Automated Abnormality Detection in Histopathological Images with Deep Learning

Lyndon Chan, Mahdi Hosseini, Konstantinos Plataniotis

Multimedia Lab, Department of Electrical & Computer Engineering, University of Toronto

The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, UNIVERSITY OF TORONTO

Introduction

Currently, the "gold standard" in diagnosing a considerable number of diseases is for pathologists to analyze a histopathological slide of a tissue biopsy under the microscope [1]. But with the advent of high-accuracy deep learning models for general image recognition and the concurrent development of high-fidelity digital pathology scanners to digitize glass slides into whole slide images (WSIs) in recent years, computer algorithms have increasingly been developed to assist pathologists in analyzing digitized slides (known as computational pathology).

Some proposed tools identify particular diseases, such as classifying cancerous breast tissue as IDC or non-IDC [2] or colon tissue as cancerous or non-cancerous [3]; others detect specific diagnostic components, such as mitotic figures in breast tissue [4] or glands in colon tissue [5].

However, these tools require labelled exemplars of specific tissue types exhibiting particular diseases, and hence cannot be generalized to handle a wide range of tissues or conditions with insufficient data. We propose to approach computational pathology from a novel direction: to predict general tissue type and infer abnormalities from low confidence predictions.

Summary

Given a set of WSI patches labelled with a variety of tissue types, we train a deep learning model to predict tissue types at the patch level. Then, we apply this trained model to predict tissue type in an unseen WSI and infer tissue abnormality at the slide level. Finally, we present preliminary results for both tasks to demonstrate the feasibility of this novel approach to slide-level abnormality detection.

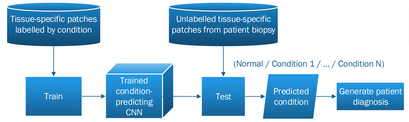


Figure 1: Conventional approach to Computational Pathology – discriminate between finite conditions

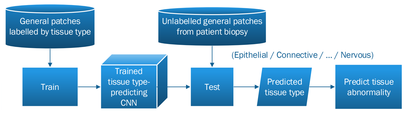
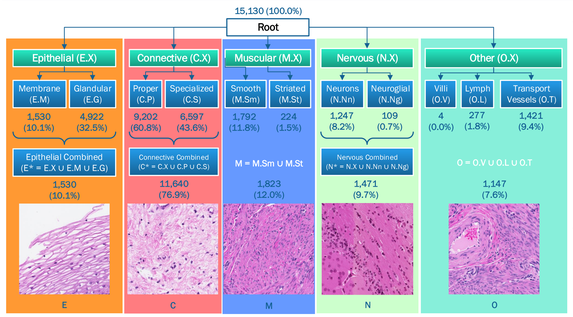


Figure 2: Our proposed approach to Computational Pathology – discriminate between normal tissue types, infer abnormality from low confidence predictions

Data



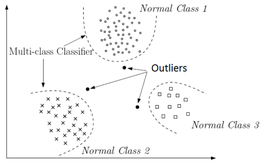
A total of 15,130 non-overlapping patches (sized 1088x1088 pixels) from 99 digitized glass slides were provided by Huron Digital Pathology, each labelled with at least one of 14 hierarchical tissue types known to exist in the gastrointestinal tract (cf. blue tree nodes in Figure 3).

Figure 3: Distribution of available labelled patches, as organized in a two-level hierarchy

Methodology

1. Obtain tissue patches, labelled with multiple tissue types, then resize to 299x299 pixels
2. Modify pre-trained Inception-V3 for 14-class multi-label tissue type classification (replace last two layers with softmax and binary-class classifier)
   a. Train in MATLAB, with 10-fold cross-validation (8:1:1 training-validation-test split), 20-epoch early stopping, fixed learning rate of 2e-5
3. Evaluate trained network's tissue type classification performance on 10-fold cross-validation test sets
   a. Obtain ROC performance analysis on test sets
   b. For novel WSI, visualize maximum-confidence labels k^* scaled by confidence score p_k across the $K = \{1,\ldots,14\}$ classes (cf. $V(x,y)$ in Eqn. 1)
4. Apply trained network on novel WSI and detect abnormality
   a. Calculate patch-level abnormality score (cf. $A(x,y)$ in Eqn. 2) and overlay on WSI

