Whole Slide Image Analysis and Detection of Prostate Cancer using Vision Transformers

Kobiljon Ikromjanov
Dept. of Digital Anti-Aging Healthcare
Inje University
Gimhae, Republic of Korea
kobiljonikromjanov@gmail.com

Subrata Bhattacharjee
Dept. of Computer Engineering
Inje University
Gimhae, Republic of Korea
subrata_bhattacharjee@outlook.com

Yeong-Byn Hwang
Dept. of Digital Anti-Aging Healthcare
Inje University
Gimhae, Republic of Korea
hyb1345679@gmail.com

Rashadul Islam Sumon
Dept. of Digital Anti-Aging Healthcare
Inje University
Gimhae, Republic of Korea
sumon39.cst@gmail.com

Hee-Cheol Kim
Dept. of Digital Anti-Aging Healthcare
Inje University
Gimhae, Republic of Korea
heeki@inje.ac.kr

Heung-Kook Choi
Dept. of Computer Engineering
Inje University
Gimhae, Republic of Korea
cschk@inje.ac.kr

***Abstract*—Prostate cancer (PCa) is the most frequently diagnosed non-skin malignancy in men and the second leading cause of fatality from cancer. The most prognostic marker for PCa is the Gleason grading system on histopathology images. Pathologists examine the Gleason grade on stained tissue specimens of Hematoxylin and Eosin (H&E) based on tumor structural growth patterns from whole slide image (WSI). According to the Gleason grading system, prostate cancers are scaled into five grades based on glandular patterns of differentiation. It varies from grade 1 (normal tumor) to grade 5 (abnormal tumor). Cancer cells that look similar to healthy cells receive a low score. Recent developments in Computer-Aided Detection (CAD) using Artificial Intelligence (AI), mainly Deep learning (DL) have brought the immense scope of automatic detection and recognition at better accuracy in adenocarcinoma like other medical diagnoses. Automated DL systems have delivered promising results from histopathological images to accurate grading of prostatic adenocarcinoma. This study aims to classify multiple patterns of images extracted from the WSI of a prostate biopsy based on the Gleason grading system. First, extract patches from the detected region of interest (ROI), then applying Vision Transformers (ViT) model for classification. Finally, the classified patches are scored and graded. The proposed deep learning model in this research will be able to assist the pathologist and other researchers to identify and treat of prostate cancer.**

Keywords— whole slide image, prostate cancer, vision transformers, artificial intelligence

I. Introduction

Deeplearning AI architectures are developed and applied to medical images, making high-precision diagnosis possible. For diagnosis, the medical images need to be labeled and standardized, before data pre-processing and training DL model. The final predicted diagnosis results can be obtained immediately and accurately. Tumor detection and classification in histopathology images are important for early diagnosis and treatment planning. Many techniques have been proposed for classifying medical image data through quantitative assessment [1-3]. However, some quantitative ways of evaluating medical images are inaccurate and require considerable computation time to analyze large amounts of data. Analytical strategies applied to AI algorithms can improve diagnostic accuracy and save time.

To identify different kinds of prostate tumors, pathologists use different screening methods. Male hormones such as testosterone cause prostate cancer to grow and survive. Like all cancers, prostatic adenocarcinoma begins once a mass of cells has grown out of control and invades other tissues. Cells become cancerous due to the accumulation of defects, or mutations, in their DNA. Mutations in the abnormal cells' DNA cause the cells to grow and divide more rapidly than normal cells do. Histological examination of tissues and the detection of cancer by physicians remains the gold standard in cancer diagnosis. The diagnosis of PCa is heavily time-consuming. In addition, it is based on subjective grading. For example, the study by Ozkan et al. reported that two pathologists disagreed about the presence of cancer in 31 of 407 baseline biopsies and that the total concordance of the accessed Gleason score was only 51.7%, describing these challenges in diagnosing the PCa consistently [4]. Therefore, the development of computer-assisted decision support tools is essential for saving time, predicting disease outcomes, and improving precision medicine for pathologists.

Automated diagnosis can reduce workloads and pathologist variability. Researchers face difficulty in studying the Gleason scoring system [5, 6]. Accurate annotations and pathological accuracy are required to train the model correctly. At present, automated computerized techniques are in high demand for medical image analysis and processing. However, we propose a ViT [7-9] model to classify the grading of PCa. We detect ROI, then patches for classification and scoring [10-12]. After scoring all the patches, overall grading is performed for each WSI which is helpful for pathologists to save time.

