18-Aug-2020

56193 (Research Article)

Construction of a Convolutional Neural Network Classifier developed by CT images for Pancreatic Cancer Diagnosis

Dear Prof. Lian-Sheng Ma

Thank you very much for your email dated 1 May 2020, and the valuable comments of the three referees and editors. Based on the comments, we have made extensive modification on the original manuscript. Here, we attached a copy of the revised manuscript with all changes HIGHLIGHTED. A list of answering every question from the referees was enclosed. We hope that the revised manuscript is acceptable for publication. We are happy to answer any further questions.

Thank you!

Yours sincerely,

Chaohui Yu, MD

Department of Gastroenterology

The First Affiliated Hospital, College of Medicine, Zhejiang University

79 Qinchun Road, Hangzhou 310003, China

Tel: +86-13957161659; Fax: +86-571-87236611

E-mail: zyyyych@zju.edu.cn

Replies to Editors' comment

Comment 1: ***Scientific quality****: Please resolve all issues in the manuscript based on the peer review report and make a point-to-point response to the issues raised in the peer review report.*

Reply: Thank you very much for your valuable comments and suggestions to improve our manuscript! The comments and suggestions by three reviewers are replied point-by-point below in the Section "**Replies to Reviewer**".

Comment 2: ***Language quality:*** *Please resolve all language issues in the manuscript based on the peer review report.*

Reply: Thank you very much for this suggestion! We have polished the language in the revised manuscript by inviting a native speaker. If language qualify still didn't match the criteria, we are happy to edit language by the suggested organization again.

Comment 3: ***Special requirements for figures:*** *Figures must be presented in the order that they appear in the main text of the manuscript (numbered as 1, 2, 3, etc.). The requirements for the figures and figure legends include: (A) All submitted figures, including the text contained within the figures, must be editable. Please provide the text in your figure(s) in text boxes; (B) For line drawings that were automatically generated with software, please provide the labels/values of the ordinate and abscissa in text boxes; (C) Please prepare and arrange the figures using PowerPoint to ensure that all graphs or text portions can be reprocessed by the editor; and (D) In consideration of color-blind readers, please avoid using red and green for contrast in vector graphics or images.*

Reply: Thank you very much for this suggestion! We have modified the figures in the PowerPoint to meet the requirements.

Comment 4: ***Special requirements for tables****: Tables must be presented in the order that they appear in the main text of the manuscript (numbered as 1, 2, 3,etc.). Please verify that the tables are referred to in the text by their respective Roman numerals and that the numbering order is correct, and format the tables. Please verify that there are no missing or multiple spaces in the text and tables, e.g. before or after parentheses, between words, or before or after symbols like +, ×, ±, <, >, ≥, and ≤. Please verify that the special words or letters in the text and tables are correct, e.g. P(uppercase),n(lowercase),via, vs (lowercase, no punctuation), in vivo, in vitro, and et al (no punctuation) are italicized.*

Reply: Thank you very much for this suggestion! We have checked tables. In addition, arterial phase images are taken before venous phase in a CT scan, logically. Thus, according to Reviewer 2's comment "the venous and arterial phase are swapped and incorrectly placed in Figure 1", we have made the correction in the Tables, too.

Page 33 , Table 3 & Page 35, Table 5: Columns "Vein Phase" and "Artery phase" are replaced by "Arterial phase"," Venous Phase".

Comment 5: ***Special requirements for references*****:** *Please provide the PubMed numbers and DOI citation numbers to the reference list and list all authors of the references. Please revise throughout. The author should provide the first page of the paper without PMID and DOI numbers. NOTE: The PMID is required, and NOT the PMCID; the PMID number can be found at https://pubmed.ncbi.nlm.nih.gov. (Please begin with PMID:) The DOI number can be found at http://www.crossref.org/SimpleTextQuery/. (Please begin with DOI: 10.**).*

Reply: Thank you very much for this suggestion! We have added PMID and DOI in the Reference list. We have attached the first page of the paper without PMID or DOI numbers, including Reference 3 (Getting from website), 14/15/16/29 (Conference paper).

