The paper titled "The value of computer-aided diagnosis in Breast Imaging Reporting and Data System ultrasound training for residents" is interesting. The AI-SONIC breast automatic detection system can help residents to quickly master BI-RADS, improve the consistency between residents and experts, and help to improve the confidence of residents in the classification of BI-RADS, which may have potential value in the BI-RADS training for radiology residents. However, there are several minor issues that if addressed would significantly improve the manuscript.

Comment 1: The font in the manuscript is inconsistent, and there are also issues with line spacing. Please carefully check and make corrections.
Reply 1: Thanks for your reminding, we have read the article carefully and adjusted the font and line spacing.
Changes in the text: The text font and line spacing have been modified in detail.

Comment 2: The abstract is not sufficient and needs further modification. The research background did not indicate the clinical needs of the research focus.
Reply 2: Thanks for your valuable comments, we have added a note on the importance of research in the background section.
Changes in the text: Page 1,line 28-30.

Comment 3: Can conventional ultrasound BI-RADS classification with AI-SONIC breast automatic detection system reduce the probability of BI-RADS biopsy? Suggest adding relevant content.
Reply 3: AI-SONIC breast automatic detection system has proved its value in previous studies (see Reference 16 for details). The AI-SONIC breast automatic detection system has a high consistency with senior radiologists and can reduce the biopsy rate of Category 4 nodules of BI-RADS. We have also added this part to the article.
Changes in the text: Page 4,line 98-99.

Comment 4: All figures are not clear enough. It is recommended to provide clearer figures again.
Reply 4: All figures are not clear enough. It is recommended to provide clearer figures again.
Changes in the text: See a separate file for the picture.

Comment 5: The introduction part of this paper is not comprehensive enough, and the similar papers have not been cited, such as "A new nomogram for predicting the malignant diagnosis of Breast Imaging Reporting and Data System (BI-RADS) ultrasonography category 4A lesions in women with dense breast tissue in the diagnostic setting, PMID: 34249630". It is

recommended to quote the article.

Reply 5: Thanks for your suggestion, we have added a reference to this article in the introduction section, see Reference 12 for details.

Changes in the text: Reference 12.

Comment 6: How to examine associations between potential predictors and kappa values? Suggest adding relevant content.

Reply 6: We are very sorry that we cannot understand the reviewer's question well. Kappa values are generally defined as the proportion of agreement among raters after the exclusion of accidental agreement. Less than 0 indicated poor agreement; 0.00–0.20 indicated slight agreement; 0.21–0.40 indicated fair agreement; 0.41–0.60 indicated moderate agreement; 0.61–0.80 indicated substantial agreement; 0.81–1.00 indicated almost perfect agreement.

Changes in the text: Page 7,line 202-205.

Reviewer B

Comment 1: First, the title needs to indicate the research design of this study such as a randomized clinical trial. Please also revise the term "value" to be accuracy, since this term is very broad.

Reply 1: Thanks for your valuable comments, we have added the research design method to the title and modified the word "value".

Changes in the text: Page 1,line 2-3.

Comment 2: Second, the abstract needs some revisions. The background needs to describe the knowledge gap and significance of this research focus. The methods need to specify the randomization method and the outcome measurements of confidence and recognition degree. The results need to first report the baseline comparability across the three groups and the accurate P values for these comparisons.

Reply 2: Thank you for your valuable comments. We have added the importance of research to the background section. Baselines for the three groups were added to the results section, along with their respective P-values.

Changes in the text: Page 1,line 28-30. Page 2,line 43-45.

Comment 3: Third, the introduction needs to have comments on the knowledge gaps of AI-SONIC for the detection of BC. If it has been proved to be accurate, what the clinical needs for this research are. Please further clarify. In addition, the authors need to analyze the potential reasons for the varying accuracy of CAD-aided BC detection and explain how the AI-SONIC system overcomes these limitations.

Reply 3: Thank you for asking these thought-provoking questions, which are really worth thinking about. AI-SONIC system detection of BC has been demonstrated in previous studies, as described in reference 16. However, whether it is valuable in the training of resident BI-RADS classification has not been reported, which is the question we want to explore in this study. The accuracy of CAD assisted BC detection is different, which I guess may be

related to factors such as the database of CAD pre-training and the different experts sketched areas of interest. This is a very interesting topic that can be explored in depth in future research. The biggest feature of the AI-SONIC system is that it can quantify the description of several aspects involved in BI-RADS classification, which is clearly presented to the trainees, and this is the source of our ideas to carry out this research.

Changes in the text: Page 4, line 92-95, 101-102.

Comment 4: Fourth, in the methodology of the main text, please describe the clinical research design and sample size estimation procedures. Due to the very small sample size, there is no need to do statistical comparisons across the three groups. In statistics, please describe the statistical methods for pairwise comparisons after ANOVA and ensure $P<0.05$ is two-sided.

Reply 4: Thank you for your valuable advice on statistics. Determining the appropriate sample size is indeed the first important issue to consider in clinical research. However, you may also know that the number of doctors in China is decreasing year by year, and it is difficult for us to recruit enough students to participate in our study in a short period of time, so the sample size of this study is relatively small. In the future research, we will recruit more students to participate in our research, so as to make our research more convincing. According to your suggestion, we have removed pairwise statistical comparisons between the three groups. For details, see Table 2 and the statistics section of the paper.

Changes in the text: Table 2, Page 7, line 200.

Comment 5: Finally, please consider to cite several related papers: 1. Xiao M, Zhao C, Zhu Q, Zhang J, Liu H, Li J, Jiang Y. An investigation of the classification accuracy of a deep learning framework-based computer-aided diagnosis system in different pathological types of breast lesions. J Thorac Dis 2019;11(12):5023-5031. doi: 10.21037/jtd.2019.12.10. 2. Suzuki K. A review of computer-aided diagnosis in thoracic and colonic imaging. Quant Imaging Med Surg 2012;2(3):163-176. doi: 10.3978/j.issn.2223-4292.2012.09.02. 3. Liu C, Chen S, Yang Y, Shao D, Peng W, Wang Y, Chen Y, Wang Y. The value of the computer-aided diagnosis system for thyroid lesions based on computed tomography images. Quant Imaging Med Surg 2019;9(4):642-653. doi: 10.21037/qims.2019.04.01.

Reply 5: Thank you for the three valuable references you provided. After careful study, we quoted them in appropriate places in the paper.

Changes in the text: Reference 13,14,21.