



---

*Research article*

## **Predicting the prognosis of HER2-positive breast cancer patients by fusing pathological whole slide images and clinical features using multiple instance learning**

**Yifan Wang<sup>1,\*</sup>, Lu Zhang<sup>1</sup>, Yan Li<sup>1,\*</sup>, Fei Wu<sup>2</sup>, Shiyu Cao<sup>3</sup> and Feng Ye<sup>3</sup>**

<sup>1</sup> School of Computer Science, Southwest Petroleum University, Chengdu 610500, China

<sup>2</sup> Institute of Applied Physics and Computational Mathematics, China Academy of Engineering Physics, Beijing 100094, China

<sup>3</sup> Institute of Clinical Pathology, Key Laboratory of Transplant Engineering and Immunology, NHC, West China Hospital, Sichuan University, Chengdu 610041, China

\* **Correspondence:** Email: [yifan.wang@swpu.edu.cn](mailto:yifan.wang@swpu.edu.cn), [liyanlvq@126.com](mailto:liyanlvq@126.com).

**Abstract:** In 2022, breast cancer will become an important factor affecting women's public health and HER2 positivity for approximately 15–20% invasive breast cancer cases. Follow-up data for HER2-positive patients are rare, and research on prognosis and auxiliary diagnosis is still limited. In light of the findings obtained from the analysis of clinical features, we have developed a novel multiple instance learning (MIL) fusion model that integrates hematoxylin-eosin (HE) pathological images and clinical features to accurately predict the prognostic risk of patients. Specifically, we segmented the HE pathology images of patients into patches, clustered them by K-means, aggregated them into a bag feature-level representation through graph attention networks (GATs) and multihead attention networks, and fused them with clinical features to predict the prognosis of patients. We divided West China Hospital (WCH) patients ( $n = 1069$ ) into a training cohort and internal validation cohort and used The Cancer Genome Atlas (TCGA) patients ( $n = 160$ ) as an external test cohort. The 3-fold average C-index of the proposed OS-based model was 0.668, the C-index of the WCH test set was 0.765, and the C-index of the TCGA independent test set was 0.726. By plotting the Kaplan-Meier curve, the fusion feature ( $P = 0.034$ ) model distinguished high- and low-risk groups more accurately than clinical features ( $P = 0.19$ ). The MIL model can directly analyze a large number of unlabeled pathological images, and the multimodal model is more accurate than the unimodal models in predicting Her2-positive breast cancer prognosis based on large amounts of data.

**Keywords:** breast cancer; HER2; multiple instance learning; deep learning; survival analysis; prognosis

---

## 1. Introduction

According to Global Cancer Statistics 2022, for women, breast cancer, lung cancer, and colorectal cancer account for 51% of all new diagnoses, with breast cancer alone accounting for almost one-third [1]. HER2-positive breast cancer is a highly heterogeneous tumor, accounting for approximately 15–20% of invasive breast cancers [2]. Before the popularization of HER2-targeted therapy, early-stage HER2-positive patients tended to have shorter recurrence times, higher rates of metastasis, and higher mortality than HER2-negative patients. With the widespread use of the monoclonal antibody trastuzumab, which targets HER2, and the small-molecule antibody molecular tyrosine kinase inhibitor lapatinib, which targets HER1 and HER2, the prognosis of HER2-positive patients has improved significantly [3–5]. However, more than 30% of patients still experience recurrence, metastasis, or death within 10 years after treatment [2].

The purpose of survival analysis is to analyze the factors that cause events such as death or recurrence in patients in a certain period after treatment, which has important clinical application. Through radiomics and genomics, many scholars have attempted to identify factors that are associated with death or metastasis in patients [6–8]. It is easier for a clinician to choose an appropriate treatment when they have a more accurate assessment of a patient's survival risk [9–11]. In recent years, with the popularization and development of digital pathological images, pathologists can obtain digital pathological slices more quickly and with higher resolution. However, predicting patient prognosis from pathological images and clinical features is still challenging for the following reasons: 1) Whole slide images (WSIs) often contain more than one billion pixels. Therefore, mainstream computers and models cannot process them directly. 2) Due to the high heterogeneity of breast cancer, patients may have several pathological slides with different characteristics. 3) There are large differences in image features and clinical features.