Comment 6: Please verify that the references are cited by Arabic numerals in square brackets and superscripted in the text, and that the numbering order is correct. There should be no space between the bracket and the preceding word or the following punctuation. When references in the text and tables are cited with author name(s), it is necessary to manually verify that the name(s) is consistent with the first author's surname in the corresponding reference list.

Reply: Thank you very much for this suggestion! We have checked the references as requested.

Comment 7: ***Special requirements for article highlights:*** *If your manuscript is an original study (basic study or clinical study), meta-analysis, systemic review, the "article highlights" section should be provided. Detailed writing requirements for "article highlights" can be found in the Guidelines and Requirements for Manuscript Revision.*

Reply: Thank you very much for this suggestion! We have added article highlights in the revised manuscript.

*Comment 8:****Ethical documents:****Please double check the accuracy of all ethical documents and verify the completeness of the documents according to the type of manuscript.*

Reply: Thank you very much for this suggestion! We have double checked.

*Comment 9:****Approved grant application form(s) or funding agency copy of any approval document(s):****If your manuscript has supportive foundations, the approved grant application form(s) or funding agency copy of any approval document(s) must be provided.*

Reply: Thank you very much for this suggestion! We have attached funding agency copy of any approval documents.

Replies to Reviewer 1

Comments to the Author

Authors developed CNN which detects pancreatic cancer. It is important and interesting, but there are several concerns to be raised.

Comment 1: Not only verbal informed consent, but explanation by literature is desired, like opt-out method.

Reply: Thank you very much for this valuable comment and suggestion!

We have added more detailed explanation about obtaining consent in the study in the revised manuscript.

Page 9, line 58-60, "Because of the retrospective study design, we verbally informed all the participants included in the study, patients who do not want their information to be shared could do so through the opt-out method."

Comment 2: As authors describe in the limitation part, only pancreatic cancer and normal pancreas were included. What about other tumors like IPMN?

Reply: Thank you very much for this suggestion! We agree that IPMN is an important pancreatic neoplasm. In the future, we will investigate the performance of our deep learning models on detecting IPMN.

Page 21, line 379-382: "…and other neoplastic lesions , e.g., intraductal papillary mucinous neoplasm (IPMN). In the future, we plan to investigate the performance of our deep learning models on detecting these disease."

Replies to Reviewer 2

Comments to the Author

Comment 1:Title. Okay. Abstract section has provided the results of binary classifier and ternary classifier. However, the conclusion is not related or does not reflect their purpose and results. Keywords. Okay. Background. Okay.

Reply: Thank you very much for this valuable comment and suggestion! Due to words in conclusion part were limited to less than 30, we have added several sentences to emphasize the important finding in the Sections "Abstract" and "Discussion".

Page 5, Section "Abstract- Conclusion": "We proposed a deep learning-based pancreatic cancer classifier trained on medium-sized datasets of CT images. It was suitable for screening purposes in general medical practice. "

Page 18, line 289-297: "In this study, we developed an efficient pancreatic ductal adenocarcinoma classifier using a CNN trained on medium-sized datasets of CT images. We evaluated our approach on the datasets in terms of both binary and ternary classifications, with the purposes of detecting and localizing mass, respectively. In the binary classifiers, the performance of plain, arterial and venous phase had no difference, its accuracy on plain scan achieved 95.47%, sensitivity achieved 91.58%, and specificity achieved 98.27%. In the ternary classifier, the arterial phase had the highest sensitivity in detecting cancer in the head of the pancreas among three phases, but it achieved only moderate performances. "

Page 21, line 394-399: "In conclusion, we developed a deep learning-based, computer-aided pancreatic ductal adenocarcinoma classifier trained on medium-sized CT images. The binary classifier may be suitable for disease detection in general medical practice. The ternary classifier could be adopted to localize the mass, with moderate performance. Further improvement in the performance of models would be required before it could be integrated into a clinical strategy."

