[BIOINFO 2023]

Confidence Estimation of a Clinical Decision Support System for Determining the Colonoscopy-Surveillance Interval

Nayeon Kim[1], Yunyol Shin[2], Yinhua Piao[3], Sun Kim[1,3,4,5], Joo Hyun Lim[6] and Jung Ho Bae[6,*]

[1]*Interdisciplinary Program in Artificial Intelligence, Seoul National University*
[2]*Devsisters Corp.*
(Yunyol Shin is now in Devsisters Corp., but did this work earlier as a SNU student)
[3]*Department of Computer Science and Engineering, Seoul National University*
[4]*Interdisciplinary Program in Bioinformatics, Seoul National University*
[5]*MOGAM Institute for Biomedical Research*
[6]*Seoul National University Hospital, Healthcare System Gangnam Center*
**Corresponding author: bjh@snuh.org*

Colorectal cancer (CRC) is the third most common cancer worldwide, and colonoscopy screening program is the gold standard for CRC prevention. After colonoscopy screening, it is crucial to recommend an appropriate colonoscopy surveillance interval to ensure early detection of CRC. Besides, excessively frequent examinations may elevate the risk of complications, such as perforation and bleeding, without significant additional benefits in managing the risk of CRC. Nonetheless, previous studies showed that endoscopists in practice either frequently overuse or underuse guideline-based surveillance intervals. One of the reasons for this misuse may be the discordance between colonoscopy and pathology reports, which encompass complex combinations of patient demographics, colonoscopy findings, and polyp characteristics. To address this issue, numerous clinical decision support systems (CDSSs) have been developed to enhance adherence to clinical guidelines during the decision-making process. Natural language processing (NLP), a computer linguistic technique which is effective for extracting information from free-text reports, has been used to develop automated systems providing guideline-based colonoscopy-surveillance interval recommendations with high accuracy. However, those NLP-based CDSSs can produce incorrect recommendations due to their susceptibility to linguistic variations emerging across time, location, and among endoscopists, which may not be adequately captured by the training data. Since CDSSs are implemented for high-risk environment such as the medical field, it is important to reduce the risk of incorrect predictions. Therefore, for reliable usage, it is necessary to detect and quantify the

uncertainty of the predictions made by CDSSs. In this study, we developed a rule-based NLP system that process both colonoscopy reports and pathology reports to generate surveillance interval recommendations. Additionally, we built a neural network-based uncertainty estimator that measures the level of out-of-distribution of the given colonoscopy and pathology reports to produce confidence scores for the CDSS's decisions. We believe that our reliable NLP-based CDSS with an uncertainty estimator can significantly enhance medical decisions and improve healthcare delivery.