In this paper, we propose an end-to-end system to predict the prognosis of patients through the fusion of pathological images and clinical features. The main contributions of this study are as follows: 1) The proposed model is based on MIL and thus does not require pixel-level annotation. The proposed model directly learns WSI representations from pathological images. 2) The proposed model fuses clinical features and image features through compact bilinear pooling (CBP), and the C-index of the fusion model is 0.06 higher than that of the single-modal model. 3) During the validation of the fusion model, stratified by lymph node status, the model still performed well.

Although traditional machine learning techniques, including support vector machines (SVMs) [12] and neural networks [13], have found wide applications in different fields, such as speckle reduction [14], image segmentation [15], and image retrieval [16], their performance needs further improvement. Deep learning can achieve this goal, and deep learning methods for pathological image processing can be divided into supervised learning and weakly supervised learning according to the labels of the data. Supervised learning typically analyzes partial regions of interest (ROIs) in WSIs, and the labels are based on the ROIs. Weakly supervised learning analyzes all or multiple WSIs, and the labels are typically based on pathology reports. Classic ROI-based methods require manual annotation by pathologists, and then patient traits are predicted by manually designing quantitative features. Cain et al. [17] established two multivariate models using radiomics features and proved that radiomics features could be used to predict the pathological complete response (pCR) to neoadjuvant therapy (NAT) in patients with triple-negative/HER2-positive breast cancer. Wang et al. [18] proposed

a computer-aided diagnosis and survival analysis system for non-small cell lung cancer (NSCLC) based on WSIs. They manually designed and extracted 166 image features for diagnosis and prognosis, with a classification accuracy of 92%. Unlike traditional learning methods, recent studies have used CNNs to perform feature extraction and classification/prediction [19]. For example, Yang et al. [20] divided the ROIs into  $512 \times 512$  pixel tiles and trained a ResNet-50-based model to predict the recurrence risk of breast cancer patients. Yu et al. [21] divided the WSIs into  $1000 \times 1000$  pixel tiles and extracted 9879 quantitative image features, which were used for prognosis prediction after verifying the validity of the features through two types of classification tasks. In our previous study [22], we used a deep fully convolutional neural network to perform end-to-end segmentation on pathological tissue slices. Our method achieved state-of-the-art segmentation accuracy on public nuclei histopathology datasets. Yan et al. [23] combined a CNN and an RNN to establish a cancer classification model, and the classification accuracy reached 92%.

However, ROIs may not capture all the information of WSIs, and pathologists must also manually annotate ROIs, which is time-consuming. WSI-based analysis methods have been proposed for the above reasons and can directly obtain the prediction results of WSIs or patients. Zhu et al. [24] proposed an effective whole-slide histopathological image survival analysis framework (WSISA) to predict patient prognosis. Li et al. [25] proposed a WSI-based framework, DeepGraphSurv, and pioneered graph convolutional networks (GCNs) for survival analysis. WSIs were extracted as a graph, and their network used graph convolutional networks to obtain a WSI representation. Yao et al. [26, 27] introduced MIL based on prior knowledge of pathology and proposed DeepAttnMISL. They clustered each WSI into multiple graphs and then aggregated them into bag representations through an attention-based MIL model. Wu et al. [28] proposed DeepGCNMIL based on this framework to optimize the model structure, which increased the C-index by 0.035. Campana et al. [29] built a cancer classification model based on MIL and trained it using a large amount of data from more than 40,000 WSIs, and their AUC was 0.98. Chen et al. [30] innovatively combined gene expression and pathological picture features to predict the WHO grade and prognosis of patients and achieved good results.

