Latent Lessons from Histopathology: A Registration-Free Cross-Modal Supervision Framework for mpMRI-based ISUP Grading

Daan Schouten[1,2], Jeroen van der Laak[1], Diederik Somford[4,5], Heidi Küsters-Vandevelde[3], Nadieh Khalili[1], Joeran S. Bosma[8], Anindo Saha[8,9], Jasper Twilt[8,9], Tiago Coelho[8], Maarten de Rooij[10], Henkjan Huisman[10], Jeroen Veltman[6,12], Derya Yakar[7,11], and Geert Litjens[1,2]

[1] Department of Pathology, Radboud University Medical Center, Nijmegen, the Netherlands daan.schouten@radboudumc.nl
[2] Oncode Institute, Utrecht, the Netherlands
[3] Department of Pathology, Canisius Wilhelmina Ziekenhuis, Nijmegen, the Netherlands
[4] Prosper Prostate Cancer Clinics, Nijmegen/Eindhoven, the Netherlands
[5] Department of Urology, Canisius Wilhelmina Ziekenhuis, Nijmegen, the Netherlands
[6] Department of Radiology, Ziekenhuisgroep Twente, Hengelo, the Netherlands
[7] Department of Radiology, University Medical Center Groningen, Groningen, the Netherlands
[8] Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, the Netherlands
[9] Minimally Invasive Image-Guided Intervention Center, Radboud University Medical Center, Nijmegen, the Netherlands
[10] Department of Medical Imaging, Radboud University Medical Center, Nijmegen, the Netherlands
[11] Department of Radiology, Netherlands Cancer Institute, Amsterdam, the Netherlands
[12] Department of Multi-Modality Medical Imaging, Technical Medical Centre, University of Twente, Enschede, the Netherlands

Abstract. Accurate prostate cancer grading is crucial for optimal disease management decisions, but current multiparametric MRI (mpMRI) based models rely on vast amounts of expert-labeled data. We propose a cross-modal latent supervision framework, where we aim to distill rich histopathology information into the mpMRI encoder by self-supervised joint pretraining on unlabeled but patient-matched mpMRI and digitized prostatectomy data. Our framework jointly optimizes a unimodal mpMRI SimCLR objective to learn general mpMRI representations and a multimodal contrastive objective to align these mpMRI embeddings with a prostatectomy embedding obtained through a histopathology foundation model (PRISM). Finetuning this mpMRI encoder with a subset of labeled data demonstrates the effectiveness of our cross-modal supervision framework, outperforming both strong supervised and self-supervised mpMRI baselines for ISUP grading with 5.9 and 4.6 percentage points in quadratic weighted kappa, respectively. Our framework is

registration-free and tolerant to data asymmetries, delivering substantial grading improvements when only 3% of mpMRI data is accompanied by its corresponding prostatectomy embedding during multimodal pretraining.

Keywords: Prostate cancer grading · Cross-modal supervision · Self-supervised learning · Multimodal learning

1 Introduction

Background. Multiparametric MRI (mpMRI) plays an increasingly central role in the diagnostic pipeline for prostate cancer (PCa) detection. Although mpMRI possesses a high sensitivity in detecting clinically significant PCa (csPCa) [6], targeted (or systematic) needle biopsy remains the gold standard in assessing the nature of the suspicious lesion, and in the case of PCa, the cancer aggressiveness through assessment of the ISUP grade group [8]. This ISUP grade group forms the basis for optimal disease management decisions, where treatment is typically reserved for csPCa [9].

To automate csPCa detection, recent years have shown a surging interest in the development of mpMRI-based csPCa detection models, culminating in the international PI-CAI challenge demonstrating superiority of current models compared to a pool of 62 radiologists [13]. Despite these encouraging results, these models typically rely on vast amounts of expert-labeled data without leveraging the abundance of unlabeled data that may be available. Moreover, extending this binary csPCa vs non-csPCa detection paradigm to exact ISUP grade group prediction could further improve optimal disease management decisions.

