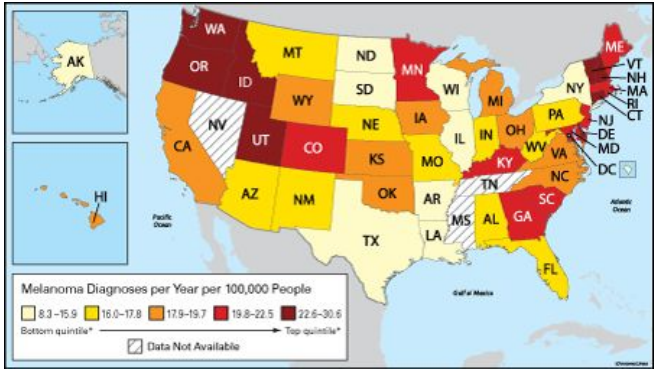
Scale-Aware Transformers for Diagnosing Melanocytic Lesions

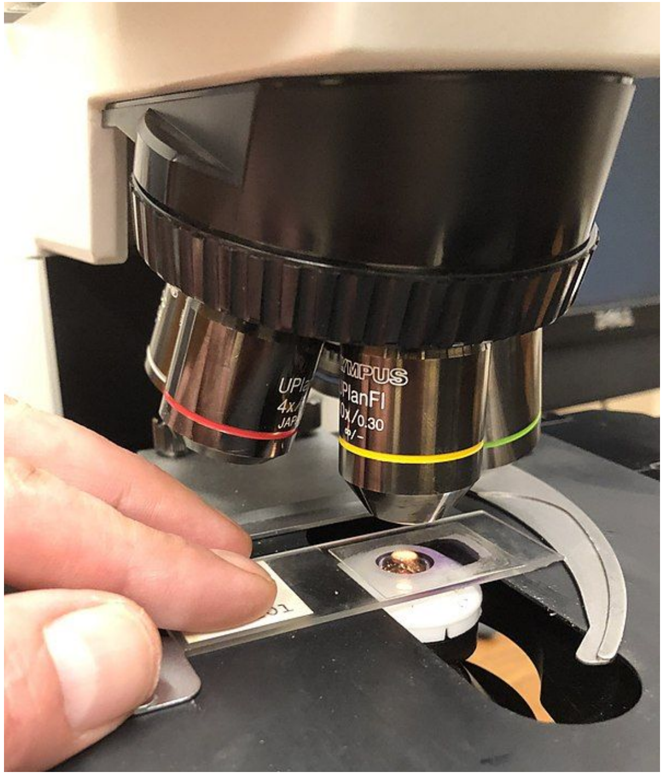
Wenjun Wu , Sachin Mehta , Shima Nofallah , Stevan Knezevich , Caitlin J. May , Oliver H. Chang , Joann G. Elmore and Linda G. Shapiro

Melanoma

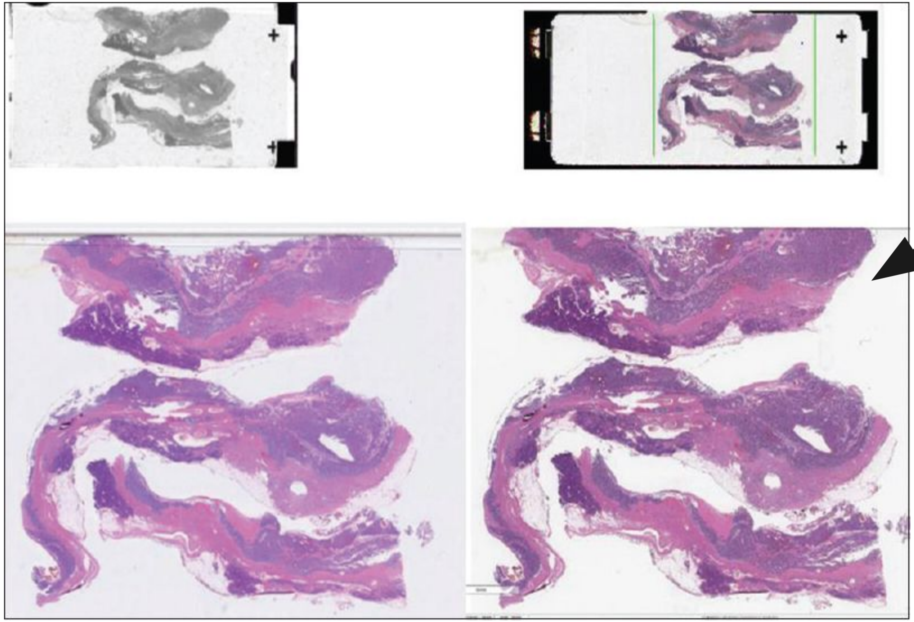
- Melanoma is the most aggressive type of skin cancer
- One of the most diagnosed cancers in the US
- Gold standard for diagnosis → visual assessment of skin biopsy by pathologists