Although there have been many related studies of WSI-based analysis, research on prognosis prediction using MIL remains limited, and due to the scarcity of patient follow-up data, there is currently no related research that focuses on the prognosis analysis of HER2-positive patients. In this study, we established an MIL framework, GATSurvMIL, that integrates clinical features and WSI representations. During pathological diagnosis, we often do not think that the prognosis of a patient is related to the entire WSI, but a portion of the WSI reflects the patient's survival status.

## 2. Materials and methods

### 2.1. Data collection and screening

We collected the clinical features and WSIs of 1069 HER2-positive patients in WCH from 2009 to 2019 and downloaded the data of 160 HER2-positive patients from TCGA [31]. The study protocol (2020-427) was approved by the Ethics Committee on Biomedical Research, WCH of Sichuan University.

The distribution of specific clinical characteristics of patients is shown in Table 1. The age of diagnosis in the two datasets has a large gap, and the overall age of TCGA diagnosis is relatively high.

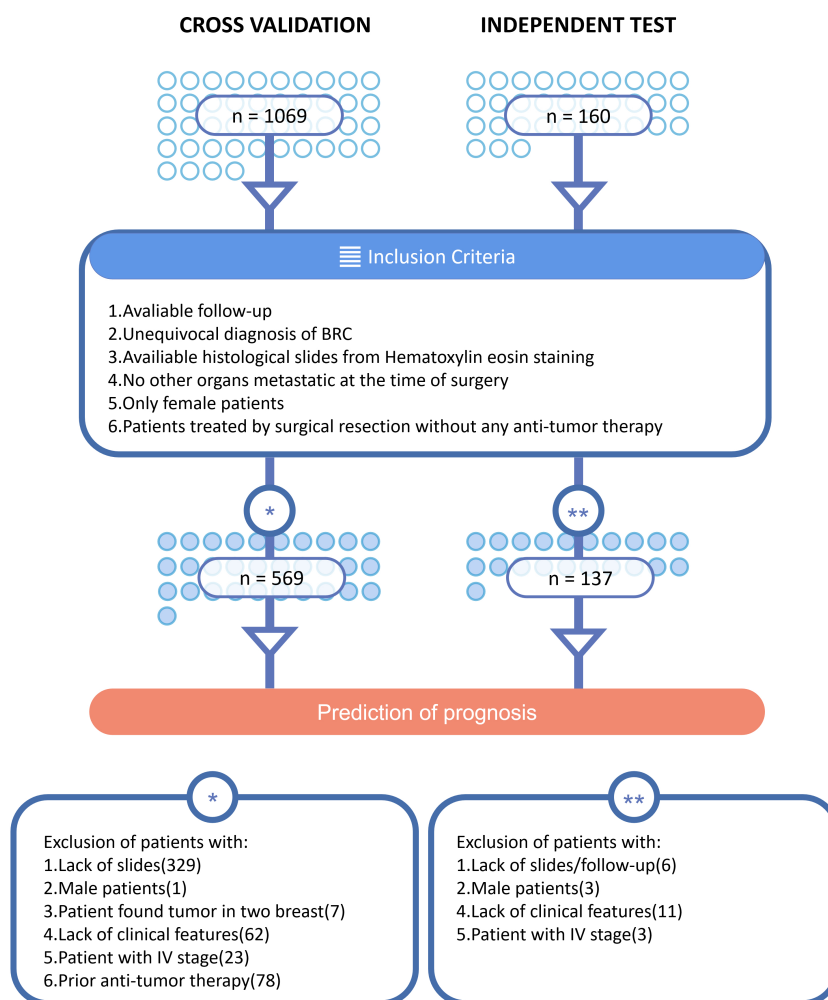
The mortality rate of TCGA is also much higher than that of WCH, but there are fewer patients with recurrence or metastasis.