Related work. Previous work going beyond the binary csPCa dichotomy typically investigates either direct ISUP grade group prediction [16] or distinguishes between the most clinically relevant ISUP grades [2,1,7,12]. A promising direction in this field is the inclusion of histopathology image data during training for multimodal feature learning [2,1]. Indeed, as the ISUP grade group is directly derived from histopathology, one may expect that teaching the mpMRI-based model to detect histopathology-style features could improve its ability to distinguish between ISUP grades. Importantly, the matched mpMRI-histopathology data is only presented to the model during training, with the final model making its predictions purely on mpMRI.

The main limitation of these models, however, relates to their requirement of (coarse) registration between modalities to enable voxel- or slice-level supervision [2,1]. Specifically, distilling information from a prostatectomy whole-slide image (WSI) to the mpMRI encoder requires matching this WSI with the corresponding slice in the mpMRI volume. Since multimodal registration is a difficult task and often requires manual input, this severely limits the scale of multimodal datasets. In addition, current cross-modal supervision methods may also impose a rigid task-specific architectural design, for example through the use

of a geometry-consistent GAN [2], limiting their general applicability. These challenges have thus far limited the introduction of a more general cross-modal supervision framework, where rich histopathology data can be distilled into the mpMRI encoder without the need for sophisticated registration methods.

Contributions. In this work, we propose a registration-free framework for cross-modal supervision, applied to the task of ISUP grading on mpMRI. Rather than chasing spatial supervision through (coarse) registration, we redesign the supervision task to occur exclusively in latent space. This not only removes the need for registration, but also elegantly unlocks flexibility in the supervisory modality, circumventing the need for online optimization of the supervisory modality's encoder through incorporation of readily available foundation models.

Our main contributions can thus be summarized as the following: i) we demonstrate the effectiveness of self-supervised pretraining vs. a strong supervised baseline, ii) we extend the unimodal pretraining paradigm to multimodal imaging data, and iii) we showcase the potential of cross-modal supervision through off-the-shelf foundation models. Our code is available at `https://github.com/DIAGNijmegen/crossmodal-supervision`.

2 Methodology

The general paradigm of cross-modal supervision in this work is inspired by CLIP [11], which leveraged natural language as supervisory modality to images. Analogously, we can treat histopathology as supervisory modality to mpMRI, and train a model with a contrastive objective to bring these modalities closer together in latent space. Our method effectively consists of two learning branches: a unimodal self-supervised branch that learns general mpMRI representations and a multimodal contrastive branch that aligns these mpMRI representations with the corresponding histopathology representation. This two-branch architecture is schematically visualized in Figure 1.

Unimodal self-supervised learning. The unimodal self-supervised branch entails a three-dimensional adaptation of the SimCLR framework [3] with two domain-specific adaptations. First, given the importance of the augmentation strategy to create the augmented views, we tailor the augmentations to 3D grayscale mpMRI data. Specifically, we employ modest affine transformations, image histogram modifications through gamma transforms and bias field perturbations. The strength of the intensity augmentations was determined empirically, aiming for clear discrepancies in per-channel image histograms across the two views to prevent shortcut learning [3]. Second, a key modification in our implementation is the use of an automatically generated prostate gland mask to generate a $256 \times 256 \times 16$ crop to ensure the presence of the prostate in the crop, preventing two distinct anatomical structures (i.e., the bladder and the rectum) from being designated as a positive pair for a given instance. The two

augmented views are then generated by sampling a random point in the prostate gland mask and performing a center crop around that point.

As each mpMRI study consists of three sequences (i.e., T2-weighted (T2w), high b-value (HBV) and apparent diffusion coefficient (ADC) map), these sequences were handled as multi-channel input to a 3D ResNet-18 model [10]. The objective for this unimodal branch was formulated as minimizing the normalized temperature-scaled cross-entropy (NT-Xent) loss [3]. Formally, for a minibatch of $N = 64$ mpMRI studies, we generate two augmented views per study, yielding $2N$ projected embeddings $\{z_k\}_{k=1}^{2N}$. For a positive pair of views (i, j) (two augmentations of the same study), the NT-Xent loss is

$$\ell_{\text{uni}}(i,j) = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau_{\text{uni}})}{\sum_{k=1}^{2N} \mathbf{1}_{[k\neq i]} \exp(\text{sim}(z_i, z_k)/\tau_{\text{uni}})}, \tag{1}$$

where $\text{sim}(\cdot,\cdot)$ denotes cosine similarity and $\tau_{\text{uni}} > 0$ is the temperature. The total loss $\mathcal{L}_{\text{uni}}$ for a batch is then computed by averaging $\ell_{\text{uni}}(i, j)$ for all positive pairs (i, j) in a batch.

