A Comparison of Histopathology Imaging Comprehension Algorithms based on Multiple Instance Learning

Adam Saunders[1], Hong-Jun Yoon[2]
[1]Department of Electrical and Computer Engineering, University of Dayton, Dayton, Ohio
[2]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

Introduction

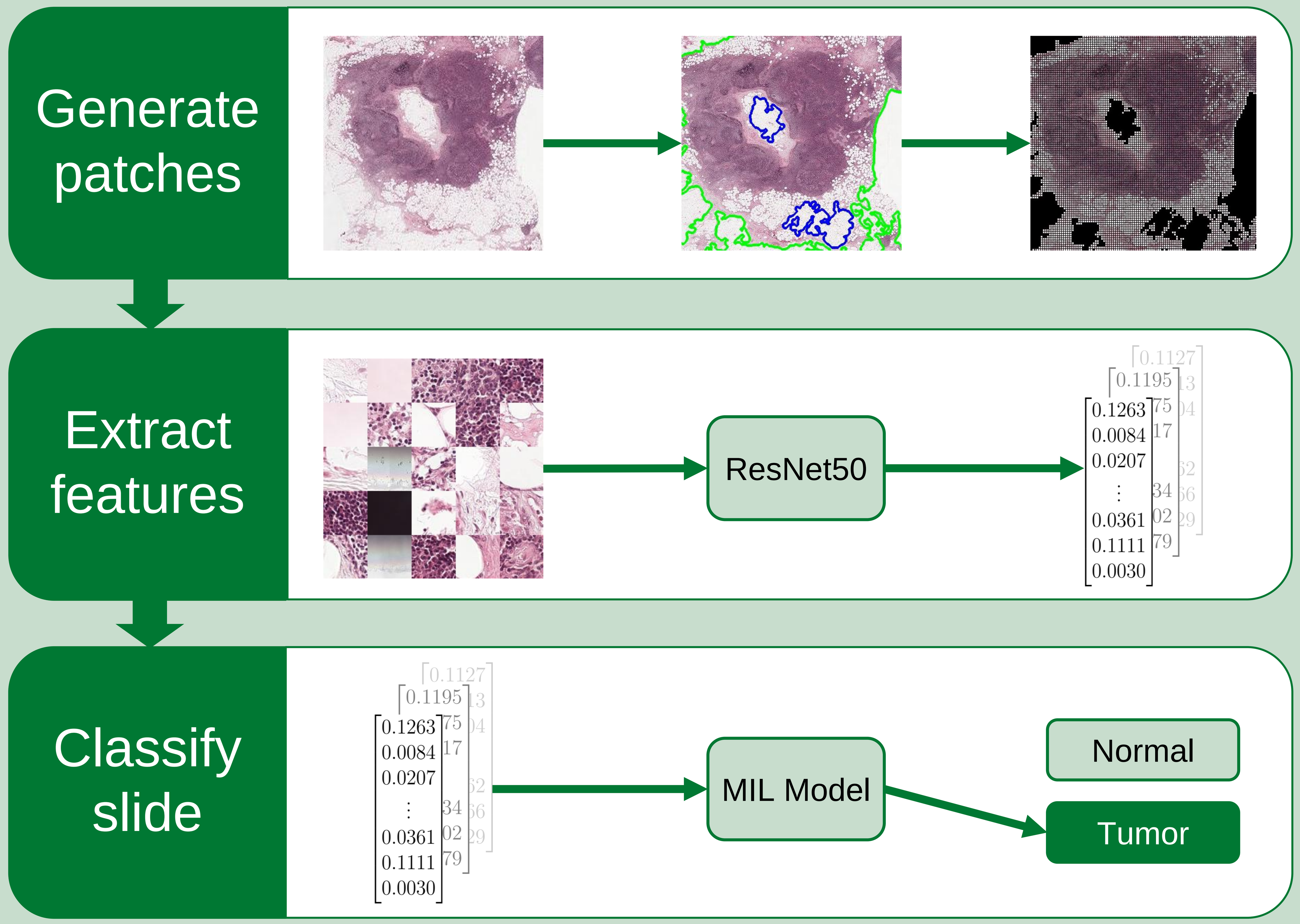
- Cancer pathology involves evaluating tissue samples on glass slides, which is tedious and error-prone
- Whole-slide imaging allows for digital analysis of tissue samples using machine learning
- One state-of-the-art method is multiple instance learning (MIL), which involves splitting the slide into many patches
- Attention-based MIL models have performed better than other MIL models [1]
- Transformer-based models, which use self-attention, have reported a high AUC [2]

We are interested in comparing MIL algorithms to find an optimal model for classifying whole-slide images. Also, we want to verify that Transformer-based algorithms achieve a higher performance than other attention-based algorithms.