**Table 1.** Distribution of clinical characteristics of patients with WCH and TCGA.

Variable	WCH Patient Series		TCGA Patient Series	
	n	%	n	%
Number of patients	1069		160	
Number of Histopathology image slides	3700		167	
Age				
< 50	590	55.19	37	23.12
≥ 50	479	44.81	123	76.88
Histological grade				
1	1	0.09	/	
2	235	22.98	/	
3	593	55.47	/	
NA	240	21.46	/	
Tumor stage				
I	207	19.36	18	11.25
II	429	40.13	95	59.38
III	356	33.30	42	26.25
IV	23	2.15	3	1.88
NA	54	5.06	2	1.24
ER				
Positive	521	48.74	121	75.63
Negative	539	50.42	38	23.75
NA	9	0.84	1	0.62
PR				
Positive	478	44.71	101	63.13
Negative	581	54.35	59	36.87
NA	10	0.94	0	0
Lymph node status (LMN)				
Positive	608	56.88	93	58.13
Negative	430	40.22	64	40.00
NA	31	2.90	3	1.87
Survival				
0	51	4.77	23	14.38
1	1018	95.23	137	85.62
Recurrence/Metastasis				
0	935	87.46	153	95.62
1	134	12.54	7	4.38

Data screening in this study was based on the criteria shown in Figure 1. All patients were identified as HER2 3+ via immunohistochemistry or HER2 gene amplification based on fluorescence

in situ hybridization (FISH). We then performed clinical pairing of the data to ensure that the clinical characteristics of the patients were evenly distributed, and all excluded patients were randomly selected.

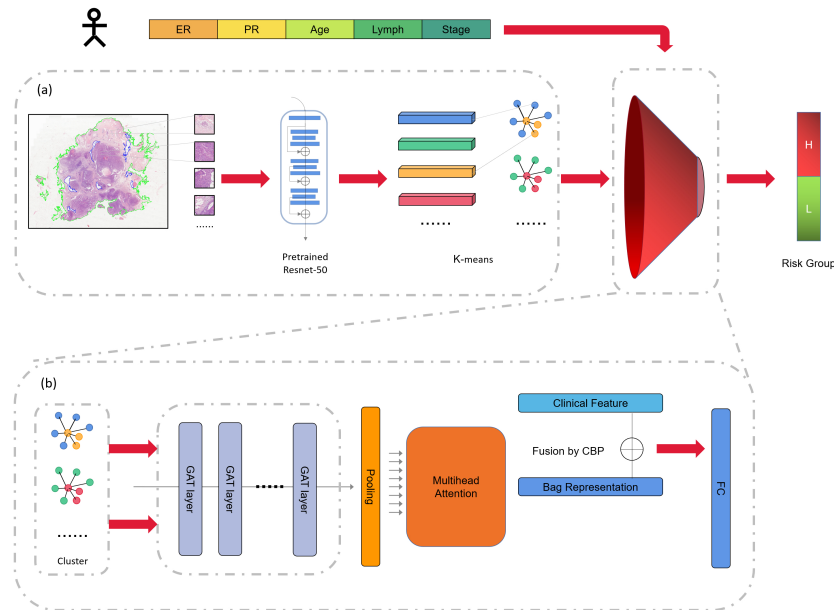


**Figure 1.** Data screening process. We screened both datasets according to the criteria in the figure, showing the specific reason and number of patients removed.

## 2.2. WSI processing

Because a WSI generally contains more than 1 billion pixels, existing machine learning models cannot directly process an entire WSI. Therefore, we cut each WSI into portions for sampling, as shown in Figure 2. First, we used an advanced tissue region segmentation framework [32] to segment each WSI's foreground from the background and then cut the foreground tissue area into patches, each with a size of  $256 \times 256$  pixels. Although patches can be used as the input of common neural networks, a WSI usually has more than 10,000 patches. Thus, we calculated the energy values of all the patches by 2D convolution and then used the 500 patches with the highest energy values to represent the WSI.

By doing this, we can reduce the computational complexity and information redundancy.



**Figure 2.** The framework of GATSurvMIL. (a) We divided the non-overlapping foreground of the WSI segmentation into patches and clustered the patch features with higher energy values. (b) The main architecture of the model, the image features are fused with clinical features after GAT and Attention modules to output risk values.

