Foundation-Model Ensembling with Vision–Language-Informed Representations for Glioblastoma Subregion Classification

Lingyi Zhao[1 [0000−0001−8544−3326]], Xiaoyi Zhu[1 [0000−0003−4738−0779]], and Khurram Shafique[1 [0000−0002−3834−324X]]

Novateur Research Solutions, Ashburn VA 20147, USA
lzhao@novateur.ai, xzhu@novateur.ai, kshafique@novateur.ai
https://novateur.ai/

Abstract. Patch-level classification of glioblastoma histopathologic subregions is complicated by severe class imbalance, limited patient diversity, and substantial morphological similarity between classes. We present a two-level ensemble that combines diversity across foundation-model architectures, training-set class distributions, and learned fusion heads. Virchow2 and UNI2-h encoders are adapted using low-rank adaptation on complementary prevalence-preserving and class-balanced datasets constructed from the same patient-disjoint five-fold partition. To complement the visual encoders with language-informed semantic cues, we incorporate CONCH image embeddings together with similarity scores to predefined pathology-relevant textual concepts. All encoder representations are integrated using structured fusion heads warm-started from the independently trained classification heads and subsequently refined through low-rank head adaptations and trainable mixing weights. Our final model averages the class probabilities produced by two fusion models: one trained on the class-balanced dataset using three LoRA-adapted encoders, comprising Virchow2 and two UNI2-h models, and another trained on the prevalence-preserving dataset using the same encoders together with the frozen CONCH representation. A targeted embedding-space router then redistributes probability mass between the white-matter and cortical-infiltration classes. On the BraTS-Path 2026 official validation set, the proposed approach achieves an accuracy of 0.9264, a Matthews correlation coefficient of 0.8976, and a macro-F1 of 0.7364. These results demonstrate the value of combining complementary training distributions, vision–language representations, and targeted embedding-level refinement for imbalanced computational pathology classification.

Keywords: Pathology foundation models · Vision–language models · Glioblastoma subregion classification.

1 Introduction

Glioblastoma is the most common malignant primary parenchymal brain tumor, with a grim prognosis and a median survival of only 12–18 months[12, 14, 15]. It

is widely infiltrative across the cerebral hemispheres and highly heterogeneous in both molecular profile and histopathologic appearance. This heterogeneity is a major obstacle to effective treatment[14, 15]. Correctly diagnosing these tumors and characterizing their heterogeneity is therefore central to selecting precise therapy. In the gold-standard histopathology workflow, this involves identifying distinct morpho-pathological features throughout digitized tissue sections[14, 10, 11, 1], including cellular tumor (CT), geographic necrosis (NC), pseudopalisading necrosis (PN), infiltration into the cortex (IC), microvascular proliferation (MP), white-matter infiltration (WM), leptomeningeal infiltration (LI), dense macrophages (DM), and presence of lymphocytes (PL). The BraTS-Path challenge advances deep-learning models that identify these tumor subregions consistently, to assist diagnosis and grading. Expert neuropathologists annotate the tissue sections, the annotated regions are divided into same-size patches, and each patch is treated as an individual case[1].

Given the limited amount of labeled data and substantial histopathologic variability, pretrained pathology foundation models provide a strong basis for this task. Self-supervised pathology foundation models [3, 16] pretrained on large collections of whole-slide images, including Virchow2 [17], UNI/UNI2-h [2], H-optimus[13], and Midnight[6], as well as vision–language models such as CONCH[9] learn transferable representations of tissue morphology that can be adapted to downstream classification tasks. However, fully fine-tuning these ViT-scale backbones is computationally expensive and prone to overfitting on the modestly sized labeled sets. We therefore use low-rank adaptation (LoRA) [5], which keeps the pretrained backbone frozen while optimizing a small set of low-rank adapter parameters, corresponding to approximately 0.1% of the model parameters. Each classifier is constructed from a frozen pathology foundation model, LoRA adapters, and a lightweight classification head.

