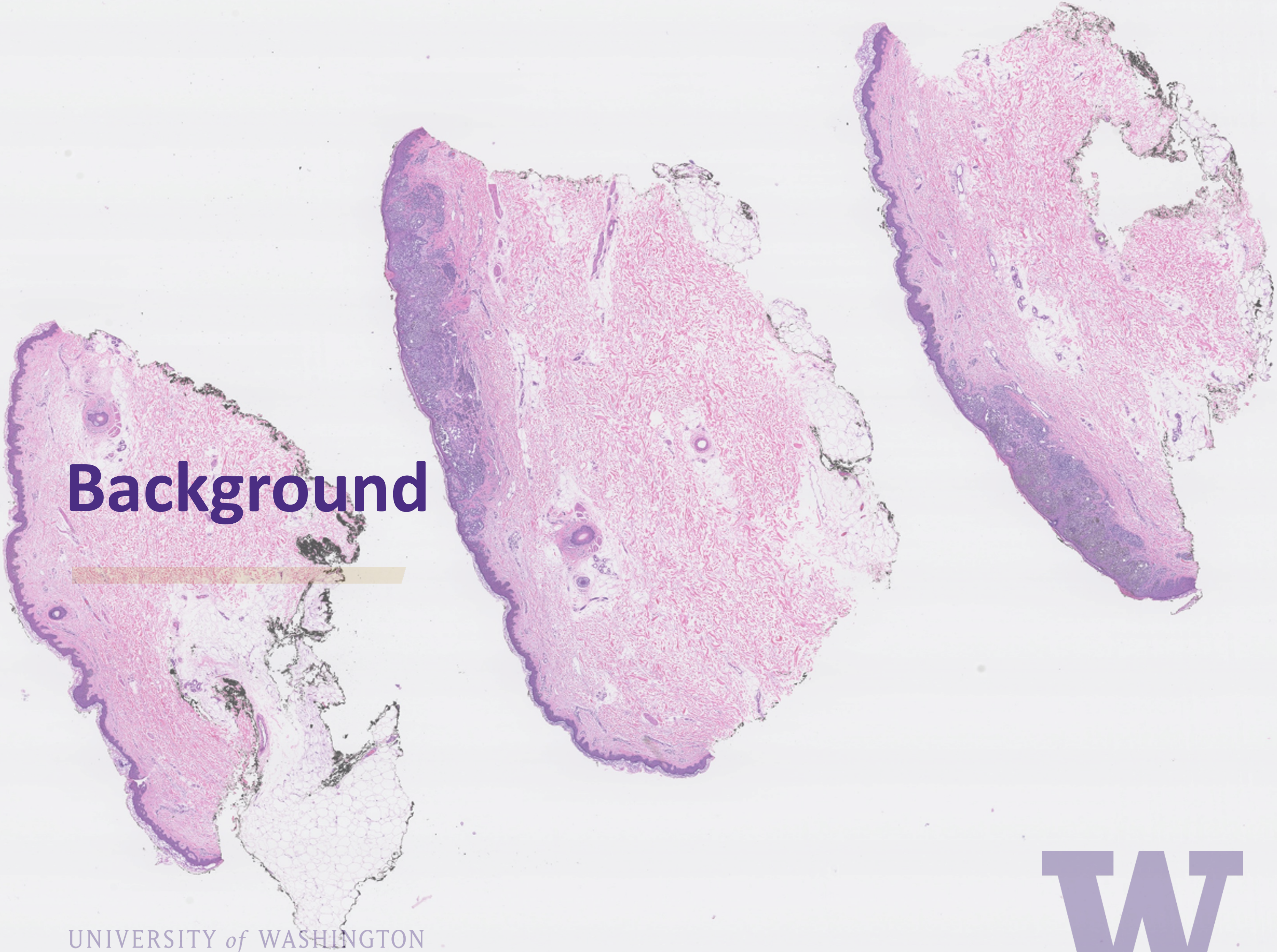
Applications of Visual Transformers for Whole Slide Skin Biopsy Image Diagnosis

Wenjun Wu

Winter 2023

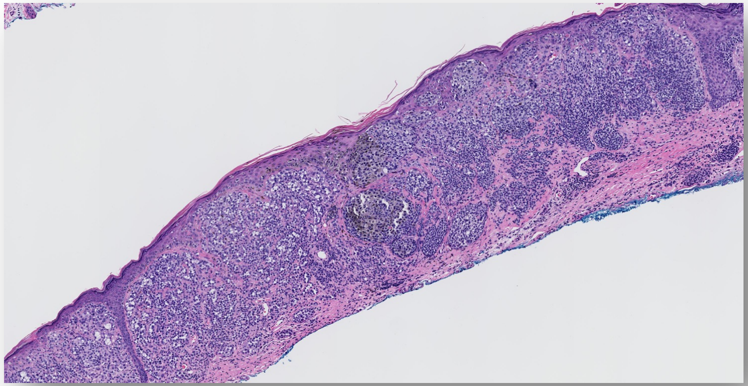
UNIVERSITY *of* WASHINGTON



Background

What is Melanoma?

- Melanoma is the most aggressive type of skin cancer.
- Melanoma occurs when UV radiation triggers DNA damages in melanocytes
- The "gold standard" for diagnosis of invasive melanoma relies on the visual assessments of skin biopsy images by pathologists.

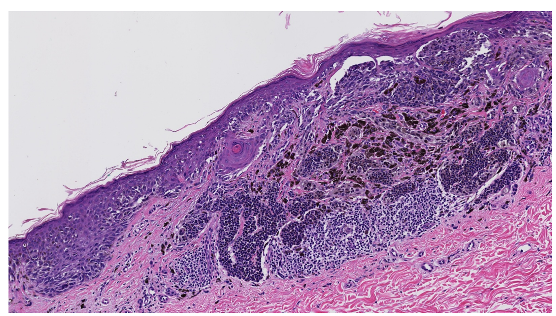


An example of an Invasive Melanoma T1b in M-Path dataset.

