Deep-IMMA: An Interpretable Immune-Aware Multimodal Deep Representation Learning for Cancer Prognosis and Recurrence Prediction

Kyeong Joo Jung[1,†], Swapna Sourav Rout[4,†], Jianwei Qiu[3], Sanghee Cho[3], Elizabeth McDonough[3], Soumya Ghose[3], Chrystal Chadwick[3], Robert West[2], James D. Brooks[2], Fiona Ginty[3], Parag Mallick[2], Kshitij Jadhav[4], Raghu Machiraju[1,*]

[1] Department of Computer Science and Engineering, Ohio State University, Ohio, USA.
[2] School of Medicine, Stanford University, CA, USA.
[3] GE HealthCare Technology and Innovation Center, 1 Research Circle, New York, USA.
[4] Koita Centre for Digital Health (KCDH), IIT Bombay, India.
† Contributed equally | * Corresponding author: machiraju.1@osu.edu

Abstract. Cancer prognosis and recurrence prediction from histopathology images has relied on morphology-focused models or genomic data. However, these approaches insufficiently capture the spatial organization and interactions between epithelial, stromal, and immune cell populations, increasingly recognized as important contributors to disease progression. We propose **Deep-IMMA** (**Deep I**mmune-Aware **M**ultimodal **M**ulti-task **A**lignment), a general multimodal learning framework that derives biologically grounded prognostic representations directly from histopathology images. Deep-IMMA introduces a two-stage architecture that encodes the tumor–immune microenvironment via structured multimodal supervision. In Stage 1, tile-level visual embeddings are learned via joint multi-task multimodal supervision from structured textual captions and spatial cell-type maps, with an auxiliary cross-modal consistency objective, yielding semantically enriched and immune-aware representations. This joint training integrates vision, text, and spatial priors while requiring only histology image at inference time. In Stage 2, these representations are aggregated using a spatially aware Multiple Instance Learning module that explicitly preserves neighborhood relationships and immune topology, improving slide-level risk discrimination. Using prostate cancer as a case study, we evaluate Deep-IMMA on a tri-modal prostate cancer histology dataset (Prostate-TriMod), comprising 717 prostate tissue microarray (TMA) core images from 212 patients. Deep-IMMA achieves a test AUROC of 0.761 (95% CI: 0.666–0.847), outperforming prognostic models based on only morphology and clinical data for recurrence prediction. Importantly, Deep-IMMA generates interpretable spatial–textual outputs, including tile-level semantic descriptions, immune-infiltration and attention heatmaps, enabling spatially grounded risk interpretation. These results demonstrate that immune-aware multimodal supervision can establish a robust, transparent and cost-effective baseline for cancer prognosis, bypassing expensive auxiliary data.

Keywords: Deep-IMMA Framework, Immune-Aware Foundation Model Adaptation, Biologically Grounded Spatial-Textual Representation Learning, Prostate Cancer Recurrence Prediction.

