Decoding the visual attention of pathologists to reveal their level of expertise

Souradeep Chakraborty[1], Rajarsi Gupta[2], Oksana Yaskiv[4], Constantin Friedman[4], Natallia Sheuka[4], Dana Perez[4], Paul Friedman[4], Gregory Zelinsky[1,3], Joel Saltz[1,2], and Dimitris Samaras[1]

[1] Department of Computer Science, Stony Brook University, NY, USA
[2] Department of Biomedical Informatics, Stony Brook University, NY, USA
[3] Department of Psychology, Stony Brook University, NY, USA
[4] Department of Pathology and Laboratory Medicine, Northwell Health Laboratories, USA

Abstract. We present a method for classifying the expertise of a pathologist based on how they allocated their attention during a cancer reading. We engage this decoding task by developing a novel method for predicting the attention of pathologists as they read Whole-Slide Images (WSIs) of prostate tissue and make cancer grade classifications. Our ground truth measure of a pathologists' attention is the x, y and z (magnification) movement of their viewport as they navigated through WSIs during readings, and to date we have the attention behavior of 43 pathologists reading 123 WSIs. These data revealed that specialists have higher agreement in both their attention and cancer grades compared to general pathologists and residents, suggesting that sufficient information may exist in their attention behavior to classify their expertise level. To attempt this, we trained a transformer-based model to predict the visual attention heatmaps of resident, general, and specialist (Genitourinary) pathologists during Gleason grading. Based solely on a pathologist's attention during a reading, our model was able to predict their level of expertise with 75.3%, 56.1%, and 77.2% accuracy, respectively, better than chance and baseline models. Our model therefore enables a pathologist's expertise level to be easily and objectively evaluated, important for pathology training and competency assessment. Tools developed from our model could be used to help pathology trainees learn how to read WSIs like an expert.

Keywords: Histopathology · Visual attention · Prostate cancer grading

1 Introduction

A pathologist reading a whole-slide image (WSI) for cancer diagnosis is a complex and specialized cognitive task requiring years of training. Radiology has long appreciated the role played by attention during cancer readings [10, 15, 17, 18], and a similar appreciation has been growing in digital pathology [3, 4, 7, 8, 14]. Being able to predict the visual attention of pathologists as they read WSIs

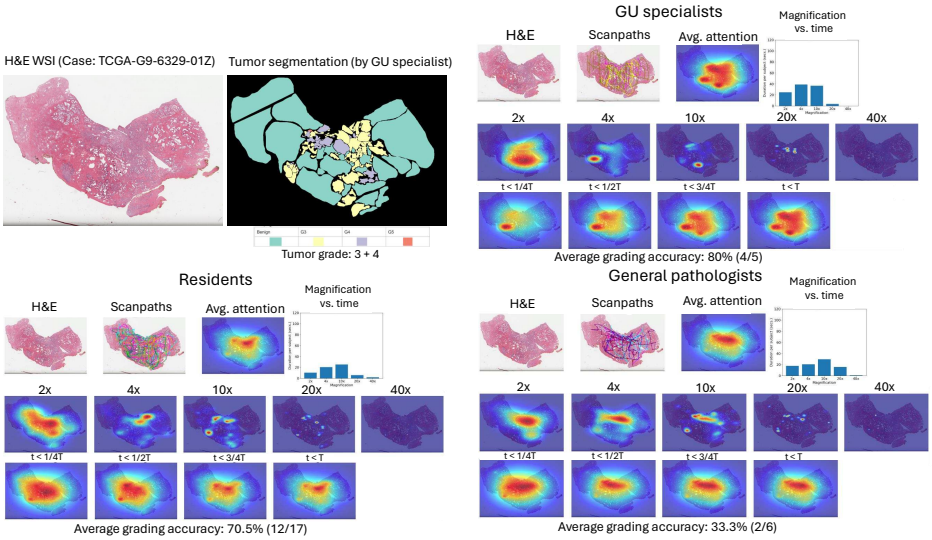

Fig. 1: Attention heatmaps computed for GU specialists (top-right) and general and resident pathologists (bottom). More detailed heatmaps are also shown for different levels of magnification and viewing durations. Upper-left: grade-level segmentation of a WSI by a GU specialist. The attention heatmaps of GU specialists correlate higher with the tumor annotations compared to the non-specialists, and the GU specialists have the highest grading accuracy.