Histology Examination



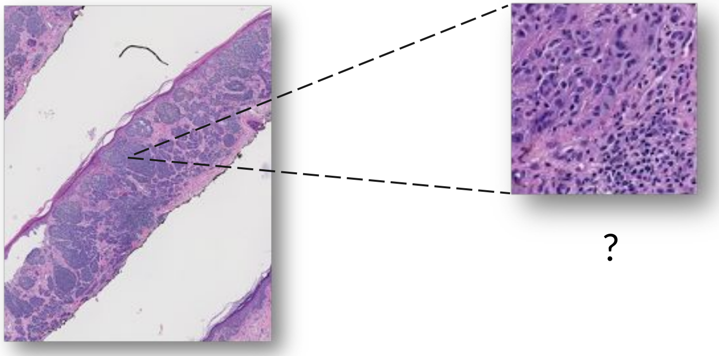
Digitized Whole slide Images (WSI)



Multiple tissues

Difficulties in diagnosis

Mixed normal and cancerous tissue

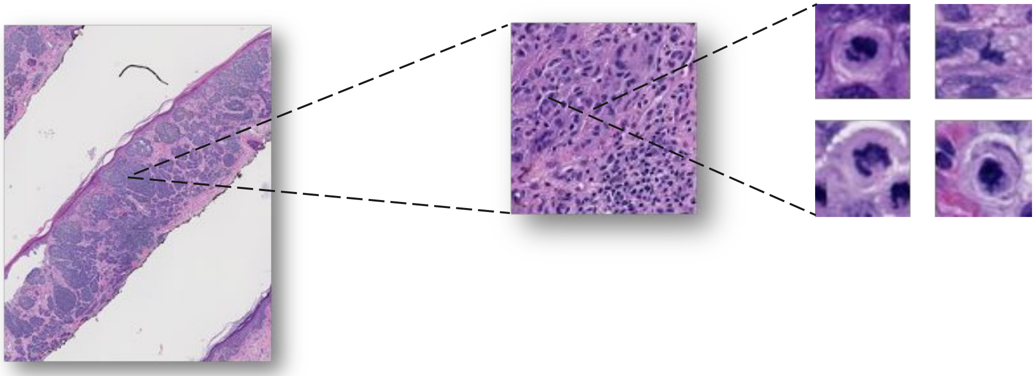


?

