

Using Natural Language Processing to Identify Laterality of Nephrolithiasis Surgical Procedures

Cosmin A. Bejan, PhD,¹ Siwei Zhang, MS,² Yaomin Xu, PhD,^{1,2} Ryan S. Hsi, MD³

¹Department of Biomedical Informatics, ²Department of Biostatistics, ³Department of Urology; Vanderbilt University Medical Center, Nashville, TN, USA

Introduction

Laterality of surgical procedures could provide important insights to better understand clinical phenotypes and patient outcomes. In the Electronic Health Record (EHR), laterality of procedural codes, such as Current Procedural Terminology (CPT) codes, can be represented by CPT modifiers. However, their utilization and validity are unknown. The purpose of this study is twofold. First, we assess the coverage of CPT modifiers expressing laterality of nephrolithiasis surgical procedures (i.e., surgery performed on the left, right, or both kidneys) and evaluate their accuracy in the EHR. Second, we develop a natural language processing (NLP) method to automatically extract such laterality attributes from clinical notes when CPT modifiers are not available.

Methods

From the de-identified EHR of Vanderbilt University Medical Center, nephrolithiasis surgical procedures were extracted using the following CPT codes: 52352, 52353, 52356, 50590, 50080, and 50081. For laterality, we identified the CPT modifiers indicating left, right, and bilateral procedures. The evaluation of CPT modifiers consisted of manually reviewing 150 randomly sampled patients with at least one nephrolithiasis surgical procedure and a corresponding CPT modifier for laterality (50 each for left, right, and bilateral).

For the procedures without CPT modifiers, a previously developed NLP system for kidney stone phenotypes was adapted to automatically extract laterality for every patient with a nephrolithiasis surgical procedure. A set of 200 patients with nephrolithiasis surgical procedures was randomly selected and, for each procedure, a laterality attribute (left, right, or bilateral) was assigned after manual review of the corresponding patient record. Pattern matching algorithms were used to analyze notes associated with each surgical procedure. Regular expressions encoding *left*, *right*, and *bilateral* or *both* were implemented to analyze the textual context of keywords including *ureteroscopy*, *ureterorenoscopy*, *shockwave*, *kidney*, *ureter*, *ureteral*, *percutaneous*, *renal*, *UPJ (ureteropelvic junction)*, *hydronephrosis*, and *hydroureteronephrosis*. The laterality was selected based on the maximum number of matching expressions. For bilateral, this condition was relaxed to a limited number of matches of the *bilateral* or *both* expressions. A grid search was performed on 50% of the data (training set) to optimize the parameters listed in **Table 1**. Evaluation was performed on the remaining 50% of the data (test set).

Results

We extracted a total of 11,537 nephrolithiasis surgical procedures. Of these, 2,276 have an assigned CPT modifier for laterality and 3,219 were performed since the date of the first modifier. The evaluation of CPT modifiers for laterality indicates a precision of 98.3%, 98.1%, and 100% for left, right, and bilateral, respectively.

From the procedures without CPT modifiers available, 311 were reviewed for laterality, out of which 155, 132, and 24 were labeled as left, right, and bilateral, respectively. **Table 1** lists the parameter values that achieved the best results on the training set for each laterality category. Overall, the best performance was achieved when analyzing notes at and +/- 2 days from the procedure date and matching the left context of the relevant keywords (e.g., *bilateral ureteroscopy*). The NLP evaluation on the test set using these values is shown in **Table 2**.

Conclusions

We demonstrated the potential of extracting laterality of nephrolithiasis procedures from notes when this information is not available in structured format. When available, CPT modifiers indicate laterality with high precision.

Table 1 Parameter values optimized by the NLP system

Parameter	Left	Right	Bilateral
Days apart from the procedure date for note selection	2	2	2
Context laterality (left, right, both)	left	left	left
Context size (characters)	20	20	10
Minimum number of matches for bilateral expressions	-	-	4

Table 2 System evaluation

Laterality	P	R	F
Left	89.94	92.26	91.08
Right	84.21	96.97	90.14
Bilateral	48.72	79.17	60.32

P: precision, R: recall, F: F1 score

Funding: This study was supported by the National Institutes of Health (NIH) grant R21DK127075.