1 Introduction

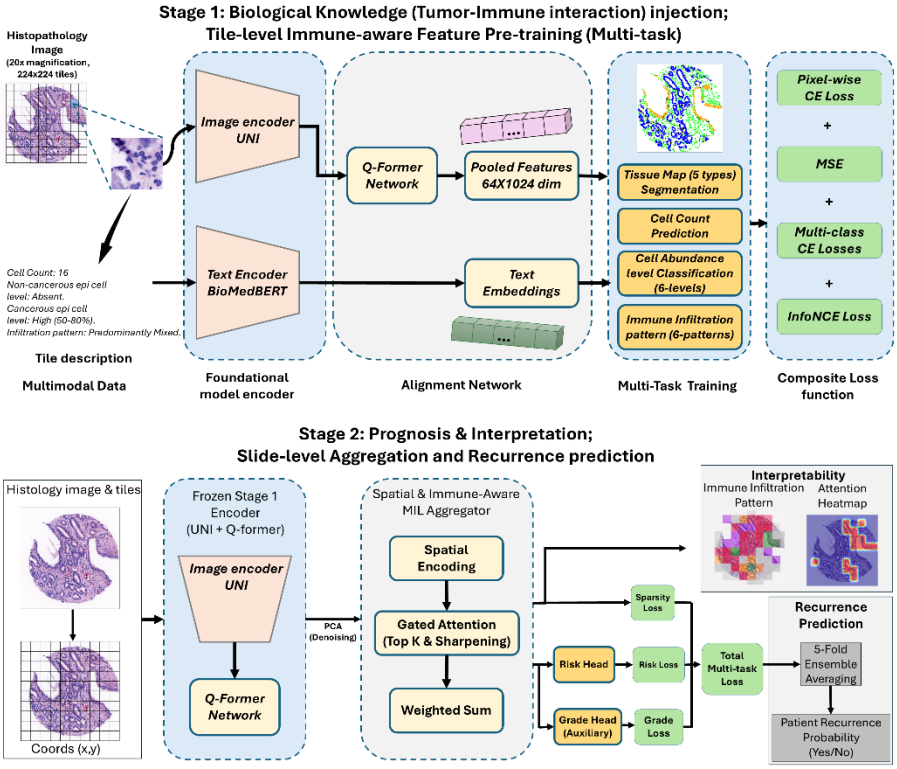

Fig. 1. Deep-IMMA architecture. Stage 1 aligns histology tiles with tissue maps and captions via multi-task multimodal learning to inject biological knowledge. The model includes DICE loss for segmentation, MSE (Mean Square Error) for cell count, Cross-Entropy (CE) losses for 6 cell type abundance levels, 6 immune infiltration phenotypes (Cold, Mixed, Compartmentalized, Hybrid, Various, and No Activity) and Information Noise Contrastive Estimation (InfoNCE) loss for text-image alignment. **Stage 2** applies spatially aware MIL for BCR prediction, producing interpretable attention and immune heatmaps.

Accurate cancer prognosis remains a clinical challenge, where predicting biochemical recurrence (BCR) is essential for individualized management. This is particularly critical in prostate cancer [1–3], where reliable BCR prediction from histopathology remains a long-term challenge despite various AI-based efforts [4,5]. Clinical tools such as CAPRA-S [6] achieve only moderate discrimination (C-index ≈ 0.73-0.77). While recent multimodal efforts combining baseline MRI with clinical covariates [7] report an AUC of 0.74, meta-analytic evidence suggests imaging/multimodal inputs can improve BCR prediction [8]. These limitations motivate representation-learning frameworks that directly encode spatially resolved immune infiltration patterns from histopathology slides. Recent pathology foundation models (FMs) (e.g., UNI, H-optimus-0, Prov-GigaPath [9-11]) and multimodal vision–language models (e.g., PathChat,

CONCH, OmiCLIP [12-14]) have improved feature transferability by aligning image and text embeddings. However, these approaches often remain morphology-centric and existing vision–language supervision is rarely structured to encode tumor–immune topology and other salient biological events explicitly, and slide-level outcomes are learned without biologically grounded inductive constraints.

We propose **Deep-IMMA** (**Deep I**mmune-aware **M**ultimodal **M**ulti-task **A**lignment), a universal two-stage immune-aware framework that adapts a pathology foundation model via joint supervision from dense tissue maps and structured tile captions using a Q-former [15] to enforce cross-modal consistency. We apply a spatially aware MIL framework [16] to aggregate embeddings, building upon the attention-based MIL foundation [17].

We evaluate Deep-IMMA as a prostate cancer case study on a tri-modal prostate cancer histology dataset (**Prostate-TriMod**) [18] (717 images/212 patients; 711 images/210 patients retained after excluding missing BCR labels). Deep-IMMA jointly infers segmentation, immune phenotypes, and tile-level descriptions and provides spatially grounded interpretability (immune-pattern maps and Top-K attention) alongside recurrence prediction. Our key contributions are: (i) **Biologically Structured Multimodal Supervision:** We introduce immune-phenotype–conditioned structured supervision, which embeds tumor–immune spatial structure directly into foundation-model adaptation; (ii) **Inductive Bias for Sparse Outcome Learning:** We demonstrate that structured multimodal pre-training improves stability and representation quality for slide-level recurrence prediction. (iii) **Unified Multi-Task Framework:** We show that segmentation, captioning, and BCR prediction can be jointly supported within a single immune-aware FM adaptation pipeline, improving interpretability without requiring expensive MRI or genomic inputs.

