STACKED MULTIMODAL FUSION OF GENETIC AND HISTOPATHOLOGY FEATURES FOR SURVIVAL PREDICTION IN BREAST CANCER

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate survival prediction in breast cancer requires capturing complementary information at both the molecular and tissue morphology levels, which single-modality models fail to represent effectively. We propose a multimodal survival prediction framework that integrates genetic expression features and whole-slide histopathology representations. Genetic embeddings are learned using a multi-head attention network designed to produce stable low-dimensional representations from high-dimensional transcriptomic profiles. Tissue morphology features are extracted from segmented whole-slide images using a fine-tuned ConvNeXt-Tiny backbone. Each modality is used to train an independent Cox proportional hazards model, producing unimodal risk estimates. These modality-specific risk scores are then combined through a stacked fusion strategy, where a meta-model learns the final survival prediction. Evaluated on the TCGA-BRCA dataset, the proposed approach achieves a concordance index of 0.666, outperforming both unimodal baselines and standard multimodal fusion strategies. To demonstrate generalizability, the framework was externally validated across three additional pan-cancer cohorts (LUAD, BLCA, KIRC). The stacked fusion consistently outperformed unimodal baselines across all cohorts with a mean cross-cancer Concordance Index (C-index) of 0.684. These results demonstrate that leveraging both molecular and histological signals jointly improves the robustness and discriminative power of patient-level survival prediction.

1 INTRODUCTION

Cancer prognosis prediction remains challenging due to the complex interplay between tissue-level morphology and molecular-level alterations. Histopathology whole-slide images (WSIs) capture tumour heterogeneity, stromal response, and mitotic activity, offering rich phenotypic cues, while transcriptomic profiles encode genetic programs driving progression. Relying on a single modality alone can limit the survival accuracy as morphologically similar patients can exhibit distinct genetic abnormalities. Additionally, high-dimensional gene expression data suffer from noise, collinearity and small sample sizes.

Deep learning models have advanced in both domains. CNN-based WSI analysis extracts discriminative visual biomarkers, and transcriptomic modelling using Attention-based Models (Chen et al., 2021; Jaume et al., 2024) to capture complex, dense biological cross-talk between genomic pathways for robust survival prediction. However, jointly leveraging both sources remains difficult. Early fusion approaches that concatenate WSI embeddings with gene-expression features (Mobadersany et al., 2018) often show moderate concordance because the genomic modality is extremely high-dimensional relative to the number of patient samples. This imbalance, combined with heterogeneous feature scales, makes simple concatenation prone to overfitting. As these methods rely on linear compatibility in the joint feature space, they frequently degrade calibration and generalization.

To address these challenges, we introduce a stacked Cox survival fusion framework that integrates ConvNeXt-Tiny WSI embeddings with Multi-Head Attention (MHA) based genomic representations through late, risk-level fusion. Recent work favors late, risk-level integration (Lee et al., 2019;

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Cheerla & Gevaert, 2019) for greater stability and interpretability. Each modality is independently mapped to a one-dimensional hazard score via a Cox model, and a meta-Cox stage learns optimal cross-modal weighting. This design offers:

1. produces more reliable risk predictions, since each modality contributes only a single, stable risk value,
2. avoids overfitting, because the fusion operates on two compact risk scores rather than thousands of raw features,
3. improves performance with limited data, yielding a higher C-index and clearer risk-group separation compared to direct feature fusion.

2 Related Work

Survival prediction traditionally relies on the Cox proportional hazards model, valued for interpretability but limited by linear assumptions on high-dimensional data. Neural network models such as DeepSurv (Katzman et al., 2018) and Dynamic-DeepHit (Lee et al., 2019) increase modelling flexibility but remain unimodal and tabular.

