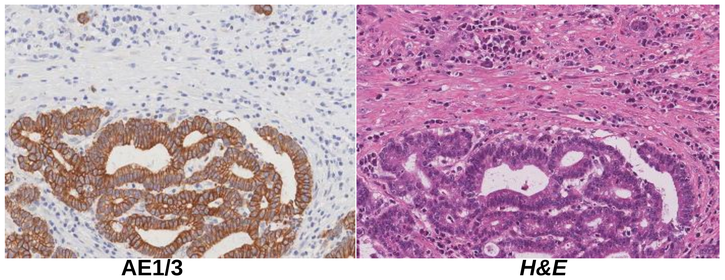
Tumor Segmentation in Colorectal Cancer using Transformers

Arijit Dutta[1], Thomas Tavolara[2], M. Khalid Khan Niazi[2]
[1]University of Pittsburgh, Department of Bioengineering, Pittsburgh, Pennsylvania
[2]Wake Forest School of Medicine, Center for Biomedical Informatics, Winston-Salem, North Carolina

INTRODUCTION

Colorectal cancer (CRC) is the third leading cancer-related cause of death in the United States with an estimated 52,980 deaths and 149,500 new cases in 2021 alone [1]. Though Hematoxylin & Eosin (H&E) staining is the recommended standard for tumor identification, it is an arduous procedure that suffers from high intra- and inter-observer variability in proper identification. Agreement improves when using pan-cytokeratin (AE1/3) staining but it remains non-standard due to its high cost and lack of availability. The objective of this study is to develop a deep learning model to accurately identify tumor regions in H&E images while utilizing AE1/3 only for development.



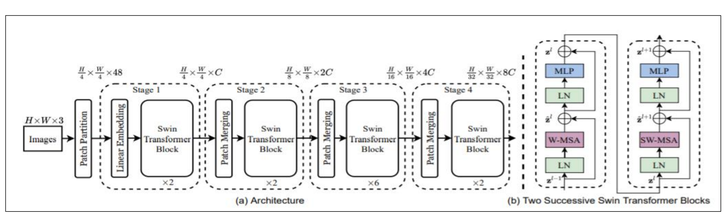
AE1/3 ***H&E***

MATERIALS AND METHODS

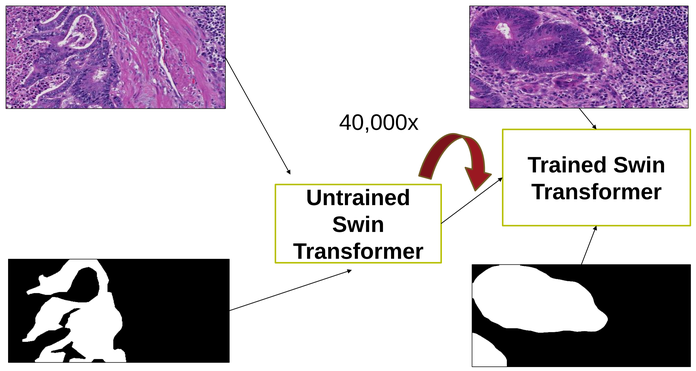
Dataset

- 648 adjacent H&E and pan-cytokeratin HPF slides from 119 patients
- Slides annotated for tumor and non-tumor regions
- 5-fold cross validation used by splitting patients into five groups

Swin transformers were used to carry out the semantic segmentation of colorectal cancer [2]. Swin transformers are a type of vision transformer which forgo the convolutions of standard Convolutional Neural Networks for natural language process-based attention. Swin transformers specifically differ from other vision transformers of if its shifted approach and ability to learn at multiple scales.



