When RNA is Missing: A Multimodal Fusion Benchmark for BCG Progression Prediction in HR-NMIBC

Saeedeh Hashemi[1], Tahlita CM Zuiverloon[2], Nadieh Khalili[1]

[1] Department of Pathology, Research Institute for Medical Innovation, Radboud University Medical Center, Nijmegen, The Netherlands
[2] Department of Urology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Nadieh.Khalili@radboudumc.nl

Abstract. High-risk non-muscle-invasive bladder cancer (HR-NMIBC) is associated with frequent progression after Bacillus Calmette-Guérin (BCG) immunotherapy, making accurate risk prediction important for treatment planning and surveillance. Current clinical diagnosis relies on pathological phenotype for staging while molecular genotype plays an important role in BCG response. Multimodal learning could therefore improve prediction by integrating histopathology, RNA-seq, and clinical data, but it remains unclear which fusion strategy is most effective in limited-size clinical cohorts. In addition, RNA-seq is often unavailable in clinical practice; designing architectures that remain robust to missing-modality conditions is crucial. In this study, we systematically compared multimodal fusion strategies for time-to-event prediction in HR-NMIBC using 272 patients with complete histopathology whole-slide images, bulk RNA-seq, and clinical data from the CHIMERA challenge. We evaluated unimodal, bimodal, and trimodal Cox survival models using concatenation, gated fusion, low-rank bilinear fusion, and cross-attention. Multimodal fusion improved performance in most configurations compared with unimodal models, with the best performance achieved by trimodal gated fusion, with a C-index of 0.725 ± 0.083. We further found that modality representation affected performance and that RNA dropout improved robustness when RNA-seq was unavailable at inference time. These findings highlight the importance of fusion design for clinically realistic multimodal survival prediction.
Keywords: Multimodal fusion, survival prediction, RNA missing modality.

1 Introduction

Cancer progression remains a persistent challenge in oncology. In high-risk non-muscle invasive bladder cancer (HR-NMIBC), standard treatment is repeat transurethral resection (ReTUR) followed by BCG immunotherapy, yet treatment failure occurs in approximately 40-50% of patients [1]. Accurate progression prediction is therefore clinically important, guiding treatment intensification and surveillance planning.

Modern oncology data captures disease biology at multiple levels: histopathology whole-slide images (WSI) reveal the tumor phenotype, bulk RNA-seq characterizes cancer genotype, and clinical records provide complementary information such as demographics, comorbidities, family history, prior treatments and other prognostic variables. In HR-NMIBC, subtype assessment in routine clinical practice is primarily based

on histopathological evaluation of WSI, while RNA-seq has also shown clear clinical value in this setting. De Jong et al. [2] defined three BCG response subtypes using whole-transcriptome sequencing of HR-NMIBC tumors, including one subtype showing an aggressive, immunosuppressive profile and consistently worse recurrence and progression outcomes across two independent cohorts. This work demonstrates that molecular profiling can meaningfully stratify BCG response beyond standard clinicopathological risk groups, motivating the integration of RNA-seq with pathology and clinical data for a more complete predictive model.

Historically, most AI models in medicine have been developed for unimodal inputs. More recently, multimodal AI approaches that integrate heterogeneous data sources have gained increasing attention, reflecting the inherently multimodal nature of clinical decision-making across medical disciplines. A central challenge is the choice of fusion strategy, how information from different modalities is combined, with approaches ranging from simple concatenation to gated mechanisms [3], bilinear interactions [4], and attention-based co-attention [5]. With the development of attention-based models [6], multimodal learning has increasingly moved beyond self-attention within a single modality toward cross-attention between modalities. Cross-attention enables information from one modality to guide the representation of another, allowing models to capture interactions between pathology, genomic, and clinical data at a more fine-grained level [5,7].

