

**Title:** Automatic detection of actionable radiology reports using bidirectional encoder representations from transformers

**Version:** 1 **Date:** 09 May 2021

**Reviewer's report:**

Natural language processing (NLP) has been proved to be effective in medical text classification. In this paper, the author used BERT in identifying free-text radiology reports including actionable findings. The author also use order information in radiology reports.

BERT had been applied in medical language processing recent years in lots of papers. This work was a simple application of BERT, which lacked innovation. Moreover, the clinical significance of this work had not been explicitly stated.

Major comments

1) The author compared BERT with statistical machine learning methods to prove the advantage of BERT. I think more comparison with other deep learning methods was needed.

2) Result in Figure 6 to 8 suggested that all three methods mainly relied on whether radiology reports contain specific expressions of recommendation, suspect, or negation. The author should discuss the clinical meanings of these top expressions, since I didn't find clinical significance of expressions like "no", "apparent" etc.

3) In the classification problem, the low positive case ratio (0.62%) of data needs to be noted. In this study, oversampling or under sampling methods were not applied to solve the data imbalance problem. In the Discussion section, the author mentioned the method by Madabushi et al. However, the author didn't apply this method to solve the data imbalance problem.

4) The dataset contained radiology reports corresponded to brain, head and neck, body and musculoskeletal. Writing emphasis and content in these different kinds of reports had big difference. The author should discuss the impact on the result since these reports were simply mixed together.