Although adapting pathology foundation models yields strong individual classifiers, no single encoder or training distribution fully resolves the severe class imbalance in this task. The two most prevalent classes, CT and NC, collectively account for approximately 47% of all patches, whereas LI, PL, and DM represent only 2.53%, 2.02%, and 1.88%, respectively. In our experiments, models trained on the prevalence-preserving distribution achieved higher accuracy and Matthews correlation coefficient (MCC) but showed lower F1 scores for the underrepresented classes. Conversely, models trained on a more balanced distribution improved performance on these underrepresented classes while modestly reducing performance on the dominant classes. These complementary behaviors motivate an ensemble that varies both the foundation-model encoder and the training-set class distribution. This design combines architectural diversity with the stronger accuracy and MCC performance of prevalence-preserving training and the improved minority-class sensitivity of balanced training.

Despite ensembling, many remaining errors were concentrated among a small number of morphologically similar class pairs, most notably WM and IC. Preliminary analysis indicated that the frozen backbone embeddings retained discriminative information for this distinction that was not fully expressed by the

multiclass prediction head. We therefore introduce a targeted post-hoc router that re-examines the embedding only when WM and IC are the two highest-probability classes and refines their relative probabilities while leaving all other class probabilities unchanged.

The limited number of labeled patches and contributing patients, particularly for rare classes, makes it difficult to learn robust class-specific representations from the challenge data alone. We therefore incorporate CONCH to complement the task-specific visual encoders with knowledge acquired through vision–language pretraining. By aligning each patch with predefined pathology-relevant text concepts, CONCH provides pretrained, language-informed semantic cues that do not rely solely on the available class labels. These concept scores are combined with the visual embedding to provide complementary information that may improve recognition of underrepresented morphologies.

In summary, our contributions are threefold. First, we construct an architecture- and prior-diverse ensemble of LoRA-adapted pathology foundation models trained on complementary prevalence-preserving and class-balanced datasets. Second, we introduce a targeted, mass-preserving WM/IC embedding-space router to refine predictions for this challenging class pair. Third, we develop warm-started fusion heads that integrate multiple encoder representations, including a frozen CONCH vision–language representation.

2 Methods

2.1 Data specification

Each sample in the dataset is a 512×512 H&E-stained image patch acquired at $40\times$ magnification, corresponding to a spatial resolution of approximately 0.25 μm/pixel and a field of view of about 128×128 μm. Each patch is assigned to one of ten histopathologic classes: CT, PN, MP, NC, IC, WM, LI, DM, PL, or none of the above (NOTA) [1]. The dataset is highly imbalanced, with common tumor and necrosis classes substantially outnumbering rare classes such as PL and LI. This imbalance, together with the morphological similarity among several subregions, motivates the complementary modeling strategies introduced below.

2.2 Patient-Disjoint Dataset Construction and Tail-Class Augmentation

To preserve the original class distribution while making use of as many available patches as possible, and to provide a complementary training set with a more balanced class distribution, we construct two datasets from the same patient-disjoint five-fold partition: a prevalence-preserving dataset and a class-balanced dataset. In both datasets, all samples from a given patient remain within a single fold. The prevalence-preserving dataset retains the relative class frequencies of the original training pool. The class-balanced dataset instead samples up to approximately 1000 patches per class in each fold, subject to the number of patches

available for each class and patient, thereby increasing the relative contribution of underrepresented classes. In both datasets, patches are allocated across contributing patients using a water-filling procedure together with a per-patient cap, limiting the influence of patients that contribute disproportionately large numbers of patches while preserving the predefined patient-level fold assignments.

To further address class-specific deficits, we apply online resampling and tiered augmentation during training. Light augmentation consisting of horizontal and vertical flips, rotations by multiples of 90°, and mild color jitter is applied to all classes. Samples from the patient-limited LI and PL classes, including all resampled copies, receive stronger augmentation with increased color jitter, random resized cropping, affine perturbation, and Gaussian blur. The resampling strategy is tailored to classes with insufficient patch counts within individual training folds.