Difficulties in learning to diagnose

Mixed normal and cancerous tissue

Feature is dependent on resolution



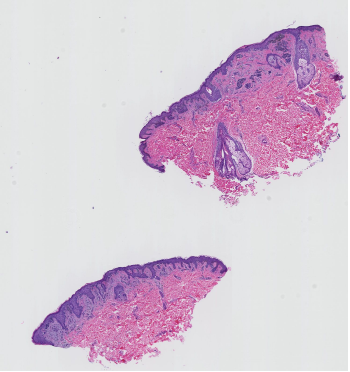
Difficulties in learning to diagnose

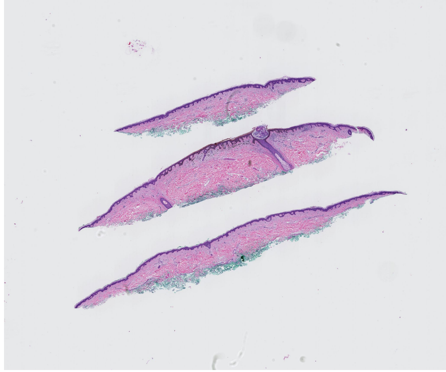
Mixed normal and cancerous tissue

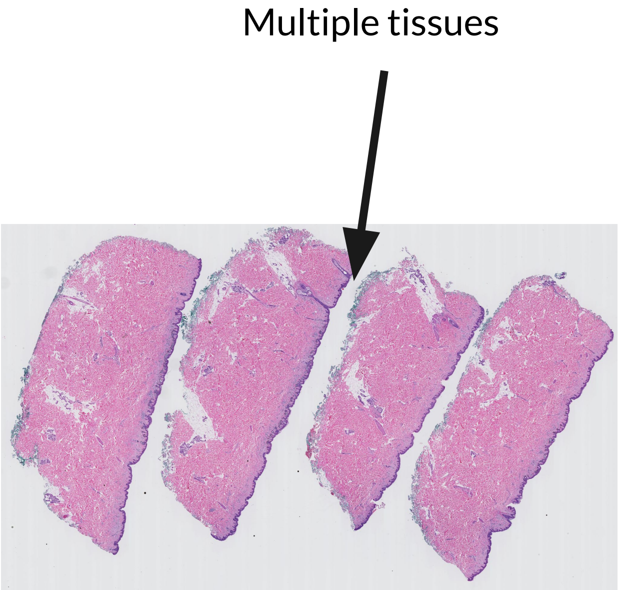
Feature is dependent on resolution

Dataset

Dataset





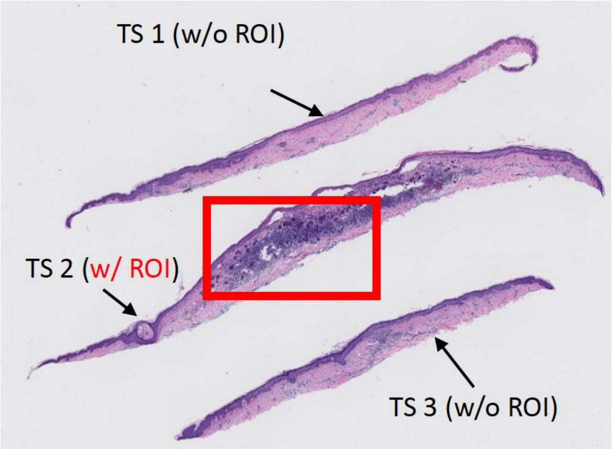


Diagnostic Category	Number of WSIs				Average WSI size (in pixels)
	Training	Validation	Test	Total	
MMD	26	6	29	61	11843×10315
MIS	25	5	30	60	9133×8501
pT1a	33	6	34	73	9490×7984
pT1b	18	6	22	46	14858×12154
Total	102	23	115	240	11130×9603

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).

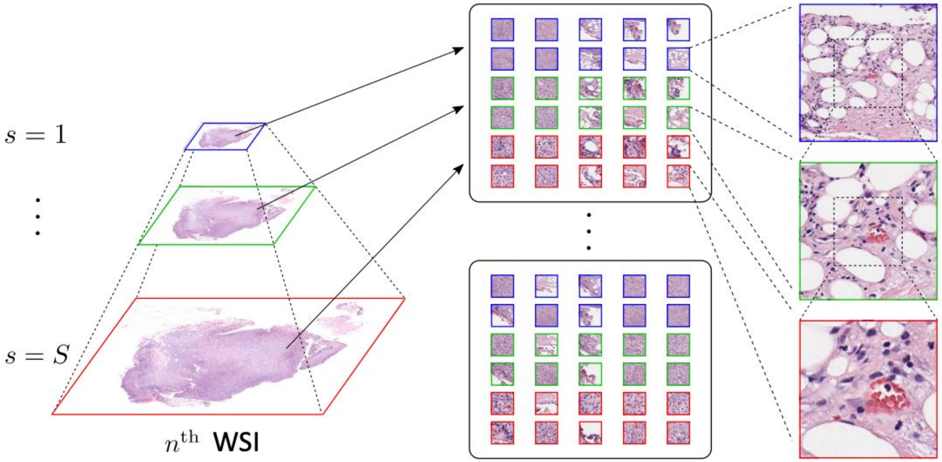
Dataset

Invasive T1a Skin Biopsy Image (or Class 3)

