomical site and laterality, and other complications and manifestations. CDI specialists now have to deal with increasing laterality and specificity of diagnoses, combination of codes and pathophysiology of diseases. Automation of CDI can leverage the expert guidelines and NLP with contextual understanding and apply additional business rules logic, helping CDI specialists identify information that may be missing from the medical record, and that could result in additional coding opportunity or understanding of the clinical complexity for accurate reimbursement. An automated system for CDI can process medical records faster and gather relevant data from entire medical record to mark out the areas which need more concentration from the CDI specialist. Due to its speed, it can be used to query the medical practitioner while the patient is still in hospital undergoing treatment.

III. Proposed Approach

A. Model

Our automated CDI system as depicted in Fig. 2 has a model which depends on the following components:

1) **Clinical marker** - a specific fact or entity detected by the NLP (like a diagnosis, or lab measurement)
2) **Marker group** - Group of clinical markers which collectively help recognizing the presence of diagnosis or procedure.
3) **Exclusion group** - Clinical markers of accurately documented specificity related to a diagnosis which when present, the query will not be suggested.

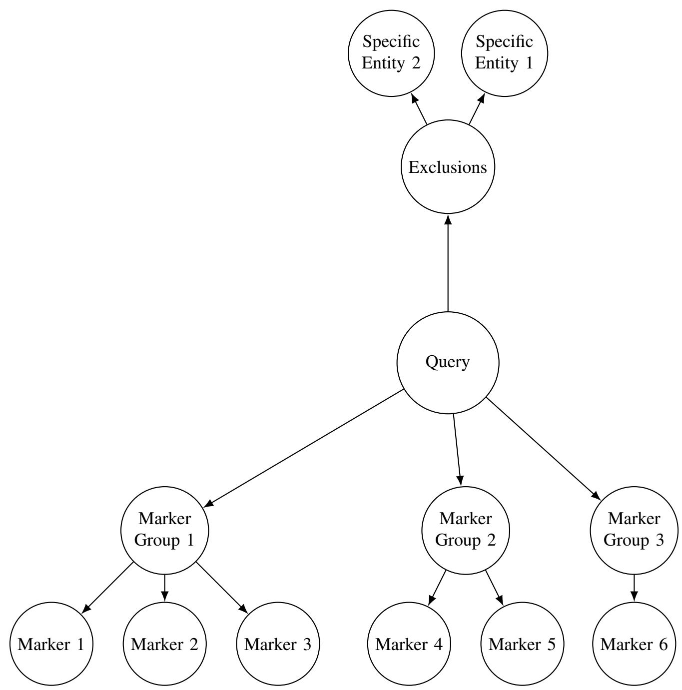

Fig. 2. CDI Query Knowledge Representation

B. Knowledge creation

Rules for automated CDI contain a combination of marker groups which when present (or absent) in the medical record, makes the basis for query suggestion. Then knowledge base is created for them with consultation of CDI experts. For generating this knowledge, important diagnoses like *Heart Failure, Pneumonia, Sepsis, Encephalopathy* etc. were analysed which are amongst the most queried diagnoses in healthcare and also affect the DRGs of the case.

Sample knowledge for Heart Failure Query is shown in Table I.

TABLE I
HEART FAILURE QUERY KNOWLEDGE

Clinical Markers		
Type	***Values***	**Description**
Diagnosis	Congestive Heart Failure (CHF)	Unspecified Heart Failure is documented
Measurements	Ejection Fraction >60 BNP >100	Measurements suggestive of diastolic heart failure
Medications	Lasix Metoprolol	For blood pressure control and fluid retention
Exclusions	Acute Diastolic CHF Chronic Diastolic CHF	Specificity for CHF found in health record

When a medical record is processed in which the markers given in the Table I are encountered, it signifies a gap in the documentation because the diagnosis of CHF in this case can be more specifically recorded. This becomes an opportunity for CDI. The documentation of CHF or Heart Failure is a prerequisite of this query suggestion. If all other markers are present but CHF is not documented, then this query will not be suggested.

C. Architecture

This CDI process starts when the first document of patient's medical record is submitted. The NLP module continues to analyze the components of the patient's clinical documentation as more data is added to the case. If the found clinical markers matches from the knowledge of a particular query, that query is suggested. If the required specificity for an earlier query is found in any later document, then that suggestion is removed. Fig. 3 displays the architecture of the current system.