II. Dataset Collection

The dataset was collected online, which is publicly available on the Kaggle PANDA challenge [13]. It has 10,616 WSI images and 10,516 corresponding WSI mask images. Radboud University Medical Center and Karolinska Institute have teamed up to organize this PANDA competition. Fig. 1 shows some examples of Radboud and Karolinska images and their annotations. However, in our experiment, we have used more than 5,000 images with their masks from Radboud University Medical Center dataset. After having a patching process on those images, we got approximately 304,000 acceptable patch images of size 256 × 256 pixels, corresponding to 5 classes: stroma, benign, score 3, score 4, and score 5. We have split about 240,000 patch images for training, 60,000 patch images for validation, and around 4,000 patch images for testing.

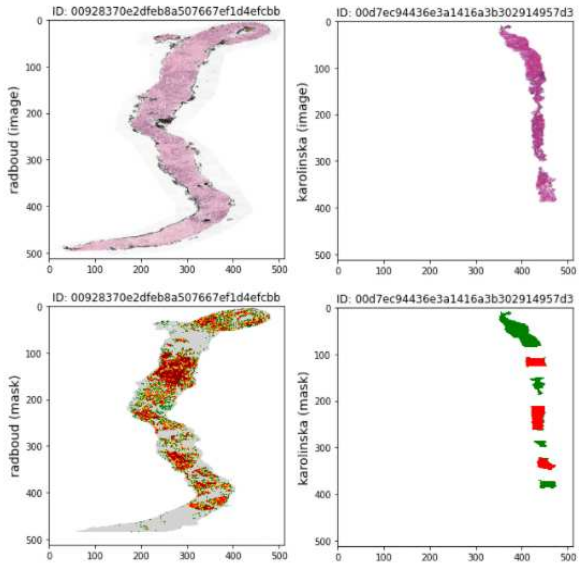

Fig. 1. Sample images and ground masks

III. METHODOLOGY

To follow up the proposed model, we applied patching for making an acceptable dataset for the model. Then, we manipulated the ViT model and get the training output according to the Gleason Scoring System. The following Fig. 2 shows the general view of the applied method for the training of the ViT model and its architecture. In the following, we will clarify patching, a ViT model, and a Gleason Score System.

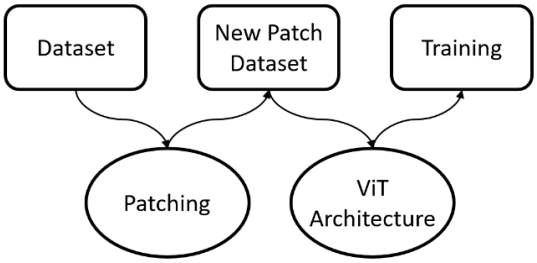

Fig. 2. The process of patching and training ViT model

A. Patching

The ability to compare image regions (patches) has been the basis of many approaches to core computer vision problems, including object, texture, and scene categorization. The WSI is also called a gigapixel image that is composed of more than 1 billion pixels [14], and it is computationally unfeasible to perform ROI-based image analysis in such a high dimensional space. Therefore, in many existing works, image analysis have been performed over small image patches. This has the advantage of making computational tasks such as learning, inference, and likelihood estimation much easier than working with gigapixel image directly. In this order, the entire WSI cannot be trained in GPU memory at once, so one solution is to select a subset of patches from the high dimensional image. In this study, we have patched all 10,516 WSIs and their ground truth mask images cikkected from the Kaggle dataset. Fig. 3 illustrates the steps for patching the WSI. Fig. 3 (a) shows 395 patches of size 256 × 256 pixels, while Fig. 3 (b) indicates the annotation done by pathologists. In Fig. 3 (c), the patches are extracted from ROI annotated in Fig. 3 (b).