In computational pathology, imaging modalities like MRI (Yadav et al., 2025) and ultrasound (Yadav et al., 2024) have been explored, yet histopathology images such as WSIs can provide fine context of the tissue morphology. Early WSI-based survival models, such as WSISA (Zhu et al., 2017), used patch-level CNNs and hence, lacked global context. Multiple Instance Learning (MIL) models, including attention-MIL (Ilse et al., 2018), CLAM (Lu et al., 2021), adversarial MIL (Liu et al., 2024), and transformer-based aggregators (Jiang et al., 2023), are reported to improve slide-level representation learning. ConvNeXt architectures (Liu et al., 2022), particularly ConvNeXt-Tiny, further strengthen patch-level encoding via Transformer-inspired design while retaining CNN locality, enabling sample-efficient WSI modelling.

For genomics, high-dimensional transcriptomic data motivate attention-based architectures such as the Transcriptome Transformer (Koo et al., 2025), and pathway-guided embeddings (Jaume et al., 2024) have shown strong performance in small-sample settings.

Multimodal survival prediction spans three fusion strategies: early feature concatenation, bilinear interaction modelling, and late fusion. Early fusion risks overfitting; bilinear methods such as Pathomic Fusion (Chen et al., 2020) capture cross-modal correlations but are computationally expensive.

The proposed stacked Cox framework builds on this direction by training separate Cox models on ConvNeXt-Tiny and MHA embeddings, then fusing the resulting risks via a secondary Cox layer. Unlike end-to-end multimodal architectures such as HIPT (Chen et al., 2022), the proposed method is lightweight, data-efficient, and achieves state-of-the-art performance on TCGA-BRCA with improved concordance and reduced variance.

3 Methodology

The proposed framework integrates two distinct modalities: histopathology and genomics into a unified survival modelling paradigm through stacked Cox fusion. The pipeline involves four key stages:

1. data preprocessing and patient alignment,
2. modality-specific feature encoding using deep learning models,
3. late fusion of risk scores through a meta-Cox model,
4. evaluation using concordance and survival stratification metrics.

The entire design ensures statistical interpretability, dimensional stability, and generalization across heterogeneous biomedical data.