This work reframes recurrence prediction as **biologically constrained representation learning**, **injecting tumor-relative immune topology** into foundation-model adaptation to support **H&E-only risk discrimination** and **spatial interpretation**.

2 Methods

We present **Deep-IMMA**, a two-stage foundation modeling framework for predicting biochemical recurrence (BCR) (**Fig. 1**). **Stage 1** learns biologically structured tile-level representations via multi-task multimodal learning across histopathology images, dense tissue maps, and structured captions (including semantic alignment). **Stage 2** aggregates these embeddings using a spatially aware Multiple Instance Learning (MIL) model with Top-K gated attention and auxiliary grade supervision [19]. This hierarchy injects immune-aware inductive bias before sparse slide-level outcome learning.

2.1 Data Representation and Statistics

We evaluate Deep-IMMA on Prostate-TriMod [18], a tri-modal prostate TMA-core dataset pairing virtual H&E/histology images with single-cell-derived tissue maps, rule-based structured captions, and clinical outcomes. After excluding core images without

biochemical recurrence (BCR) labels, 711 TMA core images were retained, with a 26.0% overall BCR rate. Recurrence demonstrates a strong correlation with Grade Group (GG), ranging from 2.5% BCR for GG1 to 51.2% BCR for GG3. To prevent data leakage, patient-level splits were used, resulting in 136 patients (462 core images) for training, 32 patients (103 core images) for validation, and 42 patients (146 core images) for testing. **Stage 1** supervision utilized 31,306 training tiles and 8,378 validation tiles. Each individual case in this cohort includes the following components (**Fig. 1, 'Multimodal Data', and 'Multi-Task Training'**): (i) **Virtual H&E/histology images** (20x), generated from DAPI and autofluorescence imaging, are partitioned into 10x10 non-overlapping 224x224 tiles (100 per core image) derived from 2040x2040 crops, ensuring that 2D coordinates are preserved (**Fig. 1, 'Histopathology Image'**); (ii) **5-class tissue maps** that define regions for cancer epithelium (blue), benign epithelium (orange), stroma (green), immune cells (black), and background (white) (**Fig. 1, '5-class Tissue Map' on top of ' Multi-Task Training'**); (iii) **Structured Captions**, rule-based summaries derived from the single-cell data, encoding cell counts, cell abundance levels, and immune infiltration phenotypes. Immune phenotype labels encode the spatial proximity and infiltration depth of immune cells relative to tumor glands: **Cold** (no immune cells nearby), **Mixed** (intra-glandular), **Compartmentalized** (peritumoral), **Hybrid** (both), **Various** (multiple non-dominant signals), and **No Activity**. (**Fig. 1, 'Tile description'**); (iv) **Clinical endpoints**, which specify the GG (3-class) and BCR status for each patient.

2.2 Stage 1: Immune-Aware Representation Learning

Stage 1 extracts biologically grounded 1024-d embeddings using a **UNI vision transformer** backbone and **Q-Former** bottleneck. The Q-Former uses 64 learnable query tokens to aggregate image tokens into a compact semantic bottleneck. For the **U-Net decoder**, skip features are extracted from UNI blocks 3, 6, and 9 and projected to 256, 128, and 64 channels, respectively. The query outputs are reshaped into an 8×8 grid for the segmentation decoder. Simultaneously, **BioMedBERT** [20] maps captions into a projected latent space (**Fig. 1, 'Alignment Network'**). Unlike standard multimodal alignment, this multi-task objective captures enriched representations by encoding critical spatial features and cellular interactions through joint infiltration and segmentation tasks: (**Fig. 1, 'Composite Loss function'**):