Comment 2: Methods. The description in this section is different from the description found in the "abstract". The authors did not provide the numbers of patients and CT images they used in this study. However, they provided the numbers in the "results" discordant from description in the "abstract".

Reply: We apologize for the confusion! We have made the corrections in the revised manuscript.

Page 9, line 63-70: "A total of 343 patients were newly pathologically diagnosed with pancreatic cancer from June 2017 to June 2018. Of the 343 subjects, 222 underwent an enhanced-CT abdomen in our hospital before surgery or biopsy. We randomly collected 190 patients who underwent abdomen enhanced-CT with normal pancreas. Thus, among the 412 enrolled subjects, 222 were pathologically diagnosed with pancreatic cancer, and the remaining 190 diagnosed as normal were included as a control group. "

Page 10, line 79-82: "Finally, datasets of 3,494 CT images obtained from 222 patients with pathological confirmed pancreatic cancer and 3,751 CT images from 190 patients with normal pancreas were collected."

Comment 3: How many of them disagree to give verbal consent?

Reply: Thank you very much for this suggestion! We have made the correction in the revised manuscript. Patients who do not want their information to be shared could do so through the opt-out method. Finally, a total of 343 patients were newly pathologically diagnosed with pancreatic cancer from June 2017 to June 2018. Of the 343 subjects, 222 underwent an enhanced-CT abdomen in our hospital before surgery or biopsy and agreed their data to be shared.

Page 9, line 58-60: Because of the retrospective study design, we verbally informed all the participants included in the study, patients who do not want their information to be shared could do so through the opt-out method.

Comment 4: Did their images obtained after biopsy if they are proven cancer?

Reply: We apologize for the confusion! We have made the correction in the revised manuscript.

Page 9, line 65-66: Of the 343 subjects, 222 underwent an enhanced-CT abdomen in our hospital before surgery or biopsy.

Comment 5: The authors should tell their readers about the sizes of tumors and the time of CT acquisition (arterial phase and venous phase) after contrast medium administration as well as the total volume of contrast medium and the rate of injection.

Reply: Thank you very much for this suggestion! We have added the tumor size and imaging technique in the revised manuscript.

Page 15, line 220-221: The median tumor size of cancer group were 3.5 cm (range, 2.7-4.3 cm).

Page 32, Table 2: "Tumor size (Median[quartile 1, quartile 3]), cm" ,"Tumor size ≤ 2"," 2 <Tumor size ≤ 4", "Tumor size> 4".

Page 9-10, line 73-78: Multiphasic CT was performed by following a pancreas protocol and using a 256-channel multidetector row CT scanner (Siemens). The scanning protocol included unenhanced and contrast material–enhanced biphasic imaging in the arterial and venous phases after intravenous administration of 100 mL of ioversol at a rate of 3 mL/sec by using an automated power injector. Images were reconstructed at 5.0-mm thickness.

Comment 6: Results. Under subheading "2", ", ….CNN model on 1,702, 2,058, and 2,037 test images in….. CNN using 186, 250, and 285 test images in,,,,". It may cause confusion because the former images are training images, while latter images are testing images.

Reply: We apologize for the confusion! The first "test images" mentioned in the sentence should be replaced by "training images". As the reproducibility

is not well supported in our study, resampling the train/validation/test split and averaging over a number of runs would definitely improve the robustness of the study. Instead of one set of "the train/validation/test", we adopt a 10-fold cross validation on our dataset and updated the evaluation results by averaging over the 10 runs in the revised manuscript. Thus, we delete the sentence described the specific number of train/validation/test images in the revised manuscript.

Page 15: ", ….CNN model on 1,702, 2,058, and 2,037 test images in….. CNN using 186, 250, and 285 test images in…" was deleted.