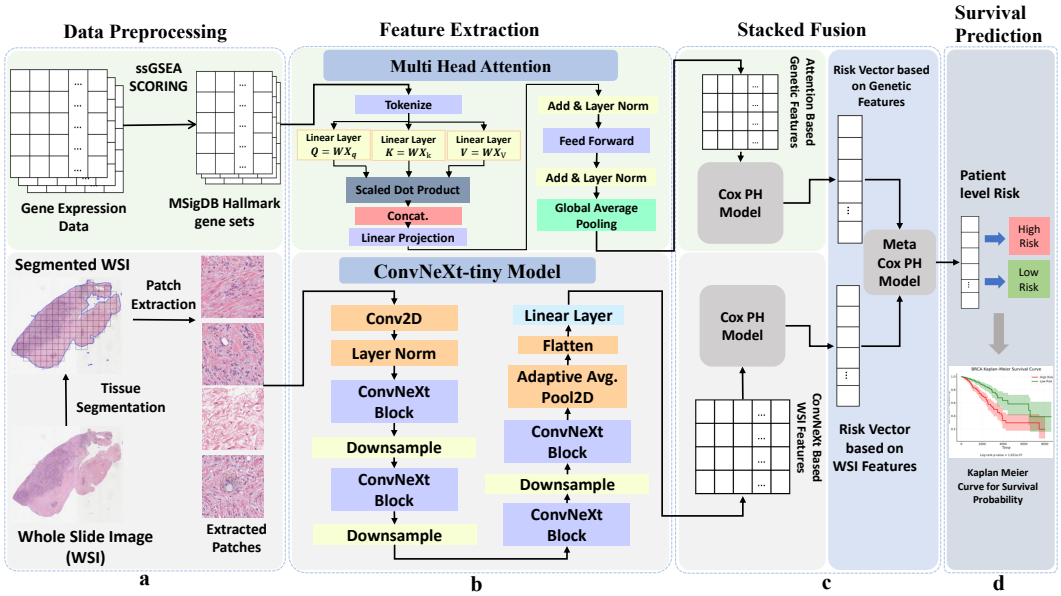

Figure 1: Schematic overview of the proposed multimodal survival pipeline. (a) Preprocessing of both modalities, including ssGSEA Scoring based dimensionality reduction for genomic data and tissue segmentation with patch extraction for WSIs. (b) Feature extraction using MHA for genomic embeddings and ConvNeXt for histopathological features. (c) Stacked Cox fusion integrating modality-specific risk scores into a unified survival model. (d) Final risk-based stratification with corresponding Kaplan–Meier survival curves illustrating high- and low-risk patient groups.

3.1 Dataset and Preprocessing

We evaluate our method on the complete TCGA-BRCA cohort comprising ∼1096 patients. Each case includes WSI, gene expression profile, and corresponding clinical survival metadata obtained from TCGA annotations, with time-to-event and censoring information for deceased and censored patients.

A patient is represented by (x_w, x_g, t, δ) where x_w = WSI-derived feature representation, x_g = genomic expression vector, t = observed time, δ = censoring indicator.

WSIs and genomic files were filtered and matched for each patient to ensure valid identifiers, consistent dimensions and cross-modal alignment. The finalized dataset maintains the event-to-censor ratio of the full cohort.

For WSIs, tissue regions are segmented using Otsu thresholding:

$$T = \text{argmin}_\tau \ \sigma^2(\tau) \tag{1}$$

where $\sigma^2(\tau)$ is the intra-class variance for threshold τ.

For genetic data, expression values are log-transformed and z-score normalized, followed by dimensionality reduction using single-sample Gene Set Enrichment Analysis (ssGSEA) using the MSigDB Hallmark gene sets before being fed into the MHA for embedding generation. ssGSEA ensures that the biologically relevant signals are present while the data is being compressed simultaneously. The preprocessing and input preparation steps for both modalities are illustrated in Fig. 1 (a).

3.2 Genomic Representation via MHA

To represent the gene expression modality, MHA is employed to learn complex genomic pathway features. As illustrated in Fig. 1 (b), the pathway-token attention mechanism models complex biological interactions and adaptively weights specific biological features. Each of the $N = 50$ pathway enrichment scores is projected into a d-dimensional embedding space, forming an input representation $X \in \mathbb{R}^{N \times d}$. This matrix is processed by a Multi-Head Self-Attention (MHSA) encoder to

adaptively model pathway interactions which allows the network to learn which biological pathways are co-expressed or co-dependent in the context of cancer survival.

For a single attention head, the input X is linearly projected into Query (Q), Key (K), and Value (V) matrices:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V \tag{2}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_k}$ are the learned projection weights, and d_k is the dimension of the keys. The attention mechanism computes a weighted sum of the values, where the weights represent the relevance of each pathway to the others. This is governed by the scaled dot-product attention equation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

The scaling factor $\frac{1}{\sqrt{d_k}}$ prevents the dot products from growing excessively large, ensuring stable gradients during back-propagation.

The MHSA layer concatenates the outputs from h parallel attention heads, allowing the model to jointly attend to information from different representation subspaces. Following standard transformer architecture, the output is passed through a feed-forward network equipped with residual connections and layer normalization.

Finally, the contextualized pathway tokens are aggregated via global average pooling to yield a single patient-level genomic vector $z \in \mathbb{R}^d$. This pooled representation is passed through a linear Cox risk head to optimize the partial likelihood loss, outputting the genomic-derived risk score r_{gen}:

$$r_{gen} = \beta_g^T z \tag{4}$$

where β_g represents the learned coefficient vector of the Cox regression layer.

3.3 ConvNeXt-Tiny-Based Histopathological Feature Extraction

For histopathology analysis, the study employs the ConvNeXt-Tiny architecture (Liu et al., 2022), a lightweight variant of the ConvNeXt family that retains the Transformer-inspired design principles of the original model. As illustrated in Fig. 1(c), ConvNeXt-Tiny preserves the hierarchical convolutional structure while incorporating large-kernel depthwise convolutions, inverted bottlenecks, and layer normalization to enhance receptive field coverage and gradient flow. Its reduced parameter count makes it more suitable for patch-level feature extraction in survival modelling, especially when patient datasets are limited.