- $\mathcal{L}_{\text{seg}}$ **(Tissue-map segmentation)**: This is used for region-level composition summaries rather than boundary-perfect delineation. Uses U-net decoder with skip connections for 5-class tissue map segmentation (pixel-wise cross-entropy).
- $\mathcal{L}_{\text{cls-i}}$ **(Immune infiltration pattern classification)**: Immune pattern classification from pooled embedding (cross-entropy).
- $\mathcal{L}_{\text{cls-c}}$ **(Cell type abundance level classification)**: Cell abundance level classification from pooled embedding (cross-entropy).
- $\mathcal{L}_{\text{reg}}$ **(Cell count regression):** Auxiliary cell count regression (MSE loss).
- $\mathcal{L}_{\text{align}}$ **(Semantic):** Vision-text alignment (InfoNCE loss).

Tasks are balanced using homoscedastic uncertainty weighting [21] to prevent negative transfer:

$$\mathcal{L}_{\text{Stage1}} = \sum_{l\in\{seg,cls-i,cls-c,reg,align\}} (\frac{1}{2\sigma_l^2}\mathcal{L}_l + \log\sigma_l),$$

where σ_i is a learnable parameter representing task-dependent noise. This allows the model to dynamically prioritize tasks during training. The resulting 1024-d embeddings are cached to serve as the foundational inputs for **Stage 2**.

2.3 Stage 2: Spatially Aware Slide-Level Prediction

Stage 2 aggregates cached embeddings utilizing a spatially aware MIL framework (**Fig. 1**, **'Stage 2'**). Embeddings undergo dimensionality reduction via Principal Component Analysis (PCA) to 512-dimensions to mitigate overfitting and are subsequently augmented with 2D sinusoidal positional encoding to preserve tissue topological structure. Prognostic regions are identified using Top-K selection ($K = 15$) and temperature-softmax ($\tau = 0.6$). These parameters were empirically optimized to balance the capture of spatial heterogeneity with the sharpening of attention on the most discriminative prognostic regions. We keep K and τ fixed across splits and confirm their contribution through ablations. (**Fig. 1, 'Spatial & Immune-Aware MIL Aggregator'**).

The model minimizes: $\mathcal{L}_{Stage2} = \lambda_r\mathcal{L}_{\text{BCR}} + \lambda_g\mathcal{L}_{\text{grade}} + \lambda_s\mathcal{L}_{\text{sparsity}}$. These components include:

- $\mathcal{L}_{\text{BCR}}$ **(BCR)**: cross-entropy with a 1:4 class-weighting ratio and label smoothing (0.1) to address the 26.0% BCR imbalance.
- $\mathcal{L}_{grade}$**(Grade)**: Categorical cross-entropy for 3-class pathological grade group prediction, serving as a biological stabilizer.
- $\mathcal{L}_{\text{sparsity}}$ **(Sparsity Regularizer)**: An entropy-based loss that forces the gated attention mechanism to concentrate on a limited set of diagnostic tiles.

Hyperparameters (λ_r=0.7, λ_g=0.2, λ_s=0.1), were optimized on training/validation only using Bayesian optimization (*hyperopt*) [22], and we report test performance with 2,000-bootstrap Confidence Intervals (CIs) (**Table 2**).

2.4 Implementation and Training

Final reporting uses a locked patient-level test set evaluated once. Hyperparameters and checkpoints are tuned via group based 5-fold stratified on training/validation splits. Both stages are optimized with AdamW (weight decay 2×10^{-1}); **Stage 2** additionally uses stochastic tile dropping (20% at data loading, plus an independent 50% per-tile mask during training) and dropout p=0.5 in the context multi-layer perceptron (MLP), training for 40 (**Stage 1, batch size 32**) and 200 (**Stage 2, batch size 4**) epochs. Decision thresholds are calibrated on validation folds using Youden's J [23] to maximize balanced risk assessment. The final checkpoint is selected via a weighted score combining recurrence and grade-group discrimination: (Area Under the Receiver Operating

Characteristic curve (AUROC) for BCR risk (AUC_r) and Grade Group (AUC_g)): $Score = \alpha AUC_r + (1-\alpha)AUC_g$, where $\alpha = 0.7$. **Implementation:** PyTorch (CUDA) on Linux, trained on an NVIDIA V100S (32GB). Ethics/IRB approval was obtained for the underlying cohort (IRB# 11612). Code and configuration files are released at: https://github.com/iamkj03/DEEP-IMMA.