Despite growing interest, many multimodal studies propose a single fusion architecture without systematically comparing alternative fusion strategies, leaving it unclear which approach best integrates pathology, RNA-seq, and clinical data under realistic conditions such as limited cohort size and missing modalities. In this work, we compare multimodal fusion strategies for time-to-progression prediction after BCG treatment in HR-NMIBC, using a subset of the CHIMERA dataset. We evaluate four fusion strategies: concatenation, gated fusion, low-rank bilinear fusion, and SurvPGC-style co-attention [5] across unimodal, bimodal, and trimodal settings with pathology, RNA-seq, and clinical data. We further assess different clinical and RNA representations and test robustness to missing RNA using modality dropout during training.

2 Dataset

We used a subset of the CHIMERA Task 3 data, including HR-NMIBC patients from Erasmus with all three modalities (WSIs, RNA-seq, clinical data). We defined our own 5-fold cross-validation splits, stratified by event status, rather than using the official CHIMERA split, to enable fair comparison across unimodal, bimodal, and trimodal models on the same 272 patients.

Histopathology. Each WSI is accompanied by a binary tissue mask outlining the tissue section, used to restrict downstream processing to tissue-containing regions.

Transcriptomics. Bulk RNA-seq data is extracted from selected tumor regions and normalized using DESeq2 (19,359 genes).

Clinical data. To avoid redundancy with information already captured by the imaging and transcriptomic modalities, we exclude clinical variables derived from WSI or RNA-seq. This leaves five clinical variables: age (years), sex (M/F), smoking status (Yes/No), number of BCG instillations (count), and ReTUR status (yes/no).

Labels. The prediction target is time-to-event progression to advanced disease, defined by two variables: progression (binary, 1 = true, 0 = false) and time_to_prog_or_FUend (float, months to progression or end of follow-up).

3 Methods

3.1 Method Overview

Our framework consists of three stages: modality-specific encoding, fusion, and survival prediction. Each modality (histopathology, RNA-seq, and clinical data) is first encoded independently into a fixed-size representation. These unimodal representations are then combined in a fusion block, which we systematically vary across four fusion strategies to compare their effect on performance.

To assess how each modality contributes to survival prediction, both individually and in combination, we evaluate models in a progressive setting: unimodal models using each modality alone, bimodal models for all pairwise combinations, and trimodal models combining all three modalities under each of the four fusion strategies.

In clinical practice, histopathology and clinical data are almost always available, while RNA-seq is frequently missing. To reflect this, we additionally train models with RNA-seq dropout during training, allowing us to evaluate the robustness of each fusion strategy to missing modalities at inference time. Figure 1 illustrates this overall pipeline, from modality-specific encoding through fusion to survival prediction.

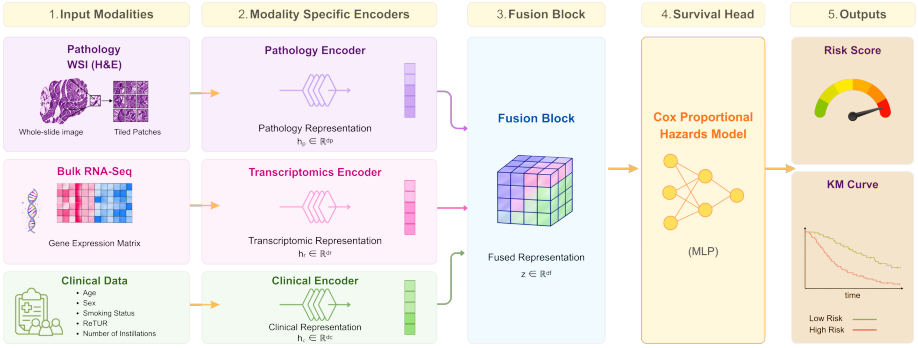

Fig. 1. Overview of the proposed multimodal survival prediction pipeline.

3.2 Modality-Specific Encoders

For each modality, we generate two types of representations depending on the fusion strategy: a single pooled embedding for embedding-based fusion, and a set of tokens for attention-based fusion.

Histopathology. WSIs are encoded using the UNI foundation model [8], producing frozen tile-level embeddings (1024-dim). For embedding-based fusion, tile-level embeddings are aggregated into a slide-level embedding using attention-based multiple instance learning (AB-MIL), which learns a single attention weight per tile, computes a weighted average of the tile embeddings, and projects the resulting vector to a 256-

dim pathology embedding. The AB-MIL attention and projection layers are trained end-to-end for each experimental configuration. For attention-based fusion, the tile-level embeddings themselves (1024-dim) are used directly as tokens.