Each WSI is divided into multiple non-overlapping patches that capture distinct histological regions. Every patch is passed through the ConvNeXt-Tiny feature extractor up to the global average pooling layer, producing a dense feature embedding per patch. The patient-level representation is obtained through aggregation across all patches. The aggregated features are standardized and reduced in dimension using PCA to mitigate redundancy and improve numerical stability for survival modelling. This step also controls overfitting given the limited number of patients relative to the feature space size.

3.4 Stacked Cox Survival Fusion

The proposed stacked Cox survival fusion integrates the two modalities at the risk-prediction level rather than the feature level. As illustrated in Fig. 1 (d), separate Cox Proportional Hazards models are initially trained for each modality:

$$h_{WSI}\left(t|x_w\right) = h_0(t)e^{\beta_w^T x_w}, \quad h_{GEN}\left(t|x_g\right) = h_0(t)e^{\beta_g^T x_g} \tag{5}$$

where $h_0(t)$ is the baseline hazard. x_w and x_g denote the ConvNeXt and MHA embeddings, and β_w and β_g are the learned coefficient vectors. Each model produces a modality-specific risk score r_{WSI} and r_{GEN}, summarizing the hazard contribution from its respective space.

A second meta-Cox model is then trained using these risk scores as predictors:

$$h_{stack}(t|r_{WSI}, r_{GEN}) = h_0(t)e^{\gamma_1 r_{WSI} + \gamma_2 r_{GEN}} \tag{6}$$

where γ_1 and γ_2 represent the relative weights of each modality's risk. This design transforms a high-dimensional multimodal fusion problem into a low-dimensional, interpretable risk-integration task. The model is trained using a penalized partial likelihood with an L_2 regularization term to ensure numerical stability and prevent overfitting.

3.5 Evaluation Metrics and Protocol

Model performance is assessed using the C-index to measure predictive discrimination, the Kaplan–Meier (KM) estimator to visualize survival stratification, and the log-rank test to quantify group separation significance. Evaluation employs stratified k-fold cross-validation to handle censored data distribution and bootstrap resampling for estimating 95% confidence intervals of the C-index. A penalizer is applied across all Cox models for consistent regularization strength.

3.6 Implementation Summary

All components are implemented in Python, using PyTorch for feature extraction and Lifelines for Cox survival modelling. Computations are executed on GPU-accelerated infrastructure to support efficient batch-level inference from WSIs. The framework outputs patient-level risk scores, modality contributions, and survival curves, providing interpretable results suitable for both prognostic evaluation and clinical translation.

4 Results and Discussion

This section discusses the experimental outcomes of the proposed framework. Hyper-parameters were tuned to optimise performance on heterogeneous cancer datasets. The model was validated across three additional pan-cancer cohorts: LUAD, BLCA, and KIRC.

4.1 Experimental Setup

The proposed model was trained and tested on the TCGA-BRCA dataset (1096 patients) and then validated on the TCGA-LUAD (453 patients), TCGA-BLCA (374 patients), and TCGA-KIRC (506 patients) datasets. Evaluation is performed using stratified 3-fold cross-validation to obtain the C-index.

4.2 Quantitative Results

The proposed stacked Cox fusion achieves robust survival prediction performance, with a mean C-index of 0.666 ± 0.040 and a statistically significant log-rank test p-value of 1.62×10^{-7}. This demonstrates strong prognostic separation between high- and low-risk groups, reflecting effective integration of histopathological and genomic risk information. These results are consistent with the comparisons reported in Table 1.

In contrast, ablation baselines yield substantially lower concordance, with the concatenation fusion model achieving 0.652 ± 0.036 and the Kronecker fusion only 0.400 ± 0.231. As further illustrated in Table 1, these results highlight the limitations of direct feature-level fusion, which tends to overfit and exhibit unstable generalization under small-sample survival settings.