Cross-modal supervision. The cross-modal supervision branch handles the distillation of histopathology information to the mpMRI encoder, ensuring that mpMRI studies with similar histopathological features are close in latent space. As patient-level paired mpMRI-prostatectomy data is only available for a small subset of cases, this objective is only applied to these multimodal cases in each batch. Each prostatectomy contains multiple WSIs which are individually embedded with the PRISM foundation model [15] and max-pooled along the slide-dimension for a single static 1280-D prostatectomy embedding used in training. This prostatectomy embedding is then passed through a projection head to map it into a shared representation space with the embeddings of the mpMRI crops. The prostatectomy embedding is then contrasted with the mpMRI embeddings through the multimodal NT-Xent loss, which is conceptually similar to the unimodal NT-Xent loss but defines positive pairs inter-modality rather than intra-modality.

Formally, let $M = 16$ denote the number of cases in the batch with paired prostatectomy data, and index these paired cases by $n \in \{1, \ldots, M\}$. Let y_n denote the projected prostatectomy embedding for paired case n, and let $z_{n,1}$ and $z_{n,2}$ denote the embeddings of the two augmented mpMRI views. For an anchor view $z_{n,v}$ with $v \in \{1, 2\}$, we define the multimodal NT-Xent loss as

$$\ell_{\text{multi}}(n,v) = -\log \frac{\exp(\text{sim}(z_{n,v}, y_n)/\tau_{\text{multi}})}{\sum_{m=1}^{M} \exp(\text{sim}(z_{n,v}, y_m)/\tau_{\text{multi}})}, \tag{2}$$

and, similar to the unimodal NT-Xent loss, average over all positive pairs $(z_{n,v}, y_n)$ to compute $\mathcal{L}_{\text{multi}}$. The final loss $\mathcal{L}$ is given by a weighted sum of unimodal and multimodal terms, $\mathcal{L} = \mathcal{L}_{\text{uni}} + \alpha\,\mathcal{L}_{\text{multi}}$, where α is a scalar loss weight. We note

again that $\mathcal{L}_{\text{multi}}$ does not receive contributions from the $N-M$ unpaired cases in the batch.

Training recipe. Our cross-modal supervision framework entails a two-step training approach, consisting of general multimodal pretraining and task-specific finetuning. The model pretraining stage consists of 30 epochs in which the unimodal and multimodal NT-Xent losses are jointly optimized using the SGD optimizer (weight decay=6×10^{-4}) using an initial learning rate of 3×10^{-3} that decayed to zero over the course of training. Following the original CLIP implementation [11], the temperature parameter in both NT-Xent loss functions is formulated as a learnable parameter and directly optimized during training to prevent tuning as hyperparameter. To combat data asymmetry, the dataloader oversamples cases with paired mpMRI-histopathology data, ensuring that each batch of 64 mpMRI scans contains 16 instances with paired data. The α for the multimodal loss component $\mathcal{L}_{\text{multi}}$ was empirically set to 0.6233.

Following pretraining the model was finetuned on the downstream tasks by adding a single linear layer as classification head. The loss function was based on the exact downstream task, using binary cross-entropy for binary csPCa detection and Soft Labels for Ordinal Regression (SORD) [4] to respect the ordinal nature of the ISUP grading task. Finetuning was performed for 60 epochs using AdamW (learning rate=1×10^{-3}, weight decay=1×10^{-4}) with a cosine annealing schedule and linear learning rate warmup of 6 epochs. To mitigate forgetting of pretraining features, the newly initialized classification head was trained with the aforementioned learning rate while the full backbone was tuned with a $10\times$ smaller learning rate. For each downstream task we found that finetuning the full encoder yielded better performance than finetuning the last few layers.