will be crucial for next generation computer-assisted clinical decision support systems, but here our focus is on pathology training with respect to attention and determining whether a trainee is allocating their attention like a specialist. Also, several studies [9, 1, 12] have shown significant variability in histopathology diagnosis, indicating challenges in the consistent interpretation of WSIs and highlighting the need for strategies to improve diagnostic accuracy and agreement among pathologists. The studies most closely related to our goal is recent work attempting to predict the cursor-based movements of a pathologist's viewport as a measure of attention during a pathology reading [7, 8]. In [8], they did this for prostate using a fine-tuned ResNet34, and in [7] a model based on a swin-tranformer was used to predict attention during multi-stage GI-NETs examination, although the former study is most relevant to ours because they also predicted differences in attention heatmaps between genitourinary (GU) specialists and general pathologists during prostate cancer grading. However, their study was limited to only five WSIs viewed by 13 pathologists. More broadly, data scarcity, both in terms of the number of pathologists and WSIs, is a problem preventing the discovery of patterns in a pathologist's attention behavior and is severely limiting the training of models to make more accurate attention predictions. More data are needed to gain deeper insights into how pathologists of different expertise allocate their attention during readings, and to develop the predictive tools that can help pathology trainees attend like specialists.

We help remedy this data scarcity problem by collecting the largest known dataset of pathologist attention to date: 43 pathologists reading 123 WSIs of prostate cancer, yielding a total of 1016 attention trajectories. With this larger dataset, we can begin to study the relationship between a pathologist's attention and their cancer grading and how the attention of a specialist differs from that of a pathology trainee. To highlight the richness of our dataset, in Fig. 1 we visualize attention heatmaps computed for pathologists having three levels of expertise in the pathology task: GU specialists, general pathologists, and residents. The top row for each shows the WSI with overlaid attention trajectory and an average attention heatmap, and the middle and bottom rows show more specific heatmaps for five magnification levels and for four quartiles of reading time. Qualitative differences in attention between pathologists having different levels of expertise appear almost everywhere you look. Residents and general pathologists clearly allocated their attention to different regions in the WSI and at different magnifications, compared to the GU specialists, factors possibly affecting grading accuracy. Also shown is a grade-level tumor segmentation of the WSI by another GU specialist, where we found a higher correlation between specialist attention and this ground truth (cross correlation scores of specialists, general, and resident pathologists are 0.452, 0.418, and 0.413 respectively).

Motivated by the above analysis, we introduce two models to solve two different pathology tasks: (a) predicting pathologist attention (ProstAttFormer), and (b) predicting expertise (ExpertiseNet), both essential technical components towards developing our AI-assisted pathologist training pipeline. ProstAttFormer is a transformer-based model designed to predict pathologists' visual attention through attention heatmaps across various magnification levels. Prior approaches [7, 8] rely on patch-wise training at a single magnification (10x). Our ProstAttFormer improves attention prediction performance (at different magnifications) by more effectively leveraging inter-patch feature correspondences via multi-headed self-attention mechanism in transformers. ExpertiseNet is a model that can predict pathologist expertise based on their attention behavior as they grade WSIs. The model leverages cumulative temporal attention heatmaps and magnification-wise attention heatmaps for classifying a pathologist as resident/general/specialist pathologist. These models will make possible an objective evaluation of a pathologists' acquisition of benchmark levels of expertise during their training. While our current dataset already allows us to improve pathologist attention prediction and shows that the way pathologists allocate their attention spatio-temporally form strong signals for predicting their expertise, gathering additional data will enable us to build a comprehensive training pipeline for pathologists that will leverage our ProstAttFormer and ExpertiseNet models to guide trainees on where and how long to focus their attention. This will accelerate their learning process and expertise development.

