BioFact-MoE: Biologically Factorized Mixture of Experts for Vision–Language Prognostic Modeling in Hepatocellular Carcinoma

Junlin Yang[1], Tian Yu[2], Nicha C. Dvornek[1,2], Yuexi Du[2], Peiyu Duan[2], Annabella Shewarega[1], Lawrence H. Staib[1,2,3], James S. Duncan[1,2,3,4], and Julius Chapiro[1]

[1] Department of Radiology & Biomedical Imaging,
[2] Department of Biomedical Engineering,
[3] Department of Electrical Engineering,
[4] Department of Statistics & Data Science
Yale University, New Haven, CT, 06510, USA
junlin.yang@yale.edu

Abstract. Hepatocellular carcinoma (HCC) is biologically heterogeneous, shaped by the interplay between hepatic functional reserve and tumor-related oncologic factors; thus, similar survival outcomes may reflect fundamentally different underlying biological processes. Prognostic modeling in HCC is informed by rich multimodal information from multiparametric MRI and radiology reports from routine clinical practice. Existing prognostic vision–language models (VLMs) learn a single entangled latent representation that blends hepatic and tumor-related factors, limiting both accuracy and biological interpretability. We present BioFact-MoE, a biologically factorized Mixture of Experts (MoE) framework that explicitly decomposes liver and tumor factors via biologically supervised experts within a residual MoE survival architecture. On a HCC cohort of N=588 patients (pretrained on 4,582 3D MRI image-report pairs), BioFact-MoE consistently improves survival prediction over all baselines across time horizons, achieving 12-, 18-, and 24-month AUCs of 75.33%, 75.85%, and 73.96%. Beyond scalar risk prediction, gated expert weights enable phenotype-aware risk stratification. Pathway-informed gating uncovers clinically meaningful treatment-associated survival heterogeneity. In held-out validation, hepatic and tumor embeddings show selective associations with liver function and tumor burden markers, respectively ($p<0.05$), without supervision. The code is available at https://github.com/jy-639/BioFact-MoE.

Keywords: Biological Factorization · MoE · Vision-Language Models · Parameter-Efficient Fine-Tuning · Prognostic Modeling

1 Introduction

Hepatocellular carcinoma (HCC) is a leading cause of cancer-related mortality [4], with prognosis governed by two interacting yet biologically distinct axes:

hepatic functional reserve and tumor-specific oncologic behavior [11]. Patients with similar survival can have different underlying drivers. In some patients, poor outcomes are driven by severe hepatic dysfunction, whereas in others they are attributable to tumor biology. Disentangling these pathways is critical for risk attribution and personalized treatment planning, yet most existing data-driven prognostic models compress both axes into a single entangled representation. Recent vision-language models (VLMs) improve radiology representation learning through image–report alignment, and finer-grained semantic grounding at the sentence, entity, and concept level has been explored [3, 18, 10, 15]. However, these approaches organize representations by semantic granularity rather than clinically grounded biological pathways, implicitly mixing hepatic and tumor signals, limiting prognostic modeling performance and biological interpretability in heterogeneous cohorts such as HCC. In clinical practice, radiology reports routinely describe hepatic functional reserve and tumor-related characteristics. Yet standard representation learning does not explicitly disentangle these factors. We therefore introduce pathway-specific structure as a domain-informed inductive bias to encourage biologically aligned representations. Parameter-efficient adaptation via LoRA [5] provides a mechanism for lightweight specialization, but a single shared adapter applies the same transformation regardless of biological heterogeneity. Recent LoRA–MoE approaches, including MoLoRA [17], LoRAMoE [2], and MixLoRA [9], learn expert routing and specialization implicitly from downstream task objectives, without incorporating prior structural differentiation among experts. Such designs are effective for multi-task adaptation but are less suited to settings where meaningful domain structure exists. In HCC prognostic modeling, biological axes are present in multimodal pretraining data yet are not explicitly encoded in survival labels.

