**Peer Review File**

Article information: https://dx.doi.org/10.21037/gs-23-417

**Reviewer A**

First of all, my major concern for this study is the small sample size of the validation sample, in particular the numbers of low nuclear grade, ER +, PR+ and HER 2+ in the validation sample. In fact, such a small sample sizes do not allow the validation. The authors need to address this concern in their revisions. Given this limitation and the AUC values lower than 0.8, I suggest the authors to tone down the current title and the current conclusion. The title needs to clearly indicate the development and validation of two prediction models.

**Reply 1:** Thank you for your comment. Because this study is based on a small-scale study of a single disease, which is also a limitation of our research, we did not conduct a multicenter study. We have discussed this in the limitations section. In addition, DCIS is an early-stage cancer with a lower mortality rate. We collected the cases confirmed by DCIS surgery in West China Hospital for 16 years thus far, but the available ultrasound data is still limited. Due to the inherent causes of the disease and limitations in diagnostic techniques, a small proportion of low nuclear grade DCIS were diagnosed after surgery, which also caused data imbalance. We have tried several different model methods, but the results are still not ideal. The AUC values of all models are less than 0.8. In the future, larger data scales and multi center research may be needed to optimize the models to improve predictive ability. We have made corresponding modifications in the title, abstract, and conclusion sections.

**Changes in the text:** see Page 1, lines 1-3, we have changed the title; Page 4, lines 70-73, we have added 'Based on small-scale datasets, our study…'; Page 19, lines 436-450, we have changed the contents about limitation part.

Second, the abstract needs some revisions. The background did not explain why the authors developed two models and compared them in a study and what the current knowledge gap is. The methods need to briefly describe the inclusion of subjects, the clinical factors collected, and how

the low nuclear grade, ER+, PR+ and HER2+ were ascertained. The results need to briefly summarize the rates of low nuclear grade, ER+, PR+ and HER2+ and AUCs in both training and validation samples for both DLR and CML models. The conclusion needs comments for the limitations of this study.

**Reply 2:** Thank you for your comment. DCIS is a strongly heterogeneous breast cancer, which has variability in ultrasonic manifestations. Some tumors have undefined boundaries, making it difficult to label and outline them. Our aim was to attempt different research methods (both "white box" and "black box" methods), specifically using deep learning, deep imaging omics, and clinical machine learning methods. Due to the advantages of deep imaging omics in tumor contour recognition, this may be suitable for the data we are studying. Clinical machine learning models can explain the importance of imaging features. Therefore, we conducted a preliminary attempt using deep imaging omics and clinical machine learning methods in this study. Due to word count limitations, we did not include specific data collection definitions such as low nuclear level, ER positive, PR positive, and HER2 positive in the abstract. However, we provided specific explanations in the methodology section.

**Changes in the text:** see Page 3, lines 37-71, we have changed and added some contents in Abstract; We have changed the relevant expression (Page 8, lines 147-154, Page 10, lines 192-195; Page 19, lines 449-450.)

Third, in the introduction of the main text, the authors need to briefly analyze why radiomics and clinical factors could predict the molecular subtype of pure DCIS, what has been known on the accuracy and limitations of the two types of models, based on radiomics and clinical factors, and explain why the authors combined the two types of predictors together to improve the predictive accuracy.

**Reply 3**: Thank you for your comment. Clinical machine learning models can use interpretable ultrasound features for model training and testing, and we can identify which features contribute significantly in these features. In contrast, deep learning or deep imaging omics models act as black boxes, and from the input to the output, our research results provide a preliminary exploration that can explain the relationship between white boxes and black boxes.

**Changes in the text:** We have changed the relevant expression and added some contents (see Page 6, lines 99-103; Page 7-8, lines 115-145).

Fourth, in the methodology of the main text, the authors need to accurately describe the clinical research design, and sample size estimation procedures. In statistics, please ensure $P<0.05$ is two-sided and provide the threshold AUC values for a good predictive model. Please also describe the calculation of sensitivity and specificity, as well as their threshold AUC values.

**Reply 4:** Thank you for your comment. Our study grouped cases after inclusion and exclusion, with each task independently and randomly distributed. We added an AUC threshold on the ROC curve. At the same time, in the methodology section, we also provided a formula for calculating evaluation indicators.

**Changes in the text:** We have added the related contents and changed the Figure 3. (see Page 11, lines 218-219; Page13, lines 294-302; Page16, lines 353-354; Page31, lines 766-769; and Figure 3.)