In summary, our main contributions are: (1) the largest known pathologist attention dataset (123 WSIs across 43 pathologists), (2) a transformer-based attention prediction model that outperforms existing models, and (3) a pathologist expertise prediction model based on their attention.

2 Dataset of Pathologist Attention and Cancer Grades

2.1 Dataset creation

Similar to [8], we used the QuIP caMicroscope, a web-based toolset for digital pathology data management and visualization [13] for recording the attention data of pathologists as they viewed WSIs of prostate cancer tissues (TCGA-PRAD dataset) for tumor grading. We collected attention data from 43 pathologists spanning resident (18), general (15), and GU specialist (10) levels of expertise from 11 separate institutions. After reading the instruction/consent screens, a pathologist (remotely located) was shown a WSI fit into their viewport (no magnification). They were free to navigate through the WSI in x,y,z as they conducted their reading and the GUI recorded their 1050×1680 viewport image at each mouse-cursor sample (20 Hz). Upon concluding their reading, the pathologist entered the tumor grade (primary and secondary) and a level of confidence in each decision into our interface. This basic procedure iterated for all the readings in the experiment.

The 123 WSIs we used for our study were selected by a team general pathologist from among 342 WSIs in the TCGA-PRAD dataset [19]. In total, the data collection resulted in 1016 attention scanpaths with 329, 158 and 529 scanpaths from residents, general, and specialist pathologists respectively. The average viewing time per slide per pathologist was 95 seconds. Additionally, a GU specialist pathologist annotated the Gleason grades on a set of 22 WSIs. We computed from these attention data an attention heatmap that captures the aggregate spatial distribution of the pathologist's attention, similar to [7, 8].

2.2 Relationship between pathologist attention and cancer grading

Multiple factors contribute to variability in cancer diagnoses [2, 4], with variability in a pathologist's attention recently added to this list [7, 8]. We extend this work by characterizing in our dataset the relationship between variability in attention during cancer readings and variability in cancer classifications. We estimate agreement in tumor grading by computing an average pairwise concordance score as:

$$Conc_{Grade}^{i,j} = 1 - \frac{\sqrt{(PG_i - PG_j)^2 + (SG_i - SG_j)^2}}{\sqrt{(PG_i - PG_j)_{max}^2 + (SG_i - SG_j)_{max}^2}}, \tag{1}$$

where, $Conc_{Grade}$ is the normalized score concordance between the primary and secondary Gleason scores of a pair of pathologists i and j. Concordance scores closer to 1 indicate better agreement. We estimate agreement in attention by computing an attention heatmap for each pathologist viewing a given WSI and then obtaining the average pair-wise cross-correlation between the different heatmaps. We hypothesize finding that variability in a non-specialists attention will lead to variability in their cancer classifications more so than specialists, who as a group will tend to agree more both on how a cancer should be graded and where they should look for it in a WSI. The pattern shown in Fig. 2 confirms our

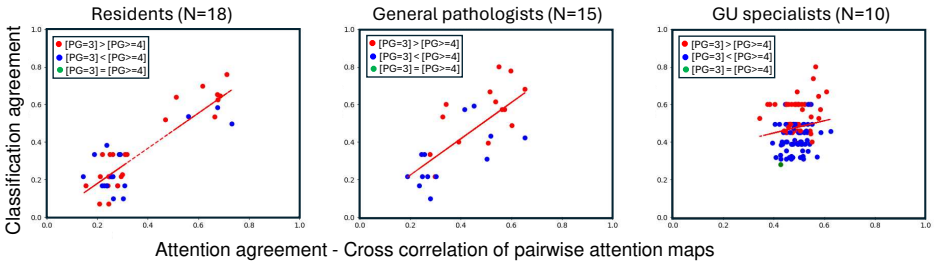

Fig. 2: Grade concordance vs. attention heatmap correlation across three groups of pathologists based on their expertise level. Each point represents a WSI. Red points indicate WSIs where more pathologists assigned a primary grade $PG = 3$ than $PG >= 4$, blue points indicate the opposite. The green point (right panel) indicates an equal number of pathologists making the different classifications.

hypothesis. Plotted is the degree of concordance in cancer classification (y-axes) against the degree of variability in attention (x-axes) for pathologists at different expertise levels. Each data point represents a WSI examined by at least two pathologists of the same expertise level. The average concordance scores were 0.32, 0.43, and 0.48 for residents (N=18), general pathologists (N=15), and GU specialists (N=10), respectively. Regression lines fit to these data show positive correlations, which highlights a positive correlation between attention variability during WSI examination and variability in tumor grading concordance.