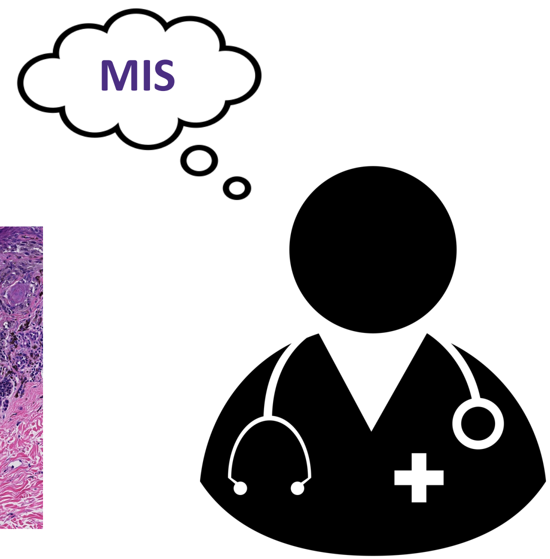
Why melanoma diagnosis?

> Unfortunately, diagnostic errors are common

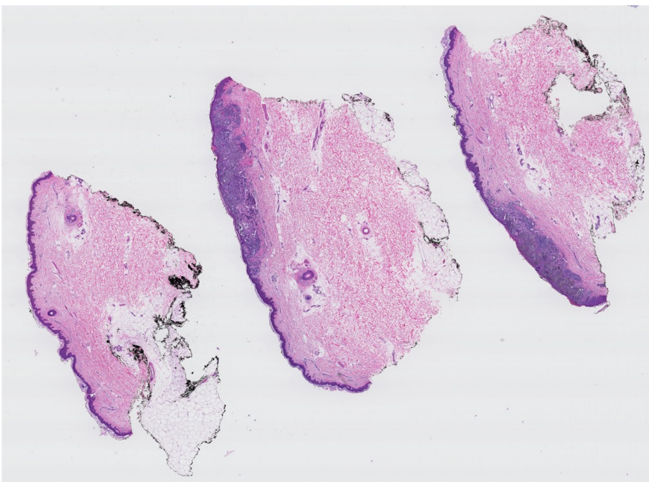
> Computer-aided diagnostic system can be a second reader and help reduce uncertainties

