**Reviewer A**

Comment 1. It would be better for figuring out that combining Table 1a & 1b into a one table.
(Left column - development data set, Right column - validation set.)

Reply 1: We have combined Table 1a & 1b into a one table as advised.
Changes in the text: page 26, line 543-545.

Comment 2. Please explain the abbreviations such as AIS, MIA and IAC below Table 1.

Reply 2: We have explained the abbreviations of AIS, MIA and IAC below Table 1 as advised.
Changes in the text: page 26, line 545.

Comment 3. Pleas add the exact unit corresponding to CEA, CA 19-9, CT value, et al. in Table

Reply 3: We have added the unit corresponding to serum biomarker and CT value as advised.
Changes in the text: page 27, line 556

Comment 4. In table 2, there is a spelling error - "Vocule sign", please correct it.
Reply 4: We have corrected it as "Vacuole sign" as advised.
Changes in the text: page 27, line 556.

Comment 5. Please describe CT characteristics in detail, especially about what "abnormal vessel" means. It would be nicer if you can add a typical CT image of patients.

Reply 5: We have described the CT characteristics of "abnormal vessel" and add a CT image of "vessel curve" in the article as advised.

Changes in the text: Page9, line 189-190; Page 30, line 574

Comment 6. The size, CT value, lobulation, spiculation, pleural retraction, and air-bronchogram, which are the factors of prediction model in your paper, are previously known as risk factors for adenocarcinoma. (AJR 2020;215:351, European Radiology 2020;30:3650). You have to add the differences with previous reports in discussion.

Reply 6: We have added the difference of our study with previous reports: "The difference of our study with previous studies were that we combined the risk factors together in one prediction model, aims to increase CT diagnosis accuracy of IAC." In our data, the combination model has a higher diagnosis AUC than diameter or CT value alone.

Changes in the text: page17, line 364-366;

**Reviewer B**

This paper compares the imaging and pathological findings of 344 pure GGNs resected during one year, creates a prediction model for IAC, and applies it to 345 pure GGNs noted during a half year at four institutions to evaluate the accuracy of the prediction model. The sample size was surprisingly large. This report is important for clarifying the nature of pure GGN lesions and should be made public.

Of the 1579 lung cancer surgeries, 304 were pure GGN, accounting for 20%. Of these, about 40% were diagnosed as IAC. 75% of the tumors larger than 10 mm, and 2 cases with tumors smaller than 6 mm were 75% were 10 mm or larger, and even 2 cases of less than 6 mm were reported to be IAC. This is an amazing rate. In our own study, about 8% of lung cancer resections were pure GGN, and 4% were diagnosed as IAC despite an average tumor diameter of 14mm. In the past literature, reports from Chinese institutions tended to have a high percentage of IAC, but the result of this study is particularly evident. Therefore, the prediction model for IAC seems to lack international applicability.

The reasons for the gap from international standards seem to be as follows:

1. The definition of GGN is different, and solid tumors are included in pure GGN.

2. The criteria for pathological diagnosis are different.

It is more likely that the criteria for pathological diagnosis are different, because the maximum CT value is -416HU even for IAC, which can be considered as a pale ground-glass shadow.

I have some comments:

Comment 1. Please present the imaging and pathological findings of a typical case.

Reply 1: We have added CT and pathological image of a typical IAC case of a pGGN ≥8 mm in the article as advised.

Changes in the text: Page14, line 297-298; Page 30, line 574

Comment 2. Please clarify the criteria for indication for surgery. More than half of the patients had pure GGNs of less than 10 mm. Such lesions usually do not require therapeutic intervention. Please describe whether the trend of increase is confirmed by follow-up observation and what percentage of pure GGNs found are operated on.

Reply 2: That is true trend of overtreatment and unnecessary operation of pGGNs in China. It is a pity that we have no accurate data of the percentage of pure GGNs found are operated on. However, one main goal of this article is to avoid the unnecessary therapy of pGGNs of AIS and MIA by increase the diagnosis accuracy of IAC.

Changes in the text: page 15, line 318-321; page 17, line 373-374.

Comment 3. What percentage of cases diagnosed as IAC were really highly malignant tumors?

How many had lymphovascular invasion, pleural invasion, or nodal metastasis? Please describe the surgical procedure for IAC and whether there have been any recurrent cases.

Reply 3: The IAC of pGGNs are mostly not highly malignant tumors. There was no nodal metastasis, only 1 case of lymphovascular invasion and 11 cases of pleural invasion (PL1) found in the 133 IAC cases in the development dataset.