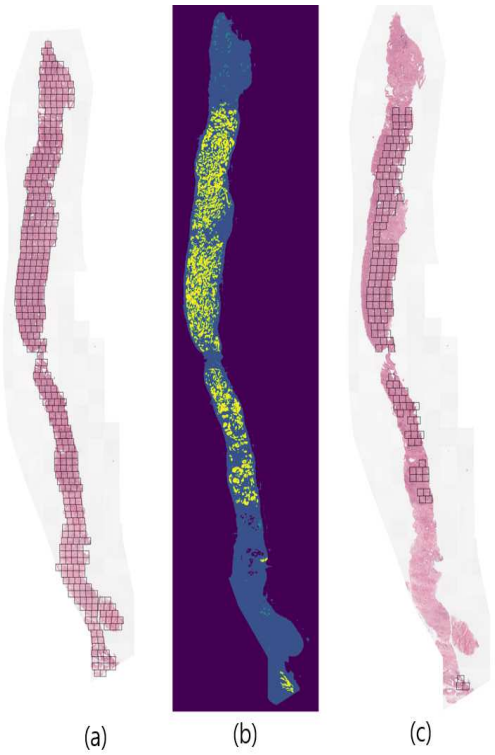

Fig. 3. The steps for patching a WSI

B. Vision Transformers

The Vision Transformer, or ViT, is a model for image classification that employs a Transformer-like architecture over image patches. An image is split into fixed-size patches, each of them is then linearly embedded, position embeddings are added, and the resulting sequence of vectors is fed to a standard Transformer encoder. The standard approach of adding an extra learnable "classification token" to the sequence is used to perform the classification.

Inspired by the ViT model for the classification of each patch, we experiment with applying the patched images from the WSI as input images. First, we split them into fixed-sized images, then flatten them. After creating lower-dimensional linear embeddings from flattened image patches, we include positional embeddings. Moreover, feeding the sequence as an input to a state-of-the-art transformer encoder, we can pre-train the ViT model with image labels, which are then fully supervised on a big dataset. Lastly, we can fine-tune the downstream dataset for image classification. Fig. 4 shows ViT architecture for classification.