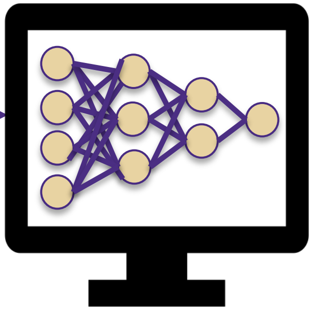


An example of Skin Biopsies of pT1a



Goal

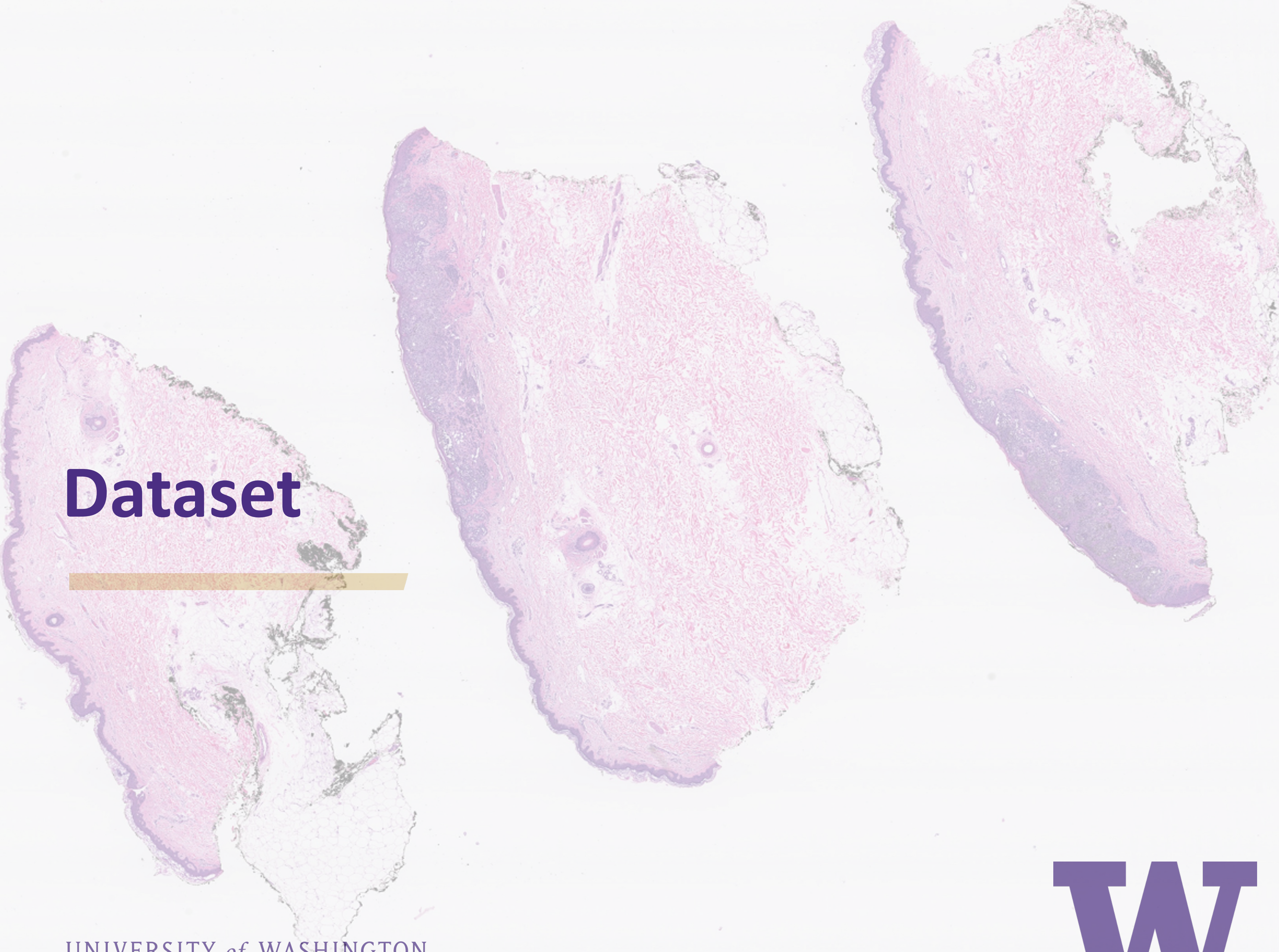




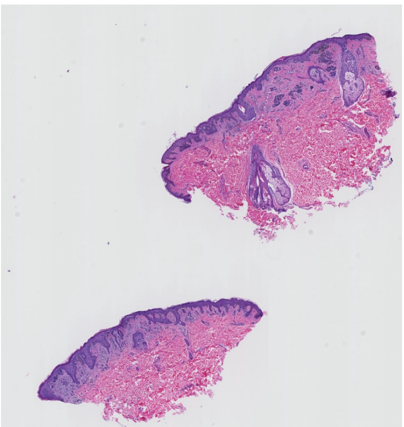
Diagnosis

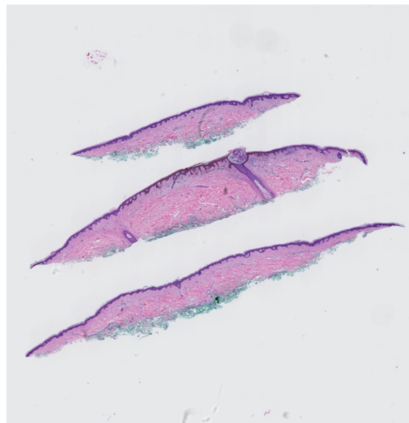
Model

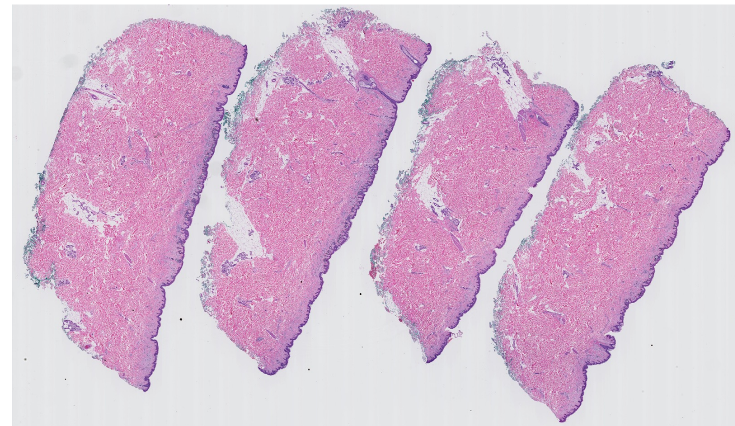
Dataset



Melanoma Dataset







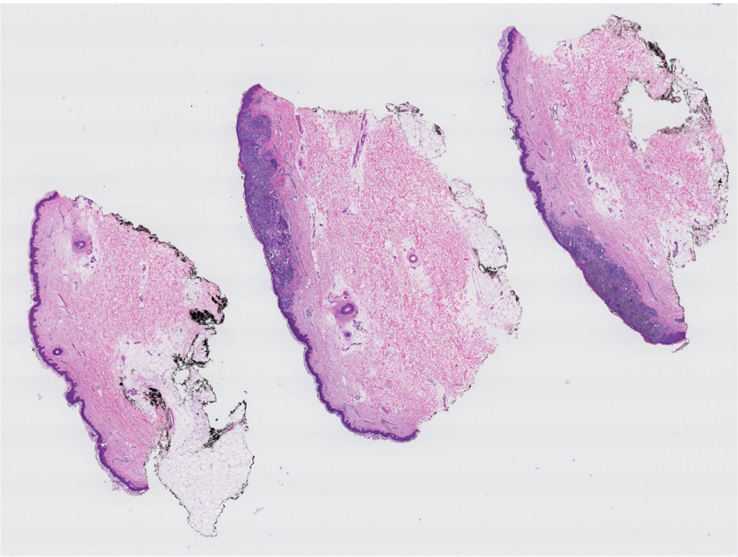
Diagnostic Category	Number of WSIs				Average WSI size (in pixels)
	Training	Validation	Test	Total	
MMD	26	6	29	61	11843×10315
MIS	25	5	30	60	9133×8501
pT1a	33	6	34	73	9490×7984
pT1b	18	6	22	46	14858×12154
Total	102	23	115	240	11130×9603

Statistics of skin biopsy whole slide image (WSI) dataset.
Diagnostic terms for the dataset used in this study are as follows: mild and moderate dysplastic nevi (MMD), melanoma in situ (MIS), invasive melanoma stage pT1a (pT1a), invasive melanoma stage ≥ pT1b (pT1b).

