veradigm®

DEVELOPMENT AND VALIDATION OF A METHOD TO EXTRACT LEFT VENTRICULAR EJECTION FRACTION DATA FROM EHR PHYSICIAN NOTES

Oguntuga AO [1], Overcash J[2], Nguyen N[1]

Veradigm®, [1]San Francisco, CA and [2]Raleigh, NC, USA

BACKGROUND + INTRODUCTION

- Left Ventricular Ejection Fraction (LVEF) is an assessment of the pumping power of the left ventricular wall of the heart.
- In the course of conducting a retrospective congestive heart failure study, the Life Science Analytics team noted that healthcare providers using a Veradigm ambulatory Electronic Health Record (EHR) platform recorded LVEF assessments in SOAP notes (Subjective, Objective, Assessment and Plan) and not in structured data elements.
- Extracting these assessments required building a Natural Language Processing (NLP) pipeline to perform Information Extraction (IE) from SOAP notes.
- During a literature review of methods to extract LVEF assessments from clinical notes, we came across rule-based NLP pipelines. Publications showing the use of machine learning pipelines were not found.

OBJECTIVES

- To build and compare the performance of rule-based and machine learning (ML) pipelines with the goal of improving extraction accuracy.

METHODS

Understanding characteristics of LVEF assessments in SOAP notes:

Components:

- Keyword or phrase or acronym synonymous to LVEF
- A percentage score
- A date indicating when the test was performed (if present in the note)

Types of LVEF assessments based on the percentage score of the assessment:

- **Ratio Percentage Value**: For example, “Prior echocardiogram in 2011 reported normal LV ejection fraction of 69% and diastolic dysfunction.” The keyword indicating an LVEF being referenced here is LV ejection fraction, the score is 69% a single (ratio) numeric value and the date is 2011.
- **Interval percentage value:** For example, “Echocardiogram on 1/1/2010 showed low normal left ventricular ejection fraction of 50-55% with mild LV hypokinesis.” The keyword indicating an LVEF being referenced here is left ventricular ejection fraction, the score is between a numeric interval of 50-55% and the date is 1/1/2010.
- **Relative percentage value:** For example, “Ischemic cardiomyopathy with E.F >55%.” The keyword indicating an LVEF being referenced here is E.F, the score is relatively greater than 55% and there is no date mentioned about when the LVEF was performed.

Process: The IE task consists of two subtasks: Named Entity Recognition (NER) and Relationship Extraction (RE). The NER subtask (rule-based in Figure 1 and machine learning based in Figure 2) identifies textual entities that indicate a provider is referring to an LVEF assessment within a note. The RE subtask then focuses on associating related individual entities that belong to the same assessment. The rule-base pipeline (Figure 3) uses Regular Expressions (RegEx) for both NER and RE, while the machine learning (ML) pipeline (Figure 4) uses a Conditional Random Field (CRF) model for NER and RegEx for RE.

Figure 1 | Steps in Rule-Based NER Module

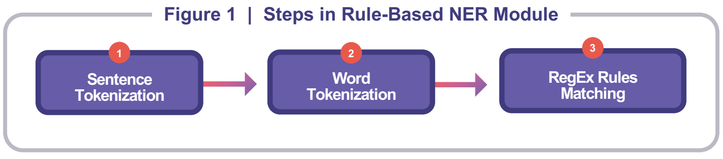


Figure 2 | Steps in CRF NER Module

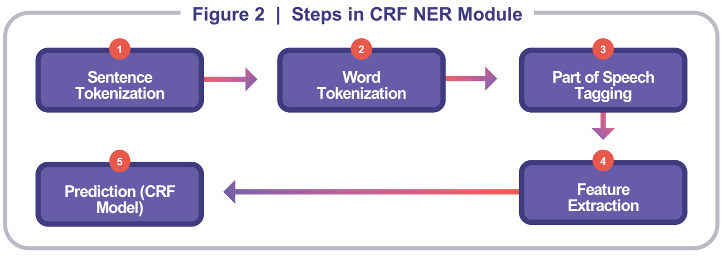


Figure 3 | Rule-Based NLP Process Flow Diagram

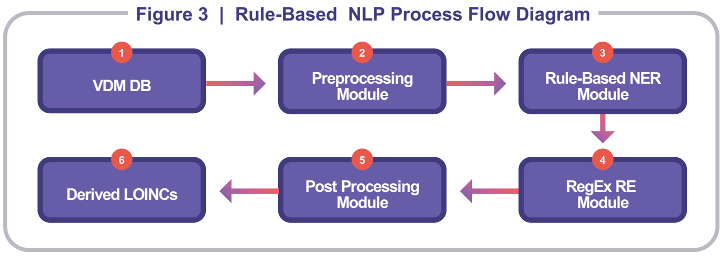
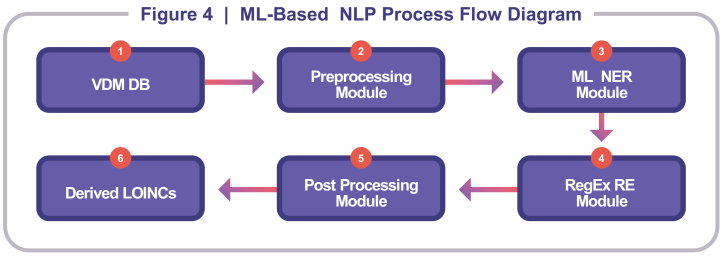