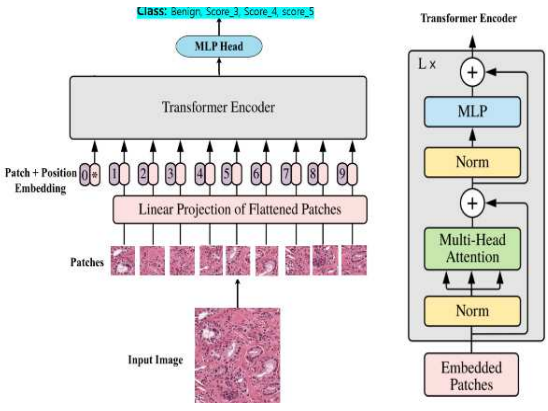

Fig. 4. ViT architecture for classification

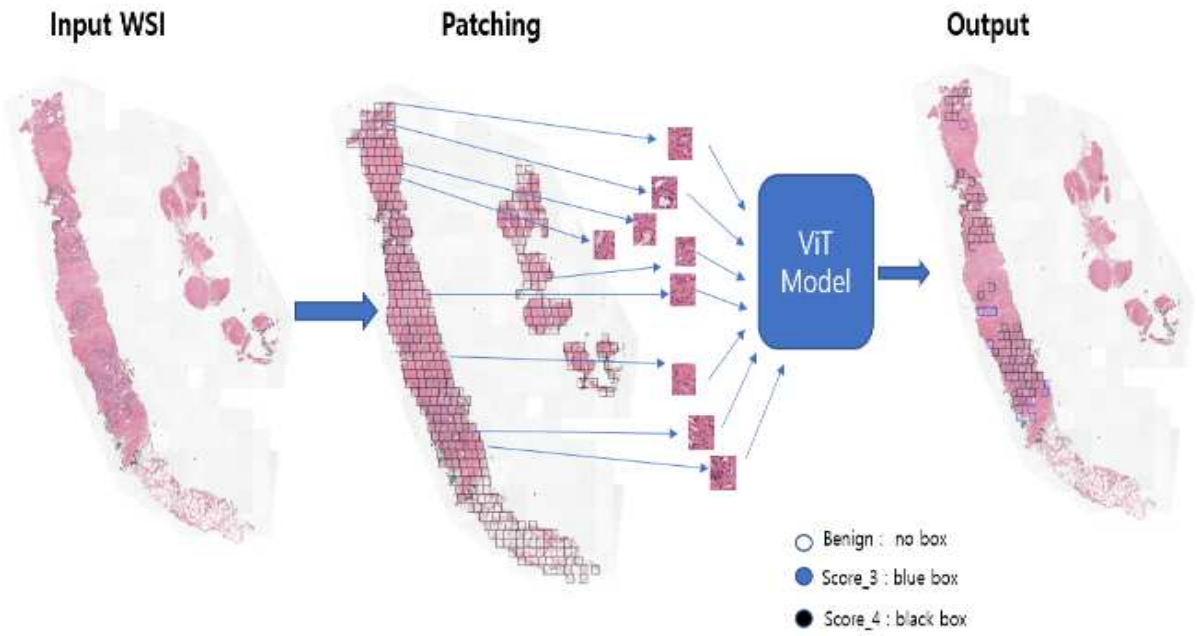

Fig. 5. Example of patching and prediction of ViT model

C. *Gleason Score System*

The Gleason Score is the grading system used to determine the aggressiveness of prostate cancer. This grading system can be used to choose appropriate treatment options. The Gleason Score ranges from 1-5 and describes how much cancer from a biopsy looks like healthy tissue (lower score) or abnormal tissue (higher score). Most cancers score a grade of 3 or higher. Since prostate tumors are often made up of cancerous cells that have different grades, two grades are assigned for each patient. A primary grade is given to describe the cells that make up the largest area of the tumor and a second grade is given to describe the cells of the next largest area. For instance, if the Gleason Score is written as 3+4=7, it means most of the tumor is grade 3 and the next largest section of the tumor is grade 4, together they make up the total Gleason Score. If the cancer is almost entirely made up of cells with the same score, the grade for that area is counted twice to calculate the total Gleason Score. Typical Gleason Scores range from 6-10. The higher the Gleason Score, the more likely it is that cancer will grow and spread quickly. Our proposed method shows the potential of AI systems for Gleason grading, but more importantly, shows the benefits of pathologist-AI synergy.

IV. Result

The ViT model evaluated the patches as stroma, benign, score 3, score 4, and score 5 according to the level of cancerous cells and differentiates score 3, 4, and 5 with three different colors which would be very helpful for the pathologists to identify the affected ROIs. In the following Fig. 5, we can see around 760 patches in the beginning, and the ViT model predicted the level of cancerous cells and differentiated a whole image as no box for the benign, blue box for score 3, and black box for score 4.

The performance measures used for model evaluated are precision, recall, and f1-score. Overall the model performed well and achieved an accuracy of 80.0%. The following Table I shows the performance measures for the ViT model. In the precision, the model achieved good scores on the stroma, benign, and score 3. In the recall and f1-score, the model predicted stroma and benign cases very well compared to score 3, 4, and 5 cases. Fig. 6 demonstrates the overall architecture of the procedure for generating the final predicted WSI image. In the final image, the pathologists can see the affected areas and its level. The model efficiency can be evaluated through the confusion matrix, shown in Fig. 7.

TABLE I. VIT ARCHITECTURE FOR CLASSIFICATION

Classes	Precision (%)	Recall (%)	F1-score (%)
Stroma	99.0	90.0	94.0
Benign	84.0	93.0	88.0
Score 3	82.0	73.0	78.0
Score 4	63.0	72.0	67.0
Score 5	74.0	71.0	72.0
Average	80.4	79.8	79.8

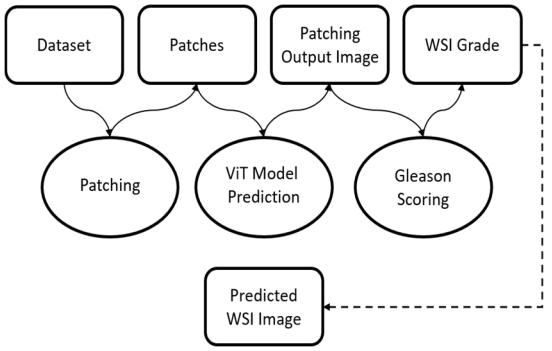

Fig. 6. The overall architecture for prediction and generating the final predicted image