Difficulties in diagnosis

Size of whole slide images (WSI)

An example image from ImageNet [500 x 375]



An example WSI at 10x [15264 x 19824]

Difficulties in diagnosis

Size of whole slide images (WSIs)

Dataset size

Diagnostic Category	Number of WSIs: Training	Validation	Test	Total	Average WSI size (in pixels)
MMD	26	6	29	61	11843×10315
MIS	25	5	30	60	9133×8501
pT1a	33	6	34	73	9490×7984
pT1b	18	6	22	46	14858×12154
Total	102	23	115	240	11130×9603

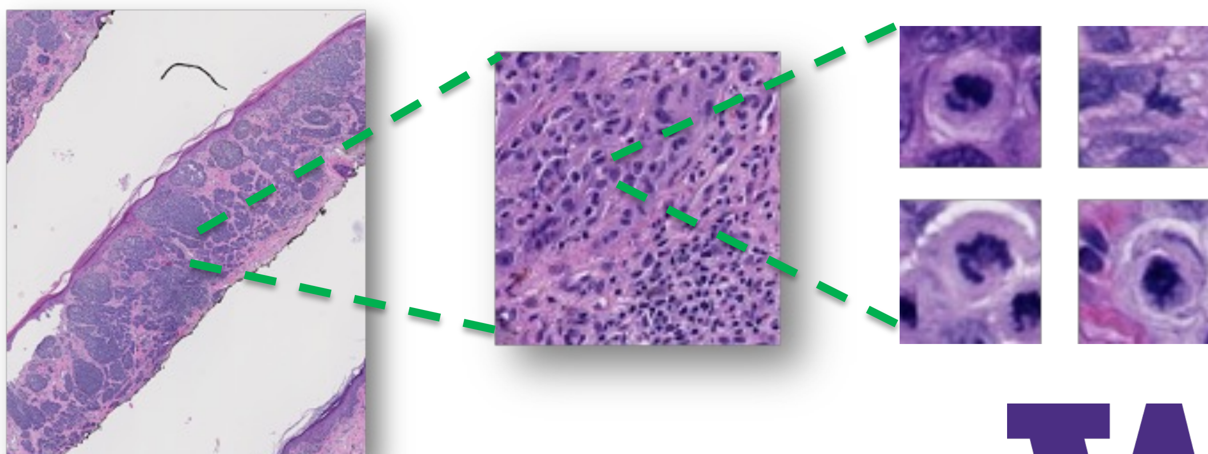
TABLE 1: Statistics of skin biopsy whole slide image (WSI) dataset. The average WSI size is computed at a magnification factor of x10. Diagnostic terms for the dataset used in this study are as follows: mild and moderate dysplastic nevi (MMD), melanoma in situ (MIS), invasive melanoma stage pT1a (pT1a), invasive melanoma stage ≥ pT1b (pT1b).

Difficulties in diagnosis

Size of whole slide images (WSIs)

Dataset size

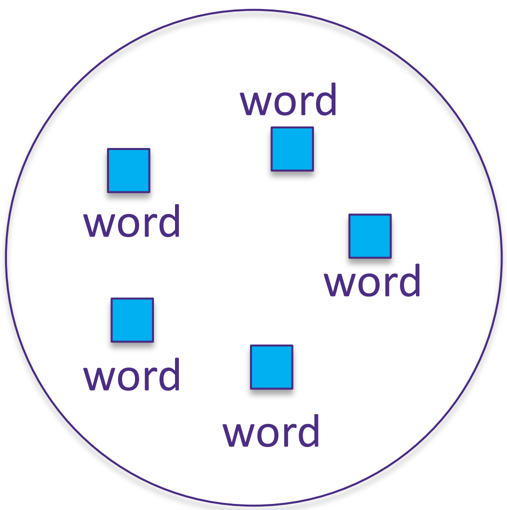
cancerous structure vs. normal structure



