Response Letter

Dear Editors and Reviewers:

Thank you for your letter and the reviewers' comments concerning our manuscript entitled "**Machine Learning-Based Gray Level Co-occurrence Matrix (GLCM) Prediction Models for Undifferentiated-type Early Gastric Cancer: a comparative analysis of five practical prediction models for evaluating lymph node metastasis"(ID: 78896)**. Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet with approval. The revised portion is marked in red in the paper. The main corrections in the paper and the responds to the Editor's and reviewer's comments are as following:

Comments to Reviewer #1:

Scientific Quality: Grade B (Very good)

Language Quality: Grade A (Priority publishing)

Conclusion: Minor revision

Specific Comments to Authors: I read this article with interest. Wei et al. found that the accurate risk stratification of UEGC patients who should underwent additional surgery depended on the added value of GLCM. Second, a new ML-based prediction model was used to identify patients and whether they have LNM. This article is well written. However, I have minor concerns. 1. Abbreviations in the table and figure should be exlained. 2. Please describe the difference among random forest model(RFC), decision tree(DT), support vector machine(SVM), eXtreme gradient boosting (XGBoost) and neural network(ANN).

Response: Thank you for your useful suggestions. We have adopted them and revised the relevant contents of the manuscript according to your suggestions.

1.Abbreviations in the table and figure should be exlained.

Response: We confirm that the abbreviations in the table and figure have been fully explained. Please refer to the revised figures and tables. Thank you.

2.Please describe the difference among random forest model(RFC), decision tree(DT), support vector machine(SVM), eXtreme gradient boosting (XGBoost) and neural

network(ANN).

Response: Thank you for your suggestions. In order to better describe the differences between the five machine learning models, we have added relevant contents in the methodology section of the article, as follows:

Based on machine learning algorithm, the commonly used iterative algorithm models are included, that is, random forest model(RFC), decision tree(DT), support vector machine(SVM), eXtreme gradient boosting (XGBoost) and neural network(ANN). Among them, RFC is an integrated method, which forms cumulative effect by integrating multiple relatively simple evaluators. It can be understood that random forest is an integrated learning tool based on DT. In addition, SVM is a kind of generalized linear classifier which classifies data binary by supervised learning. ANN is a nonlinear equation transformation output algorithm composed of input layer, hidden layer and output layer. XGboost is an additive model. In each iteration, only the sub models in the current step are optimized.(line190-198).

Comments to Reviewer #2:

Scientific Quality: Grade A (Excellent)

Language Quality: Grade B (Minor language polishing)

Conclusion: Minor revision

Specific Comments to Authors: Congratulations to the authors for the originality of their research and the robustness of their methodology. AI is gaining more and more popularity in various clinical aspects and will soon become an indispensible step in the decision-making process of complex clinical challenges, as is exactly the case with undifferentiated early gastric cancer. All the checklist points have been addressed effectively and adequately. In the attached file you can find some minor comments and suggestions. Two more things in addition: 1) Regarding cross-validation, why would you opt not to use for a k-fold method? 2) I would suggest recapitulating the core characteristics of your study (study design-prognostic vs diagnostic-classification vs regression-according to Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning 397 Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res [Internet]. 2016 Dec 1;18(12). Available from:

https://pubmed.ncbi.nlm.nih.gov/27986644/. Although not mandatory or endorsed by a formal corpus, it would make the study more legible, especially on behalf of the non-familar-with-AI clinician.

Response: Dear reviewer, thank you very much for your valuable comments and suggestions. We also carefully studied some important comments and suggestions you gave us in the attachment, and made corresponding modifications in the manuscript (shown by red marks).

Two more things in addition:

1) Regarding cross-validation, why would you opt not to use for a k-fold method?

Response: Thank you very much for your questions. In this study, we mainly rely on the principle of "bag repeatedly put back and extract", sort according to the weight of variables, and finally get the final predictor of the prediction model from the variables. As for validation, we mainly conduct internal verification by randomly dividing the training set and the verification set, which is also one of the most commonly used methods reported in machine learning related literature. In addition, we highly recognize the ten fold cross validation method proposed by you. Generally, the method is to divide the data set into ten parts, and take turns to take 9 parts as training data and 1 part as test data for testing. However, in consideration of the sample size of this retrospective study, the use of a data set with 10 aliquots may cause greater bias (for example, 526 patients, divided by 10 aliquots, each aliquot has only about 53 patients, which is not conducive to the iterative operation of machine learning). This, together with inadequate reporting of data sources and modeling process, makes research results reported in many biomedical papers difficult to interpret. Therefore, we did not choose the k-fold method. However, in the future, we will include the data of more patients as an external validation cohort. At that time, we can carry out the 10-fold cross-validation method according to your suggestions to prove the accuracy and robustness of our model.

2) I would suggest recapitulating the core characteristics of your study (study design-prognostic vs diagnostic-classification vs regression-according to Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning 397 Predictive Models in Biomedical Research: A Multidisciplinary View. J Med

Internet Res [Internet]. 2016 Dec 1;18(12). Available from: https://pubmed.ncbi.nlm.nih.gov/27986644/. Although not mandatory or endorsed by a formal corpus, it would make the study more legible, especially on behalf of the non-familar-with-AI clinician.

Response: Thank you very much for your questions. We have studied the literature reported by Wei et al, we agree with you and the authors said that a set of guidelines was generated to enable correct application of machine learning models and consistent reporting of model specifications and results in biomedical research. Herein, in this study, we have added relevant contents in the methods section, as follows:

In this study, we refer to the guide for the best use of prediction models in biomedical research proposed by Wei et al, that is, the Delphi method is used to generate the list of reported items(line198-200). We believe that according to the standard of machine learning model construction, the method and subsequent verification in this paper can withstand repeated tests.

