Medical Image Preprocessing and Convolutional Neural Network for Colorectal Cancer Image Classification

Natinai Jinsakul[1,2]; Cheng-Fa Tsai[2]*

[1]*Department of Tropical Agriculture and International Cooperation, National Pingtung University of Science and Technology, Pingtung, Taiwan*

[2]*Department of Management Information Systems, National Pingtung University of Science and Technology, Pingtung, Taiwan*

Background: Colorectal cancer is the third leading cause of cancer-related deaths in humans that has become a widespread and increased health problem. Moreover, this disease has spread to young people. However, this issue has used medical imaging and several computer science techniques; such as, artificial intelligence (AI) in terms of state-of-the-art in deep learning to assist in the process of diagnosing the disease. The objectives of this study were to employ image preprocessing and deep learning of a convolutional neural network (CNN) to classify images of colorectal cancer and compare the validation results of the image training and test between no preprocessing and the preprocessing of colorectal cancer images.

Methods: The image dataset was collected from the CT *colonography* of The Cancer Imaging Archive (TCIA) and employed images from 328 patients, who were classified into 104 cancer patients and 224 normal patients. There was a total of 425 images comprising 201 cancer images and 224 non-cancer images. All of the images used preprocessing with a center crop that emphasized specifically on the colon. The training and test section applied a CNN named NASNet for classifying and validating the indices calculation of the average accuracy, average precision, average recall and F1 measure. In addition, all of the results of the validated indices were compared with the validation results of the original images without preprocessing.

Results: The validation results of the training and test with the CNN on no image preprocessing had an average accuracy of 88.42%, average precision of 89.65%, average recall of 89.25%, and F1 measure equal to 88.46%. The validation results from using the image preprocessing training and test with the CNN for image classification had an average accuracy score of 89.41%, average precision of 89.19%, average recall of 91.48% and F1 measure equal to 89.62%.

Conclusions: The operation of the image preprocessing by using a center crop that focused specifically on the images was able increase the accuracy of the image classification; thus, the validation indices results proved to be better than using the original images.

Furthermore, the computer science method in this paper could be applied to other medical image problems and other fields relevant to image classification.

Acknowledgements: The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan, for financially supporting the research [No. MOST 107-2637-E-020-006 and No. MOST 107-2321-B-020-005].