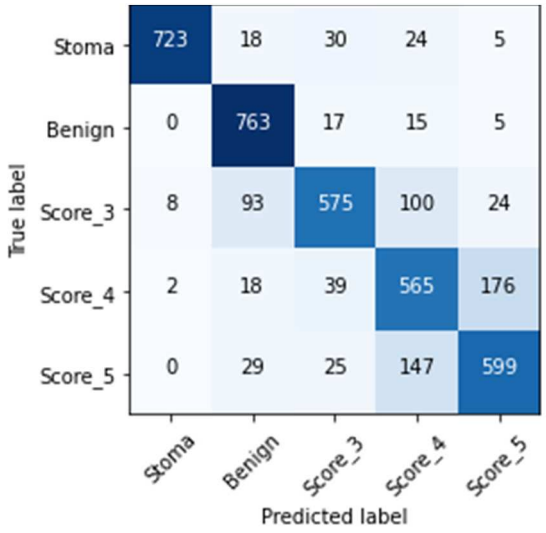

Fig. 7. Confusion matrix of ViT model

V. Discussion

The proposed patching technique and ViT model were designed to help the pathologists classify different cancer images which consist of two active approaches for a vision processing task. The sliding window approach in image processing is used to get the local information by sub-dividing the images into many blocks (may be overlapping or non-overlapping). The kernel is a small matrix act as a transformation, it is used to map the original data into modified one. An overview of the model is depicted in Fig. 4. The first layer of the ViT linearly projects the flattened patches into a lower-dimensional space. The components resemble plausible basis functions for a low-dimensional representation of the fine structure within each patch. Fig. 5 illustrates the workflow for testing and prediction WSI samples. First WSI is divided into patches then each patch is fed to ViT model for scoring. If there are cancer patch images on prediction, it counts the number of predictions: score 3, score 4, and score5 and makes bounding boxes on WSI that could be helpful for the doctors to make a better decision.

VI. Conclusion

The proposed method is patching WSIs and selecting ROI patches using ground truth images to train the ViT model. As the ViT is an attention-based model, it may give better concentration on cancer tissue areas while training and testing. This paper proposes a deep learning-based classification of multiple patterns of images extracted from the WSI of a prostate biopsy based on the Gleason grading system. The results show possibilities to assist the pathologist and other researchers to identify and treat of prostate cancer. In the future, we will develop the Mask-RCNN architecture for more improvement, train the model on a greater number of datasets, and explain the prediction of the model via different interpretable techniques.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C2008576).

References

[1] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology," *Nat. Rev. Clin. Oncol.*, 2019, doi: 10.1038/s41571-019-0252-y.

[2] B. Aygüneş, S. Aksoy, G. Cinbiş, K. Kösemehmetoglu, S. Önder, and A. Üner, "Graph convolutional networks for region of interest classification in breast histopathology," 2020, doi: 10.1117/12.2550636.

[3] S. Bhattacharjee, C. H. Kim, D. Prakash, H. G. Park, N. H. Cho, and H. K. Choi, "An efficient lightweight cnn and ensemble machine learning classification of prostate tissue using multilevel feature analysis," *Appl. Sci.*, 2020, doi: 10.3390/app10228013.

[4] T. A. Ozkan, A. T. Eruyar, O. O. Cebeci, O. Memik, L. Ozcan, and I. Kuskonmaz, "Interobserver variability in Gleason histological grading of prostate cancer," *Scand. J. Urol.*, 2016, doi: 10.1080/21681805.2016.1206619.

[5] W. Bulten *et al.*, "Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study," *Lancet Oncol.*, 2020, doi: 10.1016/S1470-2045(19)30739-9.

[6] N. Chen and Q. Zhou, "The evolving gleason grading system," *Chinese Journal of Cancer Research*. 2016, doi: 10.3978/j.issn.1000-9604.2016.02.04.

[7] M. Is, R. For, and E. At, "An image is worth 16x16 words: visual image transformer," 2021.

[8] A. Vaswani *et al.*, "Attention Is All You Need," *Adv. Neural Inf. Process. Syst.*, Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762.

[9] Y. Xia *et al.*, "Effective Pancreatic Cancer Screening on Non-contrast CT Scans via Anatomy-Aware Transformers," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2021, pp. 259–269.

[10] X. Zhu, J. Yao, and J. Huang, "Deep convolutional neural network for survival analysis with pathological images," 2017, doi: 10.1109/BIBM.2016.7822579.

[11] J. Yao, X. Zhu, J. Jonnagaddala, N. Hawkins, and J. Huang, "Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks," *Med. Image Anal.*, vol. 65, p. 101789, Oct. 2020, doi: 10.1016/j.media.2020.101789.

[12] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," *Computers in Biology and Medicine*. 2021, doi: 10.1016/j.compbiomed.2020.104129.

[13] "Prostate cANcer graDe Assessment (PANDA) Challenge", Accessed on: Oct. 10, 2021. [Online]. Available: https://www.kaggle.com/c/prostate-cancer-grade-assessment/overview/description.

[14] D. Tellez, G. Litjens, J. Van Der Laak, and F. Ciompi, "Neural Image Compression for Gigapixel Histopathology Image Analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021, doi: 10.1109/TPAMI.2019.2936841.