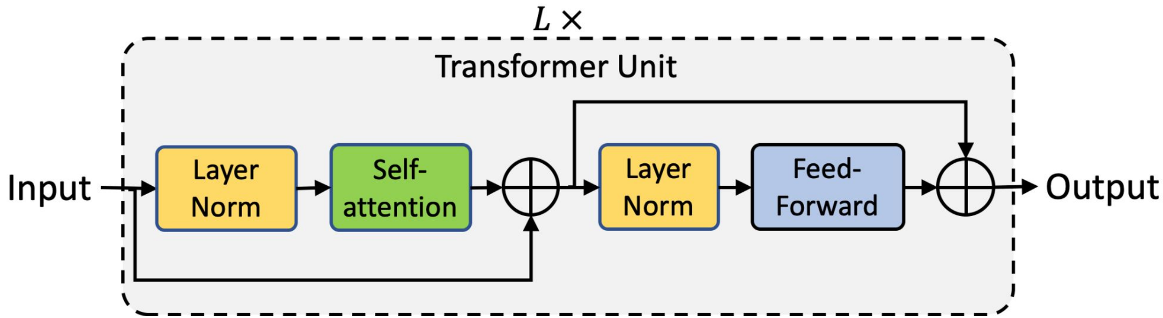


Key Idea

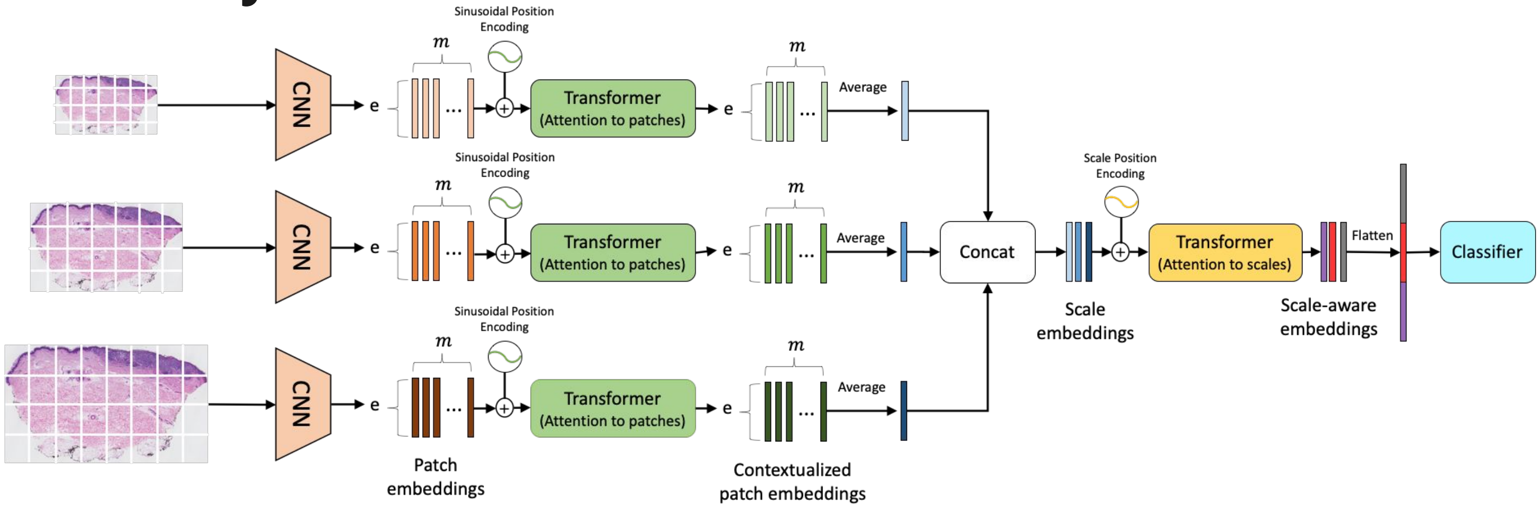
- Self-attention-based framework for classifying WSIs at multiple input scales
- A soft label assignment method to reduce ambiguities