Equally clear is that the strength of these correlations depends on the level of expertise. Correlations were strong and significant for residents ($r = 0.88, p < 0.01$) and general pathologists ($0.73, p < 0.01$) but weaker and not significantly different from 0 for specialists ($0.15, p = 0.09$). We interpret this expertise difference to mean that specialists tend to agree on where they should attend in a WSI and this agreed upon focus leads to greater agreement in classifications, but some resident and general pathologists (the clusters near 0.2 correlation) are still learning where to attend and consequently missing or misclassifying cancers.

3 Methodology

3.1 Predicting attention heatmaps

Fig. 3 shows the pipeline for our attention heatmap prediction model, ProstAttFormer. We first split an input WSI into a sequence of N patches. Next, we extract patch-wise feature embeddings using an off-the-shelf feature extractor. To capture positional information, learnable position embeddings are added to the sequence of patches to get the resulting input sequence of tokens. A transformer [16] encoder composed of several layers is applied to the sequence of tokens to generate a sequence of contextualized encodings. The sequence of patch encodings is decoded to a heatmap using the decoder (a $D \times 1$ convolutional layer, D = embedding size) that learns to map the patch-level encodings to patch-level

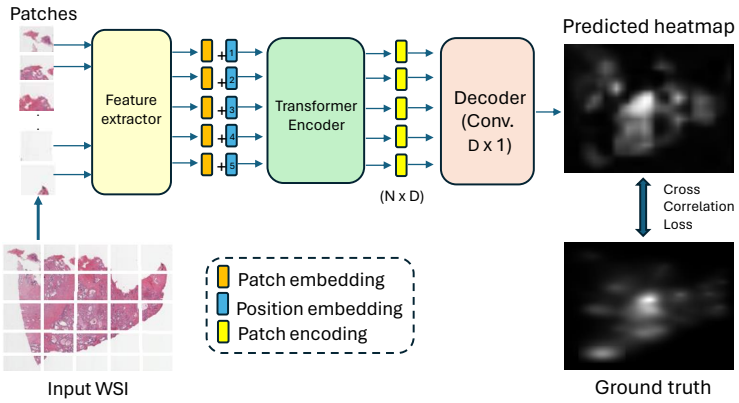

Fig. 3: ProstAttFormer, our attention prediction model that predicts pathologists attention on a WSI at different magnification levels.

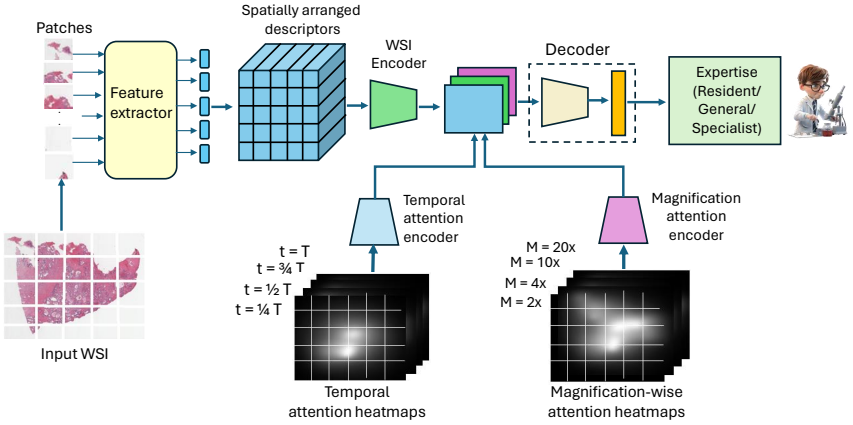

Fig. 4: ExpertiseNet, our attention based pathologist expertise prediction model.

attention scores. The final predicted heatmaps are obtained after map normalization. We used loss $L = 1 - CC(M_{Prd}, M_{GT})$ for training, where, CC = cross correlation between the predicted map M_{Prd} and the ground truth map M_{GT}.