2.3 Task and validation protocol

Model development was primarily guided by patient-disjoint five-fold cross-validation, with performance assessed from pooled out-of-fold (OOF) predictions. Official validation feedback was additionally used as a secondary criterion when deciding whether to retain subsequent modifications, as detailed in Supplementary Sec. 8. To prevent information leakage, folds are defined at the patient/slide level rather than patient/class. Because multiple classes from the same patient may appear on the same slide, splitting by patient/class can place patches with correlated slide-specific features in both training and validation folds, leading to overly optimistic performance, particularly for rare classes.

2.4 LoRA-adapted foundation models

We first construct the core ensemble members by adapting Virchow2 and UNI2-h to the ten-class classification task using low-rank adaptation (LoRA). These models serve as the primary task-specific classifiers, while additional frozen foundation-model representations are introduced in the following section.

All LoRA-adapted ensemble members use the same overall model design. Each classifier starts from a frozen pathology foundation-model backbone and is adapted using LoRA modules (r=4, α=8) inserted into the attn.qkv projection of every transformer block, resulting in approximately 0.1% trainable parameters. The final patch representation is formed by concatenating the class token with the mean-pooled patch tokens and is passed to a linear classification head.

Input preprocessing and data augmentation Each 512×512 patch is resized to 224×224 before being passed to the backbone, preserving the full field of view. Preprocessing follows the original configuration used to train each foundation model, including the corresponding normalization and interpolation settings. Augmentation is applied only to the training data, while validation uses resizing and normalization alone.

One-vs-rest focal-BCE head The classification head outputs one logit per class, and each logit is optimized using a one-vs-rest binary focal-BCE objective. For class c, patches labeled as class c are treated as positive examples, while patches belonging to any of the remaining $C-1$ classes are treated as negative examples. This formulation allows the loss to account for the substantial variation in class prevalence across the dataset.

To reduce the effect of class imbalance, positive examples for class c are weighted by $\text{pos\_weight}_c = (n_{\neg c}/n_c)^\rho$, where ρ is the positive-weight power, and n_c and $n_{\neg c}$ denote the numbers of positive and negative training examples, respectively. Raising the negative-to-positive ratio to a power of $\rho \leq 1$ moderates classes with extreme imbalance while still increasing the influence of rare positives. ρ is optimized based on empirical performance during model development.

We further use focal modulation to reduce the influence of easily classified examples. Specifically, the binary cross-entropy loss is multiplied by $(1-p_t)^\gamma$, where p_t denotes the predicted probability assigned to the correct binary target: $p_t = p_c$ for a positive example and $p_t = 1 - p_c$ for a negative example. The focusing parameter γ controls how strongly well-classified examples are downweighted, with larger values placing greater emphasis on difficult examples. We set γ based on empirical performance during model development [8]

2.5 CONCH Integration

To incorporate pathology-relevant textual concepts alongside visual morphology, we include CONCH as an additional frozen representation. CONCH (CONtrastive learning from Captions for Histopathology) [9] is pretrained on paired histopathology images and captions, aligning visual and textual representations within a shared embedding space. Leveraging this shared space, we measure the alignment between each patch and predefined concept prompts to derive complementary, semantically meaningful cues that may not be fully captured by the image embedding alone.

Specifically, we use CONCH as a frozen feature extractor, leaving both its vision and text encoders unchanged. For each patch, we concatenate the ℓ_2-normalized 512-dimensional image embedding with an 18-dimensional concept-score vector:

$$\phi_{\text{CONCH}}(x) = \left[\hat{e}(x) \;\middle|\; \hat{P}\,\hat{e}(x) \right] \in \mathbb{R}^{530}, \tag{1}$$

where $\hat{e}(x) \in \mathbb{R}^{512}$ is the normalized image embedding and $\hat{P} \in \mathbb{R}^{18\times 512}$ contains normalized text prototypes representing nine disease-related and nine background or artifact concepts. Specifically, we define 18 concepts, including 9 disease-related concepts and 9 background/artifact categories, with each described by several natural-language text descriptions. Every description is wrapped in four prompt templates, and the concept's prototype is the ℓ_2-normalized mean of the CONCH text embeddings across all its description–template combinations. The resulting concept scores, $\hat{P}\hat{e}(x)$, are cosine similarities between the normalized image embedding and the text prototypes, providing a language-informed

complement to the visual representation. For each cross-validation fold, a multiclass logistic-regression head is fitted on the corresponding training folds to map $\phi_{\mathrm{CONCH}}(x)$ to ten class logits.