4.3 Kaplan–Meier Survival Stratification

Kaplan–Meier curves for various datasets in Fig. 2 show separation between high and low-risk groups derived from the stacked Cox model, indicating effective survival ranking. The patients in each cohort were stratified into high-risk and low-risk groups based on the median predicted hazard score. The resulting curves show considerable divergence between the survival probabilities of the two groups.

The statistical significance of this separation between the low- and high-risk groups can be confirmed by the log-rank test p-values. The proposed framework achieved a strong p-value of 1.62×10^{-7} on the BRCA development set and an exceptional prognostic separation on the KIRC validation set with

Table 1: Comparison of unimodal and multimodal survival prediction models across four TCGA cohorts. The proposed stacked Cox fusion model achieves robust performance across diverse cancer types.

Method	Modality	BRCA	LUAD	BLCA	KIRC
MHA (Genomic) (Koo et al., 2025)	Unimodal	0.576 ± 0.078	0.639 ± 0.073	0.590 ± 0.049	0.649 ± 0.038
ConvNeXt (WSI) (Liu et al., 2022)	Unimodal	0.661 ± 0.036	0.628 ± 0.037	0.642 ± 0.045	0.723 ± 0.026
MaL (Ning et al., 2022)	Unimodal	0.641 ± 0.045	0.622 ± 0.041	0.635 ± 0.048	0.710 ± 0.035
Pathomic Fusion (Chen et al., 2020)	Multimodal	0.612 ± 0.048	0.605 ± 0.051	0.615 ± 0.055	0.685 ± 0.042
MCAT (Chen et al., 2021)	Multimodal	0.635 ± 0.042	0.628 ± 0.045	0.638 ± 0.049	0.710 ± 0.038
MSEN (Zhou et al., 2024)	Multimodal	0.648 ± 0.045	0.641 ± 0.052	0.645 ± 0.048	0.720 ± 0.035
SURVPATH (Jaume et al., 2024)	Multimodal	0.655 ± 0.038	0.647 ± 0.041	0.651 ± 0.046	0.732 ± 0.028
Stacked Cox (Proposed)	Multimodal	$\mathbf{0.666 \pm 0.040}$	$\mathbf{0.663 \pm 0.067}$	$\mathbf{0.660 \pm 0.077}$	$\mathbf{0.747 \pm 0.030}$