Comment 7: Because their results demonstrated that plain CT was sufficient for binary classifier, I think they have to provide data regarding the tumor sizes.

Reply: Thank you very much for this suggestion! We have add the tumor size in the revised manuscript.

Page 15, line 220-221: The median tumor size of cancer group were 3.5 cm (range, 2.7-4.3 cm).

Page 32, Table 2: "Tumor size (Median [quartile 1, quartile 3]), cm" ,"Tumor size ≤ 2"," 2 <Tumor size ≤ 4", "Tumor size> 4".

Comment 8: Discussion. The results regarding plain CT was sufficient for screening and the high diagnostic values of their CT are very interesting. The authors should explain and elaborate it in this section.

Reply: Thank you very much for this comment. According to Review 3's suggestion, we adopt a 10-fold cross validation on our dataset and updated the evaluation results by averaging over the 10 runs in the revised manuscript. The result has been a slightly changed. The sensitivity of the binary classifier was 91.58%, 94.08%, 92.28% the plain scan, arterial phase, and venous phase, respectively, with no significant differences. Compared with arterial and venous phase, plain phase had relatively same sensitivity, easier access, lower radiation. Thus, our results indicated that the plain scan alone was sufficient

for the binary classifier. We have made the correction and further explanation in the revised manuscript on Sections "Results" and "Discussion".

Page 16, line 243-249:"The difference in accuracy, specificity and sensitivity among the three phases were not significant ($\chi2$=0.346, *P*=0.841; $\chi2$=0.149, *P*=0.928; $\chi2$=0.914, *P*=0.633, respectively). Sensitivity of the model is considerably more important than its specificity and accuracy, because the purpose of a CT scan is cancer detection. Compared with arterial and venous phase, plain phase had relatively same sensitivity, easier access, lower radiation. Thus, these results indicated that the plain scan alone might be sufficient for the binary classifier. "

Page 18-19, line 316-325: "all three phases had high levels of accuracy and sensitivity, with no significant differences among three phases. This indicates potential ability of plain scan in tumor screening. Relatively same performance of sensitivity on plain phase can be explained by the size of tumor in our study and redundant information given by arterial or venous phase. In the current study, most tumors were larger than two centimeters, allowing plain scan easier to assess tumor morphology and size. In addition, there are less noisy and unrelated information in the images of the plain scan phase. Thus, it is relatively easy for our CNN model to distill pancreatic-cancer-related features from such images."

Comment 9: It is also interesting to know that the sensitivity of pancreatic head/neck cancer is very much higher at arterial phase in their study. Wouldn't the unopacified SMV at arterial phase cause confusion in detecting pancreatic head/neck tumor?

Reply: Thank you very much for this valuable comment. After adopting 10-fold cross validation on our dataset the result of ternary classifier also has been a slightly changed in the revised manuscript. Difference in the sensitivity scores of cancers in the head among the three phases was significant; the arterial phase had the highest sensitivity. However, difference in sensitivity in cancers in the tail among the three phases was not significant.

We have add the updated results and possible explanation for the interference of unopacified SMV in the revised manuscript.

Page 17, line 282-286:"The difference in the sensitivity scores of cancers in the head among the three phases was significant (χ^2=16,651, P<0.001), with the arterial phase had the highest sensitivity. However, difference in sensitivity in cancers in the tail among the three phases was not significant (χ^2=1.841, P=0.398)."

Page 19-20, line 346-359: "It is worth noting that unopacified superior mesenteric vein (SMV) at arterial phase may cause confusion in tumor detection. However, SMV has a relatively fixed position in CT image, accompanied by an artery of the same name, which may help the classifier distinguish it from tumor. Further studies in pancreatic segmentation should be imported to solve this problem. "

Comment 10: Do they mean that results of artificial intelligence are better than the results as published by other reports? What are the significances of their research? Can we use AI to help us detect and manage pancreatic cancer?