2.6 Warm-Started Ensemble Fusion Head

Ensemble members. The fusion framework draws on three LoRA-adapted pathology foundation-model encoders with jointly trained linear classification heads: Virchow2 and UNI2-h trained on the prevalence-preserving dataset, and a second UNI2-h trained on the class-balanced dataset. Selected fusion configurations additionally incorporate the frozen CONCH representation and its logistic-regression head, as described in Sec. 2.5.

Warm-started trainable fusion head. We construct the ensemble decoder as a structured two-layer linear network that operates on the concatenated representations of the participating foundation models. Depending on the configuration (Sec. 2.9), the input contains either the three LoRA-adapted encoder embeddings or these embeddings together with the frozen CONCH representation. The first layer produces ten logits for each included model and has a block-diagonal weight structure, such that each set of logits depends only on its corresponding representation. Each diagonal block, including its bias, is initialized from that model's pretrained classification head. The resulting member-specific logits are concatenated and passed to a second linear layer, L_2, whose coefficients are initialized from the optimized ensemble weights.

The pretrained member heads remain frozen during fusion-head training. For each member m, we add a low-rank update to its frozen affine head. The fusion layer L_2 and the low-rank member-specific updates are then optimized jointly, while the underlying encoders and original classification heads remain fixed. The fusion head is trained using focal binary cross-entropy. Optimization uses AdamW with a low learning rate and cosine annealing. Model selection is performed using out-of-fold MCC and macro-F1 results.

2.7 Fold-Specific Calibration

The across-fold ensemble is obtained by averaging calibrated probabilities. Specifically, for ensemble member m, let $z_{m,f}$ denote the logits produced by fold-specific model $f \in \{1, \ldots, 5\}$. We estimate a scalar temperature $T_{m,f}$ for each member-fold pair by minimizing the negative log-likelihood on the corresponding held-out out-of-fold predictions [4]. For each fold-specific model, the ten logits are divided by the fitted scalar temperature and converted to a multiclass probability distribution using the softmax function. The resulting calibrated probabilities are then averaged across folds. Calibrating before fold averaging allows each fold model to contribute probabilities on a comparable confidence scale, rather than requiring one shared temperature to account for variation across all folds and members. This calibration is applied post hoc and does not require retraining the foundation-model classifiers.

2.8 WM/IC embedding-readout routing gate

Our empirical analysis showed that the multiclass head does not fully exploit the information encoded in the backbone representation for distinguishing WM from IC. Specifically, we observed that a lightweight classifier trained on the frozen embeddings achieved better WM/IC discrimination than the corresponding multiclass logit margin obtained from LoRA-adapted foundation models. Based on this observation, we introduce a targeted embedding-space refinement that is applied only when WM and IC are the two highest-probability classes. The refinement redistributes probability within the WM/IC pair while preserving their combined probability mass and leaving all other class probabilities unchanged.

WM/IC Embedding Classifier. For ensemble member m and cross-validation fold f, we train a binary WM/IC classifier using embeddings produced by member m from the same member-specific training dataset and sampling configuration. The classifier is trained on all ground-truth WM and IC patches from the four non-validation folds. To construct this classifier, the concatenated embeddings from the encoders are first standardized and reduced to two principal components using principal component analysis (PCA). The resulting features are then used to train a binary ℓ_2-regularized logistic-regression classifier. The classifier output is subsequently calibrated using a scalar binary temperature $T^{\mathrm{b}}_{m,f}$, estimated by minimizing the binary negative log-likelihood on the validation-fold-f WM/IC samples for which the model identifies WM and IC as the two highest-probability classes. This produces $\pi_{m,f}(x) \in [0, 1]$, which represents the calibrated estimate of $P(\mathrm{IC} \mid x)$ for patch x.