We present **BioFact-MoE** to address this gap by enforcing expert specialization through biologically defined pathway inputs during unsupervised contrastive pretraining, establishing domain-informed inductive bias before survival supervision is introduced. **Contributions: (1) Biologically factorized pretraining.** Large language model (LLM)-guided report decomposition paired with anatomical patch masking trains pathway-specific LoRA adapters that independently encode hepatic and tumor representations. **(2) Improved prognostic performance.** BioFact-MoE achieves 12-, 18-, and 24-month AUCs of 75.33%, 75.85%, and 73.96%, outperforming image-only, multimodal, vision–language pretraining, and MoE baselines. **(3) Residual MoE survival architecture.** A gated residual MoE head dynamically routes patient-specific risk contributions across hepatic and tumor pathways, enabling phenotype-aware risk stratification. Exploratory analysis shows the liver-dominant low-risk subgroup benefits from transarterial chemoembolization (TACE), consistent with clinical expectations. **(4) Biological validation.** Hepatic embeddings align with Platelet-Albumin-Bilirubin (PALBI), while tumor embeddings align with bilobar disease and immunoscore (all $p<0.05$), without supervision, confirming biological specificity.

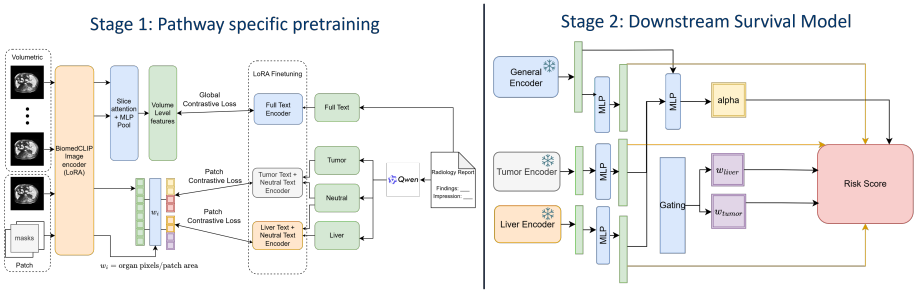

Fig. 1. BioFact-MoE framework. Both MRI volumes and radiology reports are used as input at pretraining and downstream stage. **Stage 1:** LLM-guided report decomposition supervises three pathway-specific LoRA adapters via contrastive pretraining with anatomical patch masking. **Stage 2:** Frozen pathway encoders are integrated by a residual MoE survival head with adaptive gating for Cox-based survival prediction.

2 Method

2.1 Stage 1: Pathway-Specific Biological Factorized Pretraining

BioFact-MoE framework operates in two stages (Fig. 1). In the first stage, biologically factorized pathway-specific adapters are pretrained using a contrastive vision–language objective on MRI–report pairs. We build on BiomedCLIP [19] (ViT-B/16 image encoder, PubMedBERT text encoder), with all backbone weights frozen. For volumetric MRI, we adopt a 2.5D strategy: each axial slice is encoded independently, a lightweight 1-layer self-attention transformer with learnable positional embeddings aggregate inter-slice context, and a two-layer MLP attention module produces a weighted sum of slice embeddings as the volume-level representation.

LLM-Guided Report Decomposition Each radiology report t is parsed offline by Qwen3-8B [16] into three segments: $t \rightarrow (s_{\text{liver}}, s_{\text{tumor}}, s_{\text{neutral}})$, where s_{liver} captures hepatic parenchymal quality, portal hypertension signs, and splenic findings; s_{tumor} captures tumor morphology, vascularity, and aggressiveness; and s_{neutral} retains remaining context. During pretraining, the neutral segment is concatenated to both s_{liver} and s_{tumor}, forming pathway-specific textual inputs that preserve shared clinical context while maintaining biological specialization.

Anatomical Patch Masking Organ masks from TotalSegmentator [1] weight patch tokens by organ occupancy $w_i = (\text{organ pixels in patch}) \,/\, (\text{patch area})$, so the pathway embedding $v = \sum_i w_i t_i / \sum_i w_i$ is restricted to pathway-relevant anatomy. The liver pathway uses liver and spleen masks; the tumor pathway uses liver, portal vein, and inferior vena cava (IVC).