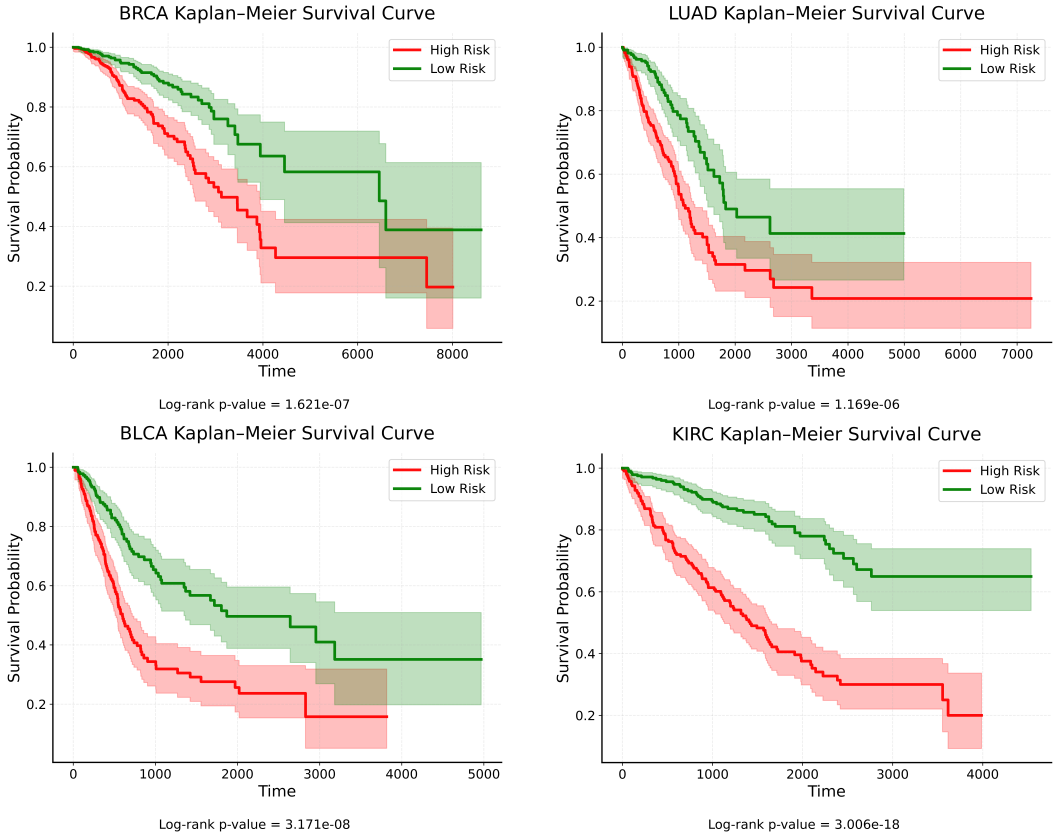

Figure 2: Kaplan–Meier survival curves comparing low-risk and high-risk groups across all four cancer cohorts. The shaded bands represent standard deviation, and the highly significant log-rank p-values demonstrate the effectiveness of the stacked fusion strategy in pan-cancer survival stratification.

a p-value of 3.01×10^{-18}. These metrics indicate the framework's ability to stratify heterogeneous patient populations into clinically meaningful prognostic categories.

4.4 Interpretation of Modality Contributions

In the stacked meta-Cox model, the learned coefficients (γ_1, γ_2) represent each modality's contribution to the overall survival risk as modality-specific weights as shown in Eq. 6. The ConvNeXt-based WSI risk generally carries a higher weight $(\gamma_1 > \gamma_2)$, indicating that morphological tissue features contribute slightly stronger prognostic cues, while the MHA-derived genomic embeddings provide complementary molecular context. Unlike feature-level fusion, this risk-space integration offers explicit, interpretable weighting and enhances clinical transparency.

Table 2: Ablation study of multimodal fusion strategies on the TCGA-BRCA dataset.

Fusion Strategy	C-index (Mean $\pm$ SD)
Kronecker Fusion	0.400 ± 0.231
Mean Fusion	0.648 ± 0.037
Concatenation Fusion	0.652 ± 0.036
Stacked Cox (Proposed)	$\mathbf{0.666 \pm 0.040}$

4.5 Ablation Study on Fusion Strategies

Ablation analysis confirms the superiority of the proposed risk-level fusion. To isolate the effect of the fusion strategy itself, all variants use identical genetic and WSI feature representations. While early fusion techniques such as concatenation and bilinear methods like the Kronecker product partially capture cross-modal relationships, they suffer from over-parametrization and weak generalization. As shown in Table 2, Kronecker, mean, and concatenation fusions achieve a lower C-index in contrast to the stacked Cox design which compresses multimodal data into scalar risk scores before integration, yielding higher concordance.

5 Conclusion and Future Work

Evaluated on the TCGA-BRCA dataset, our method delivers comparable survival prediction while avoiding highly complex, over-parametrized fusion mechanisms. By turning each data source into a single, reliable risk score, the fusion model avoids unnecessary complexity and focuses on the information that truly matters. This leads to more stable predictions and a clearer separation of patients into high- and low-risk groups compared with conventional feature-level fusion.

In contrast, simple concatenation or more complex bilinear combinations tend to create large and unstable representations, which struggle when the dataset is small. These models mix the information in such a way that it becomes difficult for a model to learn the survival relevant features. The proposed risk-level fusion strategy avoids these issues by offering a compact, robust way to combine imaging and genomic data, which boosts accuracy and reliability in survival prediction.