Changes in the text: Page 11, line 238-240

Comment 4. In the abstract, the authors said "all 344 pGGNs were pathologically confirmed to be AIS, MIA or IAC", but actually 344 lesions out of 370 pGGNs were pathologically confirmed to be lung cancer.

Reply 4: We are sorry for the description and we have corrected it as "344 of all the 370 resected pGGNs were eventually diagnosed as AIS, MIA or IAC" as advised.

Changes in the text: Page 3, line 55-57

Comment 5. You define tumor size as the average of the long and short diameters, but in the TNM classification, the long diameter is defined as the total tumor size. Why did you change it?

Reply 5: We measure the size of the nodule by the average of the long and short diameters, according to the suggestion of NCCN lung cancer screening guideline. We believe this method may be suitable for some irregularity-shape nodule.

Changes in the text: None.

Comment 6. Table 1a is completely duplicated in Table 2 and is unnecessary.

Reply 6: Thanks for the comment. We have deleted the duplicated part of Table 1 in Table 2 as advised.

Changes in the text: Page 27, line 556

Comment 7. Reference is not included.

Reply 7: We have added reference in the article.

Changes in the text: Page 20-24, line 425-522

**Reviewer C**

The authors review a set of 344 surgically resected pure GGN, looking at imaging features predictive of invasive adenocarcinoma. They constructed a model using multivariable logistic regression with high predictive power (AUC 0.910), confirmed using an independent validation dataset.

Comment 1. Overall, I think this work is well done and will contribute to the literature, and perhaps even clinical practice. The biggest flaw is patient/nodule selection. The authors need to include all of the benign surgically resected nodules in their data sets. More broadly, this analysis is limited by selection bias in which patients (nodules) were actually taken to surgery; this should be mentioned as a limitation. Additional comments below.

Reply 1: Thank you.

We don't aim to discriminate benign from malignant pGGNs. We aim to discriminate IAC in malignant pGGNs, so, we exclude the benign nodules in the development dataset.

The selection bias has been added in the limitation parapraph.

Changes in the text: Page 12, line 375-381

Specific comments:

Comment 2. Abstract

-Please be more specific about the morphologic characteristics you recorded

Reply 2: We have added the morphologic characteristics we recorded in abstract as advised.

Changes in the text: Page 3, line 60-61

Comment 3. Please clarify that the inclusion criteria were patients with lung cancers that presented as GGN, not all GGN or even all resected GGN

Reply 3: We have clarified the criteria were patients with lung cancers that present as

pGGNs. Benign pGGNs were excluded as advised.

Changes in the text: Page 3, line 55-57

Comment 4: Remove the word "besides"

Reply 4: We have removed the word "besides" as advised.

Changes in the text: Page 3, line 74

Comment 5. Please define "vacuole sign" and "vessel abnormality"

Reply 5: We have defined "vessel abnormality" as "vessel curve" or "vessel enlargement" and add typical pictures in Figure 3 as advised.

Changes in the text: Page 3, line 62-63; Page 30, line 574

Comment 6. You say "significantly higher" - was a statistical test performed, otherwise remove that word

Reply 6: We have added the statistical test of $p<0.05$ of the compared the diagnosis ability different AUC curve as advised.

Changes in the text: Page 4, line 80-81,84

Comment 7. Your conclusion is a bit misleading, since all of your patients underwent surgery... we do not know how this model would apply to nodules that did not go to surgical resection in your institution

Reply 7: Since most (about 90%) of the persistent lung pGGNs were early-stage lung cancer, our goal doesn't aim to discriminate benign from malignant pGGNs. We aim to discriminate IAC in malignant pGGNs, our conclusion is that those pGGNs evaluated as AIS/MIA may be followed up and those pGGNs evaluated as IAC may be considered resected.

Changes in the text: Page 4, line 88

Introduction

Comment 8. Needs substantial work improving the grammar/clarity of writing

Reply 8: We will be round to get the help of language polishing assist to improve the

writing. We will do it later since the time for revise the manuscript was only 3 weeks.

Comment 9. The references are not included!

Reply 9: Sorry for the miss, we have added the reference.

Changes in the text: Page 20-24, line 425-522

Comment 10. Note that even invasive adenocarcinoma with ground glass component has a very high 5-year overall survival

Reply 10: It is absolutely true that even IAC with ground glass have a very good prognosis, we have added it in the article as advised.