Pathway-Specific LoRA Three independent LoRA parameter sets (A_p, B_p), $p \in \{\text{general, liver, tumor}\}$ (rank r=4, α=8) are trained separately

via symmetric InfoNCE loss [12] $\mathcal{L} = \frac{1}{2}(\mathcal{L}_{\text{img}\to\text{txt}} + \mathcal{L}_{\text{txt}\to\text{img}})$ at τ=0.1, with AdamW at a learning rate of 1×10^{-4} for 250 epochs with a batch size 32.

2.2 Stage 2: Residual MoE Survival Modeling

All pathway encoder checkpoints (including LoRA parameters) are frozen. The general encoder produces $z_{\text{base}} \in R^{1024}$. Pathway checkpoints yield $z_{\text{liver}}, z_{\text{tumor}} \in R^{1024}$. Two expert MLPs map pathway embeddings to 256-dim representations $e_{\text{liver}}, e_{\text{tumor}}$. A core component of the MoE framework is gating function G for dynamically routing patient-specific contributions across experts. In our case, $G : R^{3072} \to R^2$ is a learned linear projection.

The gate input $g_i = [z_{\text{base}};\ \delta z_i;\ |\delta z_i|]$ concatenates the global patient context z_{base}, the signed inter-pathway discrepancy $\delta z_i = z_{\text{liver}} - z_{\text{tumor}}$, and its absolute magnitude $|\delta z_i|$, allowing the gate to jointly reason about the patient's overall profile and the degree of disagreement between hepatic and tumor representations. The resulting soft weights $w_i = [w^i_{liver}, w^i_{tumor}] = \text{softmax}(G(g_i))$ (summing to 1) reflect the relative dominance of each pathway for patient i: a large w^i_{liver} indicates a liver-dominant phenotype, while a large w^i_{tumor} indicates a tumor-driven one. These weights are later used for phenotype-aware patient stratification, enabling clinically interpretable subgroup discovery beyond a scalar risk score.

A per-patient coefficient $\alpha_i = \sigma(\text{MLP}_\alpha([z_{\text{base}}; e_{\text{liver}}; e_{\text{tumor}}]))$ modulates expert contribution, with the final risk score:

$$r_i = \underbrace{h_{\text{base}}(z_{\text{base}})}_{\text{base risk}} + \alpha_i\big(w_{i1}\,\delta r_i^{\text{liver}} + w_{i2}\,\delta r_i^{\text{tumor}}\big)\,. \tag{1}$$

The MoE survival head is optimized using the Cox partial likelihood with AdamW (learning rate 1×10^{-4}) for 50 epochs with a batch size of 16.

3 Experiments

3.1 Datasets and Data Preprocessing

Both cohorts were curated from a single institution under an IRB-approved retrospective protocol. **Pretraining:** 4,582 abdominal MRI–report pairs from 1,896 patients, comprising both HCC-confirmed patients and patients screened for HCC with negative findings, used for vision–language pretraining without survival supervision. **Downstream:** 588 HCC patients with baseline multiparametric MRI-report pairs and longitudinal follow-up for overall survival, randomly split into 400/188 for training/test. Subsets have structured biomarkers: PALBI (n=118), ALBI (n=275), Immunoscore (n=62), and bilobar disease (n=386), none used for supervision.

MRIs were resampled to isotropic resolution, normalized, and cropped to 224×224×60. TotalSegmentator [1, 7] organ masks were only used during pretraining. Reports were de-identified, retaining Findings and Impression sections.

Table 1. Time-dependent AUC (mean±std) at 12, 18, 24 months. **Bold** = best.