Transcriptomics. For embedding-based fusion, the top 2,000 genes were selected using the coefficient of variation computed on the training split of each fold and passed through a trainable MLP to obtain a fixed-size RNA embedding. For attention-based fusion, we construct tokens using three methods:

Hallmark pathway filtering: Genes were grouped using Hallmark pathway definitions. Pathways with at least 90% gene coverage were retained, and the expression values of genes within each retained pathway were encoded into an RNA token using a trainable MLP.

Biological category filtering: Genes were grouped into six predefined biological categories: transcription factors, cytokines and growth factors, cell differentiation markers, protein kinases, oncogenes, and tumor suppressors. For each category, the expression values of genes assigned to that category were encoded into one RNA token using a trainable MLP.

Foundation model embeddings: Genes were first reduced from 19,359 to 11,980 using Hallmark and Reactome pathway definitions [9,10] with at least 90% gene coverage and encoded using the pretrained scFoundation model [11], yielding four 768-dimensional RNA tokens. The scFoundation embeddings were kept frozen, while the downstream projection layer was trained end-to-end for each experimental configuration.

Clinical data. Clinical data is represented in both embedding-based and token-based forms. For embedding-based fusion, we use (i) raw clinical variables passed through an MLP, and (ii) sentence-based representations of each variable encoded with CONCH [12] and BioClinical ModernBERT [13]; the resulting text embeddings are projected to a shared space, concatenated, and mapped to a 128-dim clinical embedding. For token-based fusion, each of the five clinical variables is converted into a sentence and encoded with CONCH, yielding five tokens per patient used directly in attention-based fusion, without pooling.

3.3 Fusion Strategies

We compare four fusion strategies for combining histopathology, RNA-seq, and clinical representations.

Concatenation fusion combines modality-specific embeddings into a single vector, which is passed to an MLP Cox head for risk prediction.

Scalar-gated fusion computes an independent, patient-specific gate for each modality using a two-layer MLP with sigmoid activation. Each modality embedding is multiplied by its gate before concatenation, allowing the Cox head to adaptively emphasize or down-weight each input. Gates were not normalized across modalities; each modality was weighted independently.

Low-rank bilinear fusion models pairwise cross-modal interactions in a compact rank space [4]. Each modality pair is projected into a shared 64-dimensional latent space (rank = 64), where interactions are computed by element-wise multiplication to form an interaction embedding. For trimodal models, pathology-RNA, pathology-clinical,

and RNA-clinical interactions are concatenated and passed to the Cox head; for bimodal models, only the available pair is used.

SurvPGC-style fusion performs token-level co-attention following the general design of SurvPGC [5]. Pathology tile embeddings, clinical text embeddings, and RNA-seq tokens are projected into a shared latent space, where separate co-attention modules model RNA-pathology and clinical-pathology interactions. The resulting attended representations, including RNA tokens, pathology tokens conditioned on RNA, clinical tokens, and pathology tokens conditioned on clinical information, are mean-pooled, concatenated, and passed to an MLP Cox head.

3.4 Cox Prediction Head

The fused patient representation was passed to an MLP-based Cox prediction head similar to DeepSurv [14], consisting of layer normalization, AlphaDropout, and SELU activations. A final linear layer produced a single log-risk score r_i, where higher values indicate higher predicted risk of progression. Models were trained using the negative Cox proportional hazards partial log-likelihood [15]:

$$\mathcal{L}_{\text{Cox}} = -\frac{1}{N_{\text{events}}} \sum_{i:\delta_i=1} \left(r_i - \log \sum_{j \in R_i} e^{r_j} \right) \tag{1}$$

where r_i is the predicted log-risk score for patient i, R_i is the set of patients still at risk when patient i experiences progression, and δ_i indicates whether the progression event was observed. Censored patients remained in the risk sets until their last follow-up but did not contribute an event term because their progression time was unknown. Models were optimized using AdamW (learning rate 2×10^{-4}, weight decay 1×10^{-5}), with dropout regularization and gradient clipping (maximum norm = 5.0).