3 Experiments and Results

MRI data. We retrospectively included 6680 unique patients with 10037 MRI studies between June 2011 and November 2025 in our institution. Each study consisted of an axial T2w acquisition, an HBV acquisition (b$\geq$ 1000 s/mm^2) and an ADC map. Clinical outcome or follow-up was established for 492 studies, with targeted biopsy-confirmed ISUP grade group labels in 383/492 (77.8%) studies and 109/492 (22.2%) benign cases without suspicious MRI findings (PI-RADS $\leq$ 2) and $\geq$ 3 years of follow-up without rising PSA levels. Informed consent for this study was waived due to the use of deidentified, retrospectively collected data. In addition to this institutional cohort, we supplement our dataset with the multicenter PI-CAI Public and Private Training set spanning 7763 unique patients with 8822 MRI studies, all containing a study-level ISUP grade group label. Detailed inclusion criteria and assessment of clinical outcome in this PI-CAI set have been reported elsewhere [13]. The combined dataset thus consisted of 13075 unique patients with 18859 prostate MRI studies of which 9314/18859 (49.4%) contained a study-level ISUP grade group label. This full set was partitioned into a model development set of 11075 patients and 16561 studies

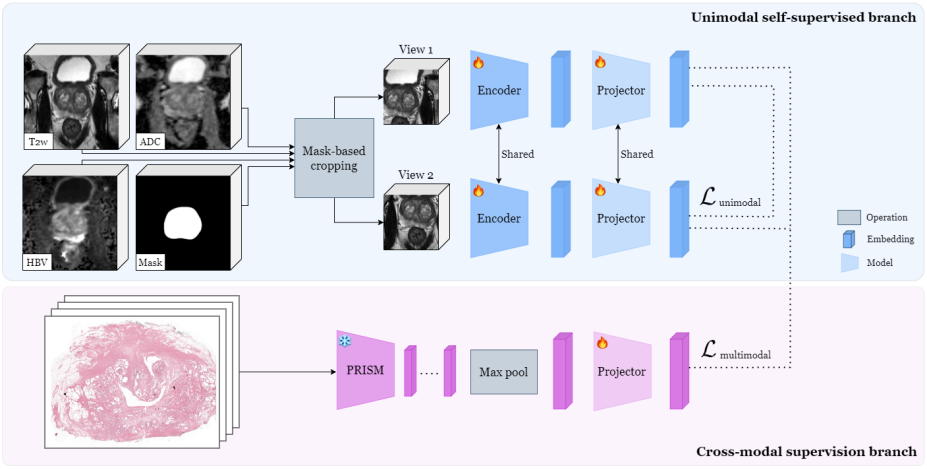

Fig. 1. Schematic overview of the proposed cross-modal supervision framework. The unimodal self-supervised branch aims to learn generally useful mpMRI representations and the cross-modal branch aims to align these representations with the corresponding histopathology representation.

(7016 ISUP labels), and an internal test set of 2000 patients and 2298 studies (2298 ISUP labels). Partitioning was performed on a patient-level basis, and the 2000 patients in the internal test set were randomly selected through a stratified sampling approach, ensuring a consistent ISUP grade group distribution with respect to the development set.

Histopathology data. For 290/11075 patients in the development set that underwent radical prostatectomy between February 2018 and June 2023, we collected and digitized the radical prostatectomy histopathology slides. All prostatectomy data originated from the institutional cohort in the development set. On average, these prostatectomies resulted in 5 (range 2-8) whole-mount slides corresponding to the different levels of the prostate, excluding the base and apex. While 189/290 prostatectomies were sectioned with the whole-mount technique, 101/290 prostatectomies contained at least one level which was further sectioned in halves/quarters. In these cases, the PythoStitcher algorithm [14] was employed to create pseudo whole-mount slides at full resolution to facilitate data harmonization across sectioning techniques. All slides were stained with hematoxylin and eosin and scanned on a 3DHISTECH P1000 scanner at 0.25 µm/pixel resolution.