Comments to Reviewer #3:

Scientific Quality: Grade A (Excellent)

Language Quality: Grade B (Minor language polishing)

Conclusion: Accept (High priority)

Specific Comments to Authors: This study focuses on the lymphatic metastasis of undifferentiated type early gastric cancer. From the perspective of clinical concerns, this is a hot spot of great concern and has always been a thorny problem faced by clinical surgeons. The author's research is of great practical value. In this study, the author proposed a new LNM prediction model through GLCM features and machine learning algorithm, which is worth learning. However, I have the following confusion about this study, as follows: 1. In this study, the author shared five machine learning models, which are: random forest classifier (RFC) model, support vector machine (SVM), eXtreme gradient boosting (XGBoost), artificial neural network (ANN), and decision tree (DT). Among them, RFC model has the best prediction efficiency, and the author also gives the corresponding candidate variables. In my opinion, the author should list clearly in Table 2, which variables are used in each model, so as to show more clearly what variables are

used in GLCM model. 2. In this study, the author emphasizes that the model obtained by machine learning algorithm is better than the traditional model, which has been demonstrated in this paper. It is suggested that the author add the relevant content of the traditional model (such as logistics regression model, and visualize it by means of nomograph) in this paper, which can more fully show the advantages and disadvantages between machine learning and traditional model (logistics model). 3. In this study, the author used CIC to visualize the optimal model, and found that the differentiation efficiency of RFC model is very good, which is very encouraging. In my opinion, the author should use other models, such as decision tree, neural network, support vector machine, gradient lifting algorithm, etc. the CIC results of these models can be presented in the attached drawings. 4. In this study, some words that need to be corrected are expressed as follows, for example: Line139, Therefore, in order to solve this thorny problem, exploring an robust tool that can predict LNM is necessary. Maybe "precise tool" is better. Line166, Two radiologists with rich experience in gastrointestinal imaging diagnosis referred to T2WI and DWI images to sketch the lesions on the ADC map respectively. In my opinion, the efficiency of artificial recognition should be compared with machine learning model, although artificial recognition may not be as good as machine learning algorithm. Similarly, the prediction efficiency of logistics model should also be shown in Table 2. In general, congratulations to the author's team. This research is of great clinical application value, and the author has a unique way to extract feature factors by using GLCM in image omics, and to use cutting-edge machine learning algorithms. These are very instructive. It is suggested that the author modify the relevant content to make the article more substantial. In short, I suggest giving priority to publishing this research.

Response: First of all, we thank the reviewers for their affirmation. Secondly, we have made the following amendments to the comments made by the reviewers, as follows:

1) In this study, the author shared five machine learning models, which are: random forest classifier (RFC) model, support vector machine (SVM), eXtreme gradient boosting (XGBoost), artificial neural network (ANN), and decision tree (DT). Among them, RFC model has the best prediction efficiency, and the author also gives the corresponding candidate variables. In my opinion, the author should list clearly in Table 2, which

variables are used in each model, so as to show more clearly what variables are used in GLCM model.

Response: Thank you for your suggestions. We have listed clearly in Table 2, which variables are used in each model, so as to show more clearly what variables are used in GLCM model. Please refer to the modified Table 2.

2) In this study, the author emphasizes that the model obtained by machine learning algorithm is better than the traditional model, which has been demonstrated in this paper. It is suggested that the author add the relevant content of the traditional model (such as logistics regression model, and visualize it by means of nomograph) in this paper, which can more fully show the advantages and disadvantages between machine learning and traditional model (logistics model).

Response: Thank you for your suggestions. In order to better show the machine learning algorithm and the traditional logistic regression algorithm, we also calculated the prediction model constructed by logistic regression in the training set and the verification set, and the results are also shown in Table 2 and supplementary Table 2. At the same time, we also made corresponding explanations in the results section, as follows:

In general, the prediction model constructed by any machine learning algorithm was better than the logistic regression algorithm in predicting LNM, which further confirmed the superiority of machine learning algorithm, especially the robustness of RFC.

3) In this study, the author used CIC to visualize the optimal model, and found that the differentiation efficiency of RFC model is very good, which is very encouraging. In my opinion, the author should use other models, such as decision tree, neural network, support vector machine, gradient lifting algorithm, etc. the CIC results of these models can be presented in the attached drawings.

Response: Thank you for your suggestions. In this study, we only show the CIC results of RFC. In order to better highlight the advantages of RFC, we also show the CIC results of the other four machine learning algorithm models. Please see the corrected Supplementary Figure 1 for details.

4) In this study, some words that need to be corrected are expressed as follows, for example: Line139, Therefore, in order to solve this thorny problem, exploring an robust tool that can

predict LNM is necessary. Maybe “precise tool” is better. Line166, Two radiologists with rich experience in gastrointestinal imaging diagnosis referred to T2WI and DWI images to sketch the lesions on the ADC map respectively. In my opinion, the efficiency of artificial recognition should be compared with machine learning model, although artificial recognition may not be as good as machine learning algorithm. Similarly, the prediction efficiency of logistics model should also be shown in Table 2.

Response: We have replaced "robust" with "precise".(line 144)

In addition, we also compared the prediction model constructed by the machine learning algorithm with the prediction ability of imaging experts. See Table 2 after correction.

We tried our best to improve the manuscript and made some changes in the manuscript. These changes will not influence the content and framework of the paper. We appreciate for Editors/Reviewers’ warm work earnestly, and hope that the correction will meet with approval.

Once again, thank you very much for your comments and suggestions.