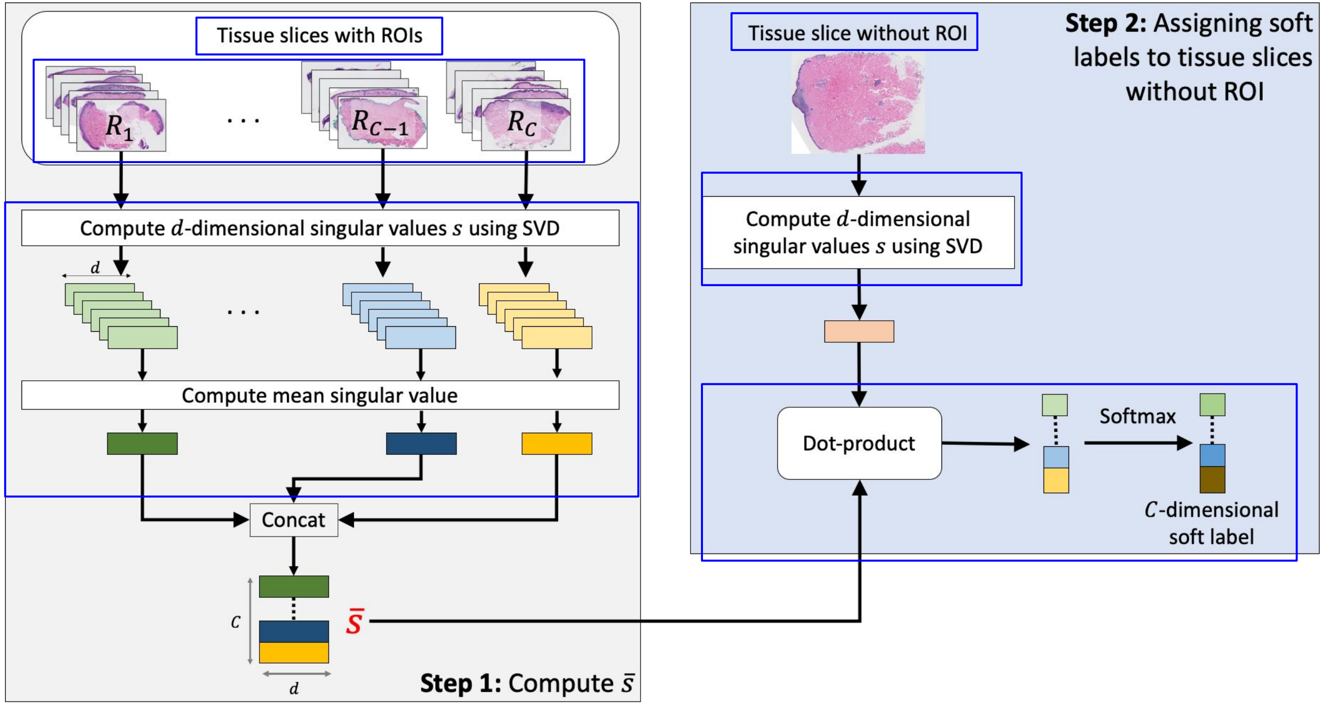
Transformer Unit



Scale-Aware Transformers for Diagnosing Melanocytic Lesions



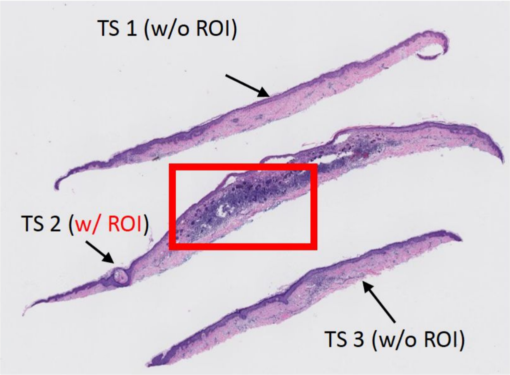
Soft labels



Tissue slices with ROIs
R_1
R_{C-1}
R_C
Compute d-dimensional singular values s using SVD
d
Compute mean singular value
Concat
C
d
$\bar{s}$
Step 1: Compute $\bar{s}$
Tissue slice without ROI
Step 2: Assigning soft labels to tissue slices without ROI
Compute d-dimensional singular values s using SVD
Dot-product
Softmax
C-dimensional soft label