3.5 Model Configurations (Unimodal, Bimodal, Trimodal)

We evaluate unimodal, bimodal, and trimodal model configurations to assess the contribution of each modality and the benefit of multimodal integration.

Unimodal models. Unimodal models were trained separately per modality: histopathology used UNI tile embeddings aggregated via AB-MIL, RNA-seq used coefficient-of-variation-filtered genes encoded by an MLP, and clinical data used either tabular MLP encoding or templated text embeddings (CONCH, BioClinical ModernBERT). Each representation was passed to a Cox head for patient-level risk prediction.

Bimodal models. Bimodal models used all pairwise combinations (histopathology-RNA, histopathology-clinical, RNA-clinical), with the same unimodal encoders fused via concatenation, gated fusion, or low-rank bilinear fusion, then passed to the Cox head. This tests whether modality pairs improve prediction and which pairs are complementary.

Trimodal models. Trimodal models integrated all three modalities: embedding-based fusion via concatenation, gated fusion, or low-rank bilinear fusion, and token-based fusion via SurvPGC-style co-attention on pathology, RNA-seq, and clinical tokens.

4 Results

All preprocessing was fold-independent to avoid leakage, with RNA imputation and feature selection statistics computed on the training fold only. Models were trained with AdamW (300 epochs, patience 40, batch size 64, lr 2×10^{-4}, weight decay 1×10^{-5}, gradient clipping 5.0). For missing-modality robustness, RNA inputs were randomly masked during training while pathology and clinical inputs remained available. Performance was evaluated using the concordance index (mean ± std across folds), with Kaplan-Meier curves for risk stratification.

4.1 Main Comparison of Unimodal, Bimodal, and Trimodal Models

This section considers raw tabular clinical data; clinical representation ablation is discussed in Section 4.4. As shown in Table 1, clinical data achieved the highest unimodal C-index, followed by RNA and pathology. Fold-level results for all configurations are provided in Supplementary Table S1.

In the bimodal setting, combining modalities generally improved performance over unimodal models, suggesting complementary predictive information across data types. The strongest bimodal result combined pathology and clinical data via low-rank bilinear fusion (0.698 ± 0.030). Pathology-RNA fusion showed the most modest improvement among modality pairs.

Across trimodal models, integrating all three modalities gave the best overall performance, with gated fusion achieving the highest C-index among all evaluated models (0.725 ± 0.083). Overall, multimodal fusion improved progression prediction in most configurations, with gated fusion providing the strongest performance across embedding-based fusion strategies.

Table 1. C-index (mean ± std) across unimodal, bimodal, and trimodal configurations using concatenation, gated, and low-rank bilinear fusion, and across raw tabular, CONCH, and BioClinical ModernBERT clinical encoders. Pathology is encoded with UNI + AB-MIL and RNA-seq with variance-filtered genes. P = pathology, R = RNA, C = clinical.

Unimodal	P			-		0.585±0.041	
	R			-		0.609±0.076	
	C			Raw tabular		0.647±0.052	
	C			CONCH		0.664±0.064	
	C			ModernBERT		0.704±0.051	
Configuration	**Modality**			**Clinical encoder**	**Concatenation**	**Gated**	**Low-rank bilinear**
Bimodal		R	P	-	0.631±0.082	0.628±0.085	0.580±0.078
	C	R		Raw tabular	0.672±0.074	0.676±0.071	0.564±0.042
	C		P	Raw tabular	0.655±0.043	0.719±0.036	0.698±0.030
	C		P	CONCH	0.624±0.021	0.623±0.067	0.593±0.041
	C	R		CONCH	0.647±0.089	0.623±0.068	0.581±0.081
	C		P	ModernBERT	0.608±0.057	0.599±0.047	0.550±0.040

	C	R		ModernBERT	0.638±0.075	0.640±0.082	0.639±0.085
Trimodal	C	R	P	Raw tabular	0.709±0.069	0.725±0.083	0.686±0.083
	C	R	P	CONCH	0.641±0.042	0.671±0.086	0.594±0.068
	C	R	P	ModernBERT	0.633±0.079	0.625±0.061	0.596±0.087

4.2 Token-based attention fusion results

Table 2 shows the performance of SurvPGC-style co-attention fusion with three RNA tokenization strategies. Biological category tokens performed best among token-based models (0.634 ± 0.050), followed by scFoundation embeddings and Hallmark pathway tokens.

