DEEP LEARNING METHODS FOR AUTOMATED GLEASON GRADING

BDD Team Gleason Grading: Yujie Zhao, Ananya Tandri, Ruitao Hu, Yuxin Du

Project Mentors: Zhenzhen Wang, Adam Charles

Clinical Background [1][2][3]:

- Prostate cancer is the #2 cancer of men worldwide
- Gleason Grading is a reliable method of determining the severity of prostate cancer and planning treatment
- Estimation of Gleason grades requires expert pathologists → problem for low-resource settings

Problem Statement:

Deep learning image analysis can be a useful tool, but previous models are

- Non-generalizable
- Trained on small datasets
- Don't account for differences in pathologist determination.

Method:

Preprocess
WSI → disjoint 256 x 256 pixel Patches

↓

Model Construction
Pretrained DenseNet + MLP

↓

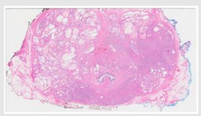
Model Training
+ Data augmentation to prevent overfitting

↓

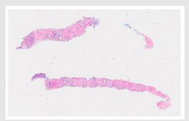
Postprocess
Predicted Logits → Risk heatmap

Dataset:

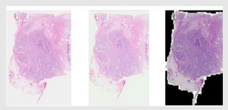
AGGC 2022 dataset [4]
Ground truth provided by multiple pathologists



Subset 1: Whole Mount Images



Subset 2: Biopsy Images



Subset 3: Whole Mount, Multiple Scanners

Conclusions:

We developed a deep learning model to automatically identify all five Gleason patterns with an accuracy of 0.74

Future Direction:

- Determine generalizability of methods/model to other cancer classification tasks
- Determine clinical utility of final product, refine model and outputs accordingly

Results:

① Subset-dependent model performs better than Subset-independent model

	Training Resources	Accuracy	Averaged
Subset-dependent	Subset1	0.67	0.74
	Subset2	0.79	
	Subset3	0.75	
Subset-independent	All Subset	0.66	0.66

② Morphological transformation further improved results

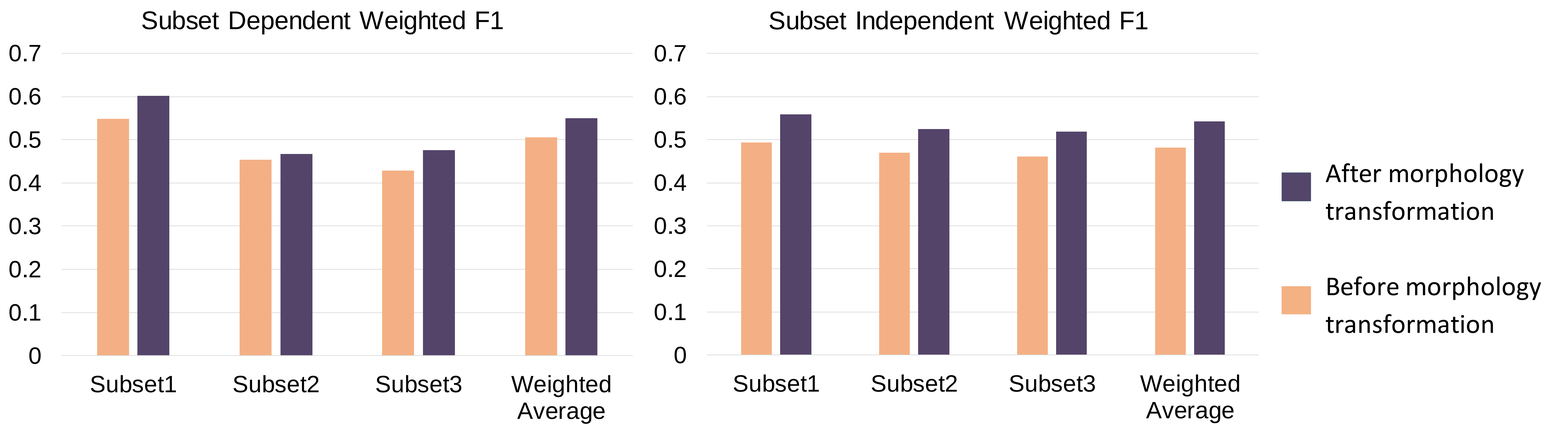
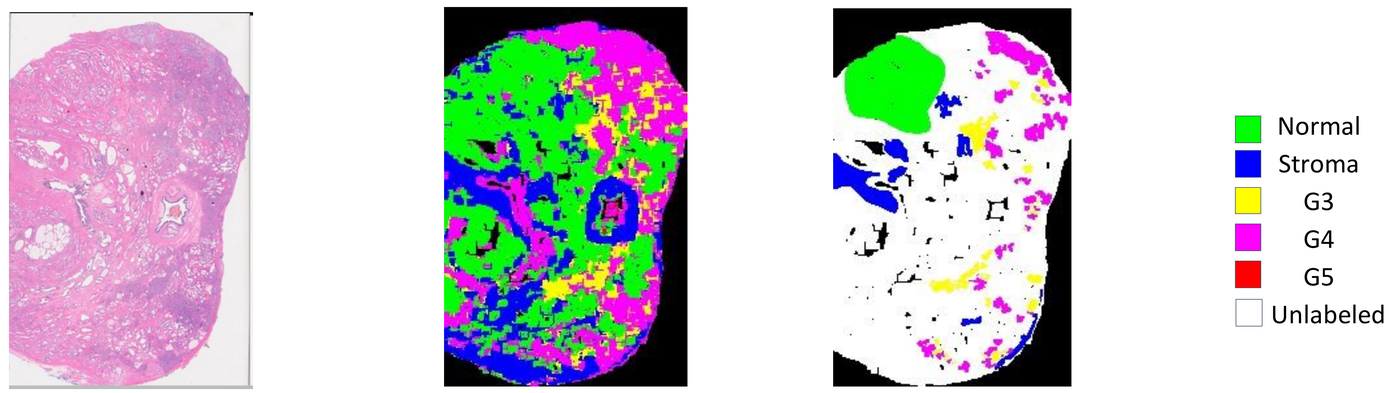


Figure 1: Comparison of weighted F1 with and without morphology transformation.
F1=2×Precision×Recall/(Precision+ Recall); Precision=TP/(TP+FP); Recall=TP/(TP+FN)
Subset-wise Weighted F1= 0.25 * F1_G3 + 0.25 * F1_G4 +0.25 * F1_G5 +0.125 * F1_Normal +0.125 * F1_Stroma
Total Weighted Average F1 = 0.6* weighted F1_subset_1 + 0.2* weighted F1_subset_2 + 0.2* weighted F1 _subset_3

③ Predicted heatmap shows good alignment with ground truth



Input WSI | Predicted heatmap | Ground truth

Figure 2: An example of the predicted heatmap in comparison with ground truth. Heatmap opaqueness was adjusted by confidence level, where more transparent regions indicate lower confidence level.

[1] Egevad et al, BJU International, 2002, [2] Chen et al., Virchows Archiv, 2016 [3] Karimi et al., IEEE Biomedical Health, 2019 [4] Automated Gleason Grading Challenge, MICCAI, 2022