3.2 Attention guided pathologist expertise prediction

We introduce ExpertiseNet (Fig. 4), a convolutional network that learns to classify the pathologist expertise based on how they have allocated attention across time and across different magnification levels. Specifically, this model accepts: (1) frozen ViT feature descriptors from a self-supervised learning model, (2) four cumulative temporal attention heatmaps computed for four fractions of the reading duration (1/4th, 1/2th, 3/4th, and entire viewing time), and (3) four attention heatmaps computed for 2x, 4x, 10x and 20x magnifications. This network is trained via the weighted Cross-entropy (CE) loss.

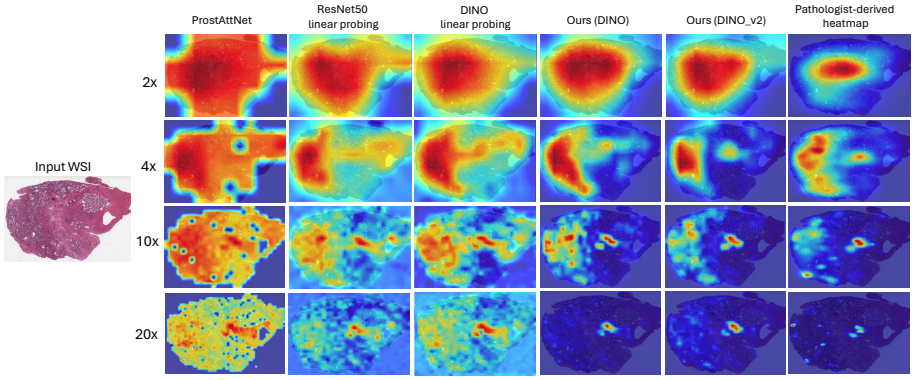

Fig. 5: Comparison of attention heatmap predictions from three baselines and our ProstAttFormer, which outperformed the others across all magnifications.

Implementation details: We used the ViT-S model (embedding size $D = 384$) trained using DINO for extracting the WSI patch features (frozen while training). For heatmap prediction, we input grids of variable sizes to our network for different magnifications - 10×10 for 2x, 20×20 for 4x, 50×50 for 10x and 60×60 for 20x. Our transformer encoder contains $n_l = 12$ layers with $n_h = 8$ attention heads. For ExpertiseNet, we used a grid size corresponding to 20x magnification. The WSI encoder consists of a *conv(D,16,1)* layer (D = embedding size). The magnification and temporal attention map encoders consist of a *conv(4,16,1)* layer each. The decoder consists of a *AvgPool2d (k=3, stride=2)* layer followed by a *conv(48,1,1)* layer and an $fc(256, 3)$ layer for final class prediction.

4 Results

4.1 Qualitative Evaluation

In Fig. 5, we qualitatively compared the attention heatmaps predicted by our model (with DINO [6] and DINO-v2 [11] features as input) with three baseline models: (1) frozen Resnet50 encoded features + linear probing using a 2048×1 convolutional layer as decoder, (2) frozen DINO encoded features + linear probing using a 384×1 convolutional layer as decoder, (3) ProstAttNet [8] on a test WSI instance from our dataset. Our ProstAttFormer produces more accurate attention heatmap predictions compared to the baselines at all magnifications.