Experiments. We validate our framework in two settings, namely ISUP grading and binary csPCa detection. We frame the ISUP grading task as a multiclass ($n = 6$) classification task, aiming to distinguish between each of the five ISUP grade groups and the benign cases. For this task we report performance through

Table 1. ISUP grade group classification performance on the internal test set. Best results per metric are shown in **bold**. SSL refers to self-supervised pretraining.

	Internal Test Set ($n = 2298$)	
Method	QWK (↑)	MAE (↓)
Supervised	0.617	0.611
Unimodal SSL	0.630	0.602
Multimodal SSL	**0.676**	**0.541**

the quadratic weighted kappa (QWK) and mean absolute error (MAE) in grade prediction. For the binary csPCa detection task we consider all ISUP grade group ≥ 2 cases to be csPCa, while considering both indolent cancers (ISUP = 1) and benign cases to be non-csPCa. To assess the performance of our method, we report the AUROC, accuracy, sensitivity, and specificity.

We pretrain our base model once using the cross-modal supervision framework, and finetune this model specifically for each of the two tasks. We compare the performance of our framework against a baseline 3D ResNet-18 model trained from scratch on the same labeled data using the same setup. In addition, we compare against a self-supervised mpMRI-only baseline that effectively ablates the multimodal training objective while being trained on the same unlabeled mpMRI dataset. We note that the exact same 3D ResNet-18 model and, where possible, hyperparameter setup was used across the proposed framework and baseline methods to ensure an apples-to-apples comparison.

Prostate cancer grading. The ISUP grading performance for each method is summarized in Table 1. The results on the internal test set indicate that while self-supervised pretraining on mpMRI data already unlocks a modest performance boost compared to the supervised baseline, our cross-modal supervision framework delivers the largest boost to grading performance.

Prostate cancer detection. The results for the binary csPCa detection task are summarized in Table 2 and visualized in Figure 2. Similar to the ISUP grading task, leveraging self-supervised unimodal pretraining on the unlabeled mpMRI data notably boosts classification performance. Adding cross-modal supervision on top of this self-supervised baseline further boosts performance.

Ablation study. In our ablation study (Table 3) we investigate the use of different slide-level foundation models (Prov-GigaPath [17] and TITAN [5]) and different pooling methods for the conversion of multiple WSI-level embeddings into a single patient-level representation. For each ablation, we retrain the mpMRI encoder with the same multimodal setup and the same hyperparameters, and subsequently finetune it for ISUP grading.

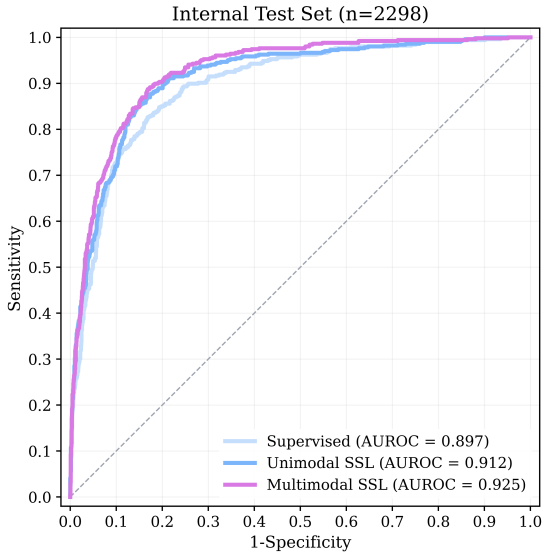

Fig. 2. ROC curve comparison of the two baseline methods and our proposed multimodal framework for binary csPCa detection.

Table 2. Binary csPCa detection performance on the internal test set. Best results per metric are shown in **bold**. SSL refers to self-supervised pretraining.