Category	Method	12m	18m	24m
Image-only	ResNet-Cox	62.74 ± 1.39	64.28 ± 0.97	63.26 ± 0.96
Multimodal survival	Image+Report Fusion	64.69 ± 0.99	65.68 ± 1.06	64.19 ± 1.13
Vision–language	BiomedCLIP [19]	65.71 ± 5.87	67.97 ± 0.64	66.91 ± 2.31
	GLoRIA [6]	73.20 ± 0.76	72.13 ± 0.88	70.00 ± 1.22
	Entity Alignment	74.27 ± 1.86	72.05 ± 1.52	70.26 ± 2.39
MoE	SparseMoE	72.56 ± 2.37	73.31 ± 1.71	71.58 ± 0.87
Ours	**BioFact-MoE**	**75.33 ± 2.60**	**75.85 ± 1.45**	**73.96 ± 3.04**

3.2 Baselines and Metrics

We compare four categories of baselines: image-only survival modeling, multimodal fusion, state-of-the-art biomedical VLMs, and MoE architectures. Image-only and multimodal fusion baselines are trained solely on the downstream survival cohort, while VLMs and MoE models first leverage large-scale image–report pretraining. Performance is evaluated using time-dependent AUC at 12, 18, and 24 months based on three independent runs with different random seeds. We report the mean and standard deviation across the three runs at each time horizon.

Image-only survival A pretrained 2D ResNet processes axial MRI slices independently. Slice-level features are averaged to obtain a patient-level representation, which is optimized under the Cox proportional hazards objective [8].

Multimodal survival Image and Report Fusion follows standard late-fusion multimodal modeling [14]. A pretrained 2D image encoder processes individual axial MRI slices, which are averaged to obtain a patient-level embedding. A pretrained text encoder extracts a global report embedding. Both embeddings are concatenated and fed into a Cox regression head for survival prediction [8].

Vision-language fine-tuning. BiomedCLIP [19] is fine-tuned end-to-end for survival by appending a Cox regression head to the joint image–report embedding. GLoRIA [6] extends global contrastive learning with local alignment, contrasting image sub-regions with individual report words; a Cox head is applied to patch- and global-level features. Finally, we implement an entity-aligned variant that exploits organ segmentation masks to explicitly supervise alignment between report-extracted anatomical entities and their corresponding segmented regions, providing a more robust grounding signal.

MoE SparseMoE implements a standard sparsely-gated MoE layer [13]. Given the frozen BiomedCLIP fused embedding, a learned gating network produces routing logits over E=4 experts. We select only the two highest-scoring experts and combine them using the normalized gate weights as output.

3.3 Survival Performance

BioFact-MoE outperforms all baselines at 12, 18, and 24 months (Table 1). Based on predicted risk scalar, we can train a decision tree model on the training

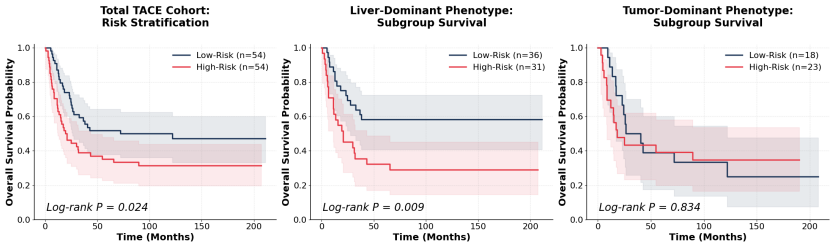

Fig. 2. Phenotype-Aware Stratification: Beyond scalar risk, gate weights stratify patients by dominant biological axis. Among patients receiving the same TACE treatment, liver-driven and tumor-driven subgroups show significantly different survival trajectories, capturing heterogeneity invisible to clinical staging alone.

Table 2. Biological validation of pathway specialization by pathway-specific linear probe of available clinical biomarkers.