WM/IC probability reallocation. The calibrated binary estimate is then incorporated into the multiclass prediction by adjusting only the relative probabilities assigned to WM and IC. Specifically, the refinement is applied only when WM and IC are the two highest-probability classes. This probability mass is redistributed between WM and IC according to the calibrated binary estimate $\pi_{m,f}(x)$. In this way, the refinement changes only the relative probabilities of WM and IC, while preserving their combined probability mass and leaving all other class probabilities unchanged.

2.9 Fusion-head ensembling

Beyond ensembling individual foundation-model encoders within a single fusion head, we introduce a second level of ensembling across fusion heads. Rather than relying on a single learned fusion strategy, we combine the predictions of two complementary heads that reuse the same three pathology-encoder embeddings but differ in their training distributions, optimization objectives, and use of the CONCH concept representation.

Concretely, our final predictor is an equal-weight ensemble of two members, M1 and M2. M1 applies a warm-started fusion head to the concatenated embeddings of the three LoRA-adapted encoders (Sec. 2.4). Its fusion head is trained

on the class-balanced training split, with its low-rank adaptations and mixing weights selected to maximize macro-F1. M2 uses the same warm-started architecture but trains its fusion head on the prevalence-preserving dataset and incorporates the frozen CONCH representation as a fourth input block (Sec. 2.5). Each member is temperature-calibrated separately and passed through the WM/IC routing gate before their final predictions are averaged.

3 Results

3.1 Experiment setup

All experiments were implemented in PyTorch 2.11.0 (CUDA 12.8) using the HuggingFace PEFT library (v0.15.2) for LoRA adaptation, and conducted on NVIDIA RTX A6000 GPUs (48 GB). The LoRA-adapted foundation-model members were trained for 5 or 10 epochs (Supplementary Sec. 7) using AdamW with an initial learning rate of 1×10^{-5}, weight decay of 0.01, and a batch size of 64 (LoRA rank 4, $\alpha = 8$, applied to the attention projections). The learning rate was controlled using a cosine-annealing schedule. Model checkpoints were selected using a configuration-specific held-out metric (Supplementary Sec. 7.3). The source code for the submitted inference container is available at https://github.com/lingyigt/bratspath-2026-inference.

3.2 Evaluation protocol

Evaluation followed a patient-disjoint five-fold cross-validation protocol. For each fold-specific model, hyperparameters and calibration temperatures were determined using the corresponding out-of-fold validation predictions. Ensemble coefficients were subsequently selected using the pooled out-of-fold predictions from all five folds. Official validation performance was evaluated through the BraTS-Path challenge evaluation pipeline [7].

3.3 Cross-validation results

Table 1 reports the local five-fold cross-validation results for the two ensemble members and their equal-weight combination, based on pooled out-of-fold predictions from patient-disjoint folds, together with the performance of the final M1+M2 ensemble on the official validation set.

The two members exhibit complementary strengths: M2 achieves higher accuracy (0.8240), MCC (0.7845), and AUROC (0.9365), whereas M1 achieves higher macro-recall (0.6417) and macro-F1 (0.6391). Their equal-weight combination provides the best overall balance, achieving the highest macro-F1 (0.6429) and specificity (0.9785) among the three local configurations while maintaining near-best accuracy, MCC, and AUROC.

On the official validation set, all six metrics exceed the corresponding local cross-validation values. In particular, accuracy increases from 0.8218 to 0.9264,

MCC from 0.7823 to 0.8976, and macro-F1 from 0.6429 to 0.7364. These results demonstrate strong performance of the final ensemble on the official validation set. The higher absolute scores may also reflect differences between the local and official evaluation distributions, including class prevalence, patient composition, and overall case difficulty.

Table 1. Local five-fold cross-validation and official validation results. Bold indicates the best local cross-validation result per metric.

Model	Acc.	MCC	AUROC†	macro-F1	Recall	Spec.†
M1	0.8074	0.7670	0.9282	0.6391	**0.6417**	0.9775
M2	**0.8240**	**0.7845**	**0.9365**	0.6330	0.6109	0.9781
M1+M2	0.8218	0.7823	0.9337	**0.6429**	0.6322	**0.9785**
M1+M2 (official)	0.9264	0.8976	0.9540	0.7364	0.7481	0.9899

† AUROC and specificity are macro-averaged for local cross-validation.