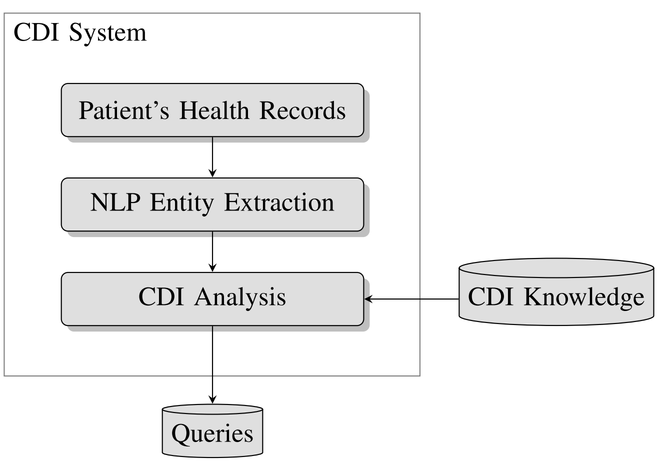

Fig. 3. CDI System Architecture

D. Detailed Process

The following steps are performed in the automatic CDI process:

1) **Entity Extraction** - Natural Language Processing (NLP) is used to extract structured information from unstructured text data of patient documentation using the system in [8]. The extracted entities comprise of the medical concepts (problem, procedure, medical device, anatomical structure, lab value, test etc.) along with their time status (if it is related to patients history or present condition), the negation value (concept is present or not) and associated measurement (if its a lab value, body measurements or test).
2) **Filtering** - This structured data is filtered based on predefined sections or type of the document. For example, it can be decided whether to consider Chest Imaging report to extract clinical markers for a specific query.
3) **Exclusions** - The entities are matched against the exclusions for each query to check for the presence of required level of documentation. If the specificity is present, the query is removed from possible suggestions. The medical entities are also checked to see if they already contain the required specificity. For example, if the medical record contains all the documentation to support Heart Failure query (as given in Table I), but also contains the diagnosis of Acute diastolic heart failure, then the Heart Failure query will not be suggested.
4) **Find probable query opportunities** - The CDI Module aggregates the entities into clinical markers in the Marker Groups which are checked against the rules in the knowledge to detect if they are part of any query suggestion. All such queries are marked as having a possibility of suggestion.

Fig. 4 displays the flowchart of CDI algorithm for the case of Heart Failure detection query.

IV. RESULT ANALYSIS

The system was tested with two hospitals' documents, Hospital 1 and Hospital 2. We recognized the following top queries used by these hospitals based on usage frequency:

1) Altered Mental Status / Encephalopathy
2) Heart Failure

Table II shows performance of the CDI System for these queries.

TABLE II
CDI PERFORMANCE ON TOP 2 QUERIES

Query	**Precision**	**Recall**	**F1**
Altered Mental Status	47.692	45.588	46.616
Heart Failure	52.777	67.256	59.143

Table III shows the impact of the system for the two hospitals.

TABLE III
CDI IMPACT

Hospital	**Coverage Improvement**	**CMI Increase**
Hospital 1	23.6%	6.04%
Hospital 2	27.4%	4.59%

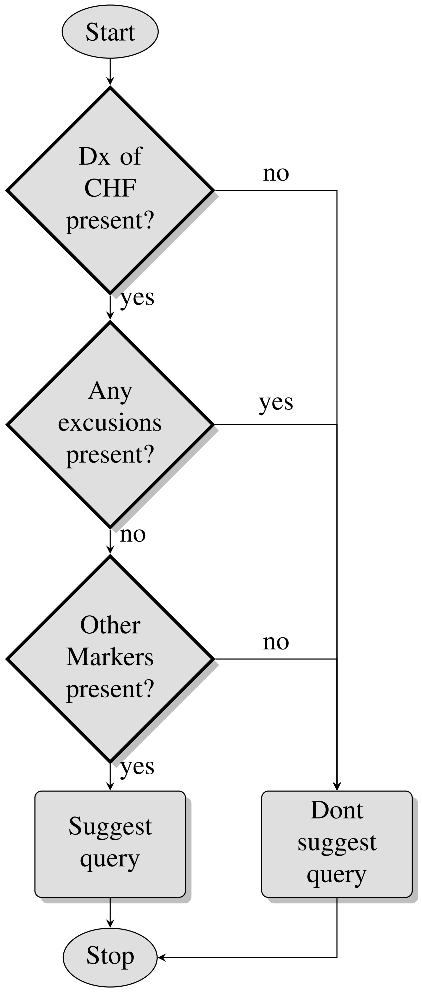

Fig. 4. CDI Process Flowchart for Heart Failure

The system is able to filter out cases in which accurate documentation level was already present and did not need any further attention from CDI specialist.