Reply: Thank you very much for this valuable comment and suggestion! Our study developed a deep-learning-aided decision support for pancreatic ductal adenocarcinoma diagnosing, which can help doctors make the decision, reduce their workload, but cannot replace the function of them. So far, many CNN applications to evaluate organs have been reported, we have added their work in the revised manuscript.

Page 20-21, line 363-374: "A CNN also has potential applications for pancreatic cancer, mainly focusing on pancreas segmentation by CT[28-29]. Our work concentrates on the detection of pancreatic cancer, and the results demonstrate that on a medium-sized dataset, an affordable CNN model can achieve comparable performance on pancreatic cancer diagnosis and can be helpful as an assistant of the doctors. Another interesting work by Liu et al, adopted the Faster R-CNN model, which is more complex and harder to train

and tune, for pancreatic cancer diagnosis. [30]. Their model was mixed images with different phases with an AUC 0.9632, while we trained three classifiers for the plain scan, arterial phase, and venous phase, respectively. Our results indicate that the plain scan, which has easier access and lower radiation, is sufficient for the binary classifier, with an AUC 0.9653.

Comment 11: Illustration and tables. Figure 1. Normal pancreas tail/body; the venous and arterial phase images are swapped and incorrectly placed. Cancers at Head/Neck; kidney enhancement pattern does not match a good venous images. If the authors have better images, please replace it.

Reply: We apologize for the confusion! We have made the correction in Figure 1.

Comment 12: References: Reference #14 and #16 are duplicate.

Reply: We apologize for the confusion! We deleted the duplicated Reference 16.

Replies to Reviewer 3

Comments to the Author

The authors present the application of a simple CNN architecture for diagnostic purposes in pancreatic cancer, for both case/control classification and ternary tumor absence/localization. Dataset size is reasonable, with 15-17 scans per participant. Performances of the CNN are also compared with 10 gastroenterologists and 15 trainees. Very interesting the careful prevention of data leakage by including images of the same patient only in one of train/test/validation sets. The work is interesting, but there are a number of issues that need to be addressed:

Comment 1: The paper is quite outdated: a large number of studies dealing with using AI on CT scans for pancreatic cancer diagnosis have been published in the literature in the last few years. The authors need to update the reference list, discuss methods and results of the published papers, and compare them with the results reported in the submitted manuscript.

Reply: Thank you very much for this valuable comment and suggestion! We have added several interesting works and compared with our work in the revised manuscript.

Page 20, line 363-364: "A CNN also has potential applications for pancreatic cancer, focusing on pancreas segmentation by CT[28-29]. "

Page 20, line 368-370: Another interesting work by Liu et al, adopted the Faster R-CNN model, which is more complex and harder to train and tune, for pancreatic cancer diagnosis[30].

Comment 2: Further, the authors need to stress the scientific points that mark the difference between the submitted paper from similar studies.

Reply: Thank you very much for this valuable comment and suggestion! We have made the corrections in the revised manuscript.

Page 20, line 365-368: "Our work concentrates on the detection of pancreatic cancer, and the results demonstrate that on a medium-sized

dataset, an affordable CNN model can achieve comparable performance on pancreatic cancer diagnosis and can be helpful as an assistant of the doctors."

Page 20-21, line 370-374: "Their model was mixed images with different phases with an AUC 0.9632, while we trained three classifiers for the plain scan, arterial phase, and venous phase, respectively. Our results indicate that the plain scan, which has easier access and lower radiation, is sufficient for the binary classifier, with an AUC 0.9653."

Comment 3: The proposed architecture is quite basic and simple - much more complex CNNs (and, in general, alternative DL solutions) have appeared, with better outcome: the authors should consider planning some advancements on the proposed solution.

Reply: Thank you very much for this suggestion! We have made the correction in the revised manuscript.