3 Experiments and Results

3.1 Experimental Design and Baselines

We evaluate (1) **Stage 1** backbone suitability for immune-aware features (T1–T3), (2) whether spatially aware MIL improves BCR prediction (E0–E3), and (3) the necessity of key Stage-2 components (E4-E7). Stage 1 compares UNI with Prov-GigaPath (T2) and H-optimus-0 (T3). **Stage 2** ablates alignment (E2), spatial encoding (E4), Top-K (E5), auxiliary grade (E6), and sparsity regularization (E7) against the full model (E3) and baselines (E0: Logistic Regression, E1: ABMIL). Beyond predictive performance, we emphasize biologically grounded interpretability on histology (Section 3.4).

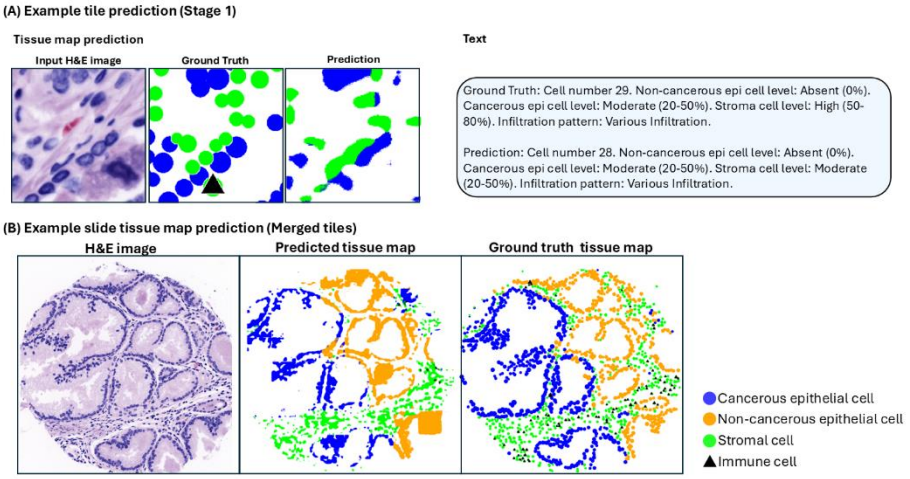

Fig. 2. Stage 1 qualitative results. (A) Example tile-level predictions: predicted 5-class tissue maps and structured captions compared with ground truth. (B) Slide-level (merged tiles) tissue map prediction; left— histopathology/virtual H&E image; middle—predicted tissue map; right—ground truth.

3.2 Representation Quality (Stage 1)

Table 1 summarizes **Stage 1** performance. UNI (T1) was selected as the optimal foundational encoder as it attained the lowest total loss and superior downstream embeddings for clinical tasks (E3, **Table 2**). Furthermore, the high IM AUC of 0.82 indicates that Deep-IMMA successfully learns complex immune infiltration phenotypes, effectively encoding spatial cell interplay into the representations. Qualitatively, **Fig. 2A** shows that the predicted tissue maps and structured captions reflect the major features

of the ground truth. **Fig. 2B** shows preserved global arrangement and structural flow of cell-type components across the slide (core image). The low DICE (≈0.22) indicates that the segmentation branch should not be interpreted as a standalone high-accuracy segmentation model; rather, $\mathcal{L}_{\text{seg}}$ serves as an auxiliary spatial/semantic anchor for tissue composition and tumor–immune topology.

Table 1. Tile-level performance on the test set (Stage 1). Loss: weighted sum of losses, including alignment. **Seg DICE**: Segmentation DICE measure. **IM AUC**: Immune infiltration pattern AUROC. **Abun Acc**: Cell type abundance accuracy average. **Count MSE**: Cell Count MSE.