Table 2 reports the per-class F1 scores of the two ensemble members and their equal-weight combination. The ensemble members exhibit complementary performance profiles. Member M2 performs better on several high-frequency classes, achieving the highest F1 scores for CT (0.849), IC (0.807), PN (0.792), and NOTA (0.993), whereas M1 performs better on less prevalent classes, including WM (0.576) and LI (0.145). By leveraging their complementary predictions, the equal-weight combination achieves the most balanced class-wise performance, ranking first or tied for first on four classes while remaining competitive with the better individual member on most others. This class-wise complementarity explains the combination's highest overall macro-F1 in Table 1. In contrast, LI and PL remain challenging for all three configurations, with F1 scores no greater than 0.15. This persistent weakness is consistent with the small number of contributing patients for these classes in the training data.

Table 2. Per-class F1 on the local five-fold cross-validation results.

Model	CT	DM	IC	LI	MP	NC	PL	PN	WM	NOTA
M1	0.822	0.696	0.804	**0.145**	0.513	0.925	0.139	0.783	**0.576**	0.988
M2	**0.849**	0.675	**0.807**	0.127	0.469	0.924	**0.143**	**0.792**	0.550	**0.993**
M1+M2	0.846	**0.697**	0.806	0.139	**0.519**	**0.926**	**0.143**	0.789	0.573	0.991

3.4 Ablation Studies

We performed three controlled ablations under the same five-fold cross-validation protocol, holding all other components fixed and, where retraining was required,

applying the same training recipe. Table 3 reports optimized probability averaging of the same three encoders in place of learned fusion, M2 without CONCH, and M1+M2 without the WM/IC router. Additional ablations, router diagnostics, dataset statistics, CONCH prompts, and reproducibility details are provided in the supplementary material.

Table 3. Controlled ablations on local five-fold cross-validation.

Configuration	Acc.	MCC	AUROC†	macro-F1
Prob. average	0.819 (−0.003)	0.779 (−0.003)	0.926 (−0.008)	0.638 (−0.005)
M2–CONCH‡	0.821 (−0.003)	0.780 (−0.004)	0.928 (−0.008)	0.634 (+0.001)
M1+M2–router	0.818 (−0.004)	0.777 (−0.005)	0.932 (−0.002)	0.639 (−0.004)

Parentheses show changes from the matched model in Table 1. ‡M2–CONCH is referenced to M2; the other rows to M1+M2. †AUROC is macro-averaged.

4 Discussion

We developed a two-level ensemble approach that combines diversity across foundation-model encoders, training distributions, and fusion heads. On the BraTS-Path 2026 official validation set, the submitted model achieved an accuracy of 0.9264, an MCC of 0.8976, and a macro-F1 of 0.7364, demonstrating strong performance on both prior-sensitive and class-balanced metrics.

The local cross-validation results suggest that training-distribution and fusion-head diversity can provide useful complementarity beyond simply combining different encoder architectures. The CONCH representation additionally provides pathology-relevant semantic cues through image–text similarity scores. Fold-specific calibration and WM/IC routing offer complementary refinements at the decoder level. Calibration accounts for confidence differences among independently trained fold models, whereas the router re-examines the embedding only when WM and IC are the two highest-probability classes and redistributes their combined probability mass without affecting the remaining classes.

The principal limitation is the limited patient representation of the rare classes, particularly LI and PL. Resampling and augmentation can increase the number and diversity of training views derived from existing patches but cannot introduce missing inter-patient variation, which likely contributes to the low F1 scores for these classes. Resizing patches from 512×512 to 224×224 may also remove fine-scale features needed to distinguish morphologically similar subregions. Future work will prioritize expanding patient representation for rare classes and evaluating higher-resolution or multiscale representations. Overall, the results show that complementary training priors, vision–language representations, and specialized fusion heads can be combined efficiently to improve imbalanced glioblastoma subregion classification.

