We thank the reviewer for the helpful comments.

Reviewers' comments:  
Reviewer #3: The authors have improved the manuscript to make it clear. However, there're still some points that I'd like to see further improvements.  
- I still have doubts about the authors' claim "Some writing conventions in Chinese clinical notes make the density of PHI much lower than that in English clinical text. (Section 1)". Please provide concrete evidence such as statistics or reference to other paper for this claim. Otherwise, please remove this claim from the paper.

We removed the claim from the paper.

- How many clinical notes did you consult to come up with the 6 regular expressions in Table 1. Are the notes completely different from the notes in the 'raw set' and the 'dense set' described in Section 2.3.2? How much time did the process of regular expression compilation take? Is it still more efficient to use dense PHI corpus if we take time for regular expression compilation into consideration?

We sampled 50 clinical notes from different departments to come up with the regular expressions, and we add this information to the new manuscript. Unfortunately, we did not record the time spent on regular expression compilation. However, we proved in the paper that rules compiled on a relatively small size of data set can be used with high sensitivity on larger data sets. In this study, we only used a training data of 300 notes, but the regular expression would still be valid if the training data gets larger. From this standpoint, the method saves more time as the training data gets larger.

- In Section 3.1.2, "two experts" annotated raw set and dense set. Did each of the expert annotate raw set and dense set separately? In that case, the result in Table 3 is not a fair comparison, since we need to take individual annotator's performance difference into consideration.

Both of them annotated raw set and dense set. We clarified this in the new manuscript  
- In Section 3.2.2. Were the radiology reports filtered out to identify PHI sentences first? Can you report performance with/without such filtering?

Thanks for pointing it out. We add more information to the section to make it clear.   
- Can you change the graphical abstract further? Please differentiate "data" and "process" in the diagram. Also, please differentiate "training/development" process and "actual de-identification process", as you have done for Figure 1.

We have updated the new version.  
- Please consider moving the following sentence from Section 2.3.2 to Section 2.4.2.  
- "In our system, a typical representation schema for … and 'S' denotes that a character itself is a PHI."  
- In Section 2.3.2, in "The remaining dense PHI corpus was evenly split and then …", did you mean "the remaining 200 notes"?

Yes. We have made it more clear in the new version.

- In Table 5, please report overall performance.

We have added it in the Table 5.

- Please check English of the followings (please clarify or correct grammar errors):  
- Section 1: awareness of protecting personal information have been raising in healthcare

Corrected.

- Section 1: that may represent patient privacy

Fixed.  
- Section 2.1: while machine learning finds the PHI but requires more annotation effort

Corrected.  
- Section 2.3.1: it is still possible to other features, mostly lexical ones, to create rules

Corrected.  
- Section 2.3.1: subsequence supervised learning

Corrected.  
- Section 4.1: Machine learning is usually successfully in sequence-labeling tasks.

Fixed.