Next, we used the pretrained ResNet-50 [33] to extract the features of the color-normalized patches. The strong representational ability of ResNet-50 has been demonstrated in many previous related studies [26]. Finally, we used K-means clustering to cluster all patch features from a patient into several phenotypes, which have different predictive powers for patients' clinical outcomes.

### 2.3. Feature aggregation

After WSI processing, each patient was represented by 10 phenotype groups, unlike the Mi-FCN used by Yao et al. [26], which ignores intranode connections and only performs a simple aggregation. The GCN has a strong learning ability for graph data [34, 35] that far surpasses CNNs in node classification tasks. Based on the GCN, GAT introduces an attention mechanism that enables the ability to process dynamic graphs. Therefore, we used GAT [36] to learn the relationship between patches within a phenotype group. Figure 2 shows the proposed embedded GAT module, which contains multiple GAT layers and performs nonlinear transformations among GAT layers through ReLU. We determined the edge index among the patches using `knn_graph` and initialized them as a dynamic graph before the phenotype enters the model. The output is a feature vector for the phenotype group.

The data from each patient contains graph embedding vectors for multiple phenotypes, and we aggregated these vectors through a multihead attention network [37, 38]. The attention function is a method to map a query and a set of key-value pairs to an output. Multihead attention projects queries

and key-value pairs into multiple projection spaces to learn richer semantic information and can speed up model training through parallel learning:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^o \quad (2.1)$$

where  $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$ , and the projections are the parameter matrices  $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_q}$ ,  $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$ ,  $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ , and  $W^o \in \mathbb{R}^{h d_d \times d_{\text{model}}}$ .

#### 2.4. Feature fusion

Clinical characteristics play a crucial role in identifying patients for clinical treatment. Through data comparison, the selected clinical features are the intersection of two datasets: {ER, PR, Age, Lymph, Stage}. Common simple feature fusion methods primarily include concatenation, element dot product, and element summation. In this study, a series of experiments were conducted to investigate the efficacy of different feature fusion techniques. The GATSurvMIL model, which is based on image features, was employed as a baseline for comparative analysis. Based on the data presented in Table 2, it is evident that the utilization of these uncomplicated techniques has proven to be ineffective in enhancing the model's performance. In fact, these methods may even impede the model's learning capacity.

**Table 2.** Comparison of C-Index with Other Methods.

Method	fold 1	fold 2	fold 3	means (std)
CBP	0.641	0.765	0.596	0.668 (0.0713)
Concat	0.565	0.712	0.496	0.591 (0.1560)
Linear Add	0.702	0.597	0.513	0.604 (0.1339)
Outer Product	0.621	0.717	0.531	0.623 (0.1315)
base	0.716	0.589	0.541	0.609 (0.1323)

Thus, we used CBP to fuse features. Original bilinear pooling typically encounters problems when the dimension of fused features is too high. The authors also used principal component analysis (PCA) to reduce the dimensions, but the results are not significant. CBP uses the idea of a linear kernel machine to solve the problem of overdimensionality. We thus fused a patient's image features and clinical features using CBP and finally input the fused features into a fully connected layer to obtain the risk value of the patient.

#### 2.5. Loss function

We used the following loss function for backpropagation to update the model parameters:

$$L(\mathbf{o}_i) = \sum_i e_i \left( -\mathbf{o}_i + \log \sum_{j:t_j \geq t_i} \exp(\mathbf{o}_j) \right) \quad (2.2)$$

where  $\mathbf{o}_i$  is the predicted risk of the  $i$ -th patient,  $t$  is the patient's event occurrence time or follow-up time, and  $e$  is the patient's state. This function can be maximized via network parameter learning to obtain the largest partial likelihood estimate. Compared with that of the simple Cox loss function, the loss function can represent the overall concordance more accurately.