Changes in the text: Page 6, line 127-128

Methods

Comment 11. What were the diagnoses for the remaining 370-344 GGN?

Reply 11: The rest 26 pGGNs were mainly infection disease. The benign cases were excluded in the article.

Changes in the text: none

Comment 12. Smoking status should be collected and tested in the model

Reply 12: In our data, few patients (only about 5%) were smokers. More and more non-smoker lung cancer identified with lung pGGNs was a trend in Asia. So, we didn't test the smoking status as a risk factor if IAC in the article.

Changes in the text: none

Comment 13. Again, need to clarify the morphologic features. please use standard Fleischner terminology

Reply 13: Thank you, but we didn't find the standard Fleischner terminology of lung nodule in the article "Fleischner Society: Glossary of Terms for Thoracic Imaging, 2008, Radiology. Only one term "cavity" was similarly to the "vacuole sign" of our article. But we think "vacuole sign" was smaller than "cavity" in some cases. So, we didn't change it.

Changes in the text: None

Results

Comment 14. What does "as high as 38.7%" mean?

Reply 14: Sorry for the indistinct description, we have modified it and delete "as high as" as advised.

Changes in the text: Page 11, line 237

Comment 15. The results of morphologic features should include percentages in AIS/MIA as well as those given for invasive adenocarcinoma

Reply 15: The results of morphologic features in AIS/MIA were shown in Table2. We mainly focus on the difference of IAC with AIS/MIA, so we didn't show it in the text.

Changes in the text: None

Comment 16. The authors compare AUC values but do not perform a statistical test. Please report p-values for comparison of the AUCs

Reply 16: There is a statistical difference of the prediction model. We have added it in the text as advised.

Changes in the text: Page 41, line 301; Page 15, line 313

Discussion

Comment 17. Limitations paragraph is needed

Reply 17: Thanks for in remind. We have added limitations paragraph as advised.

Changes in the text: Page 19, line 375-381

Figures

Comment 18. Fig 2 - what are the error bars? Why do they appear to cross zero despite significant p-value?

Reply 18: That is because the nodule size was the main risk factor for IAC and the HR of nodule size is 11.49 in the picture. It makes the HR of other risk factor seems not much obvious, but they were still with statistics difference with HRs range from 1.898-3.282.

Changes in the text: None

Comment 19. Fig 3 - please make the ROC graphs square

Reply 19: Thanks for the remind, we have modified this picture as advised.

Changes in the text: Page 31, line 589-590

**Reviewer D**

This study tries to address multiple parameters including CT characters and tumor makers in prediction of invasion adenocarcinoma lesion.

Comment 1.

Introduction:

There is increasing trends of non-smoking related lung adenocarcinoma spectrums lesions in Asian population. please try to address this point with associated references here.

--- Assessment of selection criteria for low-dose lung screening CT among Asian ethnic groups in Taiwan: from mass screening to specific risk-based screening for non-smoker lung cancer, Clinical lung cancer 2016.

--- Propensity score analysis of lung cancer risk in a population with high prevalence of non-smoking related lung cancer. BMC pulmonary medicine, 2017.

--- Prognostic effect of implementation of the mass low-dose computed tomography lung cancer screening program: a hospital-based cohort study, European Journal of Cancer Prevention, 2020

--- Natural history of persistent pulmonary subsolid nodules: long-term observation of different interval growth, Hear, lung and circulation 2019

Reply 1: We have added these references and introduced the trend of non-smoking related lung adenocarcinoma which was detected as GGNs in CT screening as advised.

Changes in the text: Page 6, line 118-120

Comment 2.

Method:

Please try to address more clearly about statistical method of training and validation

model developing and model calibration, statistical method.

Please add prediction nomogram for model prediction.

Reply 2: We have addressed about the statistical method of training and validation model developing and model calibration or statistical method as advised.

Changes in the text: Page 13-15, line 283-289; 309-310

Comment 3.

Result and discussion

Please try to address the diagnostic performance of developed model's sensitivity and specificity for IPA prediction. In addition, previous studies showed that In pure and heterogeneous GGNs, the lesion size with cut-of value of $\geqq 10$ mm, the type of heterogeneous GGN, and the higher HU value with cut-of value of $\geqq -493$ were the optimal diagnostic threshold for IPA lesions prediction with high NPV, which could help to rule out IPAs. Please try to discuss the diagnostic performance between the previous literature and the current result.

--- Reference: Semiquantative Visual Assessment of Sub-solid Pulmonary Nodules $\leqq 3$ cm in Differentiation of Lung Adenocarcinoma Spectrum, Scientific report 2017.