ID	Stage-1 Variant	Loss	Seg DICE	IM AUC	Abun Acc	Count MSE
T1	Full model (S1)	**69.13**	**0.22**	**0.82**	**0.71**	**12.35**
T2	Backbone: Prov-GigaPath	75.6	0.21	0.75	0.6	14.26
T3	Backbone: H-optimus-0	72.16	0.23	0.83	0.72	11.09

Table 2. Performance comparison of Deep-IMMA against clinical and architecture baselines (Stage 2). Values in parenthesis represent the 95% confidence intervals (CI) calculated via 2,000 bootstrap iterations. Abbreviations: AUROC: Area Under the Receiver Operating Characteristic curve, AP: Average Precision, GG: Grade Group, Sens: Sensitivity, Spec: Specificity. Sparsity reg: sparsity regularization. "-" indicates without variant.

ID	Stage-2 Variant	AUROC	AP	F1 (pos)	Sens.	Spec.
E0	Logistic Regression	0.599 (0.49-0.71)	0.396 (0.27-0.56)	0.426 (0.3-0.55)	0.541 (0.38-0.7)	0.648 (0.55-0.73)
E1	ABMIL	0.61 (0.5-0.71)	0.394 (0.26-0.54)	0.346 (0.21-0.47)	0.368 (0.21-0.53)	0.73 (0.65-0.81)
E2	- Stage1 alignment	0.607 (0.47-0.74)	0.513 (0.36-0.66)	0.442 (0.3-0.57)	0.447 (0.29-0.61)	0.796 (0.72-0.87)
E3	Full model	**0.761 (0.67-0.85)**	**0.574 (0.42-0.71)**	**0.553 (0.41-0.68)**	**0.553 (0.39-0.71)**	**0.843 (0.77-0.91)**
E4	- coordinates	0.593 (0.48-0.7)	0.43 (0.29-0.57)	0.366 (0.23-0.49)	0.395 (0.24-0.55)	0.732 (0.65-0.81)
E5	- Top-K selection	0.447 (0.33-0.57)	0.326 (0.21-0.47)	0.27 (0.13-0.4)	0.263 (0.13-0.41)	0.759 (0.68-0.83)
E6	- auxiliary GG	0.583 (0.46-0.7)	0.443 (0.3-0.6)	0.366 (0.24-0.49)	0.447 (0.29-0.6)	0.648 (0.56-0.74)
E7	- sparsity reg	0.575 (0.47-0.68)	0.314 (0.22-0.46)	0.386 (0.25-0.51)	0.447 (0.29-0.61)	0.694 (0.6-0.78)
E8	T2 features	0.532 (0.42-0.64)	0.313 (0.22-0.47)	0.214 (0.08-0.35)	0.158 (0.05-0.28)	0.889 (0.83-0.95)
E9	T3 features	0.386 (0.29-0.49)	0.245 (0.17-0.36)	0.259 (0.15-0.37)	0.368 (0.21-0.53)	0.482 (0.39-0.57)

3.3 Recurrence Prediction and Ablation (Stage 2)

Deep-IMMA (E3) achieved AUROC of 0.761 (**Table 2**), surpassing the linear baseline (E0: AUROC 0.598). While E4, E6, and E7 all contribute to performance, the sharpest decline in E5 confirms that Top-K selection is the most critical factor for BCR prognosis under weak supervision. Performance declines in E1 and E2 confirm that capturing spatial organization and tumor-immune interactions—beyond simple morphology—are the primary drivers of prognostic accuracy. Finally, the underperformance of E8 and

E9 suggests that optimizing the synergy between FM architectures and domain-specific supervision is left for future work.

3.4 Interpretability: Beyond Pixel Accuracy

Interpretability is visualized through representative slide-level immune infiltration maps and Top-15 attention heatmaps (**Fig. 3**). The recurrence case (**Fig. 3A**) exhibits higher prediction confidence (attention score), visualized by darker immune activation intensity, compared to the non-recurrence case (**Fig. 3B**). Spatially, high-attention regions (Red) often overlap with "Mixed" (green) immune infiltration zones, suggesting that the model tends to prioritize complex intra-glandular tumor–immune interactions for risk assessment. Phenotypic distributions also shift with severity; the low-grade, non-recurrence case displays distinct "Cold" (blue) and "Compartmentalized" (orange) patterns, whereas the high-grade recurrence case shifts toward "Mixed" pattern, reflecting differential immune interactions. These spatial patterns provide qualitative model evidence connecting recurrence prediction to tumor–immune dynamics.