Soft labels

**Invasive T1a Skin Biopsy Image
(or Class 3)**



Hard Label (one-hot encoding)

TS 1	0	0	1	0
TS 2	0	0	1	0
TS 3	0	0	1	0

Label smoothing (smoothing=0.1)

TS 1	0.033	0.033	0.9	0.033
TS 2	0.033	0.033	0.9	0.033
TS 3	0.033	0.033	0.9	0.033

Constrained label smoothing

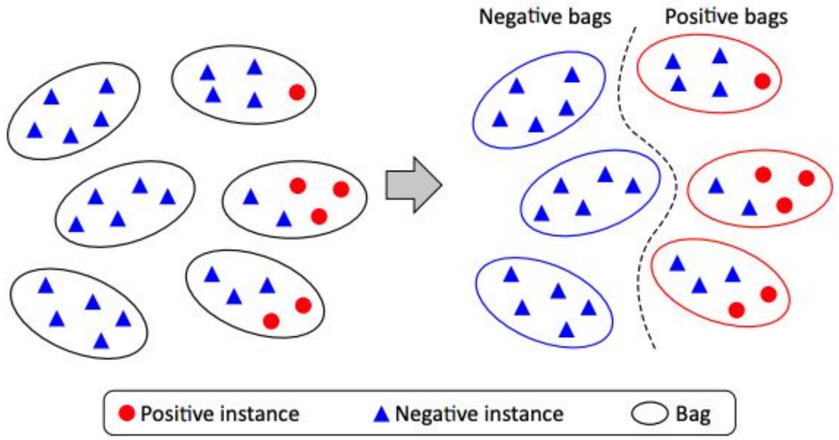
TS 1	0.5	0.5	0	0
TS 2	0	0	1	0
TS 3	0.5	0.5	0	0

Soft labels (ours)

TS 1	0.54	0.46	0	0
TS 2	0	0	1	0
TS 3	0.28	0.72	0	0

Baseline Methods

- Patch-based classification
- Weighted feature aggregation
- ChikonMIL
- MS-DA-MIL
- Streaming CNN

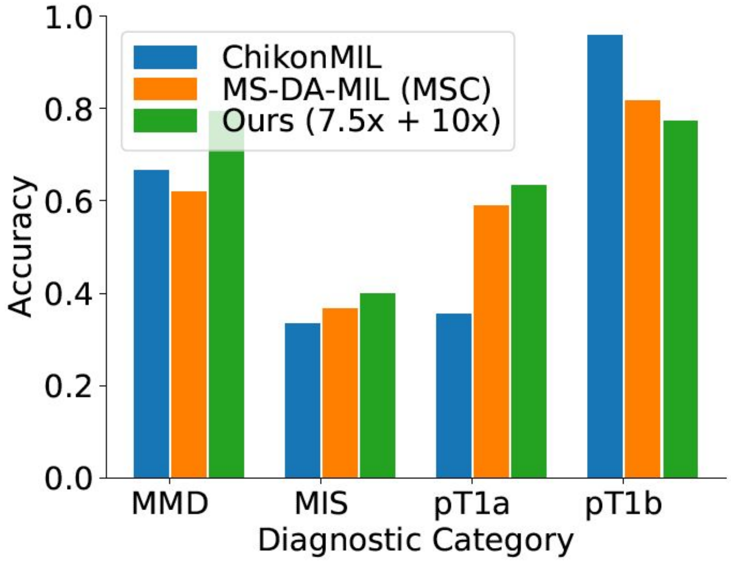


Experimental Result: baseline methods

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 ($10\times$). MSC denotes multiple input scales ($7.5\times$, $10\times$, $12.5\times$). MSC* denotes multiple input scales ($10\times$, $20\times$)