4.2 Quantitative Evaluation

We quantitatively evaluate model performance using three metrics [5]: Cross Correlation (CC), Normalized Scanpath Saliency (NSS), and KL-Divergence (KLD). In Tab. 1, we compare the 5-fold cross validation performance of the different baseline models with our models on 25 test H&E WSIs at different magnification levels. Our models trained using the DINO and DINO-v2 feature descriptors

outperform the baseline models by a significant margin by all metrics. In Tab. 2, we compare the attention prediction performance between our ProstAttFormer model trained on specialist data and our model trained on non-specialist (residents and general pathologists) data. We test these models on 17 H&E WSIs (with tumor annotations from a GU specialist) at different magnifications. We find that our model trained on specialists' data performs better than our model trained on non-specialist data on the 4x, 10x and 20x magnifications (the most commonly used for Gleason grading). These results suggest that non-specialist pathologists might benefit from training on the attention behavior of specialists.

Table 1: Comparison of 5-fold cross-validation performance between baseline models and our models on 25 test H&E WSIs at different magnifications. We evaluated ProstAttNet [8] and PathAttFormer [7] only at 10x following their original implementations.

Model	CC_{Attn}	NSS_{Attn}	KLD_{Attn}
Frozen ResNet50+Dec.	0.498 ± 0.214	0.748 ± 0.307	0.383 ± 0.023
Frozen DINO+Dec.	0.486 ± 0.192	0.705 ± 0.275	0.397 ± 0.026
Ours (w/ DINO)	$\mathbf{0.560} \pm 0.199$	$\mathbf{0.836} \pm 0.290$	0.362 ± 0.070
Ours (w/DINO-v2)	0.551 ± 0.149	0.829 ± 0.202	$\mathbf{0.348} \pm 0.022$

(a) 2x

Model	CC_{Attn}	NSS_{Attn}	KLD_{Attn}
Frozen ResNet50+Dec.	0.636 ± 0.067	1.106 ± 0.190	0.512 ± 0.151
Frozen DINO+Dec.	0.595 ± 0.067	1.014 ± 0.207	0.539 ± 0.141
Ours (w/ DINO)	$\mathbf{0.668} \pm 0.079$	1.175 ± 0.268	0.402 ± 0.071
Ours (w/DINO-v2)	0.666 ± 0.074	$\mathbf{1.181} \pm 0.264$	$\mathbf{0.397} \pm 0.062$

(b) 4x

Model	CC_{Attn}	NSS_{Attn}	KLD_{Attn}
Frozen ResNet50+Dec.	0.682 ± 0.018	1.510 ± 0.242	0.820 ± 0.249
Frozen DINO+Dec.	0.659 ± 0.027	1.436 ± 0.236	0.860 ± 0.253
ProstAttNet [8]	0.262 ± 0.017	0.883 ± 0.138	2.666 ± 0.562
PathAttFormer [7]	0.294 ± 0.014	0.924 ± 0.145	2.513 ± 0.520
Ours (w/ DINO)	$\mathbf{0.739} \pm 0.029$	$\mathbf{1.711} \pm 0.360$	0.473 ± 0.068
Ours (w/DINO-v2)	0.738 ± 0.029	1.710 ± 0.362	$\mathbf{0.473} \pm 0.055$

(c) 10x

Model	CC_{Attn}	NSS_{Attn}	KLD_{Attn}
Frozen ResNet50+Dec.	0.372 ± 0.042	1.910 ± 0.277	2.361 ± 0.503
Frozen DINO+Dec.	0.365 ± 0.062	1.892 ± 0.271	2.369 ± 0.511
Ours (w/ DINO)	0.417 ± 0.065	$\mathbf{2.266} \pm 0.368$	1.741 ± 0.349
Ours (w/DINO-v2)	$\mathbf{0.419} \pm 0.062$	2.264 ± 0.377	$\mathbf{1.731} \pm 0.341$

(d) 20x

Table 2: Comparison of our attention prediction model ProstAttFormer trained on specialist vs. non-specialist (general pathologists and residents) data, based on attention-tumor overlap on 17 test H&E WSIs at different magnifications.