Table 2. Token-based trimodal results using SurvPGC-style co-attention with UNI pathology tokens, CONCH clinical tokens, and three RNA tokenization strategies. Values are C-index mean ± std.

RNA Tokenization	Hallmark pathway	Biological category	scFoundation
C-index	0.611±0.036	0.634±0.050	0.627±0.032

4.3 Missing RNA robustness

Since RNA-seq is not available in clinical practice, we evaluate robustness to missing RNA using dropout during training (probabilities 0.30 and 0.50). During training, RNA-seq was randomly masked for a subset of samples, while histopathology and clinical inputs remained available. The masking was implemented according to the fusion strategy: RNA embeddings were set to zero for concatenation and gated fusion, RNA-related interaction terms were removed for low-rank bilinear fusion, and RNA tokens were masked for SurvPGC-style co-attention. This exposed the models to both complete and RNA-missing inputs, encouraging them to rely less on RNA-seq and enabling inference using only histopathology and clinical data.

Table 3 reports C-index under complete-RNA and missing-RNA inference. Fold-level results are provided in Supplementary Table S2. Across fusion strategies, degradation after masking RNA was generally small, with several configurations showing no decrease or slightly higher performance under missing-RNA inference, such as low-rank bilinear fusion at 0.30 dropout (drop = -0.018). Among embedding-based models, low-rank bilinear fusion with 0.30 dropout achieved the highest missing-RNA performance (0.720 ± 0.036). Notably, SurvPGC-style models trained with RNA dropout achieved higher C-index values than their non-dropout counterparts in Table 2.

Table 3. RNA missing-modality robustness: C-index (mean ± std) under complete and missing RNA at inference, by training dropout rate.

Fusion Method	Dropout	Complete RNA	Missing RNA	Δ C-index
Concatenation	0.30	0.708±0.061	0.708±0.031	0.000
	0.50	0.728±0.030	0.706±0.052	0.023
Gated	0.30	0.710±0.083	0.695±0.058	0.014
	0.50	0.672±0.066	0.679±0.053	-0.007

Low-rank bilinear	0.30	0.701±0.097	0.720±0.036	-0.018
	0.50	0.714±0.059	0.705±0.056	0.009
Surv_pathway	0.30	0.667±0.036	0.659±0.032	0.008
	0.50	0.668±0.050	0.652±0.036	0.016
Surv_biocategory	0.30	0.682±0.067	0.671±0.056	0.011
	0.50	0.692±0.049	0.663±0.034	0.029
Surv_scf	0.30	0.680±0.023	0.694±0.047	-0.014
	0.50	0.672±0.057	0.705±0.035	-0.033

4.4 Effect of clinical representation on fusion performance

Table 1 compares raw tabular data, CONCH, and BioClinical ModernBERT embeddings as clinical encoders. In the clinical-only setting, BioClinical ModernBERT achieved the highest unimodal C-index, followed by CONCH and raw tabular features. In multimodal fusion, however, raw tabular features consistently performed better, achieving the highest trimodal C-index across almost all fusion strategies compared to CONCH and BioClinical ModernBERT. Within each encoder, multimodal integration consistently improved performance for raw tabular features, whereas CONCH showed only limited benefit, with only one trimodal configuration outperforming the corresponding clinical-only model. In contrast, BioClinical ModernBERT consistently performed worse in both bimodal and trimodal settings than the corresponding clinical-only model.

4.5 Risk stratification analysis

Kaplan-Meier analysis was performed using pooled out-of-fold risk scores. Patients were stratified into high- and low-risk groups using the median predicted risk score. Figure 2 compares the Kaplan-Meier curves of the unimodal models with the best-performing trimodal gated fusion model. The trimodal gated fusion model showed the clearest separation between the high- and low-risk groups, followed by the clinical-only model. The pathology-only model showed weaker but still statistically significant separation, while the RNA-only model showed the weakest statistical separation despite a comparatively higher C-index.