Architecture of Swin transformer model [2]

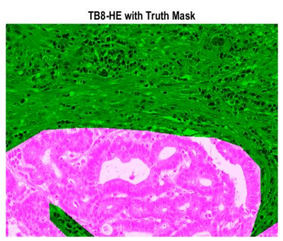


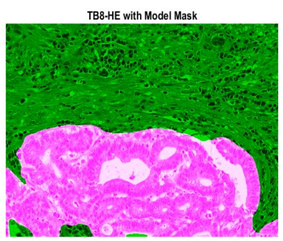
Prior to training, five groups were created by splitting patients into said groups. For each model, a unique selection of four groups were set for training and the remaining for testing. This yielded five models. To train the Swin transformer model, standard H&E images from the training set were fed into the model along with their ground truth binary masks. These masks were split into classes being labeled either tumor (white regions) or non-tumor (black regions). Each model was trained for 40,000 iterations with it seeing two different H&E images and their respective ground truths per iteration.

After training, H&E images from the testing set were sent through the model. The model produced a predicted binary mask of what it believes are the tumor and non-tumor regions. These model produced masks of the testing set were then compared to their respective ground-truths.

RESULTS AND DISCUSSIONS

The two metrics used for results were pixel accuracy and intersection over union(IoU). Pixel accuracy is conducted by. IoU is measured by calculating the number of pixels, in each class, that overlap in the model mask and ground truth mask over the sum of the total number of pixels, in each class, of both masks. Below is an example of both masks (ground truth on left and model on right) of an image overlayed on the same H&E image. The pink is the tumor while the green is in the non-tumor.



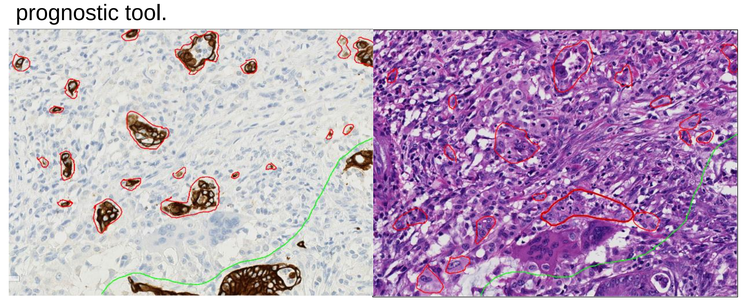


	Accuracy		IoU	
	Tumor	**Non-Tumor**	**Tumor**	**Non-Tumor**
Average	95.73% ± 1.71%	99.03% ± 0.31%	92.77% ± 1.25%	97.77% ± 0.31%
Model 1	94.58%	99.37%	92.51%	97.59%
Model 2	96.89%	98.53%	92.51%	97.59%
Model 3	93.28%	99.19%	90.93%	97.13%
Model 4	96.94%	98.96%	93.89%	98.00%
Model 5	96.96%	99.08%	93.99%	98.20%

Above is the quantitative results of each model, for each class, as well as the average and standard deviation of all five models. The accuracy of both tumor and non-tumor classes were at least 93% with an IoU for both classes of at least 90%.

CONCLUSIONS/FUTURE DIRECTIONS

The Swin Transformer was accurate in identifying tumor regions within H&E-stained images. This study shows the potential that Swin Transformers have for tumor identification and how they may be able to minimize intra- and inter- pathologist variability. In future studies, we will utilize these Swin Transformer models for tumor budding detection. Tumor buds are collections of 3-4 tumor cells which are circled in red below. Identification of these buds will serve as a very important prognostic tool.



ACKNOWLEDGEMENTS

Funding for this project was provided by the Wake Forest Center for Biomedical Informatics.

REFERENCES

[1] C. H. Song et al. Nuclear Factor Erythroid 2-related Factor 2 Knockout Suppresses the Development of Aggressive Colorectal Cancer Formation Induced by Azoxymethane/Dextran Sulfate Sodium-Treatment in Female Mice. *Journal of Cancer Prevention* (2021) 26.1, 41

[2] Z. Liu, Y. Lin, Y. Cao, H. Hu et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *arXiv* (2021) 2103.14030