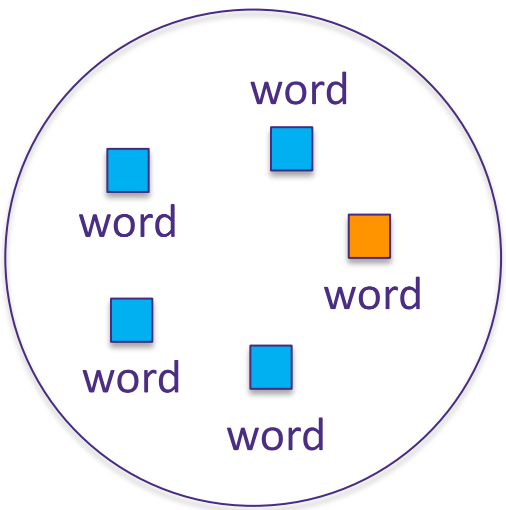
Related Work

Related Work

> Multiple Instance Learning



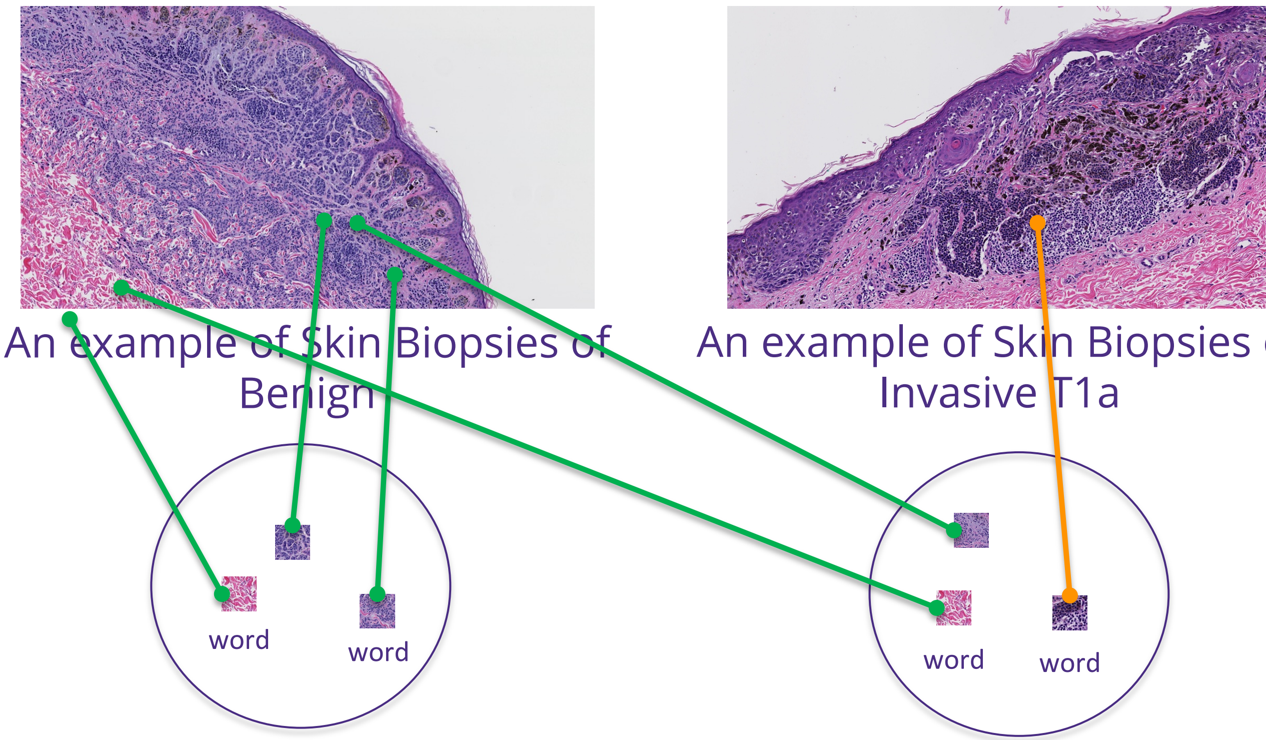
Negative Bag



Positive Bag

Related Work

> Multiple Instance Learning



An example of Skin Biopsies of Benign

An example of Skin Biopsies of Invasive T1a

Negative Bag

positive Bag

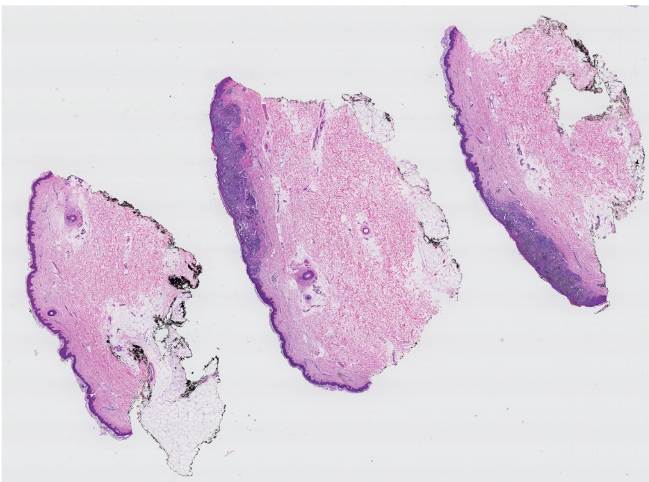
Related Work

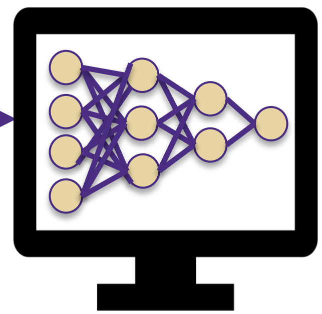
> **Multiple Instance Learning (MIL)**

- \+ reduce high computational cost
- \+ effective in learning instance/bag-wise representation
- \- Does not allow long-range/global feature interaction
- \- Prone to label ambiguity/noise