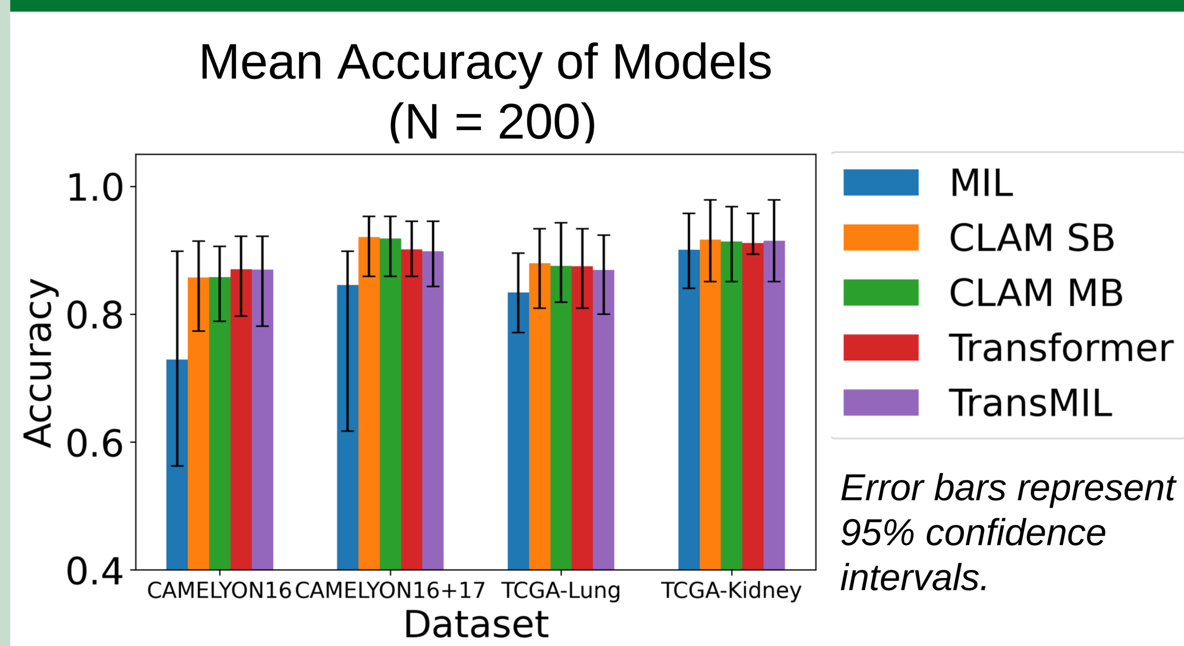
Methods

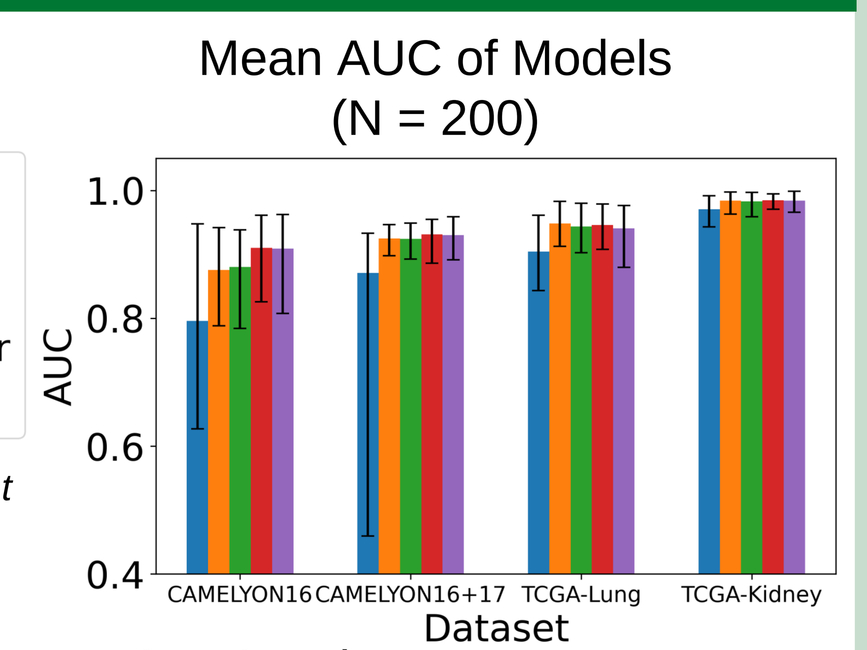
- We trained and tested five MIL models using three datasets: CAMELYON16+17 (breast cancer), TCGA-Lung (lung cancer), and TCGA-Kidney (kidney cancer)
- We estimated the mean accuracy and AUC using bootstrapping over 200 folds