### 3. Results

#### 3.1. Implementation details

We set the initial learning rate to 0.0001 and the maximum number of epochs to 100. We updated the gradient and adjusted the step size with the Adam optimizer with a weight decay of 0.0005. We randomly selected 20% of the data in the training set as the validation set to control the early stopping of the model. We tested the model performance and tuned the hyperparameters using 3-fold cross-validation. The  $k$  in `knn_graph` was set to 10, and the number of GAT layers was set to 3. We then verified the generalizability of the model using the TCGA external test set. We trained two other models with the same data partition: 1) GATSurvMIL only with image features and 2) a Cox model only with clinical features.

#### 3.2. Model prognostic results

We used the C-index as a measure of the prognostic performance of the model. We built an OS (overall survival)-based model with death as the patient outcome. The model achieved a mean C-index of 0.668 for 3-fold cross-validation, and the C-index was 0.726 with the TCGA test set. Both the fitting effect and generalization effect of the model were markedly improved compared with those when using only imaging or clinical features.

**Table 3.** Prognostic analysis of the models' C-index.

Model		C-index				
		fold 1	fold 2	fold 3	means (std)	Test_TCGA
OS	Image	0.716	0.589	0.541	0.609 (0.1323)	0.574
	Clinical	0.740	0.616	0.436	0.597 (0.1248)	0.673
	Fusion	0.641	0.765	0.596	0.668 (0.0713)	0.726
DFS	Image	0.551	0.578	0.695	0.608 (0.0625)	0.526
	Clinical	0.601	0.657	0.614	0.624 (0.0240)	0.405
	Fusion	0.701	0.600	0.636	0.645 (0.0415)	0.617
	Image (LMN-)	0.505	0.603	0.858	0.655 (0.1486)	0.746
	Clinical (LMN-)	0.418	0.671	0.446	0.511 (0.1132)	0.459
	Fusion (LMN-)	0.516	0.611	0.862	0.663 (0.1461)	0.732
	Image (LMN+)	0.686	0.738	0.503	0.642(0.1011)	0.503
	Clinical (LMN+)	0.430	0.541	0.570	0.514 (0.0603)	0.508
	Fusion (LMN+)	0.677	0.780	0.434	0.630 (0.1453)	0.656

We built a DFS (disease-free survival)-based model with recurrence, metastasis, or death as the patient outcome. The 3-fold validation cross-average C-index of the model was 0.645, and the C-index with the TCGA test set was 0.617. Because lymph node metastasis is an important factor affecting treatment, we separately modeled patients with lymph node status (LMN) according to lymph node stage. The performance of the two models are similar to that of the image-only model. The specific results are shown in Table 3.



### 3.3. Comparison of the C-index with other MIL models

We reproduced the experimental results of DeepAttnMISL [26] and DeepGCNMIL [28]. We then used the same parameters to test the performance of the proposed model. We also fused each patient's pathological features using CBP before the output layer of the model. The results in Table 4 demonstrate that the cross-validation results of our model (mean = 0.668) are much better than those of DeepAttnMISL (mean = 0.606) and DeepGCNMIL (mean = 0.592). The cross-validation standard deviation of our model was 0.0713, which is more stable than that of the other models.

**Table 4.** Comparison of C-Index with other methods.

Model	fold 1	fold 2	fold 3	means (std)	Test_TCGA
GATSurvMIL	0.641	0.765	0.596	0.668 (0.0713)	0.726
DeepAttnMISL	0.608	0.784	0.426	0.606 (0.1459)	0.714
DeepGCNMIL	0.694	0.804	0.279	0.592 (0.2260)	0.573