Figure 4 | ML-Based NLP Process Flow Diagram



Data: We manually annotated 2924 sentences containing LVEF results, pulled from de-identified SOAP notes on the Veradigm EHR platform. 1424 of the sentences were used for model tuning (training and testing) and 1500 for validation.

Tools: Natural Language Toolkit version 3.4.4 was used for tokenization and part of speech tagging. The CRF NER model was trained using Sklearn CRF Suite version 0.3.0. Custom python scripts were written to preprocess the notes after tokenization at the sentence and word level before being fed into the CRF model. For the RE stage, a custom python script with RegEx was written to associate individual entities identified during the NER stage to their corresponding LVEF assessment counterparts. All python libraries used in this pipeline are compatible with Python 3.6.0 and higher versions.[1]

Performance metrics: Precision, recall and F-1 accuracy score based on the validation set were determined for both pipeline methods.

RESULTS

Table 1 | ML Pipeline outperforms Rule-Based Pipeline

PIPELINE	N SIZE	PRECISION	RECALL	F1
Rule-Bases	1500	0.95	0.81	0.87
Machine Learning	1500	0.95	0.94	0.94

Table 2 | ML Pipeline Performance with the three types of LVEF assessments

ASSESSMENT TYPE	N SIZE	PRECISION	RECALL	F1
Ratio Percentage Value	500	0.95	0.93	0.94
Interval Percentage Value	500	0.95	0.93	0.94
Relative Percentage Value	500	0.96	0.95	0.93

DISCUSSION

Patterson et al., used a rule-based pipeline to extract LVEF assessments from general clinical notes, echocardiogram reports, and radiology reports that yielded F-1 accuracy scores of 0.872, 0.842 and 0.877 respectively.[2] Wagholikar et al., used a rule-based pipeline approach that achieved a 1.0 accuracy,[3] but their analysis focused exclusively on structured records from echocardiogram reports. In contrast, our work focused on SOAP notes which are unstructured in nature.

Our rule-based pipeline did not show significant improvement over results for rule-based pipelines from previous studies. However, the machine learning pipeline showed a significant gain in performance.

While the machine learning pipeline maintained the same precision performance as the rule-based pipeline, recall score from the machine learning pipeline was 13 percentage points higher than that of the rule-based pipeline. The significant improvement of recall raised the F-1 accuracy score from 0.87 to 0.95.

CONCLUSION

A literature review of academic work done on extracting LVEF assessments from clinical documents showed only rule-based methods had been tried with varying degrees of success on echocardiogram and radiological reports. We did not find any work that has specifically been attempted on SOAP notes from an ambulatory EHR platform.

The abundance of LVEF assessments being recorded in unstructured SOAP notes on the Veradigm EHR platform instead of in structured data sources presented an opportunity not only to determine how a rule-based pipeline might perform but also to compare its performance to that of a ML pipeline.

Our rule-based pipeline yielded an F-1 accuracy score in line with work done by others in the field, but the machine learning pipeline demonstrated improved recall performance versus the rule-based pipeline, resulting in a higher F-1 accuracy score. This machine learning methodology can also be used to develop extraction pipelines for other clinical data textual entities found in unstructured EHR notes.

Valuable patient clinical data is more fully appreciated through application of machine learning to otherwise difficult to access provider entries in electronic medical records.

REFERENCES

1. Python Release Python 3.6.0 [Internet]. Python.org. [cited 2019 Mar 12]. Available from: https://www.python.org/downloads/release/python-360/
2. Patterson OV, Freiberg MS, Skanderson M, J Fodeh S, Brandt CA, DuVall SL. Unlocking echocardiogram measurements for heart disease research through natural language processing. BMC Cardiovasc Disord. 2017 12;17(1):151.
3. Wagholikar KB, Fischer CM, Goodson A, Herrick CD, Rees M, Toscano E, et al. Extraction of Ejection Fraction from Echocardiography Notes for Constructing a Cohort of Patients having Heart Failure with reduced Ejection Fraction (HFrEF). J Med Syst. 2018 Sep 25;42(11):209.

International Society for Pharmacoeconomics and Outcomes Research (Virtual) ; Bay Lake, FL, USA; May 18 – 20, 2020