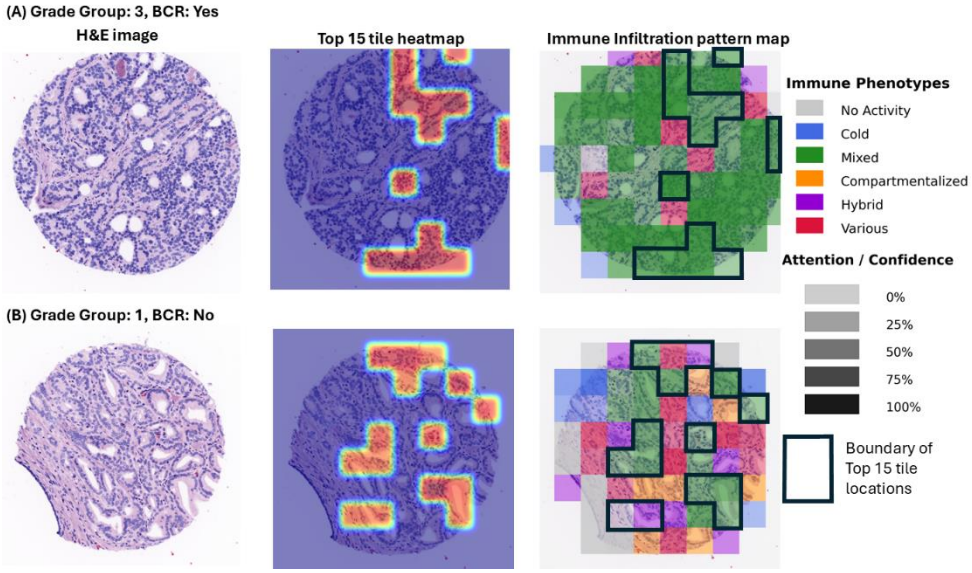

Fig. 3. Interpretability of recurrence prediction (Stage 2). (Column 1-3) Slide-level virtual H&E, Top-15 attention heatmaps, and immune infiltration pattern heatmaps. (A) Recurrence case (Grade Group 3, BCR=Yes). (B) Non-recurrence case (Grade Group 1, BCR=No).

4 Discussion and Conclusion

We presented Deep-IMMA, an immune-aware foundation modeling framework, and demonstrated it for prostate cancer BCR prediction from histological images alone (AUROC 0.761), without requiring MRI or genomic data.

Our findings suggest that tumor-relative immune topology constitutes a dominant prognostic signal that can be encoded through structured multimodal supervision. Our

framework simultaneously generates tile-level text captions, slide-level tissue maps, immune infiltration pattern maps, and attention heatmaps to bridge the gap between computational outputs and clinical outcomes.

Limitations include single-cohort evaluation, use of virtual H&E rather than conventional H&E validation, binary BCR classification rather than time-to-event modeling, and qualitative rather than pathologist-validated interpretability. However, it supports personalized management by localizing high-risk regions and providing interpretable, spatial evidence for recurrence risk.