	2x			4x			10x			20x		
Model	CC_{Seg}	NSS_{Seg}	KLD_{Seg}	CC_{Seg}	NSS_{Seg}	KLD_{Seg}	CC_{Seg}	NSS_{Seg}	KLD_{Seg}	CC_{Seg}	NSS_{Seg}	KLD_{Seg}
Ours (Specialist)	0.285	**1.032**	2.487	**0.406**	**1.263**	**2.186**	**0.582**	**1.851**	**1.584**	**0.592**	**2.619**	**1.382**
Ours (Non-Specialist)	**0.314**	1.027	**2.418**	0.386	1.253	2.250	0.561	1.814	1.690	0.566	2.310	1.563

Table 3: Pathologist expertise classification (3-way, resident/general/specialist pathologists) using temporal heatmaps, magnification-specific heatmaps, and their combination with 5-fold cross-validation on our dataset. Combining both heatmap types yields the best results.

Model	Accuracy	F1-score	AUC score
Random	0.333 ± 0.000	0.333 ± 0.000	0.5000 ± 0.000
ExpertiseNet (w/ Temporal heatmaps)	0.676 ± 0.035	0.630 ± 0.032	0.796 ± 0.007
ExpertiseNet (w/ Magnification heatmaps)	0.731 ± 0.021	0.680 ± 0.015	0.837 ± 0.020
ExpertiseNet (w/ Temporal + Magnification heatmaps)	$\mathbf{0.732 \pm 0.017}$	$\mathbf{0.696 \pm 0.010}$	$\mathbf{0.845 \pm 0.005}$

In Tab. 3 we ablate from the input to our model of pathologist expertise prediction, ExpertiseNet, either the temporal attention heatmaps or the magnification-specific heatmaps and report the effect on 5-fold cross validation classification performance (3-way, resident/general/specialist pathologist) compared to the original model that combined the two types of heatmaps. We re-

port the classification accuracy, F1-score and the AUC score of classification, with higher values for all metrics indicating better performance. ExpertiseNet performs best when a combination of both cumulative temporal heatmaps and magnification heatmaps are input to the model.

5 Conclusion

We introduced two models, ExpertiseNet and ProstAttFormer. ExpertiseNet classifies a pathologist's expertise based on their allocation of attention during cancer readings. ProstAttFormer predicts pathologists' attention as they read WSIs of prostate tissues and make cancer grade classifications. These models make it possible to predict the visual attention heatmaps of pathologists performing Gleason grading, enabling an objective evaluation of their expertise. Our models can therefore assist in pathology training and competency assessment, offering the potential for trainees to learn how to read WSIs like an expert. Future work will need to overcome the challenge of inter-observer variability in attention behavior, which impedes training more predictive pathologist models.

Acknowledgments. This work was supported by a seed grant from the Stony Brook University Office of the Vice President for Research (1150956-3-63845), NSF grants IIS-2212046 and IIS-2123920, and grants UH3-CA225021, U24-CA215109, and U24-CA180924 from the NCI and NIH.

Disclosure of Interests. Joel Saltz and Rajarsi Gupta are co-founders of Chilean Wool LLC.

References

1. Allison, K.H., Reisch, L.M., Carney, P.A., Weaver, D.L., Schnitt, S.J., O'Malley, F.P., Geller, B.M., Elmore, J.G.: Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel. Histopathology **65**(2), 240–251 (2014)
2. Bombari, D., Mora, B., Schaefer, S.C., Mast, F.W., Lehr, H.A.: What was i thinking? eye-tracking experiments underscore the bias that architecture exerts on nuclear grading in prostate cancer. PLoS One **7**(5), e38023 (2012)
3. Brunyé, T.T., Drew, T., Kerr, K.F., Shucard, H., Weaver, D.L., Elmore, J.G.: Eye tracking reveals expertise-related differences in the time-course of medical image inspection and diagnosis. Journal of Medical Imaging **7**(5), 051203–051203 (2020)
4. Brunyé, T.T., Mercan, E., Weaver, D.L., Elmore, J.G.: Accuracy is in the eyes of the pathologist: the visual interpretive process and diagnostic accuracy with digital whole slide images. Journal of biomedical informatics **66**, 171–179 (2017)
5. Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., Durand, F.: What do different evaluation metrics tell us about saliency models? IEEE transactions on pattern analysis and machine intelligence **41**(3), 740–757 (2018)
6. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9650–9660 (2021)