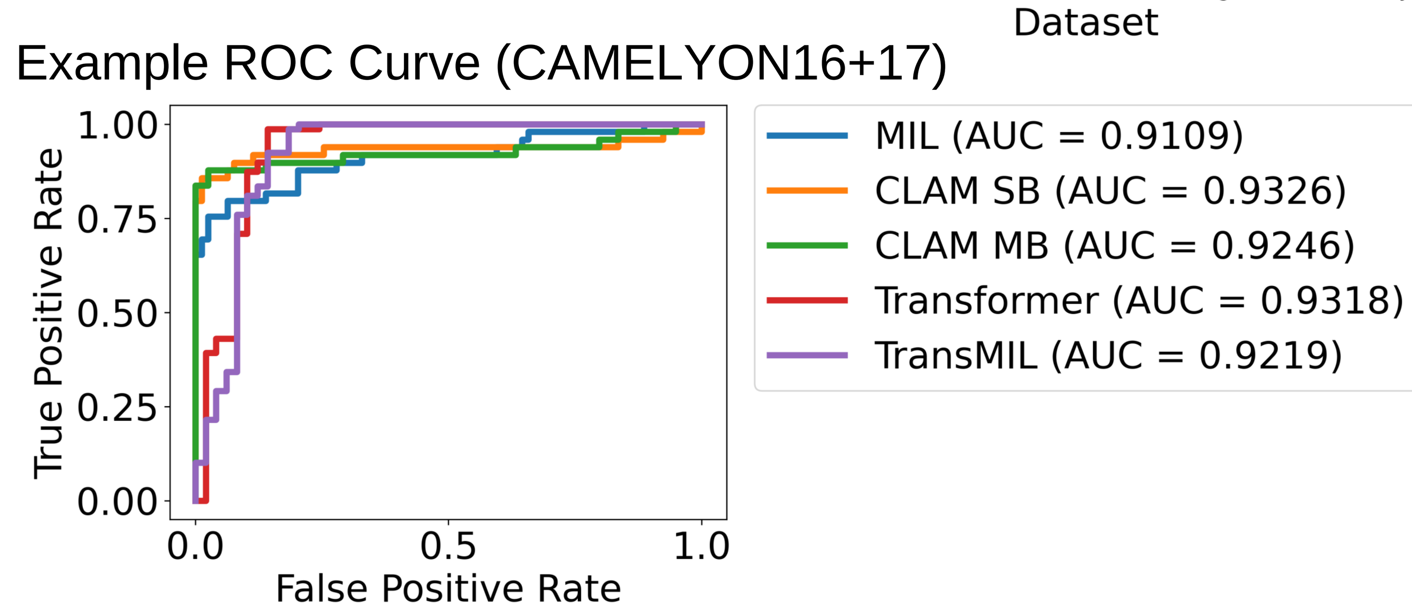
Model	Aggregation Operation	Trainable Parameters
MIL	max pooling	525,826
CLAM SB [1]	single-branch attention	790,791
CLAM MB [1]	multiple-branch attention	791,084
Transformer	Transformer	2,628,114
TransMIL [2]	Transformer with positional encoding	2,672,146



Results







Conclusions

- Attention-based MIL models performed better than the standard MIL model
  - On average, the attention-based MIL models achieved a higher accuracy and AUC than standard MIL models
- No attention-based MIL model clearly outperformed the others
- Transformer-based models did not achieve a higher accuracy and AUC than other attention-based models
  - Also, the Transformer used a lot of memory, which made training difficult
  - Positional encoding did not have a large effect on accuracy or AUC
- Therefore, for whole-slide imaging tasks, we recommend an attention-based MIL model
  - However, for large datasets, we do not recommend Transformer-based models due to the memory required for training
- Further research could focus on improving the Transformer by reducing memory usage

References

[1] M.Y. Lu et al. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biom. Eng.*, **5**, 555-570.
[2] Z. Shao et al., TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification. *35th Conference on Neural Information Processing Systems.*

Further Information

Please contact saundersa5@udayton.edu or yoonh@ornl.gov.

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships program. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Oak Ridge National Laboratory
U.S. Department of Energy | Office of Science
University of Dayton
Oak Ridge Institute for Science and Education — Shaping the Future of Science