Acknowledgement. Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (NIH) under Award Number R01CA249899 (PM). Funder website: https://www.cancer.gov/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Morgan, T.M., Boorjian, S.A., Buyyounouski, M.K., et al.: Salvage Therapy for Prostate Cancer: AUA/ASTRO/SUO Guideline Part I: Introduction and Treatment Decision-Making at the Time of Suspected Biochemical Recurrence after Radical Prostatectomy. J Urol. 211, 509–517 (2024).
2. Morgan, T.M., Boorjian, S.A., Buyyounouski, M.K., et al.: Salvage Therapy for Prostate Cancer: AUA/ASTRO/SUO Guideline Part II: Treatment Delivery for Non-metastatic Biochemical Recurrence After Primary Radical Prostatectomy. J Urol. 211, 518–525 (2024).
3. Morgan, T.M., Boorjian, S.A., Buyyounouski, M.K., et al.: Salvage Therapy for Prostate Cancer: AUA/ASTRO/SUO Guideline Part III: Salvage Therapy After Radiotherapy or Focal Therapy, Pelvic Nodal Recurrence and Oligometastasis, and Future Directions. J Urol. 211, 526–532 (2024).
4. Leo, P., Janowczyk, A., Elliott, R., et al.: Computer extracted gland features from H&E predicts prostate cancer recurrence comparably to a genomic companion diagnostic test: a large multi-site study. npj Precis. Onc. 5, 35 (2021).
5. Pinckaers, H., van Ipenburg, J., Melamed, J., et al.: Predicting biochemical recurrence of prostate cancer with artificial intelligence. Commun Med 2, 64 (2022).
6. Cooperberg, M.R., Hilton, J.F., Carroll, P.R.: The CAPRA-S score: A straight-forward tool for improved prediction of outcomes after radical prostatectomy. Cancer. 117, 5039–5046 (2011).
7. Simon, B.D., Harmon, S.A., Merriman, K.M., et al.: A multimodal automated deep learning-based model for predicting biochemical recurrence of prostate cancer following prostatectomy from baseline MRI, Presurgical clinical co-variates. Clin Imaging. 126, 110579 (2025).

8. Ling, C., Tao, N., Maimaitiyimin, A., et al.: Diagnostic systematic review and meta-analysis of machine learning in predicting biochemical recurrence of prostate cancer. Sci Rep. 15, 28378 (2025).
9. Chen, R.J., Ding, T., Lu, M.Y., et al.: A General-Purpose Self-Supervised Model for Computational Pathology, http://arxiv.org/abs/2308.15474, (2023).
10. Charlie, S., Rodolphe, J., Felipe, L.-L., et al.: H-optimus-0, https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0, (2024).
11. Xu, H., Usuyama, N., Bagga, J., et al.: A whole-slide foundation model for digital pathology from real-world data. Nature. 630, 181–188 (2024).
12. Lu, M.Y., Chen, B., Williamson, D.F.K., et al.: A multimodal generative AI copilot for human pathology. Nature. 634, 466–473 (2024).
13. Lu, M.Y., Chen, B., Williamson, D.F.K., et al.: A visual-language foundation model for computational pathology. Nat Med. 30, 863–874 (2024).
14. Chen, W., Zhang, P., Tran, T.N., et al.: A visual–omics foundation model to bridge histopathology with spatial transcriptomics. Nat Methods. 22, 1568–1582 (2025).
15. Li, J., Li, D., Savarese, S., Hoi, S.: BLIP-2: bootstrapping language-image pre-training with frozen image encoders and large language models. Proceedings of the 40th International Conference on Machine Learning. pp. 19730–19742. JMLR (2023).
16. Shao, Z., Bian, H., Chen, Y., et al.: TransMIL: transformer based correlated multiple instance learning for whole slide image classification. Proceedings of the 35th International Conference on Neural Information Processing Systems. pp. 2136–2147. Curran Associates Inc. (2021).
17. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based Deep Multiple Instance Learning, http://arxiv.org/abs/1802.04712, (2018).
18. Jung KJ, Qiu J, Cho S, et al: Multi-Scale Tri-Modal Histology Dataset Integrating Tumor Morphology, Immune Patterns, and Clinical Outcomes. bioRxiv; p. 2026.05.15.725535. doi:10.64898/2026.05.15.725535. (2026).
19. Lu, M.Y., Williamson, D.F.K., Chen, T.Y., et al.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng. 5, 555–570 (2021).
20. Gu, Y., Tinn, R., Cheng, H., et al.: Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. ACM Trans. Computer. Healthcare. 3, 2:1–2:23 (2021).
21. Alex, K., Yarin, G., Roberto, C.: Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7482–7491. (2018).
22. Bergstra, J., Yamins, D., Cox, D.: Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. Proceedings of the 30th International Conference on Machine Learning. pp. 115–123. PMLR (2013).
23. Youden, W.: Index for rating diagnostic tests. Cancer. 3, 32–35 (1950).