Biomarker	n	Classes	Liver AUC	Tumor AUC	Δ	p	Direction
PALBI	118	3	65.4	56.4	9.00	<0.05	Liver > Tumor
Immunoscore	62	4	54.4	59.3	4.9	<0.05	Tumor > Liver
Bilobar disease	386	2	59.7	63.7	4.1	<0.05	Tumor > Liver

set to achieve high-risk and low-risk stratification on the test set. In addition, gate weights provide patient-level biological interpretation. Patients are classified as liver-driven ($w_{liver} \geq 0.5$) or tumor-driven based on the dominant expert pathway. Despite comparable clinical stage, TACE-treated patients in the test set demonstrate marked heterogeneity in survival outcomes. As shown in Fig. 2, while all patients receiving TACE benefit with log-rank p=0.024, the liver-dominant phenotype subgroup show improved survival under TACE with log-rank p=0.009, while the tumor-dominant phenotype subgroup show no difference in terms of survival with log-rank p=0.834.

Exploratory treatment analysis is shown in Fig. 3. Hepatic low-risk patients show stronger survival association with TACE, consistent with clinical expectations. In tumor-dominant patients, no significant treatment interaction was detected (all $p > 0.3$); note subgroup sizes in the tumor-dominant TACE and non-TACE arms were limited (all $n \leq 12$ per arm).

3.4 Biological Validation of Pathway Specialization

Table 2 provides evidence that the learned pathway embeddings capture biologically meaningful signals aligned with clinical biomarkers. Linear probes (5-fold cross-validation) indicate PALBI, a marker of hepatic functional reserve, is more accurately predicted from frozen liver-pathway embeddings than tumor-pathway embeddings ($\Delta = 9.0$, $p < 0.05$), confirming that the liver expert has specialized toward hepatic biology. Conversely, both tumor-related variables, immunoscore and bilobar disease extent, are better captured by the tumor-pathway embeddings ($\Delta = 4.9$ and 4.1 respectively, $p < 0.05$), validating that the tumor expert

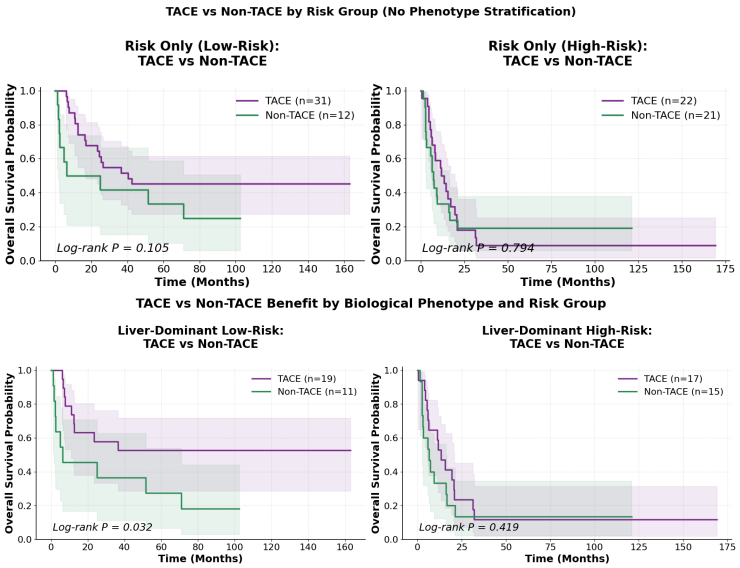

Fig. 3. In an exploratory treatment analysis, patients identified as liver low-risk by the hepatic pathway showed the greatest predicted benefit from TACE, consistent with clinical expectations.

has acquired complementary tumor-specific representations. This cross-validated biological alignment provides evidence that BioFact-MoE achieves genuine factorization rather than redundant parallel encoding.

3.5 Ablation Studies

Table 3 evaluates the role of biological grounding from both text and images across training and inference. **Random Text Split** and **Swapped Conditioning** disrupt biologically meaningful report assignment, while **No Text Segmentation** removes textual pathway supervision entirely. **No Anatomical Masking** eliminates organ-aware visual grounding. **Full Report at Inference** keeps pathway-specific text segmentation during downstream training but substitutes the full unsegmented report at inference. **Full Report** uses the unsegmented report for both downstream training and inference. The inductive bias introduced by biological grounding from both text and images is vital for expert specialization in BioFact-MoE.