### 3.4. Multivariate Cox analysis

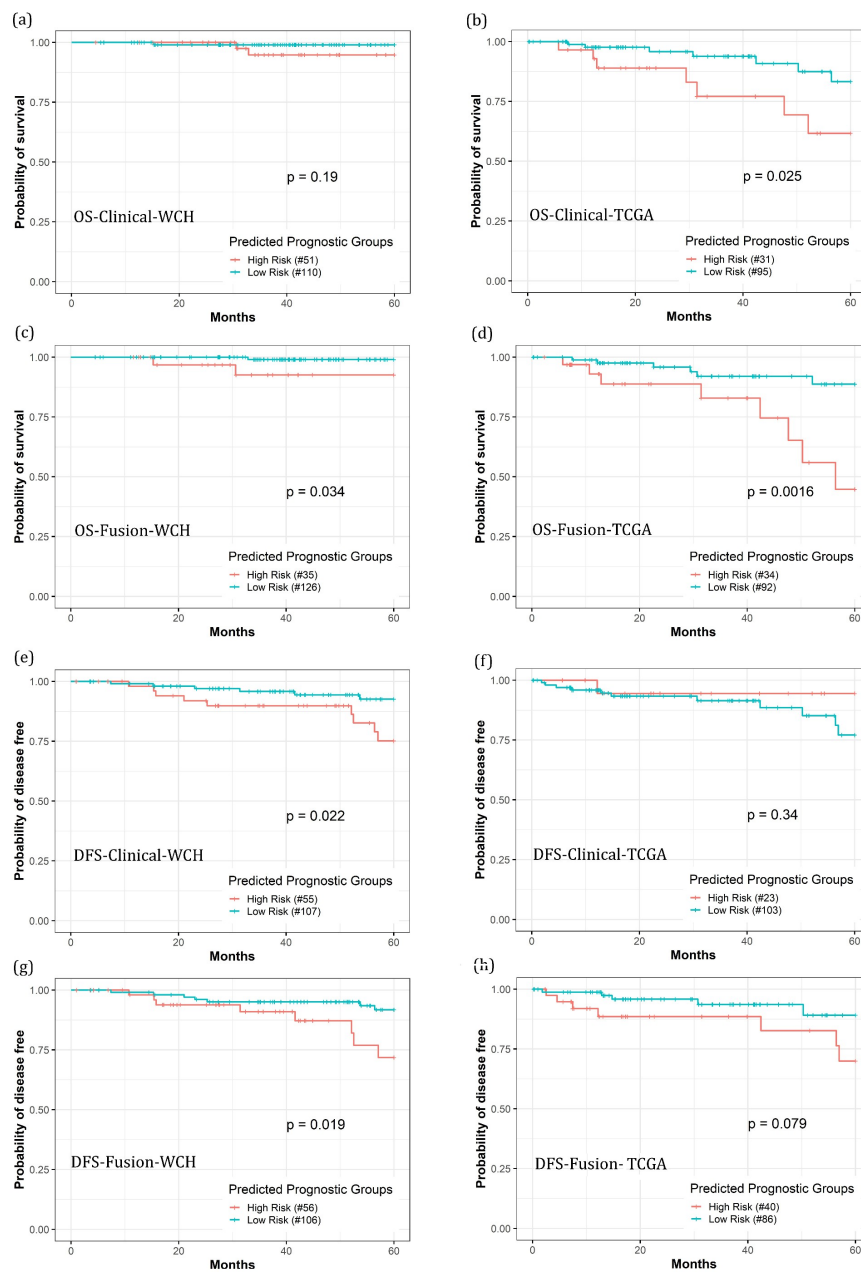
We used the image-only model prediction results and clinical characteristics as the input of the multivariate Cox model to analyze the primary factors affecting patient survival and prognosis. The results are shown in Table 5. The results of the Cox model based on OS and DFS are similar, in which the image model predicted value and tumor stage are strongly associated with patient prognosis. The hazard ratio (HR = 4.15) of the predicted value indicates that the imaging factor has a stronger impact on the prognosis of the patient.