Future work may involve models based on non-proportional hazards survival modelling and missing modality scenarios. Additional modalities such as inter-cellular interactions can be incorporated in the future allowing deeper biological insights into the cancer micro-environment.

AI use statement

In this work, we did not use generative AI tools for any part of the research itself, such as developing the method, designing experiments, or analysing results. We have used generative AI only for a few minor writing and support tasks such as checking grammar and improving the readability of some sentences, finding instructions for downloading the dataset from the GDC Data Portal, and locating some related papers for the literature review. All text edits, download steps, and papers found this way were checked by the authors before being used. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

See the ICLR 2027 AI Policy for Authors for more details.

References

Anika Cheerla and Olivier Gevaert. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics*, 35(14):i446–i454, 2019.

Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, and Faisal Mahmood. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE transactions on medical imaging*, 41 (4):757–770, 2020.

Richard J Chen, Ming Y Lu, Wei-Hung Weng, Tiffany Y Chen, Drew FK Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction

in gigapixel whole slide images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4025, 2021.

Richard J Chen, Chengkuan Chen, Yicong Li, Tiffany Y Chen, Andrew D Trister, Rahul G Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16144–16155, 2022.

Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pp. 2127–2136. PMLR, 2018.

Guillaume Jaume, Anurag Vaidya, Richard Chen, Drew Williamson, Paul Liang, and Faisal Mahmood. Modeling dense multimodal interactions between biological pathways and histology for survival prediction, 2024. URL https://arxiv.org/abs/2304.06819.

Shuai Jiang, Arief A Suriawinata, and Saeed Hassanpour. Mhattnsurv: Multi-head attention for survival prediction using whole-slide pathology images. *Computers in biology and medicine*, 158:106883, 2023.

Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1):24, 2018.

Bonil Koo, Inyoung Sung, Sangseon Lee, and Sun Kim. Transcriptome transformer: improving patient survival prediction via multitask learning of transcriptomic and clinical features. *Briefings in Bioinformatics*, 26(6):bbaf628, 2025.

Changhee Lee, Jinsung Yoon, and Mihaela Van Der Schaar. Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*, 67(1):122–133, 2019.

Pei Liu, Luping Ji, Feng Ye, and Bo Fu. Advmil: Adversarial multiple instance learning for the survival analysis on whole-slide images. *Medical Image Analysis*, 91:103020, 2024.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

Pooya Mobadersany, Safoora Yousefi, Mohamed Amgad, David A Gutman, Jill S Barnholtz-Sloan, José E Velázquez Vega, Daniel J Brat, and Lee AD Cooper. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*, 115(13):E2970–E2979, 2018.

Zhenyuan Ning, Zhangxin Zhao, Qianjin Feng, Wufan Chen, Qing Xiao, and Yu Zhang. Mutual-assistance learning for standalone mono-modality survival analysis of human cancers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7577–7594, 2022.

Agnesh Chandra Yadav, Maheshkumar H Kolekar, and Mukesh Kumar Zope. Resnet-101 empowered deep learning for breast cancer ultrasound image classification. In *BIOSTEC (1)*, pp. 763–769, 2024.

Agnesh Chandra Yadav, Maheshkumar H Kolekar, and Mukesh Kumar Zope. Modified recurrent residual attention u-net model for mri-based brain tumor segmentation. *Biomedical Signal Processing and Control*, 102:107220, 2025.

Chenyi Zhou, Hualiang Wang, Xiaomeng Li, Wanlu Liu, and Zuozhu Liu. Multimodal survival ensemble network: Integrating genomic and histopathological insights for enhanced cancer prognosis. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2330–2334. IEEE, 2024.

Xinliang Zhu, Jiawen Yao, Feiyun Zhu, and Junzhou Huang. Wsisa: Making survival prediction from whole slide histopathological images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7234–7242, 2017.