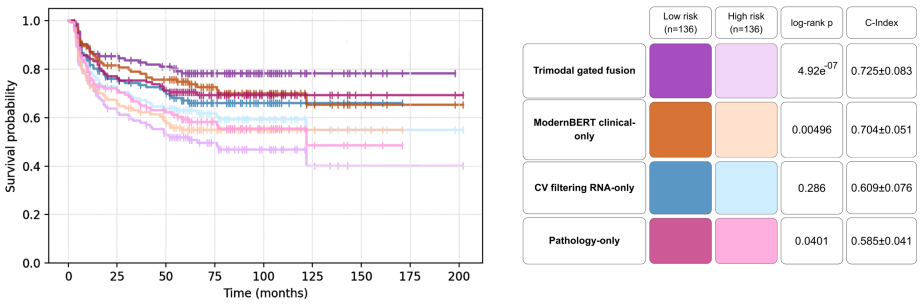

Fig. 2. Kaplan-Meier survival curves for the best-performing trimodal gated fusion model and the best-performing unimodal model per modality, stratified by the median pooled out-of-fold predicted risk score. The table reports risk group sizes, log-rank p-values, and C-index. Numbers at risk are provided in Supplementary Table S3.

5 Discussion and Conclusion

Multimodal fusion improved recurrence prediction in most configurations, indicating that pathology, RNA-seq, and clinical data carry complementary information. Gated fusion's advantage likely comes from letting the model weight each modality's contribution rather than treating all inputs equally, as concatenation does. Low-rank bilinear fusion's inconsistency across modality pairs suggests interaction-based fusion is more sensitive to noise, representation quality, and sample size than simpler embedding-based approaches, an important consideration in a cohort of this size.

Token-based co-attention underperforming embedding-based fusion likely reflects that learning stable cross-modal token interactions requires more data than was available. RNA dropout did improve these models' performance over non-dropout counterparts, which may suggest a regularizing effect beyond its missing-modality role.

Performance degradation after masking RNA at inference was generally small, and in some configurations, missing-RNA C-index slightly exceeded complete-RNA C-index. We do not interpret this as evidence that removing RNA improves prediction: given the small cohort, fold-level variability, and absence of a paired significance test, these differences may reflect noise or altered regularization dynamics from RNA masking during training. Rather, the results show that RNA-dropout training preserves performance when RNA is missing, not that missing RNA is advantageous.

The clinical representation results suggest text-derived embeddings suit clinical-only prediction but may introduce noise or redundancy when combined with high-dimensional pathology and RNA features in a limited cohort, favoring simpler tabular representations for fusion.

This study has several limitations: cohort size may affect estimate stability and disadvantage attention-based models; formal statistical testing across configurations was not performed, so small C-index differences should be read cautiously; the clinical feature set was limited; no external cohort was used for validation; and UNI, while effective, is not the most recent pathology foundation model.

In conclusion, fusion strategy, clinical representation, and missing-modality training substantially affect multimodal survival prediction in HR-NMIBC. Gated embedding-based fusion performed best overall, and RNA-dropout training improved robustness to missing RNA, supporting its use in clinically realistic settings. Future work should explore larger cohorts, external validation, newer pathology encoders, and more data-efficient attention-based fusion.

Acknowledgments. This work was financed by the Sectorplan Medical Sciences of the Ministry of OCW in the Netherlands. We thank the organizers of the CHIMERA Challenge for providing the dataset used in this study.

