DEPARTMENT of BIOSTATISTICS EPIDEMIOLOGY & INFORMATICS
DBEI & CCEB RESEARCH DAY #2022ResearchDay

Attention-based multiple instance learning with self-supervision to predict microsatellite instability in colorectal cancer from histology whole-slide images

Jacob S. Leiby[1], Jie Hao[1], Gyeong Hoon Kang[2,3,4], Ji Won Park[3,4,5], Dokyoon Kim, PhD[1,6]

[1]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, PA 19104, USA; [2]Department of Pathology, Seoul National University Hospital, Seoul, South Korea; [3]Seoul National University College of Medicine, Seoul, South Korea, South Korea; [4]Cancer Research Institute, Seoul National University, Seoul, South Korea; [5]Department of Surgery, Seoul National University Hospital, Seoul, South Korea; [6]Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA

Perelman School of Medicine UNIVERSITY of PENNSYLVANIA

Background

- Microsatellite instability (MSI) is a state of genetic hypermutation caused by defects in the mismatch repair system. It occurs in roughly 15% of all colorectal cancers [1]. It is clinically relevant as tumors expressing this pattern have shown the highest response rates to immunotherapies as well as improved overall survival [2]. MSI is determined using genetic analyses; however, these analyses are often limited to larger tertiary care centers and may add additional time and cost during diagnosis.
- Histology whole-slide images (WSIs) are routinely collected and are a rich source of data for extracting tumor characteristics, including MSI. Given their size, they first are tiled into smaller image patches to be analyzed. However, due to intra-tumor heterogeneity, not every region of the slide is relevant to the outcome.
- Fully-supervised image analysis models have limitations, as they treat each region of the slide as informative to outcome. Additionally, many are pretrained on natural images and therefore do not map medical images to informative embeddings.
- In this project, I propose to use a multiple instance learning framework with the additional use of self-supervised learning in order to overcome intra-tumor heterogeneity and improve representation learning.

Method

- The proposed model consists of a feature extraction network, an attention sub-network, attention-weighted pooling, and a classifier network (Fig. 1). It learns on a bag of tiles at a time. The learnable attention scores $\boldsymbol{a}$ and feature maps $\boldsymbol{f}$ are associated with each tile in the bag, and are then aggregated into a single representation, $\bar{f}$, by summing the attention-weighted feature maps. This is passed into a classifier to predict the outcome for the entire bag.
- In addition to the binary cross-entropy loss, we include a contrastive learning loss:

$$\mathcal{L}_{\text{con}} = -\frac{1}{N}\sum_{i,j=1}^{N} \log \frac{\exp\left(\text{sim}(f_i,\bar{f}_j)/\tau\right)}{\sum_{k=1}^{B} \mathbb{1}_{[k\neq j]}\exp\left(\text{sim}(f_i,\bar{f}_k)/\tau\right)}$$

which urges the model to learn representative embedding mapping.

- The model is trained and validated in two datasets.

Accepted in IEEE Engineering in Medicine & Biology Conference, 2022

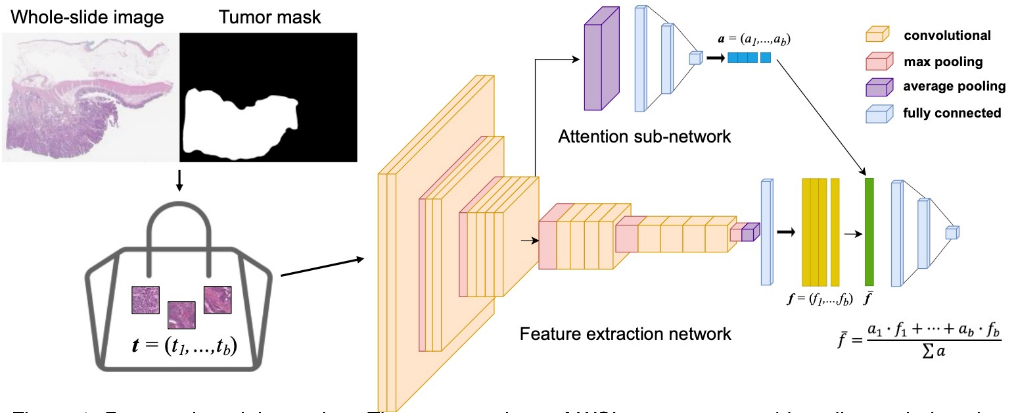

Figure 1. Proposed model overview. The tumor regions of WSIs are segmented into tiles and placed into bags as model input. The model consists of a feature extraction network, an attention sub-network, attention-weighted aggregation $\bar{f}$, and a classifier network.

Result

- The two datasets in this study include an internal and external dataset. The internal dataset is divided into training and testing subsets. The external dataset is used as external validation.
- The results are validated with five-fold cross-validation, in which the training subset was split into five folds. The model is trained on four folds, with one used as a validation set for hyperparameter tuning. The results shown in Table 1 are the performance metrics from the internal test set (not used in model training or tuning).
- We retrained the model on the full training dataset and examined the performance on the external validation set (PAIP) shown in Table 2.
- In both scenarios, we see improvement in model performance when adding the attention-based pooling ('Attention') and the contrastive loss function ('Contrastive').
- The learned attention weights are able to provide visual interpretation of model decisions (Fig 2.). We see that the tiles with the highest attentions are of known pathologic markers of MSI.

TABLE I. FIVE-FOLD CROSS-VALIDATION PERFORMANCE

Model	AUC	AUPRC
VGG19 Baseline	0.822 (0.02)	0.624 (0.05)
ResNet18 Baseline	0.828 (0.01)	0.689 (0.01)
VGG19 + Attention	0.861 (0.01)	**0.729 (0.02)**
VGG19 + Attention + Contrastive	**0.864 (0.01)**	0.690 (0.04)

TABLE II. EXTERNAL DATASET PERFORMANCE

Data	Model	AUC	AUPRC
PAIP	VGG19 Baseline	0.770	0.512
	ResNet18 Baseline	0.686	0.434
	VGG19 + Attention	0.795	0.629
	VGG19 + Attention + Contrastive	**0.876**	**0.793**

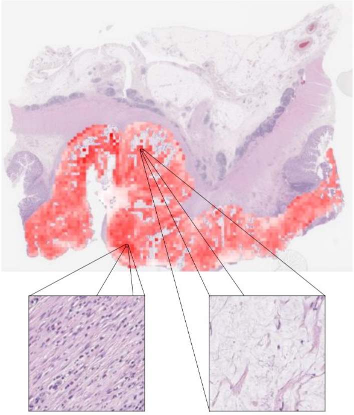

Figure 2. Heatmap showing tile attention scores for a well-predicted MSI patient, darker colors show higher attention weights assigned to those regions. Example top ranked tiles show features associated with MSI.

Conclusion

- In this study, we applied multiple instance and self-supervised learning techniques to predict MSI from WSIs. We found that our proposed method performed better than fully-supervised benchmark models.
- Additionally, we are able to visualize model decisions through attention scores. Our model was able to identify known features relevant to MSI including tumor infiltrating lymphocytes and the presence of excess mucin.

References

[1] Boland, C. R. & Goel, A. Microsatellite Instability in Colorectal Cancer. Gastroenterology, Elsevier BV, 2010, 138, 2073-2087.e3
[2] Le, D. T. et al. . PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. New England Journal of Medicine, Massachusetts Medical Society, 2015, 372, 2509-2520