	Internal Test Set ($n = 2298$)			
Method	AUROC	Accuracy	Sensitivity	Specificity
Supervised	0.897	0.815	0.842	0.807
Unimodal SSL	0.912	0.836	0.870	0.826
Multimodal SSL	**0.925**	**0.843**	**0.887**	**0.831**

4 Conclusion

In conclusion, we have presented a self-supervised multimodal pretraining framework, extending the more common unimodal pretraining paradigm through incorporation of multimodal imaging data. We show that the supervisory modality encoder can be an off-the-shelf foundation model, effectively distilling this foundation model into the modality encoder of interest. Unlike previous methods, our cross-modal supervision occurs exclusively in latent space, circumventing the need for multimodal image registration. Furthermore, our framework demonstrates strong supervision capabilities when only small amounts of multimodal data are available, enabling its application to asymmetrical multimodal datasets. Applied to prostate cancer grading and detection, our framework outperforms both strong supervised and self-supervised unimodal methods. Future work will

Table 3. ISUP grade group classification performance on the internal test set. Best results per metric are shown in **bold**.

		Internal Test Set ($n = 2298$)	
Foundation Model	WSI Pooling	QWK (↑)	MAE (↓)
Prov-GigaPath [17]	Max	0.660	0.579
TITAN [5]	Max	0.667	0.573
PRISM [15]	Mean	0.663	0.570
PRISM [15]	Attention	0.672	**0.530**
PRISM [15]	Max	**0.676**	0.541

investigate the online optimization of the supervisory modality encoder to facilitate deeper multimodal integration.

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

References

1. Bhattacharya, I., Seetharaman, A., Kunder, C., Shao, W., Chen, L.C., Soerensen, S.J., Wang, J.B., Teslovich, N.C., Fan, R.E., Ghanouni, P., Brooks, J.D., Sonn, G.A., Rusu, M.: Selective identification and localization of indolent and aggressive prostate cancers via CorrSigNIA: An MRI-pathology correlation and deep learning framework. Medical Image Analysis **75**, 102288 (Jan 2022). https://doi.org/10.1016/j.media.2021.102288
2. Bhattacharya, I., Shao, W., Li, X., Soerensen, S.J.C., Fan, R.E., Ghanouni, P., Brooks, J.D., Sonn, G.A., Rusu, M.: ArtHiFy: Artificial histopathology-style features for improving MRI-based prostate cancer detection. In: Medical Imaging 2024: Computer-Aided Diagnosis. vol. 12927, pp. 441–447. SPIE (Apr 2024). https://doi.org/10.1117/12.3006879
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A Simple Framework for Contrastive Learning of Visual Representations. In: Proceedings of the 37th International Conference on Machine Learning. pp. 1597–1607. PMLR (Nov 2020)
4. Díaz, R., Marathe, A.: Soft Labels for Ordinal Regression. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4733–4742. IEEE, Long Beach, CA, USA (Jun 2019). https://doi.org/10.1109/CVPR.2019.00487
5. Ding, T., Wagner, S.J., Song, A.H., Chen, R.J., Lu, M.Y., Zhang, A., Vaidya, A.J., Jaume, G., Shaban, M., Kim, A., Williamson, D.F.K., Robertson, H., Chen, B., Almagro-Pérez, C., Doucet, P., Sahai, S., Chen, C., Chen, C.S., Komura, D., Kawabe, A., Ochi, M., Sato, S., Yokose, T., Miyagi, Y., Ishikawa, S., Gerber, G., Peng, T., Le, L.P., Mahmood, F.: A multimodal whole-slide foundation model for pathology. Nature Medicine **31**(11), 3749–3761 (Nov 2025). https://doi.org/10.1038/s41591-025-03982-3