Related Work

> End-to-End Learning



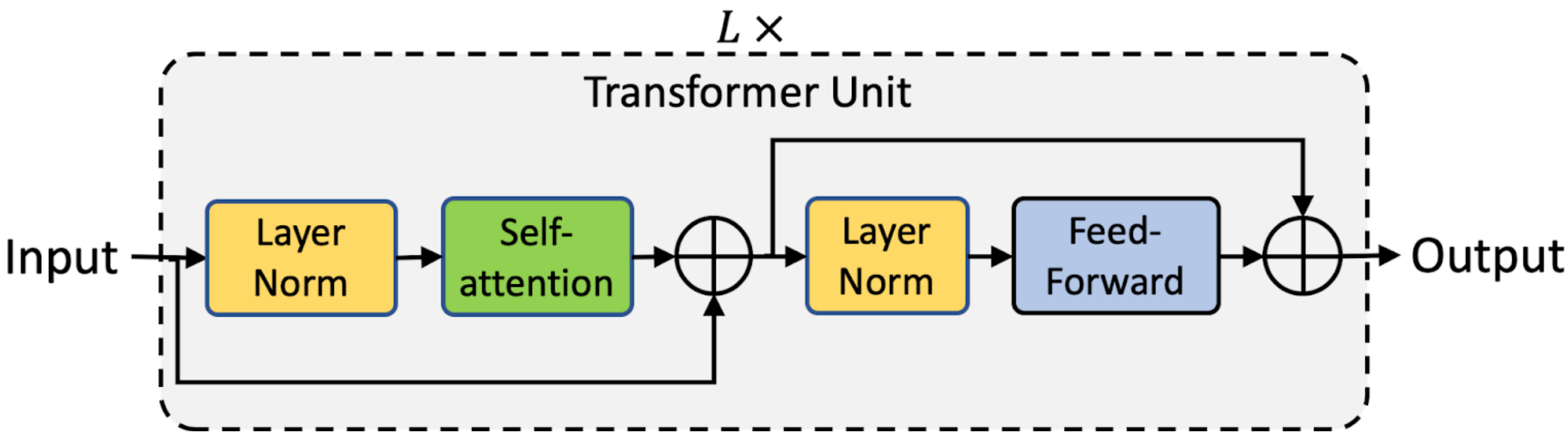


Diagnosis

Model

Related Work

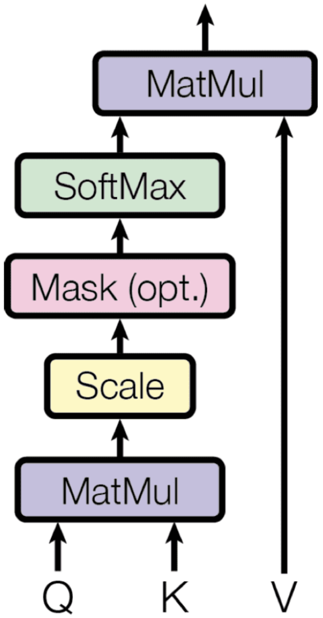
> Visual Transformers



Vaswani *et al.* Attention is all you need. *Advances in neural information processing systems, 30*

Related Work

> **Self-attention**

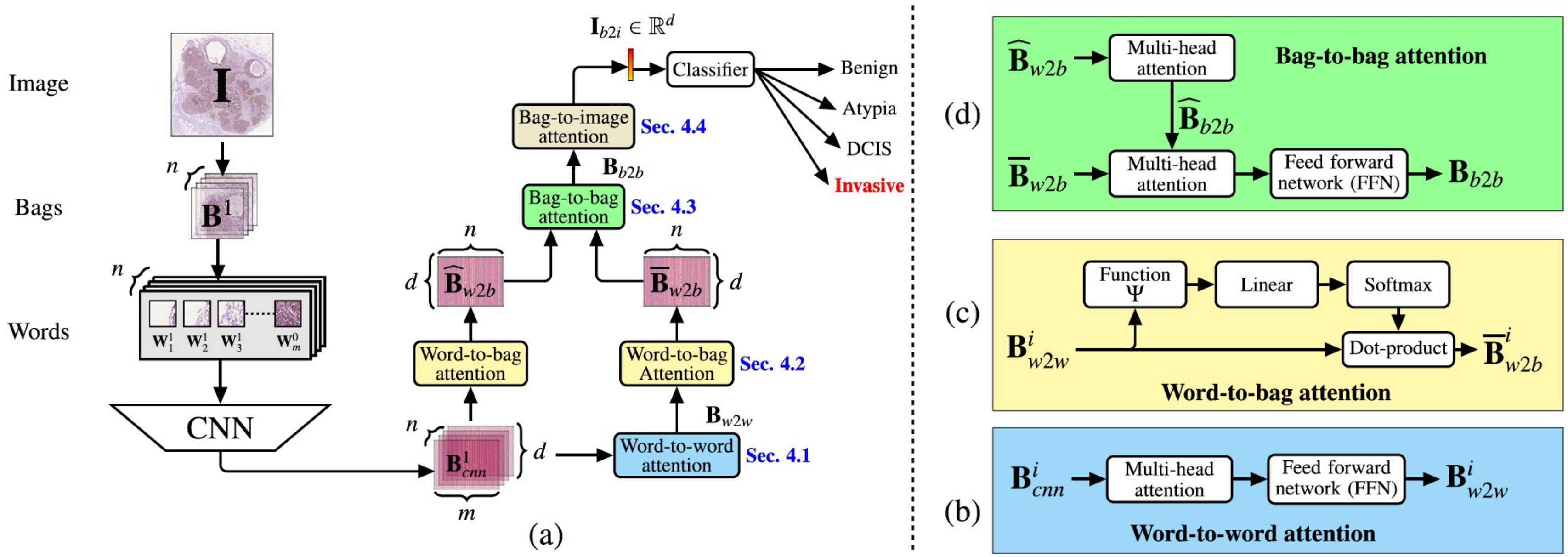


Scaled Dot-Product Attention

Vaswani *et al.* Attention is all you need. *Advances in neural information processing systems, 30*

Our Work

Holistic Attention Network (HATNet)

