04145
A natural language processing system for characterization of radiology text reports

04. Diagnostic microbiology

4i. Bioinformatics tools & pipelines

George Zhao [1], Yuting Xing [2], Bernard Hernandez Perez [1, 2], Timothy Miles Rawson [2, 3], Richard C Wilson [2, 3], Damien Keng Yen Ming [2, 3], Pantelis Georgiou [1], Alison H Holmes [2, 3]

[1]Centre for Bio Inspired Technology, Imperial College London - London (United Kingdom), [2]Centre for Antimicrobial Optimisation, Imperial College London - London (United Kingdom), [3]NIHR HPRU in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London - London (United Kingdom)

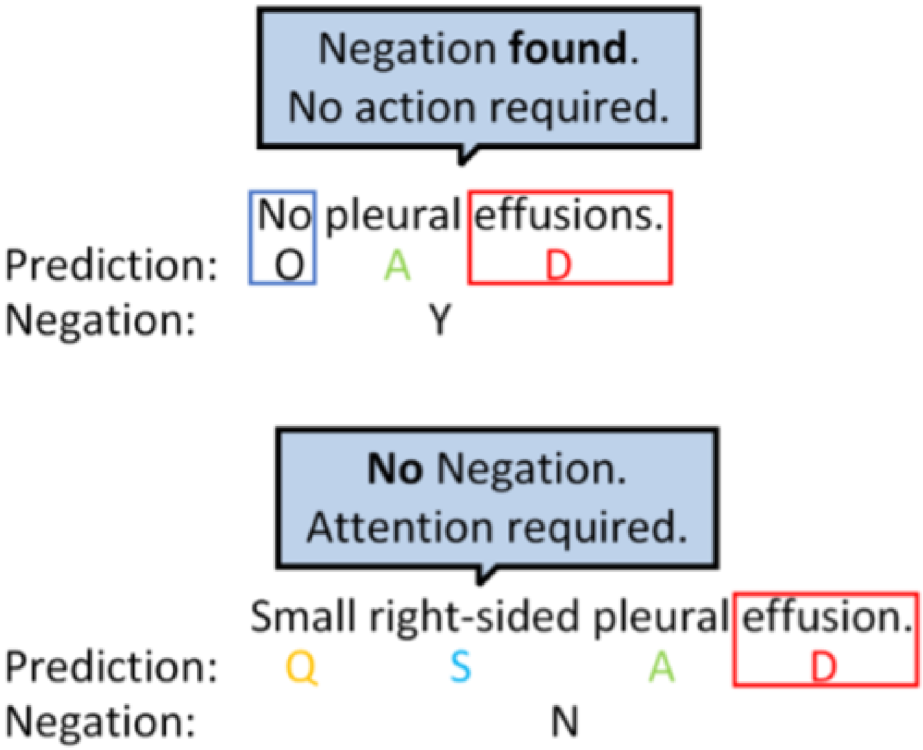
Background The development of clinical decision support system (CDSS) in infection management was greatly accelerated by the emergence and integration of electronic health records (EHRs) which store valuable patient information as unstructured text. These reports might include patient's history, symptoms, description of findings from medical examination, diagnosis, and/or prescribed medication among others. The unstructured nature of text reports presents two main drawbacks: (i) finding or reviewing information is difficult and time consuming and (ii) limits the potential secondary use of this information to enhance machine learning based CDSS. We evaluated the ability of natural language processing to detect and extract clinical entities of interest to automatically label radiology text reports.

Methods Data used consists of over 10000 computerized tomography and chest radiology text reports from the NHS Trust produced by radiologists. The system developed consists of an unsupervised deep learning based natural language processing model to automatically detect and extract clinical entities of interest. The system leverages a fast biomedical concept extraction tool quickUMLS from which it inherits the entity labels (see Figure 2) and a pre-trained biomedical language representation model BioBERT which has been fine-tuned to extract contextual information from radiology text reports to support clinical named entity recognition.

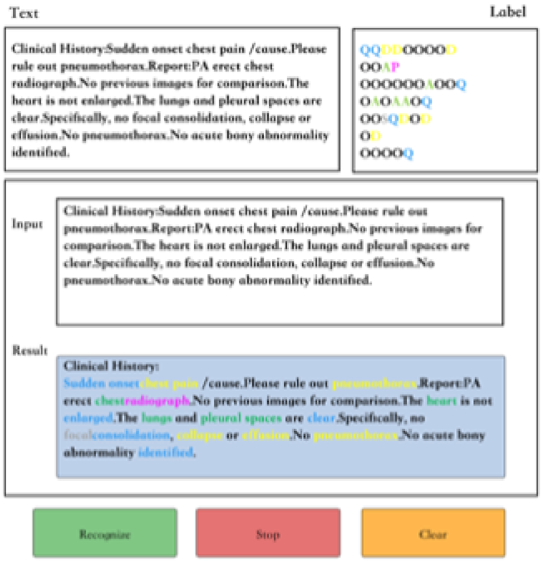
Results The performance (see Figure 3) shows that the proposed method can extract medical terms including disease, body part and location, phenotype, diagnostic procedure, or medical device. In addition, it includes a method to detect negation descriptions (see Figure 1) and contextual information which are essential and challenging in understanding the semantic meaning. Results in Figure 3 show that our system performs best on identifying the location (such as organ and body location) and the associated qualitative concept, achieving 87% and 83% in terms of F1 score respectively.

Conclusions This study presents a system for entity recognition to characterize chest radiology reports to automatically extract and present most important findings to clinicians, which could improve the efficiency of further diagnosis and treatment, thus, facilitate infection management at point of care. Work is underway to complete the development of the user interface and its integration into EPiC IMPOC, a point-of-care web-based CDSS.

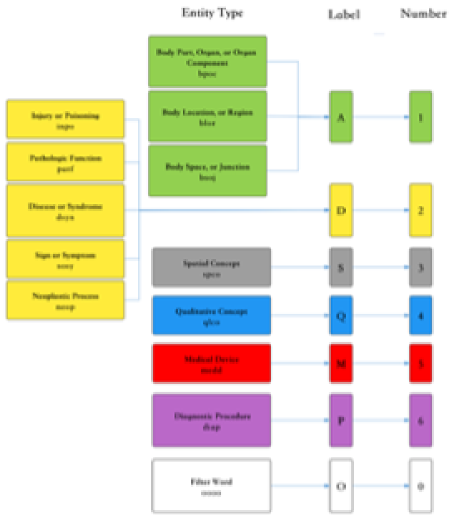
Example of negation detection



Integration into the CDSS and definition of labels



Proposed schematic for integration into the clinical decision support system



Labels of entity

Performance of the proposed system

Label	Precision	Recall	F1-score	Support
A	0.7816	0.9793	0.8693	3431
D	0.5429	0.8870	0.6735	2070
S	0.7215	0.9663	0.8261	3172
Q	0.5505	0.9360	0.6933	3080
M	0.6583	0.8155	0.7285	645
P	0.6407	0.9554	0.7670	672
Micro Average	0.6491	0.9420	0.7686	13070
Macro Average	0.6493	0.9232	0.7596	13070
Weighted Average	0.6614	0.9420	0.7741	13070

Strict Accuracy	
Unbalanced	0.8483
Balanced	0.9064

Keyword 1
natural language processing
Keyword 2
biomedical named entity recognition
Keyword 3
clinical decision support system

Conflicts of interest

Do you have any conflicts of interest to declare?
I have no potential conflict of interest to report