Figure 4: Abnormalities are defined as low-confidence class estimates [6]



$$V(x,y) = k^*(x,y)\, p_{k^*(x,y)}(x,y), \quad k^*(x,y) = \arg\max_{k \in K} p_k(x,y) \qquad \text{(Eqn. 1)}$$

$$A(x,y) = 1 - N(x,y), \quad N(x,y) = \max_{k \in K} p_k(x,y) \qquad \text{(Eqn. 2)}$$

Results

1) Tissue Type Classification

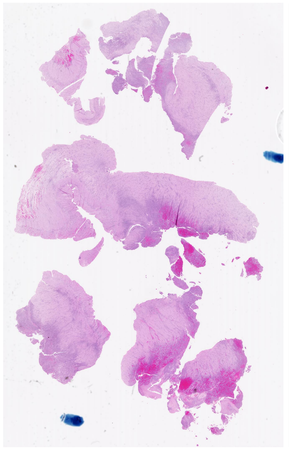


Figure 5: Novel WSI #PP48

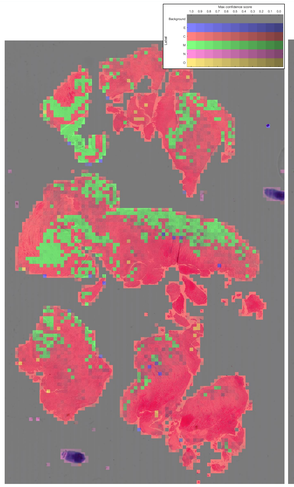


Figure 6: Visualization of maximum-confidence labels, scaled by confidence score, $V(x,y)$, for top level only, overlaid on WSI #PP48

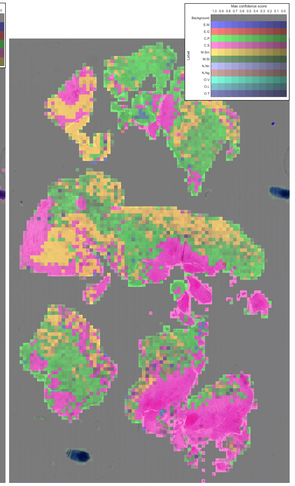


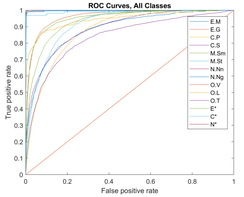
Figure 7: Visualization of maximum-confidence labels, scaled by confidence score, $V(x,y)$, for second level only, overlaid on WSI #PP48

Results (continued)

	TOTAL	E.M	E.G	C.P	C.S	M.Sm	M.St	N.Nn	N.Ng	O.V	O.L	O.T	E*	C*	N*
Threshold	-	0.81%	26.16%	61.29%	41.83%	2.58%	0.94%	8.17%	0.02%	0.00%	0.07%	1.00%	21.56%	79.58%	9.73%
TPR	85.52%	77.97%	89.31%	83.44%	78.76%	88.11%	96.43%	99.28%	98.17%	0.00%	85.92%	74.81%	88.90%	86.02%	98.71%
TNR	91.83%	83.97%	90.95%	76.33%	82.56%	90.38%	99.81%	99.70%	99.09%	100.00%	92.63%	80.31%	82.98%	81.69%	99.62%
FPR	8.17%	16.03%	9.05%	23.67%	17.44%	9.62%	0.19%	0.30%	0.91%	0.00%	7.37%	19.69%	17.02%	18.31%	0.38%
FNR	14.48%	22.03%	10.69%	16.56%	21.24%	11.89%	3.57%	0.72%	1.83%	100.00%	14.08%	25.19%	11.10%	13.98%	1.29%
PPV	74.74%	35.37%	82.63%	84.55%	77.74%	55.17%	88.52%	96.72%	43.85%	N/A	17.85%	28.26%	78.52%	94.00%	96.54%
ACC	90.44%	83.36%	90.42%	80.65%	80.91%	90.11%	99.76%	99.66%	99.08%	99.97%	92.50%	79.80%	85.42%	85.02%	99.53%
F1	79.77%	48.66%	85.84%	83.99%	78.25%	67.86%	92.31%	97.98%	60.62%	0.00%	29.57%	41.02%	83.39%	89.83%	97.61%
AUC	-	0.886	0.964	0.881	0.892	0.955	0.992	1.000	0.999	0.500	0.945	0.839	0.935	0.916	0.998