6. Drost, F.J.H., Osses, D., Nieboer, D., Bangma, C.H., Steyerberg, E.W., Roobol, M.J., Schoots, I.G.: Prostate Magnetic Resonance Imaging, with or Without Magnetic Resonance Imaging-targeted Biopsy, and Systematic Biopsy for Detecting Prostate Cancer: A Cochrane Systematic Review and Meta-analysis. European Urology **77**(1), 78–94 (Jan 2020). https://doi.org/10.1016/j.eururo.2019.06.023
7. Duran, A., Dussert, G., Rouvière, O., Jaouen, T., Jodoin, P.M., Lartizien, C.: ProstAttention-Net: A deep attention model for prostate cancer segmentation by aggressiveness in MRI scans. Medical Image Analysis **77**, 102347 (Apr 2022). https://doi.org/10.1016/j.media.2021.102347
8. Epstein, J.I., Egevad, L., Amin, M.B., Delahunt, B., Srigley, J.R., Humphrey, P.A., Committee, t.G.: The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. The American Journal of Surgical Pathology **40**(2), 244 (Feb 2016). https://doi.org/10.1097/PAS.0000000000000530
9. European Association of Urology: EAU Guidelines. Edn. presented at the EAU Annual Congress Madrid 2025 (2025)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition (Dec 2015). https://doi.org/10.48550/arXiv.1512.03385
11. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning Transferable Visual Models From Natural Language Supervision (Feb 2021). https://doi.org/10.48550/arXiv.2103.00020
12. Rajagopal, A., Westphalen, A.C., Velarde, N., Simko, J.P., Nguyen, H., Hope, T.A., Larson, P.E.Z., Magudia, K.: Mixed Supervision of Histopathology Improves Prostate Cancer Classification From MRI. IEEE Transactions on Medical Imaging **43**(7), 2610–2622 (Jul 2024). https://doi.org/10.1109/TMI.2024.3382909
13. Saha, A., Bosma, J.S., Twilt, J.J., van Ginneken, B., Bjartell, A., Padhani, A.R., Bonekamp, D., Villeirs, G., Salomon, G., Giannarini, G., Kalpathy-Cramer, J., Barentsz, J., Maier-Hein, K.H., Rusu, M., Rouvière, O., van den Bergh, R., Panebianco, V., Kasivisvanathan, V., Obuchowski, N.A., Yakar, D., Elschot, M., Veltman, J., Fütterer, J.J., de Rooij, M., Huisman, H., PI-CAI consortium: Artificial intelligence and radiologists in prostate cancer detection on MRI (PI-CAI): An international, paired, non-inferiority, confirmatory study. The Lancet. Oncology **25**(7), 879–887 (Jul 2024). https://doi.org/10.1016/S1470-2045(24)00220-1
14. Schouten, D., van der Laak, J., van Ginneken, B., Litjens, G.: Full resolution reconstruction of whole-mount sections from digitized individual tissue fragments. Scientific Reports **14**(1), 1497 (Jan 2024). https://doi.org/10.1038/s41598-024-52007-5
15. Shaikovski, G., Casson, A., Severson, K., Zimmermann, E., Wang, Y.K., Kunz, J.D., Retamero, J.A., Oakley, G., Klimstra, D., Kanan, C., Hanna, M., Zelechowski, M., Viret, J., Tenenholtz, N., Hall, J., Fusi, N., Yousfi, R., Hamilton, P., Moye, W.A., Vorontsov, E., Liu, S., Fuchs, T.J.: PRISM: A Multi-Modal Generative Foundation Model for Slide-Level Histopathology (May 2024). https://doi.org/10.48550/arXiv.2405.10254
16. de Vente, C., Vos, P., Hosseinzadeh, M., Pluim, J., Veta, M.: Deep Learning Regression for Prostate Cancer Detection and Grading in Bi-Parametric MRI. IEEE Transactions on Biomedical Engineering **68**(2), 374–383 (Feb 2021). https://doi.org/10.1109/TBME.2020.2993528

17. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., Xu, Y., Wei, M., Wang, W., Ma, S., Wei, F., Yang, J., Li, C., Gao, J., Rosemon, J., Bower, T., Lee, S., Weerasinghe, R., Wright, B.J., Robicsek, A., Piening, B., Bifulco, C., Wang, S., Poon, H.: A whole-slide foundation model for digital pathology from real-world data. Nature **630**(8015), 181–188 (Jun 2024). https://doi.org/10.1038/s41586-024-07441-w