Experimental Result: baseline methods



Experimental Result: soft label

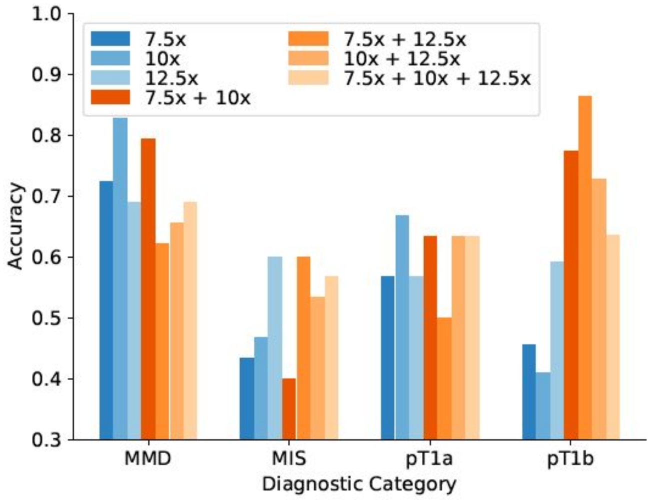
Method	Accuracy	Specificity	AUC
Hard labels	0.50	0.83	0.73
Label smoothing	0.50	0.83	0.71
Constrained label smoothing	0.56	0.85	0.77
Soft labels (Ours; Section III-C)	**0.60**	**0.87**	**0.77**

Comparison of the performance of different labeling methods.

Experimental Result: single vs. multiple input 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



(b) Class-wise accuracy of ScAtNet

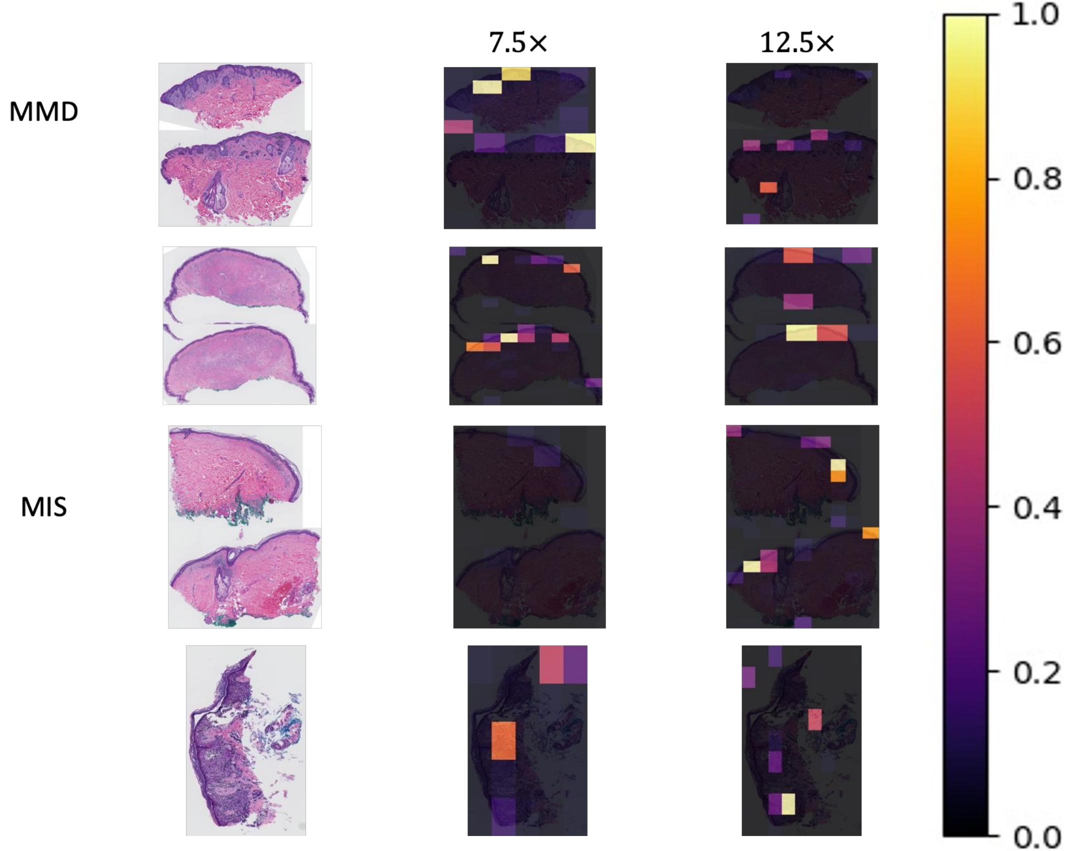
Experimental Result: pathologists performance

