

## Address on Review Comments (Paper 18)

### REVIEWER #1

1. For the comment on context on the overall significance of the problem, we further demonstrated the significance and usage context of this task in the introduction.
2. For the comment on lack of detail on the nature of the "diagnostic criteria" texts, we added a definition of diagnostic criteria and illustrated with a diagnostic criteria text of diabetes mellitus as an example in Appendix. So the reader could know about their length and linguistic complexity.
3. For the comment on where the diagnostic criteria texts came from? We explain our 4 data sources at the part **3 Results**.
4. About the numbers in the paper, such as 218 criteria, 44 criteria, 78 criteria. Those mean the count number of criteria. 44 is the number of full text diagnostic criteria, and 218 is the number of individual criteria which extract from full text diagnostic criteria, as well 78 criteria is the number of individual criteria in laboratory test category.
5. Please clarify the tasks that were evaluated is a classification task or an annotation task ? The evaluation executed on the mapping task based on mapping rules created on these two models, but the annotation task is one of the important preprocess process to structuralize textual criteria data, so as the mapping could automatically performed based on data model mapping rules.
6. For the comment on some citations are not provided. We supplemented the citations missed in last version to help reader find further information.
7. For the comment on some concepts are not defined (e.g. CUIs in 2.1). We supplemented one core definition about Diagnostic Criteria in this paper, but for those abbreviations common as usage in medical field, we just added the full name of the term.

### REVIEWER #2

1. About comment on the "sub-optimal" performance on symptoms. We have demonstrated that some studies compared the performance of the frequently used clinical NLP tools in **1 Introduction** part, and the results showed that cTAKES scored best in both performance and usability.
2. For the comment on lack of "big picture" sense of the task and the approach, see REVIEWER #1 question 2.
3. For the comment on some of the terminology would benefit from an example to build intuition. We added 4 examples to help reader understand overall task and build intuition, including diagnostic criteria text example, HQMF representation example and cTAKES annotation in SignSymptomMention and LabMention.
4. For the comment on where were the free text diagnostic criteria gathered from, see REVIEWER #1 question 3.
5. For the comment on other terminology should be defined, see REVIEWER #1 question 7.
6. For the comment on what are the Code, Value, Unit and Operator columns in Table 3. For the laboratory test criteria, for example, Thrombocytopenia (platelets <100,000 cells/mm<sup>3</sup>), Code, Value is the standard term code and expression of annotated laboratory test term *platelets*, such as SNOMED CT code and normalized expression. Unit are the recognized measurement *cells/mm<sup>3</sup>* and Operator is the recognized symbol <.
7. For the comment on the manuscript would benefit from an additional check of the grammar. Two authors rechecked the writing of revision version.

### REVIEWER #3

1. About comment on the difference between "concept mapping" and "instance mapping" is not clear. We revised the names of two mapping levels into datatype-level and data-level for clear understand.
2. For the comment on where the criteria data collected from?, see REVIEWER #1 question 3.
3. For the evaluation comments, see REVIEWER #1 question 5. Plus, just the mapping rules are manually created, all other processes are automatically.
4. About the example comments, see REVIEWER #1 question 2 and REVIEWER #2 question 3.
5. About the application of our study, we supplemented some backgrounds and explanations of this study, in the future the implementation system would support computerized clinical diagnosis.\_