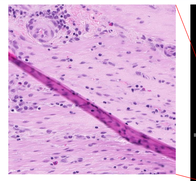
Table 1: Quantitative classification performance of the 10-fold cross-validated trained network, for all 14 tissue types, using optimal thresholds (as determined by Youden's index)

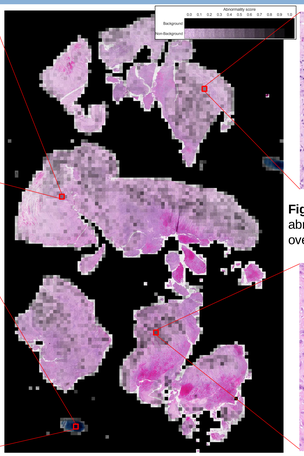
Figure 8: ROC curve, plotted for all 14 tissue types



2) Abnormality Detection

Abnormal Patch #2: folding artifact in tissue





Abnormal Patch #1: tissue appears misshapen (smooth muscle and loose connective tissue?)

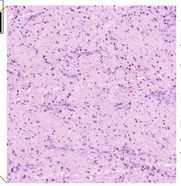
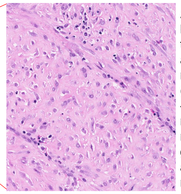


Figure 9: Visualization of abnormality score, $A(x,y)$, overlaid on WSI #PP48

Abnormal Patch #4: blue ink marking on glass slide

Abnormal Patch #3: tissue appears misshapen (transverse-cut smooth muscle with mysterious gaps around nuclei?)



Discussion and Future Directions

The preliminary results are promising for both tissue type prediction and abnormality detection. Closer inspection of the labelled patches indicates that some patches are inconsistently labelled, thus penalizing the model for learning correct predictions. And the abnormality detection demonstration indeed detects spatially-coherent regions of seemingly abnormal tissue appearance. We plan to retrain our model with improved patch labels validated by collaborating pathologists.

References

[1] L. Rorke, "Pathologic diagnosis as the gold standard", *Cancer*, vol. 79, no. 4, pp. 665-667, 1997.

[2] A. Cruz-Roa, A. Basavanhally, F. Gonzalez, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski and A. Madabhushi, "Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks", in *Medical Imaging*, 2014.

[3] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai and E. Chang, "Deep Learning of Feature Representation with Multiple Instance Learning for Medical Image Analysis", in *IEEE International Conference on Acoustic, Speech and Signal Processing*, 2014.

[4] D. Cireşan, A. Giusti, L. Gambardella and J. Schmidhuber, "Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013, pp. 411-418.

[5] W. Li, S. Manivannan, S. Akbar, J. Zhang, E. Trucco and S. McKenna, "Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks", in *IEEE Symposium on Biomedical Imaging*, Prague, 2016.

[6] S. Upadhyaya and K. Singh, "Classification Based Outlier Detection Techniques", *International Journal of Computer Trends and Technology*, vol. 3, no. 2, pp. 294-298, 2012.