

View Author Feedback

Paper ID

2689

Paper Title

Bayesian Neural Networks for Uncertainty Estimation of Imaging Biomarkers

AUTHOR FEEDBACK QUESTIONS

1. Your rebuttal is addressed to the Area Chairs only. Reviewers will not see it and will not be able to change their reviews. The goal of the rebuttal is to inform the Area Chairs of major misunderstandings, in your opinion, in the reviewers' assessment, or of incorrect statements in the reviews. An effective rebuttal focuses only on major critiques. It is not helpful to try to address every minor point in the reviews. By prioritizing and focusing on the major concerns, and by grouping multiple reviewer comments that generally pertain to the same issue into a few major categories, you are demonstrating to the Area Chairs that you understand the high level messages that were provided in the reviews. Please summarize or rephrase the criticism before you address it, and clarify to which comment(s) you are responding. While the room for rebuttal is limited, if properly utilized by condensing the response down to the essentials, this is an effective way to let the Area Chairs know that you understood the reviewer's concerns and have valid answers to the questions raised in the reviews, or to establish that certain reviewer comments were false or unsubstantiated. An effective rebuttal addresses reviewers' criticisms by explaining where in the paper you had provided the requisite information, perhaps further clarifying it. Do not promise to expand your paper to address all the questions raised by the reviewers, as you will not be able to change your article substantially, and in all likelihood you don't have sufficient room to add to the paper. These promises are likely not to be taken seriously. A good rebuttal is polite; being confrontational does not bring any added value to the paper. However, if you feel you have received a review that was not courteous, or made false or unsubstantiated arguments that you can succinctly refute, you should point this out. (See "Rebuttal Guide" <https://miccai2020.org/en/PAPER-SUBMISSION-GUIDELINE.html#rebuttal-guides>) Please enter your rebuttal below (4000 chars max)

We thank the reviewers for the constructive critique. They consider our paper "a good idea of [...] integrating segmentation uncertainty in follow-up tasks" and that the "comparison of different uncertainty generation methods is applaudable". Finally, our "results clearly demonstrate the effectiveness of the proposed method". We address the points summarized by the AC in the following.

NOVELTY and RELATED WORK

R1 raises questions about the novelty of our work. Our article is about the integration of segmentation uncertainty in follow-up analyses. We do not claim that we are the first to consider this problem, but that it "has not yet been well studied". In [14], uncertainty from MC-Dropout has been used for instance reweighting in group analyses. In our article, we take a more principled approach to this problem by:

- introducing the first approach for disease classification with uncertainty,
- proposing novel statistical models for the integration of uncertainty in group analyses,
- evaluating 4 different Bayesian neural networks for estimating the uncertainty.

We thank R1 for referring us to Mehta et al., MICCAI UNSURE, 2019. We would be happy to include it in a final version, but we would also like to emphasize that it does not impact the novelty of our work. First, they use uncertainty maps in a second neural network that again works in the image space, in contrast to the extraction of biomarkers in our work. Methods and application are therefore completely different. Second, the paper only uses MC dropout, while we compare 4 different uncertainty networks.

STATISTICAL SIGNIFICANCE

R1 mentions that our "results demonstrate clear usefulness of uncertainty in increasing diabetes classification accuracy", but that we have not performed significance testing. Following the suggestion, we have performed the

Friedman test with a pairwise Nemenyi post hoc test to assess statistical significance between disease prediction models at significance level 0.05. We note a significant improvement concerning the base model for all uncertainty methods with IoU in combination with Variable and Interaction. Further, MC-Dropout (Variable IoU) and probabilistic U-Net (Interaction IoU) significantly improve over the manual model. These results could be easily integrated into the final version of Table 2.

AVERAGING PREDICTIONS

R2 questions whether averaging the segmentation samples to obtain the final segmentation for computing the liver volume makes sense for the probabilistic and hierarchical U-Net. The reviewer is correct that both approaches were developed to generate diverse segmentations to model different raters. However, as discussed in the paper, we only have annotations from a single rater, so that the networks do not output segmentations that are structurally highly varying (different modes). Hence, computing the average is reasonable in this scenario.

RANKING

As this is the first comparison of 4 different techniques for uncertainty estimation, R1 asked about their ranking. For regression, we see the best performance for MC-Dropout with IoU and Variable. For classification, this setting has the second-best result, only topped by the probabilistic U-Net with Interaction. This difference can be attributed to the Interaction term in the classification setting, which cannot be used in regression as the volume is the dependent variable.

ALEATORIC - EPISTEMIC

We agree with R3 that the different Bayesian networks either focus on epistemic or aleatoric uncertainty. Our methods do not rely on a specific form of uncertainty, in fact, one of your main contributions is to investigate which type is better suited for follow-up analyses. Moreover, we agree that results might improve by calibrating the aleatoric uncertainty, but we only included methods in their original implementation, as we were interested in integrating the uncertainty of existing methods into follow-up analysis; not to introduce a new method for uncertainty estimation.