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

Code Availability. https://github.com/DIAGNijmegen/multimodal-survival-fusion

References

1. Jaromin M, Konecki T, Kutwin P. Revolutionizing Treatment: Breakthrough Approaches for BCG-Unresponsive Non-Muscle-Invasive Bladder Cancer. Cancers. 2024 Jan;16(7):1366. doi:10.3390/cancers16071366
2. de Jong FC, Laajala TD, Hoedemaeker RF, Jordan KR, van der Made ACJ, Boevé ER, van der Schoot DKE, Nieuwkamer B, Janssen EAM, Mahmoudi T, Boormans JL, Theodorescu D, Costello JC, Zuiverloon TCM. Non-muscle-invasive bladder cancer molecular subtypes predict differential response to intravesical Bacillus Calmette-Guérin. Sci Transl Med. 2023 May 24;15(697):eabn4118. doi: 10.1126/scitranslmed.abn4118. Epub 2023 May 24. PMID: 37224225; PMCID: PMC10572776.
3. Arevalo, J., Solorio, T., Montes-y-Gómez, M. and González, F.A., 2017. Gated multimodal units for information fusion. arXiv preprint arXiv:1702.01992.
4. Keshvarikhojasteh, H., Pluim, J.P. and Veta, M., 2026. Attention-Based Multimodal Survival Prediction with Cross-Modal Bilinear Fusion. arXiv preprint arXiv:2605.13897.
5. Hou, J., Zhang, R., Xie, Y. et al. Multimodal deep learning for cancer prognosis prediction with clinical information prompts integration. npj Digit. Med. 9, 76 (2026). https://doi.org/10.1038/s41746-025-02257-y
6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
7. G. Jaume, A. Vaidya, R. J. Chen, D. F. K. Williamson, P. P. Liang and F. Mahmood, "Modeling Dense Multimodal Interactions Between Biological Pathways and Histology for Survival Prediction," 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2024, pp. 11579-11590, doi: 10.1109/CVPR52733.2024.01100.
8. Chen, R.J., Ding, T., Lu, M.Y. et al. Towards a general-purpose foundation model for computational pathology. Nat Med 30, 850–862 (2024). https://doi.org/10.1038/s41591-024-02857-3
9. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417-425. doi: 10.1016/j.cels.2015.12.004. PMID: 26771021; PMCID: PMC4707969.
10. Marija Milacic, Deidre Beavers, Patrick Conley, Chuqiao Gong, Marc Gillespie, Johannes Griss, Robin Haw, Bijay Jassal, Lisa Matthews, Bruce May, Robert Petryszak, Eliot Ragueneau, Karen Rothfels, Cristoffer Sevilla, Veronica Shamovsky, Ralf Stephan, Krishna Tiwari, Thawfeek Varusai, Joel Weiser, Adam Wright, Guanming Wu, Lincoln Stein, Henning Hermjakob, Peter D'Eustachio, The Reactome Pathway Knowledgebase 2024, Nucleic Acids Research, Volume 52, Issue D1, 5 January 2024, Pages D672–D678, https://doi.org/10.1093/nar/gkad1025
11. Hao, M., Gong, J., Zeng, X. et al. Large-scale foundation model on single-cell transcriptomics. Nat Methods 21, 1481–1491 (2024). https://doi.org/10.1038/s41592-024-02305-7
12. Lu MY, Chen B, Williamson DFK, Chen RJ, Liang I, Ding T, Jaume G, Odintsov I, Le LP, Gerber G, Parwani AV, Zhang A, Mahmood F. A visual-language foundation model for computational pathology. Nat Med. 2024 Mar;30(3):863-874. doi: 10.1038/s41591-024-02856-4. Epub 2024 Mar 19. PMID: 38504017; PMCID: PMC11384335.
13. Sounack, T., Davis, J., Durieux, B., Chaffin, A., Pollard, T.J., Lehman, E., Johnson, A.E.W., McDermott, M., Naumann, T., Lindvall, C.: BioClinical ModernBERT: A State-of-the-Art Long-Context Encoder for Biomedical and Clinical NLP. arXiv preprint arXiv:2506.10896 (2025). https://arxiv.org/abs/2506.10896
14. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018 Feb 26;18(1):24. doi: 10.1186/s12874-018-0482-1. PMID: 29482517; PMCID: PMC5828433.
15. Cox, D. R. "Regression Models and Life-Tables." Journal of the Royal Statistical Society. Series B (Methodological) 34, no. 2 (1972): 187–220. http://www.jstor.org/stable/2985181.