7. Chakraborty, S., Gupta, R., Ma, K., Govind, D., Sarder, P., Choi, W.T., Mahmud, W., Yee, E., Allard, F., Knudsen, B., et al.: Predicting the visual attention of pathologists evaluating whole slide images of cancer. In: International Workshop on Medical Optical Imaging and Virtual Microscopy Image Analysis. pp. 11–21. Springer (2022)
8. Chakraborty, S., Ma, K., Gupta, R., Knudsen, B., Zelinsky, G.J., Saltz, J.H., Samaras, D.: Visual attention analysis of pathologists examining whole slide images of prostate cancer. In: 2022 IEEE 19th International symposium on biomedical imaging (ISBI). pp. 1–5. IEEE (2022)
9. Elmore, J.G., Nelson, H.D., Pepe, M.S., Longton, G.M., Tosteson, A.N., Geller, B., Onega, T., Carney, P.A., Jackson, S.L., Allison, K.H., et al.: Variability in pathologists' interpretations of individual breast biopsy slides: a population perspective. Annals of internal medicine **164**(10), 649–655 (2016)
10. Gandomkar, Z., Tay, K., Ryder, W., Brennan, P.C., Mello-Thoms, C.: icap: an individualized model combining gaze parameters and image-based features to predict radiologists' decisions while reading mammograms. IEEE transactions on medical imaging **36**(5), 1066–1075 (2016)
11. Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Howes, R., Huang, P.Y., Xu, H., Sharma, V., Li, S.W., Galuba, W., Rabbat, M., Assran, M., Ballas, N., Synnaeve, G., Misra, I., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: Dinov2: Learning robust visual features without supervision (2023)
12. Ronen, S., Al-Rohil, R.N., Keiser, E., Jour, G., Nagarajan, P., Tetzlaff, M.T., Curry, J.L., Ivan, D., Middleton, L.P., Torres-Cabala, C.A., et al.: Discordance in diagnosis of melanocytic lesions and its impact on clinical management: a melanoma referral center experience with 1521 cases. Archives of Pathology & Laboratory Medicine **145**(12), 1505–1515 (2021)
13. Saltz, J., Sharma, A., Iyer, G., Bremer, E., Wang, F., Jasniewski, A., DiPrima, T., Almeida, J.S., Gao, Y., Zhao, T., et al.: A containerized software system for generation, management, and exploration of features from whole slide tissue images. Cancer research **77**(21), e79–e82 (2017)
14. Sudin, E., Roy, D., Kadi, N., Triantafyllakis, P., Atwal, G., Gale, A., Ellis, I., Snead, D., Chen, Y.: Eye tracking in digital pathology: identifying expert and novice patterns in visual search behaviour. In: Medical Imaging 2021: Digital Pathology. vol. 11603, pp. 253–262. SPIE (2021)
15. Tourassi, G., Voisin, S., Paquit, V., Krupinski, E.: Investigating the link between radiologists' gaze, diagnostic decision, and image content. Journal of the American Medical Informatics Association **20**(6), 1067–1075 (2013)
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
17. Venjakob, A., Marnitz, T., Mahler, J., Sechelmann, S., Roetting, M.: Radiologists' eye gaze when reading cranial ct images. In: Medical imaging 2012: Image perception, observer performance, and technology assessment. vol. 8318, pp. 78–87. SPIE (2012)
18. Wang, S., Ouyang, X., Liu, T., Wang, Q., Shen, D.: Follow my eye: Using gaze to supervise computer-aided diagnosis. IEEE Transactions on Medical Imaging **41**(7), 1688–1698 (2022)
19. Zuley, M.L., Jarosz, R., Drake, B.F., Rancilio, D., Klim, A., Rieger-Christ, K., Lemmerman, J.: Radiology data from the cancer genome atlas prostate adenocarcinoma [tcga-prad] collection. Cancer Imaging Arch **9** (2016)