The rationale behind our basic architecture is that our dataset is of medium size. We also tried complex architectures, such as the Residual Network, but the experiment results show that our basic architecture can outperform the complex Residual Network and requires much less training and testing time. Another example is that Liu et al, adopted the Faster R-CNN model for pancreatic cancer diagnosis. The time of this complex model for processing one CT image is 10 times of that of our basic architecture, while the AUC of this model is not higher than ours.

Therefore, we consider that a more complex model may be not necessarily a better model. For a medium-sized dataset, a basic model is an effective, helpful and affordable choice for assisting doctors in pancreatic cancer diagnosis.

Comment 4: Overfitting control in training is quite limited: why dropout has not been used?

Reply: Thank you very much for this suggestion! We have added dropout in our architecture and revised the description and evaluation results of our CNN model.

Page 10-11, line 105-110: "Following the work by Srivastava N *et al*[15], a dropout rate of 0.5 is used between the average-pooling layer and the fully connected layer to avoid overfitting and increase the performance. We also tried SpatialDropout[16] between each max-pooling layer and its following convolutional layer, but found such dropouts resulted in performance degradation. Therefore, we do not use SpatialDropout."

Comment 5: Reproducibility is not well supported: resampling the train/validation/test split and averaging (with confidence intervals) over a number of runs would definitely improve the study robustness.

Reply: Thank you very much for this suggestion! We have conducted a 10-fold cross validation on our dataset and updated the evaluation results by averaging over the 10 runs.

Page 12-13, line 160-165: "A 10-fold cross-validation process was used to evaluate our techniques. We randomly divided the images in each phase into 10 folds, 8 of which were used to do the training, 1 fold was the validation set, and the remaining one was used to test the model. The entire process was repeated 10 times, and each fold will be used as the test set once. The average performance was recorded."

Comment 6:The achieved results are not impressive, both in absolute terms, in comparison with those reached by the panel of gastroenterologists/trainees and in comparison with those published in other papers (moreover, I would recommend using Matthews Correlation Coefficient as the elective classifier performance metric)

Reply: Thank you very much for this suggestion!

We have compared our work with Liu *et al*, and found that our basic architecture can outperform the complex Faster R-CNN model. Their model

was mixed images with different phases with an AUC 0.9632, while we trained three classifiers for the plain scan, arterial phase, and venous phase, respectively. Our results indicate that the plain scan, which has easier access and lower radiation, is sufficient for the binary classifier, with an AUC 0.9653.

Although our CNN model cannot stably outperform gastroenterologists, it can also be useful for assisting them in making more accurate diagnosis. In addition, in clinic, larger volume of CT has been done in a single day, however, it took several minutes for a doctor to read the image for a single person, which is time-consuming. The CNN model can process images much faster and is less prone to fatigue. Our classifier might be suitable for screening purposes in general medical practice. Doctors will focus mainly on meaningful information with the help of our classifier.

Page 19, line 331-337: "We executed our model using a Nvidia GeForce GTX 1080 GPU when performing classifications; its response time was approximately 0.02 seconds per image. Compared with the 10 seconds average reaction time required by physicians, although our CNN model cannot stably outperform gastroenterologists, the CNN model can process images much faster and is less prone to fatigue. Thus, binary classifiers might be suitable for screening purposes in general medical practice."

Page 20-21, line 368-374: "Another interesting work by Liu *et al*, adopted the Faster R-CNN model, which is more complex and harder to train and tune, for pancreatic cancer diagnosis[30]. Their model was mixed images with different phases with an AUC 0.9632, while we trained three classifiers for the plain scan, arterial phase, and venous phase, respectively. Our results indicate that the plain scan, which has easier access and lower radiation, is sufficient for the binary classifier, with an AUC 0.9653."

Comment 7: Many sentences in the paper are written in non-standard English - proofreading by a native English speaker is strongly recommended

Reply: Thank you very much for this suggestion! We have polished the

language in the revised manuscript. If language qualify still didn't match the criteria, we are happy to edit language again by the suggested organization. Thank you.