Confidence Calibration using a Bayesian Variational Inference-based approach for Colorectal Polyp Classification

Citation for published version (APA):

Document status and date:
Published: 27/07/2022

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 20. Sep. 2022
Confidence Calibration using a Bayesian Variational Inference-based approach for Colorectal Polyp Classification

Nikoo Dehghani¹, Thom Scheeve¹, Quirine E. W. van der Zander²,³, Ayla Thijssen²,³, Ramon-Michel Schreuder¹, Ad A. M. Mascleel², Erik J. Schoon³,⁴, Fons van der Sommen¹,⁴ and Peter H. N. de With⁵

¹ Eindhoven University of Technology, Eindhoven, The Netherlands
² Maastricht University Medical Center+, Maastricht, The Netherlands
³ Catharina Hospital Eindhoven, Eindhoven, The Netherlands
⁴ Eindhoven Artificial Intelligence Systems Institute, Eindhoven, The Netherlands
⁵ GROW, School for Oncology and Reproduction, Maastricht, The Netherlands

MOTIVATION
The aim of this research is to develop a well-calibrated Computer-Aided Diagnosis (CADx) system for classification of Colorectal Polyps (CRPs) with the use of Bayesian variational inference.

PROBLEM STATEMENT
- Colorectal Polyps (CRPs) are potential precursor lesions of Colorectal cancer (CRC). An improved optical diagnosis of these polyps during the colonoscopy procedure is essential for an appropriate treatment strategy.
- Computer-Aided Diagnosis (CADx) systems can play a crucial role as a second opinion for physicians to improve the overall medical diagnosis performance.
- Deep neural network-based systems, often tend to be over-confident about their decision. These systems can output high-probability predictions, even though, they are not confident about them.
- A reliable CADx system should be able to provide accurate and well-calibrated classification confidence.

DATASET

The endoscopic data used in the conducted experiments in this study are collected at the Catharina Hospital Eindhoven (the Netherlands) and the Maastricht University Medical Center+ (the Netherlands), and the Queen Alexandra Hospital in Portsmouth (United Kingdom).

- Included polyp pathologies: Hyperplastic polyps (HP), Adenomas (ADs), Sessile Serrated Lesions (SSLs) and Adenocarcinomas
- Included imaging modalities: White-Light Endoscopy (WLE), Blue-Light Imaging (BLI), Linked Color Imaging (LCI), and i-Scan modality in Modes 1, 2, and 3.
- In total 2,287 endoscopic images (1,836 pre-malignant, 451 benign)
- The independent test set contains 258 images, while the training set and the validation set include 1,624 and 405 images, respectively.

RESULTS
Experiments performed on both approaches to evaluate their classification and calibration performance using the CRP dataset. Calibration performance is evaluated using the Reliability diagrams of both models. A reliability diagram shows the accuracy as a function of the predicted confidence of samples, by grouping predictions into bins, based on their predicted confidence. Quantitative evaluation of the confidence calibration is done using the Maximum Calibration Error (MCE) across all the bins.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC*</th>
<th>MCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian</td>
<td>87.21</td>
<td>92.04</td>
<td>70.18</td>
<td>0.91</td>
<td>0.1827</td>
</tr>
<tr>
<td>Deterministic</td>
<td>84.88</td>
<td>88.06</td>
<td>73.68</td>
<td>0.90</td>
<td>0.4289</td>
</tr>
</tbody>
</table>

DISCUSSION
With the use of the Bayesian variational layers, we were able to produce calibrated classification confidence without extra post-processing steps.

The presented results convincingly demonstrate that this Bayesian approach can provide better calibrated decisions than the deterministic approach, while having a similar, or in cases, improved classification performance.

A future perspective of this research is the study of different prior distribution initialization.