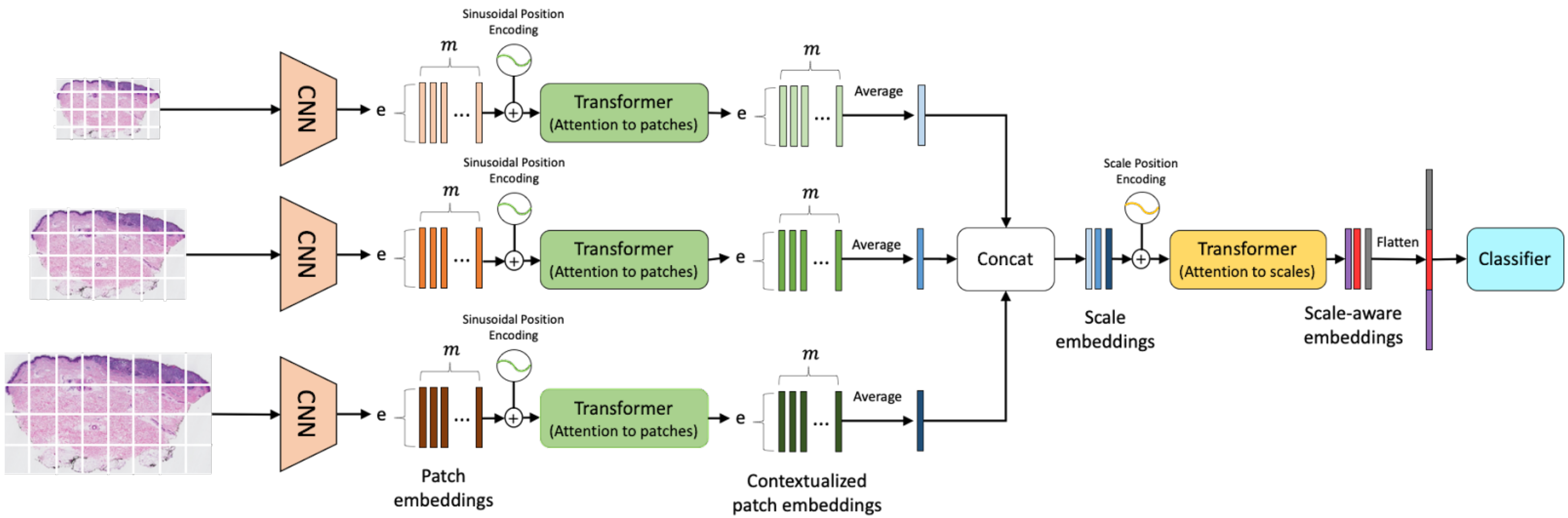


Mehta *et al.* End-to-End Diagnosis of Breast Biopsy Images with Transformers." Medical Image Analysis (2022): 102466.

HATNet (on a breast dataset)

- Outperforms CNN-based methods by a large margin
- Significant overlap between top bags, words and annotations of clinical biomarkers
- Learned representations from clinically relevant tissue structures without any supervision

Scale-Aware Transformer Network (ScAtNet)



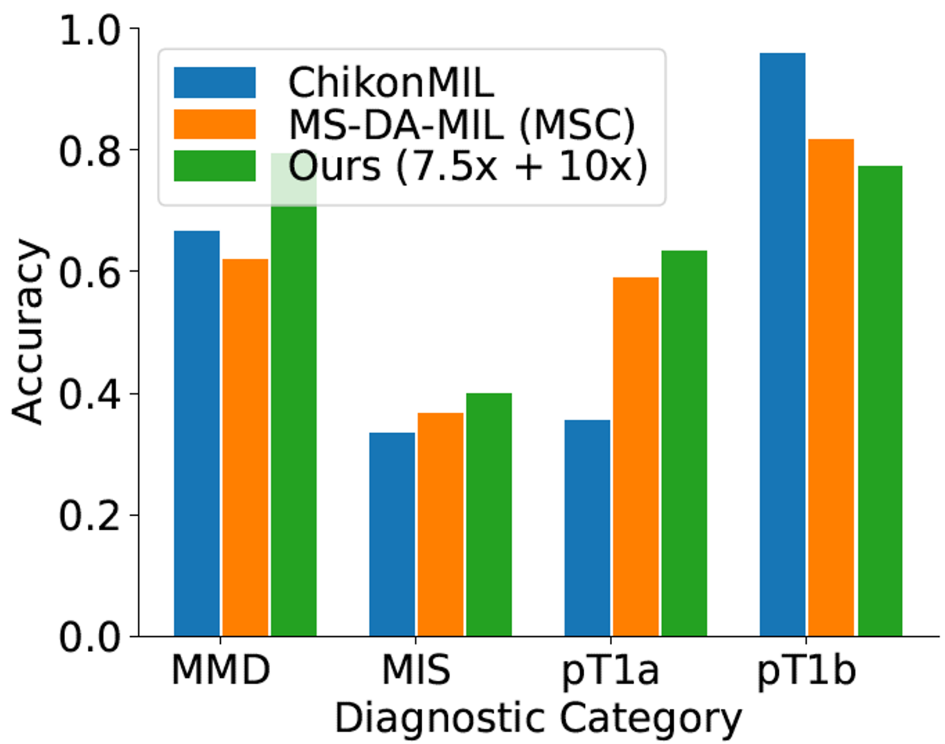
Wu *et al.* Scale-aware transformers for diagnosing melanocytic lesions. IEEE Access, 9:163526–163541, 2021.

Experimental Result: baseline methods

Row #	Method	Accuracy	F1	Sensitivity	Specificity	AUC
R1	Patch-based (SSC)	0.35	0.35	0.35	0.79	0.67
R2	Patch-based (MSC)	0.40	0.40	0.40	0.80	0.68
R3	Penultimate-weighted (SSC)	0.44	0.44	0.44	0.81	0.67
R4	Hypercolumn-weighted (SSC)	0.43	0.43	0.43	0.43	0.67
R5	Streaming CNN (SSC)	0.32	0.32	0.32	0.77	0.58
R6	ChikonMIL (SSC)	0.56	0.56	0.56	0.85	0.74
R7	MS-DA-MIL (SSC)	0.49	0.49	0.49	0.83	0.68
R8	MS-DA-MIL (MSC*)	0.58	0.58	0.58	0.86	0.75
R9	`ScAtNet` (SSC)	0.60	0.60	0.60	0.87	0.77
R10	`ScAtNet` (MSC)	**0.64**	**0.64**	**0.64**	**0.88**	**0.79**

TABLE 2: Comparison of overall performance with state-of-the-art WSI classification methods across different metrics on the test set. Here, SSC denotes single input scale (10×). MSC denotes multiple input scales (7.5×, 10×, 12.5×). MSC* denotes multiple input scales (10×, 20×)