Reply 3: We have added this reference in the article as advised. The IAC threshold in this reference was very similar to our study. In our data, multi-factor prediction model may have a higher AUC than nodule size or CT value alone.

Changes in the text: Page 17, line 364-366

Comment 4. Recent studies have demonstrated the radiomic feature have superior diagnostic performance over the conventional features, please try to discuss more about clinical feature, radiomic and tumor maker in prediction of early adenocarcinomas spectrum lesions.

Reference: A comparative study to evaluate CT-based semantic and radiomic features in preoperative diagnosis of invasive pulmonary adenocarcinomas manifesting as subsolid nodules, Scientific report 2021

Reply 4: We have added this reference in the article and discuss the advantage of radiomic in the clinical use as advised.

Changes in the text: Page 7, line 134-136


Comment 5.

Grammatical errors:

Page 11 -Corrected grammatical mistake in the "the high risk for IAC" to "the high risk for IAC." (line 12)

Reply 5: We have corrected the grammatical mistake as advised.

Changes in the text: Page 16, line 349


Comment 6.

Page 12 -Corrected grammatical mistake in the "subjectivity of radiologist" to "subjectivity of the radiologist" (line 3)


Reply 6: We have corrected the grammatical mistake as advised.

Changes in the text: Page 18, line 378


**Reviewer E**


This paper tried to establish a discriminantion model to differentiate AIS/MIA from IA based on thoracic CT images. The proposed model was composed of nodule size, max CT value, lobulation, spiculation, pleura indentation, vacuole sign, and vessel abnormality. While the presented prediction performance is reasonably high, two fundamental problems requires further justification before this model can be considered practically useful.

Comment 1. Sample distribution problem:

About three quarters of the nodules >= 10mm were IACs. This sample distribution made the nodule size along achieve an AUC of 0.891. This sample distribution might cause a serious bias of the proposed model. In our experience with a sample of more than 300 pGGNs, the percentage of the nodules >= 10mm was about 40%. I am not saying which kind of distribution is correct. My concern is the prediction model clearly depends on the sample distribution and the high prediction performance is strongly related to the sample distribution of the two data sets used in this study.

Reply 1: It is true that the IAC distribution were relatively high in the nodules >= 10mm in our study, compared with other studies. So, we choose the cut off of nodule size was 8mm in our prediction model, not 10mm. In our development dataset, the total IAC rate of the nodules >= 8mm was 59%, which was lower than the nodules >= 10mm. This may reduce the influence of the distribution bias.

Changes in the text: None

Comment 2. Annotation of lobulation, spiculation, pleura indentation, vacuole sign, and vessel abnormality.

While the nodule size is small, annotation of nodule morphology characteristic is quite subjective. The inter-observers' variation is expected to be high. It is suggested that the inter-observers' difference is provided for each of these five characteristics in each of three size groups. Furthermore, it is suggested that the influence of the inter-observers' difference on the prediction performance be analyzed.

Reply 2:

It is really important to take into consideration of the inter-observers' variation in the CT image reading. So, in the development dataset, we choose two attending doctors of Zhongshan Hospital to read imaging together, and the disagreements were discussed until reaching consensus. But unfortunately, we didn't record the different opinions of the two radiologists in the five characteristics of CT image, so we can't analyze the inter-observers' difference now in the development dataset.

However, we have partly tested the influence of inter-observers' difference in the validation dataset. Though the CT image reading of the validation dataset were finished by four different doctors of four hospitals, AUC in the validation dataset was only slightly lower than the development dataset. This result may partly show the stability of the prediction model.

Changes in the text: None

**Reviewer F**

The authors are very unique in that they examined only pure GGN excluding part-solid nodule and solid nodule among lung nodules, and their results has great clinical

significance for clinicians, who work to detect and treat lung cancers at an early stage. In particular, the fact that 39% of the analyzed pure GGN was IAC is shocking, because it was reported that the conventional pure GGNs are indolent type adenocarcinoma, and this article disagrees with the hypothesis of the radiographic and pathologic correlation, and the hypothesis of Over-diagnosis on pure GGN. In addition, using the Logistic regression model to create an IAC prediction model that is useful for early decision-making at younger age is extremely epoch-making and can be said to be suitable for publishing. In the future, I would like to wait for a follow-up report on the relationship with smoking volume and survival analysis. Thank you.

Please correct the following points:

Comment 1.