The analysis shows that system is able to suggest appropriate queries but also triggers many false positives which are then rejected (not used) by the CDI specialist. The possible reasons behind the incorrect query suggestion are:

1) **Incorrect NLP detection** - During entity extraction, NLP may have completely missed some entities or detect their temporal status, negation or relationships incorrectly. Table IV shows such examples.

TABLE IV
INCORRECT NLP DETECTION EXAMPLES

Sentence	***Correct***	***Detected***
Questionable heart failure.	Negation: Yes	Negation: No
Past Medical History: CHF, Hypertension	CHF Status: As History	CHF Status: Present
Respiratory failure, cardiac murmur	Respiratory Failure	Cardiac Failure (CHF)

2) **Complex CDI scenarios** - In some cases, the scenarios are very complex to be captured correctly and suggested. For example, in the following text:
1. Drug overdose
2. Acute encephalopathy due to 1.
Here, the fact that *encephalopathy* is associated with *drug overdose* is not determined correctly by the system and query for *unspecified encephalopathy* may be suggested.
3) **Incorrect Knowledge** - It is possible that the knowledge created for automation of CDI is insufficient or incorrect.

V. CONCLUSION

In this paper, an automated Clinical Documentation Improvement system is described. In the approach, NLP is used to find clinical markers of certain diseases and detect the completeness of the documentation. If any gaps are found between required specificity of those diagnoses and the documentation, then a query is suggested automatically. The system filters out cases which do not need any improvement and allows CDI specialists to focus on the flagged cases. This approach has resulted in decrease in time to check the cases and increased the documentation compliance.

In future, this system can be expanded to include more advanced techniques for better accuracy of query suggestion. The knowledge base of the system can be improved by leveraging the domain relationships and also depending on the user requirements for a specific query. Historical clinical data and CDI user actions on the suggested queries can be used in statistical model learning to fine tune and customize the system.

REFERENCES

[1] A. L. Towers, "Clinical documentation improvementa physician perspective: Insider tips for getting physician participation in cdi programs," *Journal of AHIMA*, vol. 84, no. 7, pp. 34–41, 2013.
[2] A. of Clinical Documentation Improvement Specialists (ACDIS), "Cdi roadmap: Phase 1, goals and objectives of cdi," www.hcpro.com/acdis/cdi_roadmap_phase1.cfm, 2014.
[3] A. H. I. M. A. (AHIMA), "The clinical documentation improvement toolkit," 2014, http://library.ahima.org/doc?oid=300359, Last accessed on 2018-08-24.
[4] K. T, B. P, B. M, Y. T, and the Medical Informatics Committee of the American College of Physicians, "Clinical documentation in the 21st century: Executive summary of a policy position paper from the american college of physicians," *Annals of Internal Medicine*, vol. 162, no. 4, pp. 301–303, 2015. [Online]. Available: +http://dx.doi.org/10.7326/M14-2128
[5] A. of Clinical Documentation Improvement Specialists (ACDIS), "Cdi week 2017: Industry overview survey," 2017, https://acdis.org/system/files/cdi-week/39056_CDI_Week_Industry_Survey_Report_2017.pdf, Last accessed on 2018-05-30.
[6] W. H. Organisation, "Internation classification of diseases, icd," 2018, http://www.who.int/classifications/icd/en/, Last accessed on 2018-05-30.
[7] C. for Medicare & Medicaid Services, "Icd-10," 2018, https://www.cms.gov/Medicare/Coding/ICD10/index.html, Last accessed on 2018-08-24.
[8] P. Pathak, P. Patel, V. Panchal, S. Soni, K. Dani, A. Patel, and N. Choudhary, "ezdi: a supervised nlp system for clinical narrative analysis," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015, pp. 412–416.