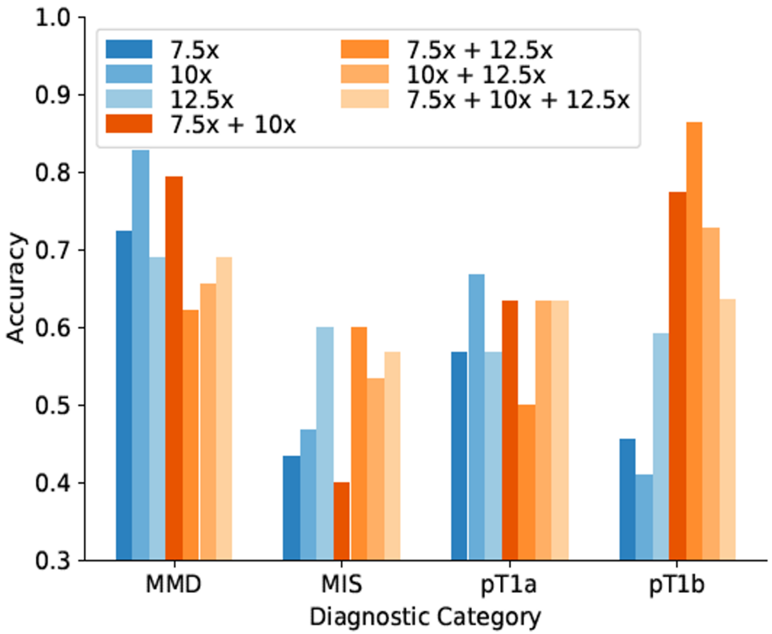
Experimental Result: baseline methods



Experimental Result: single vs. multiple input scales

Input scales 7.5×	Input scales 10×	Input scales 12.5×	Accuracy	F1	Sensitivity	Specificity	AUC
✓			0.55	0.55	0.55	0.85	0.75
	✓		0.60	0.60	0.60	0.87	0.77
		✓	0.61	0.61	0.61	0.87	0.78
✓	✓		0.64	0.64	0.64	0.88	0.79
✓		✓	0.63	0.63	0.63	0.88	0.80
	✓	✓	0.63	0.63	0.63	0.88	0.79
✓	✓	✓	0.63	0.63	0.63	0.88	0.79

(a) Overall performance of `ScAtNet`



(b) Class-wise accuracy of `ScAtNet`

ScAtNet

- > Outperforms MIL and CNN based methods
- > Achieves comparable performance to 187 practicing U.S. pathologists
- > Saliency analysis shows that ScAtNet learns to weigh features from different scales

Input scales			Accuracy	F1	Sensitivity	Specificity	AUC
7.5×	10×	12.5×					
✓			0.55	0.55	0.55	0.85	0.75
	✓		0.60	0.60	0.60	0.87	0.77
		✓	0.61	0.61	0.61	0.87	0.78
✓	✓		0.64	0.64	0.64	0.88	0.79
✓		✓	0.63	0.63	0.63	0.88	0.80
	✓	✓	0.63	0.63	0.63	0.88	0.79
✓	✓	✓	0.63	0.63	0.63	0.88	0.79

(a) Overall performance of `ScAtNet`

Wu *et al.* Scale-aware transformers for diagnosing melanocytic lesions. IEEE Access, 9:163526–163541, 2021.

Limitations

- Limited study on whole slide skin biopsy images (lack of public datasets)
- Limited in-house dataset size
- Mostly binary classification
  - This study covers the full spectrum of melanocytic skin biopsy
- Small test set
  - We have independent test set of 115 WSIs (50%)

Future Work

- Other types of image and cancer
- Learnable scale
- Wider range of scales
- Interpreting choice of scale, class and diagnosis accuracy
- Comparing viewing behavior with pathologists

Acknowledgement

Research reported in this study was supported by grants R01CA200690 and U01CA231782 from the National Cancer Institute of the National Institutes of Health, 622600 from Melanoma Research Alliance, and W81XWH-20-1-0798 from the United States Department of Defense.

Advisor:
Dr. Linda Shapiro

PI:
Dr. Joann Elmore

Pathologists:
Dr. Stevan Knezevich
Dr. Caitlin May
Dr. Oliver Chang
Dr. Mojgan Mokhtari

Collaborators:
Shima Nofallah
Dr. Sachin Mehta

Reference

[1] P. Chikontwe, M. Kim, S. J. Nam, H. Go, and S. H. Park, “Multiple instance learning with center embeddings for histopathology classification,” in International Conference on Medical Image Computing and Computer- Assisted Intervention. Springer, 2020, pp. 519–528.

[2] C. Mercan, B. Aygunes, S. Aksoy, E. Mercan, L. G. Shapiro, D. L.Weaver, and J. G. Elmore, “Deep feature representations for variable-sized regions of interest in breast histopathology,” IEEE Journal of Biomedical and Health Informatics, 2020.

[3] E. Mercan, L. G. Shapiro, T. T. Brunyé, D. L. Weaver, and J. G. Elmore, “Characterizing diagnostic search patterns in digital breast pathology: scanners and drillers,” Journal of digital imaging, vol. 31, no. 1, pp. 32–41, 2018.

[4] H. Pinckaers, W. Bulten, J. Van der Laak, and G. Litjens, “Detection of prostate cancer in whole-slide images through end-to-end training with image-level labels,” IEEE transactions on medical imaging, vol. PP, March 2021.

[5] N. Hashimoto, D. Fukushima, R. Koga, Y. Takagi, K. Ko, K. Kohno, M. Nakaguro, S. Nakamura, H. Hontani, and I. Takeuchi, “Multi-scale domain-adversarial multiple-instance cnn for cancer subtype classification with unannotated histopathological images,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3852–3861.

[6] Elmore et al., “Diagnostic concordance among pathologists interpreting breast biopsy specimens,” JAMA, 2015.

[7] J. G. Elmore, R. L. Barnhill, D. E. Elder, G. M. Longton, M. S. Pepe, L. M. Reisch, P. A. Carney, L. J. Titus, H. D. Nelson, T. Onega et al., “Pathologists’ diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study,” Bmj, vol. 357, 2017.

[8] K. H. Allison, L. M. Reisch, P. A. Carney, D. L. Weaver, S. J. Schnitt, F. P. O’Malley, B. M. Geller, and J. G. Elmore, “Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel,” Histopathology, vol. 65, no. 2, pp. 240–251, 2014.

Q&A

UNIVERSITY *of* WASHINGTON