References

1. Bakas, S., Thakur, S.P., Faghani, S., Moassefi, M., Baid, U., Chung, V., Pati, S., Innani, S., Baheti, B., Albrecht, J., et al.: BraTS-Path challenge: Assessing heterogeneous histopathologic brain tumor sub-regions. arXiv preprint arXiv:2405.10871 (2024). https://doi.org/10.48550/arXiv.2405.10871
2. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F.K., Jaume, G., Chen, B., Zhang, A., Shao, D., Song, A.H., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**, 850–862 (2024). https://doi.org/10.1038/s41591-024-02857-3
3. Chen, R.J., Lu, M.Y., Williamson, D.F.K., Mahmood, F.: Towards a general-purpose foundation model for computational pathology. Nature Reviews Bioengineering (2024). https://doi.org/10.1038/s44222-024-00207-5
4. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 1321–1330 (2017)
5. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (2022)
6. Kaplan, D., Grandhi, R.S., Lane, C., Warner, B., Abraham, T.M., Scotti, P.S.: Training state-of-the-art pathology foundation models with limited data. In: Medical Image Computing and Computer Assisted Intervention (MICCAI) (2025), https://github.com/kaiko-ai/midnight
7. Karargyris, A., Umeton, R., Sheller, M.J., Aristizabal, A., George, J., Wuest, A., Pati, S., Kassem, H., Zenk, M., Baid, U., et al.: Federated benchmarking of medical artificial intelligence with medperf. Nature machine intelligence **5**(7), 799–810 (2023)
8. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2999–3007 (2017). https://doi.org/10.1109/ICCV.2017.324
9. Lu, M.Y., Chen, B., Williamson, D.F.K., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., Parwani, A.V., Zhang, A., Mahmood, F.: A visual-language foundation model for computational pathology. Nature Medicine **30**(3), 863–874 (2024). https://doi.org/10.1038/s41591-024-02856-4
10. McGenity, C., et al.: Artificial intelligence in digital pathology: A systematic review and meta-analysis of diagnostic test accuracy. npj Digital Medicine **7**(1) (2024). https://doi.org/10.1038/s41746-024-01106-8
11. Morales, S., Engan, K., Naranjo, V.: Artificial intelligence in computational pathology: Challenges and future directions. Digital Signal Processing **119**, 103196 (2021). https://doi.org/10.1016/j.dsp.2021.103196
12. Price, M., et al.: CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the united states in 2017–2021. Neuro-Oncology **26**(Supplement 6), vi1–vi85 (2024). https://doi.org/10.1093/neuonc/noae145
13. Scalbert, M., Saillard, C., Peeters, T., Gonzalez, L., Valter, D., Llinares-López, F., Mariet, Z.E., Jenatton, R.: H-optimus-1: A foundation model for computational histopathology. In: Proceedings of the American Association for Cancer Research Annual Meeting 2026. vol. 86, p. LB174 (2026). https://doi.org/10.1158/1538-7445.AM2026-LB174
14. Schaff, L.R., Mellinghoff, I.K.: Glioblastoma and other primary brain malignancies in adults: A review. JAMA **329**(7), 574–587 (2023). https://doi.org/10.1001/jama.2023.0023

15. Sung, J.Y., Hwang, K.: Glioblastoma: Epidemiology, molecular pathogenesis, diagnosis, management, and therapeutic resistance. Molecular Biomedicine **7**(1), 63 (2026). https://doi.org/10.1186/s43556-026-00467-8
16. Xiong, C., Chen, H., Sung, J.J.Y.: A survey of pathology foundation model: Progress and future directions. In: Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI). pp. 10751–10760 (2025). https://doi.org/10.24963/ijcai.2025/1193
17. Zimmermann, E., Vorontsov, E., Viret, J., Casson, A., Zelechowski, M., Shaikovski, G., Tenenholtz, N., Hall, J., Klimstra, D., Yousfi, R., Fuchs, T., Fusi, N., Liu, S., Severson, K.: Virchow2: Scaling self-supervised mixed magnification models in pathology. arXiv preprint arXiv:2408.00738 (2024). https://doi.org/10.48550/arXiv.2408.00738