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

George Zhao1#, Yuting Xing2#, Bernard Hernandez1,2\*, Timothy Miles Rawson2,3, Richard Wilson2,3, Damien K Ming2,3, Pantelis Georgiou1, Alison H Holmes2,3

1 Centre for Bio Inspired Technology, Imperial College London, United Kingdom

2 Centre for Antimicrobial Optimisation, Imperial College London, United Kingdom

3 NIHR HPRU in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London, United Kingdom

**Word count: 347**

**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.

**Conclusion**

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.

**Keywords:** natural language processing, biomedical named entity recognition, Clinical decision support system

Diagram, text

Description automatically generated

Figure 1: Example of negation detection

A picture containing diagram

Description automatically generated

Figure 2: Proposed schematic for integration into the clinical decision support system (left) and definition of entity labels (right).

Table

Description automatically generated

Figure 3: Performance of the proposed system.