# Because there is no reference in the downloaded PDF_file, add it. Similarly, there is no Excel (supplemental material 1) file, so please add it.

Reply 1: Thank you. We have added the reference in the revised version. The IAC calculate excel template has been uploaded to the manuscript system.

Changes in the text: Page 20-24, line 425-522

Comment 2.

# About serum tumor marker, cyfra 211 → cyfra 21-1, CA199 → CA 19-9

Reply 2: Thanks for the correction, we have corrected it as advised.

Changes in the text: Page 8, line 174-176; Page 12, line 252; Page 16, line 332; Page 27, line 556;

**Reviewer G**

The authors developed and validated the prediction model for distinguishing IAC from AIS-MIA in pure GGNs. The manuscript is well written overall, and the study design is carefully constructed. I have some comments as below.

Comments

Comment 1. Abstract: Please describe the study purpose in the Background.

Reply 1: Thanks, we have added the study purpose in the Background as advised.

Changes in the text: Page3, line 51-52

Comment 2. Methods, Study design, line 4-5: The authors said "The IAC prediction model was successfully developed according to the training dataset." I think this sentence is study results, not method, and should be move into the Results section.

Reply 2: Thanks for the suggestion, we have removed it into results section as advised.

Changes in the text: Page7, line 147

Comment 3. Methods, P7, line 3: How to measure the maximal CT values of the tumor with vessels excluded? How did you place the ROI on the tumor? Did you use software? Please explain how to interpret CT images by adding an example of figure so that journal readers can easily understand.

Reply 3: We use ROI on the tumor to measure the maximal CT values of the tumor manually. We have added one typical measure method in Figure 3.

Changes in the text: Page 14, line 297-298; Page 30, line 574-581

Comment 4. Methods, P7, line 6: What is the "vessel abnormality"? Please describe the definition of this CT finding.

Reply 4: Vessel abnormality include "vessel curve" and "vessel enlargement" in our studies. We have added a typical CT image in the article as advised.

Changes in the text: Page3, line 62-63; Page 30, line 574-581

Comment 5. Were CT images enhanced or non-enhanced?

Reply 5: CT images were non-enhanced.

Changes in the text: Page 9, line 181

Comment 6. How about interobserver agreement about image evaluation?

Reply 6:

It is really important to take into consideration of the interobservers' variation in the CT image reading. Thus, in the development dataset, we choose two attending doctors of Zhongshan Hospital to read imaging together, and the disagreements were discussed

until reaching consensus. But unfortunately, we didn't record the different opinions of the two radiologists, so we can't analyze the inter-observers' difference now in the development dataset.

However, we have partly tested the influence of inter-observers' difference in the validation dataset. Though the CT image reading in the validation dataset were finished by different doctors of four hospitals, AUC in the validation dataset was only slightly lower than the development dataset. This result may partly show the stability of the prediction model.

Changes in the text: none

Comment 7. Please describe CT scanners used in this study.

Reply 7: We have added the CT scanner model in the method part as advised.

Changes in the text: Page 9, line 182-183

Comment 8. Methods, Diagnosis of AIS, MIA and IAC, line 7: "MIA often shows an invasive extent of <0.5 cm." Is the word "often" necessary?

Reply 8: Thank you, the word "often" was indeed unnecessary, we have deleted it as advised.

Changes in the text: Page 10, line 204

Comment 9. Results, Validation of the IAC prediction model, line 4-6: How about a sensitivity and a specificity when cut off values set to IAC possibility of 0.5?

Reply 9: In the validation dataset, the sensitivity and a specificity were 70.8% and 86.2% when cut off values set to IAC possibility of 0.5.

Changes in the text: None

Comment 10. Please add examples of CT images of tumors that showed high and low possibilities of IAC, respectively.

Reply 10: We have added CT images of two typical pGGNs of one tumor of IAC with high calculated IAC possibility of 0.48 and another tumor of AIS with low calculated IAC possibility of 0.02.

Changes in the text: Page 30, line 574-581

Comment 11. Discussion: The authors may want to write more about study limitation.

Reply 11: We have added one paragraph of limitation as advised.

Changes in the text: Page 18, line 375-382


Comment 12. I cannot find reference section. Please add this.

Reply 12: We have added reference section.

Changes in the text: Page 20-24, line 425-522


Comment 13. Figure 2a: I think the term, "Hazard ration" should be changed into odds ration because the authors used logistic regression model, not Cox regression model.

Reply 13: Thank for the correction. We have amended it in the picture as advised.

Changes in the text: Page 29, line 557