Table 4 isolates the contribution of each pathway component. **Base Only** removes both pathway-specific experts and retains only the base survival head, establishing the performance floor of general image-report fusion without biological factorization. **Liver Only** and **Tumor Only** retain a single expert pathway and remove the other, quantifying the individual prognostic value of hepatic and tumor-specific representations. **Joint Learning (No MoE)** trains

Table 3. Ablation study of biological grounding across training and inference.

Stage	Model Variant	12-mo AUC	18-mo AUC	24-mo AUC
Baseline Pretraining	No Text Segmentation	71.52 ± 1.91	70.63 ± 2.43	68.98 ± 3.01
	Random Text Split	68.16 ± 2.02	69.31 ± 1.64	67.84 ± 0.87
	Swapped Text Conditioning	69.28 ± 2.85	70.32 ± 0.87	69.60 ± 1.79
	No Anatomical Masking	71.67 ± 2.23	72.42 ± 0.60	69.97 ± 0.71
Downstream	Full Report at Inference	72.27 ± 1.47	70.19 ± 2.02	68.99 ± 1.97
	Full Report	74.93 ± 1.38	73.35 ± 1.11	71.42 ± 1.89
Ours	**BioFact-MoE**	**75.33 ± 2.60**	**75.85 ± 1.45**	**73.96 ± 3.04**

Table 4. Ablation study of pathway modeling contributions.

Model Variant	12-mo AUC	18-mo AUC	24-mo AUC
Base Only	65.71 ± 5.87	67.97 ± 0.64	66.91 ± 2.31
Liver Only	68.24 ± 1.00	68.85 ± 1.79	67.84 ± 1.83
Tumor Only	68.26 ± 2.03	70.28 ± 0.21	69.48 ± 3.53
Liver & Base	68.61 ± 0.81	69.40 ± 1.62	68.51 ± 2.08
Tumor & Base	71.27 ± 0.12	72.70 ± 0.09	70.75 ± 2.64
Joint Learning (No MoE)	71.31 ± 1.64	71.65 ± 0.86	71.27 ± 2.02
Liver & Tumor (No Base)	72.71 ± 0.78	75.22 ± 1.69	73.13 ± 2.22
BioFact-MoE	**75.33 ± 2.60**	**75.85 ± 1.45**	**73.96 ± 3.04**

liver, tumor, and base pathways jointly with equal contribution and no mixture routing, testing whether adaptive expert weighting via MoE is necessary beyond simple multi-pathway fusion.

We observe that Tumor & Base slightly outperforms Liver & Base, possibly because the shared pathway already captures substantial liver-related global signals, while tumor-specific features benefit more from dedicated specialization. Notably, Liver & Tumor (No Base) achieves performance comparable to the full BioFact-MoE at 18- and 24-month horizons, with only modest differences at 12 months. This suggests that biologically specialized pathways capture most of the prognostic signal, while the base pathway mainly provides additional robustness rather than driving performance.

Overall, these results indicate that biological factorization is the primary contributor, and the residual MoE design serves to enhance stability rather than act as the main source of performance gain.

4 Conclusion

We presented BioFact-MoE, a biologically factorized MoE framework for vision-language prognostic modeling in hepatocellular carcinoma. By explicitly decomposing radiology reports into liver- and tumor-specific pathways and aligning pathway-specific LoRA adapters with anatomically masked visual representations, BioFact-MoE learns biologically structured multimodal embeddings without requiring biomarker supervision. A residual MoE survival head dynamically routes patient-specific risk contributions across both pathways, improving sur-

vival prediction while enabling phenotype-aware stratification. Biological validation confirms that pathway embeddings align with clinically recognized hepatic and tumor markers, supporting our factorization as an emergent property of the pretraining design.

Several limitations limited further investigation. First, the LLM-guided report decomposition relies on the capability of the LLM; errors in pathway assignment may introduce noise into the biological factorization signal. Second, structured clinical biomarkers used for biological validation are sparsely available, limiting the statistical power of pathway specialization analysis.