Finally, please cite several related papers:

1. Fuentes-Sánchez C, González-San Segundo C. Can we avoid treatment in patients with low-risk ductal carcinoma in situ? Ann Breast Surg 2023;7:33.

2. Bonev VV. Ductal carcinoma in situ: a comprehensive review on current and future management for the surgeon and non-surgeon. AME Surg J 2021;1:27.

3. O'Keefe TJ, Harismendy O, Wallace AM. Histopathological growth distribution of ductal carcinoma in situ: tumor size is not "one size fits all". Gland Surg 2022;11(2):307-318. doi: 10.21037/gs-21-599.

4. Zhu M, Pi Y, Jiang Z, Wu Y, Bu H, Bao J, Chen Y, Zhao L, Peng Y. Application of deep learning to identify ductal carcinoma in situ and microinvasion of the breast using ultrasound imaging. Quant Imaging Med Surg 2022;12(9):4633-4646. doi: 10.21037/qims-22-46.

**Reply 5:** Thank you for your suggestion. We have added the above references.

**Changes in the text:** We added the References (pages 21-31).

**Reviewer B**

The paper titled "Ultrasound deep learning radiomics and clinical machine learning models can predict low nuclear grade, ER, PR and HER2 receptor status in pure ductal carcinoma in situe" is interesting. The DLR and CML models developed by using RadImageNet pretrained network and machine leaening methods can accurately predict the low nuclear grade, ER+, PR+, and HER2+ DCIS lesions so that patients can benefit from hierarchical and personalized treatment However, there are several minor issues that if addressed would significantly improve the manuscript.

1) The abstract is not sufficient and needs further modification. The research background did not indicate the clinical needs of the research focus.

**Reply 1:** Thank you for your valuable suggestion. We have revised the content of the abstract section.

**Changes in the text:** We have changed the Abstract contents (see Page 3-4, lines 33-75.)

2) How is it feasible to use artificial intelligence technology based on ultrasound radiomics for differential diagnosis of DCIS? Suggest adding comparative analysis.

**Reply 2**: Thank you for your comment. DCIS is a heterogeneous type of breast cancer, with various imaging manifestations, most of which are non-mass manifestations, and it is difficult to determine the boundary. Although we did not use it, we used ultrasound features extracted by doctors and focused our research on deep learning techniques and interpretable clinical machine learning models for analysis. Traditional imaging omics is not the focus of our research, and in the future, traditional imaging omics techniques can be included for comparative analysis.

**Changes in the text:** We have added the relevant expressions (see Page 8, lines 141-143).

3) What are the biggest strengths and weaknesses of this research model? What is the biggest problem faced? Suggest adding relevant content.

**Reply 3:** Thank you for your comment. We have described the advantages, disadvantages, and challenges faced by these two models in the introduction section. Our advantage in research is to use current advanced technology to automatically extract contours, apply interpretable machine

learning as a control, and conduct comparative analysis from different perspectives. The disadvantage is that single center research has a small amount of data and heterogeneity of data. In the future, further requirements for data homogeneity are needed, incorporating multi-center research and further standardized procedures. Additionally, integrating multiple modalities of imaging examinations may optimize the clinical application.

**Changes in the text:** We have modified and added the relevant contents (see Page 7-8, lines 123-146; see Page 19-20, lines 453-465).

4) The models should be further optimized through larger datasets or external validation.

**Reply 4**: Thank you for your comment. Single center research is indeed our weakness, and we have discussed this issue within the limitations section. In the future, larger amounts of data and multi-center studies are needed.

**Changes in the text**: We have added the related contents (see Page 4, lines 73-74; Page 19-20, lines 446-452).

5) The introduction part of this paper is not comprehensive enough, and the similar papers have not been cited, such as "Progress on deep learning in digital pathology of breast cancer: a narrative review, J Gland Surg, PMID: 35531111". It is recommended to quote the article.

**Reply 5:** Thank you for your suggestion. We have added the above reference.

**Changes in the text:** We added the Reference 63. (page 31)

What guidance can this study provide for the early diagnosis and adjuvant treatment in DCIS? It is suggested to add relevant contents.

**Reply 6:** Thank you for your comment. The DLR and CML models developed may potentially screen out low-grade cases for imaging supervision. In addition, these models may help screen for ER, PR, or HER2 positive patients for further stratification.

**Changes in the text:** We have changed and added some contents (see Page 17, lines 392-393; Page 17, lines 399-400).