Diagnostic Category	Accuracy		F1		Sensitivity		Specificity	
	PG	Ours	PG	Ours	PG	Ours	PG	Ours
MMD	0.92	0.79	0.71	0.75	0.92	0.79	0.76	0.89
MIS	0.46	0.40	0.49	0.44	0.46	0.40	0.85	0.84
pT1a	0.51	0.65	0.62	0.63	0.51	0.65	0.95	0.84
pT1b	0.72	0.77	0.72	0.74	0.78	0.77	0.97	0.92
Overall	0.65	0.64	0.65	0.64	0.65	0.64	0.88	0.88

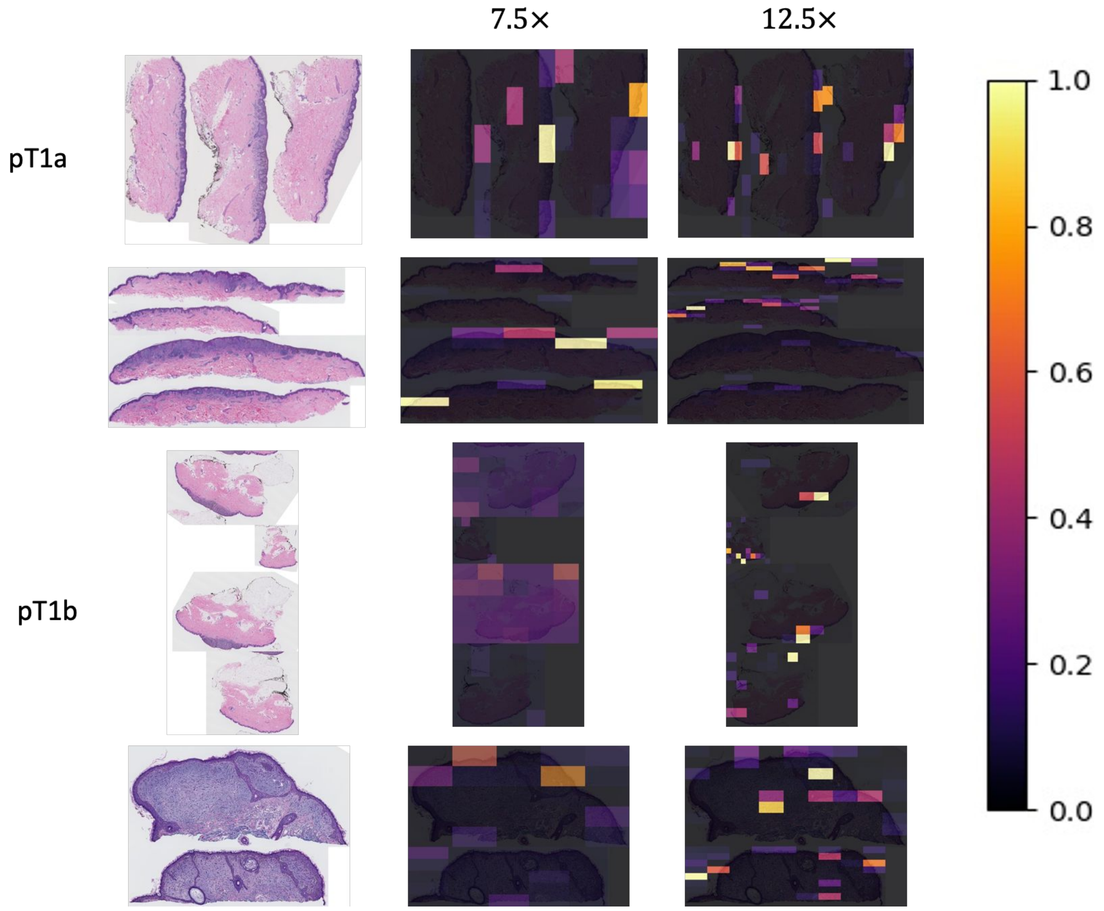
Comparison of ScAtNet with pathologists' (PG) performance.

Discussion

- 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%)
- Saliency analysis shows that different input results in different attentions



MMD
MIS
7.5×
12.5×
1.0
0.8
0.6
0.4
0.2
0.0


7.5×
12.5×
pT1a
pT1b
1.0
0.8
0.6
0.4
0.2
0.0

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

References

[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.

References

[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.