Acknowledgments. This work was supported by the R01CA206180 grant from 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. Akinci D'Antonoli, T., Berger, L.K., Indrakanti, A.K., Vishwanathan, N., Weiss, J., Jung, M., Berkarda, Z., Rau, A., Reisert, M., Küstner, T., et al.: Totalsegmentator mri: robust sequence-independent segmentation of multiple anatomic structures in mri. Radiology **314**(2), e241613 (2025)
2. Dou, S., Zhou, E., Liu, Y., Gao, S., Shen, W., Xiong, L., Zhou, Y., Wang, X., Xi, Z., Fan, X., et al.: Loramoe: Alleviating world knowledge forgetting in large language models via moe-style plugin. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1932–1945 (2024)
3. Du, Y., Onofrey, J.A., Dvornek, N.C.: Multi-view and multi-scale alignment for contrastive language-image pre-training in mammography. In: International Conference on Information Processing in Medical Imaging. pp. 247–262. Springer (2025)
4. Foglia, B., Turato, C., Cannito, S.: Hepatocellular carcinoma: latest research in pathogenesis, detection and treatment (2023)
5. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. Iclr **1**(2), 3 (2022)
6. Huang, S.C., Shen, L., Lungren, M.P., Yeung, S.: Gloria: A multimodal global-local representation learning framework for label-efficient medical image recognition. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3942–3951 (2021)
7. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
8. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC medical research methodology **18**(1), 24 (2018)
9. Li, D., Ma, Y., Wang, N., Ye, Z., Cheng, Z., Tang, Y., Zhang, Y., Duan, L., Zuo, J., Yang, C., et al.: Mixlora: Enhancing large language models fine-tuning with lora-based mixture of experts. arXiv preprint arXiv:2404.15159 (2024)

10. Li, Y., Lai, H., Zhou, X., Ming, S., Ma, W., Wei, W., Zhou, S.K.: More performant and scalable: Rethinking contrastive vision-language pre-training of radiology in the llm era. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 348–357. Springer (2025)
11. Liu, P.H., Hsu, C.Y., Hsia, C.Y., Lee, Y.H., Su, C.W., Huang, Y.H., Lee, F.Y., Lin, H.C., Huo, T.I.: Prognosis of hepatocellular carcinoma: assessment of eleven staging systems. Journal of hepatology **64**(3), 601–608 (2016)
12. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
13. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538 (2017)
14. Willis, R., Bakos, J.: Exploring fusion strategies for multimodal vision-language systems. arXiv preprint arXiv:2511.21889 (2025)
15. Wu, Y., Liu, Y., Yang, Y., Yao, M.S., Yang, W., Shi, X., Yang, L., Li, D., Liu, Y., Yin, S., et al.: A concept-based interpretable model for the diagnosis of choroid neoplasias using multimodal data. Nature communications **16**(1), 3504 (2025)
16. Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al.: Qwen3 technical report. arXiv preprint arXiv:2505.09388 (2025)
17. Zadouri, T., Üstün, A., Ahmadian, A., Ermiş, B., Locatelli, A., Hooker, S.: Pushing mixture of experts to the limit: Extremely parameter efficient moe for instruction tuning. arXiv preprint arXiv:2309.05444 (2023)
18. Zhang, M., Wu, X., Luo, H., Wang, F., Lv, Y.: Medground: Bridging the evidence gap in medical vision-language models with verified grounding data. arXiv preprint arXiv:2601.06847 (2026)
19. Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., Rao, R., Wei, M., Valluri, N., Wong, C., Tupini, A., Wang, Y., Mazzola, M., Shukla, S., Liden, L., Gao, J., Crabtree, A., Piening, B., Bifulco, C., Lungren, M.P., Naumann, T., Wang, S., Poon, H.: A multimodal biomedical foundation model trained from fifteen million image–text pairs. NEJM AI **2**(1) (2024). https://doi.org/10.1056/AIoa2400640, https://ai.nejm.org/doi/full/10.1056/AIoa2400640