Natural Language Processing to Screen for Kawasaki Disease

Chun-Nan Hsu[1], Son Doan[1], Juan D Chaparro[1,4], Cleo K Maehara[1], Ruiling Liu[2], Sisi Lu[3], Erika K Berry[4], Divya Chhabra[1], Adriana H Tremoulet[4] and Jane C Burns[4].

[1]*Biomedical Informatics, University of California, San Diego, La Jolla, CA, United States;*
[2]*Biomedical Informatics, University of Texas, Houston, TX, United States;*
[3]*Computer Science, University of Pittsburgh, Pittsburgh, PA, United States and* [4]*Pediatrics, University of California, San Diego, La Jolla, CA, United States.*

Introduction

Delayed diagnosis of Kawasaki disease (KD) may lead to serious cardiac complications. Until a simple method to identify possible KD patients from the medical record is created for KD, children will continue to experience delays in diagnosis and suffer potentially preventable morbidity and mortality.

Objective

To create a natural language processing (NLP) tool for early and rapid detection of subjects with a high suspicion for KD from text in clinical notes in the electronic medical record.

Materials and Methods

The study was conducted using emergency department (ED) notes from patients seen at Rady Children's Hospital-San Diego between 2010 and 2014.

Study Subjects
KD: Patients diagnosed between illness day 1 (first day of fever) and 10 who met American Heart Association (AHA) KD criteria
Febrile Children (FC): Patients with fever for at least three days with one or more KD-compatible clinical signs and no serious underlying medical condition

KD Tagger Development
The main component of KD-NLP is a KD tagger, which identifies KD signs from the clinical text using three steps **(Table 1 and Figure)**:

1. Lexicon look-up: Search for synonyms of KD AHA guideline clinical criteria using 28,580 keywords from the Unified Medical Language System (UMLS) dictionary
2. Pattern look-up: Search for strings of words that express KD signs
3. Negation detection: Exclude tag if "no" or "absent" linked to a KD sign

Physicians with KD expertise annotated 22 ED notes by tagging each mention of KD clinical signs for KD-NLP development and training.

Table 1: KD Signs and Tags.

Extremity changes	"… swollen hands and red feet …"
Polymorphous exanthema	"… rash in the chest …"
Conjunctival injection	"… eyes are red …"
Oral changes	"… had a strawberry tongue …"
Cervical lymphadenopathy	"… pts left side of neck swollen …"

Outcome Measures
High suspicion for KD: Fever and ≥3 KD clinical signs
Low suspicion for KD: Fever and one or two KD clinical signs
We used standard sensitivity, specificity, positive predicted value (PPV) and F-measure for evaluation.

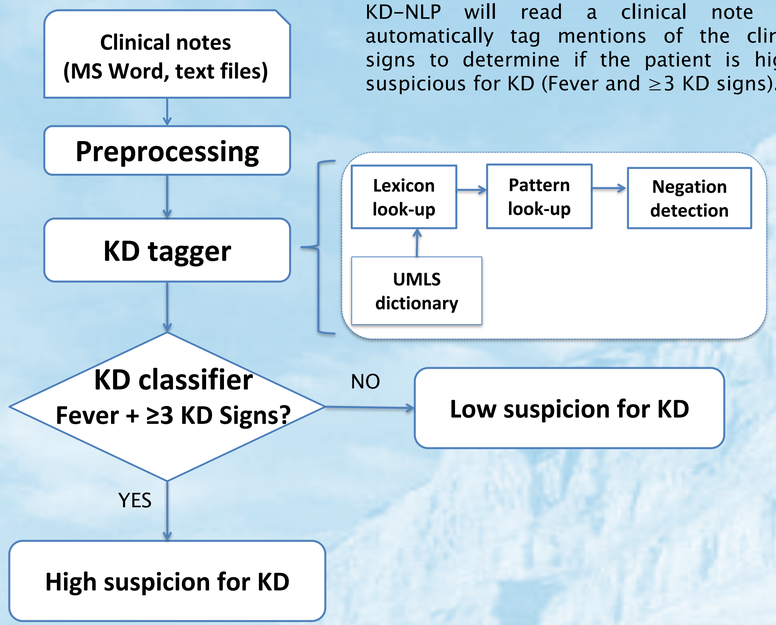

Figure: The pipeline of the KD-NLP tool. KD-NLP will read a clinical note and automatically tag mentions of the clinical signs to determine if the patient is highly suspicious for KD (Fever and ≥3 KD signs).

Validation
166 ED notes (64 KD subjects and 102 FC subjects with at least 1 KD sign) were used for validation.

Results

Table 2: Performance of KD tagger for each KD sign.

	PPV	Sensitivity	F-measure
EXTREMITY_CHANGES	71.43	65.57	68.38
POLYMORPHOUS_EXANTHEMA	84.21	98.46	90.78
ORAL_CHANGES	84.04	86.81	86.40
CONJUNCTIVAL_INJECTION	96.49	99.10	97.78
CERVICAL_LYMPHADENOPATHY	82.18	96.51	88.77
Micro-average	**85.69**	**91.91**	**88.87**

Table 3: Performance of KD-NLP.

KD-NLP Classification	Classification by Medical Expert		Total
	High Suspicion for KD	Low Suspicion for KD	
High suspicion for KD	101	12	113
Low suspicion for KD	8	45	53
Total	109	57	166

specificity: 78.95%) in identifying subjects with high suspicion for KD.

Discussion

KD-NLP was developed to identify subjects with a high suspicion for KD who should undergo further evaluation, such as laboratory testing, an echocardiogram, and referral to a major pediatric center. In this retrospective study, KD-NLP had a high sensitivity (>92%) for finding subjects with a high suspicion for KD who may require additional evaluation.

Evaluation of the errors made by KD-NLP revealed that misspelling, unusual syntax, and hypotheticals in the text can be challenging for the tool. These can be addressed in a future release by adding keywords, patterns and a spelling checker.

Conclusion

KD-NLP is effective for detection of pediatric patients in whom the diagnosis of KD should be considered. The next step is to incorporate this tool into the electronic health record system and prospectively test how this tool performs in a real-world clinical setting.

References and Funding

[1] Newburger JW, Takahashi M, Gerber MA, *et al.* Diagnosis, treatment, and long-term management of Kawasaki disease: A statement for health professionals from the Committee on Rheumatic Fever, Endocarditis and Kawasaki Disease, Council on Cardiovascular Disease in the Young, American Heart Association. *Circulation* 2004;**110**:2747-71. doi:10.1161/01.CIR.0000145143.19711.78

This research was supported by the Patient-Centered Outcomes Research Institute (PCORI) Contract CDRN-1306-04819 and iDASH NCBC through the NIH Grant U54 HL108460 to the University of California, San Diego