**Table 5.** Comparison of C-Index with other methods.

Variable	OS		DFS	
	P_value	Hazard Ratio (95% CI)	P_value	Hazard Ratio (95% CI)
ER	0.65186	0.72(0.17–2.94)	0.53722	1.36(0.51–3.59)
PR	0.30022	0.45(0.11–2.00)	0.30819	0.60(0.22–1.61)
Age	0.63668	1.17(0.61–2.22)	0.17309	0.75(0.49–1.14)
Stage	0.02797	2.59(1.11–6.04)	0.01801	1.98(1.12–3.48)
Image score	0.00466	4.15(1.55–11.12)	0.00007	2.99(1.74–5.12)

### 3.5. Kaplan–Meier survival curves

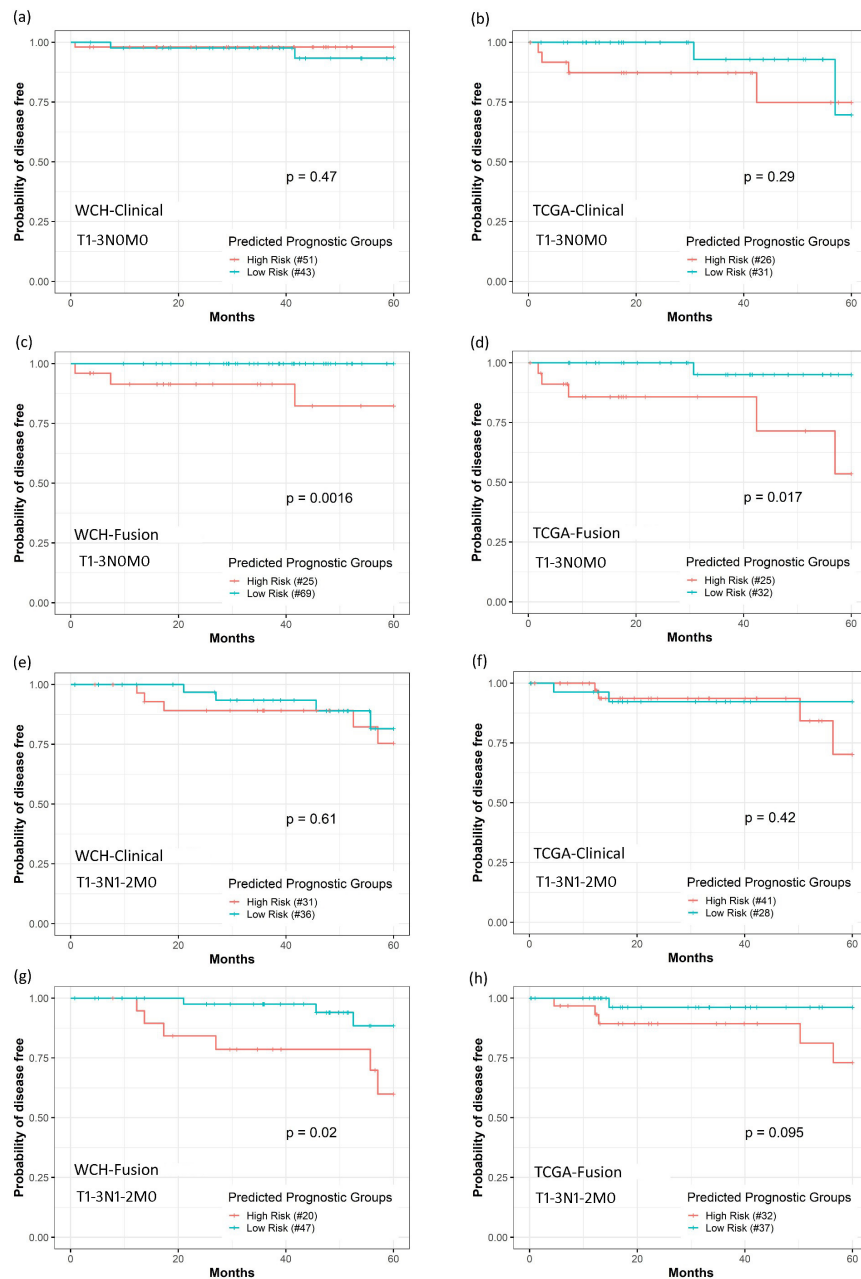
Currently, clinicians assess a patient's prognostic risk based on the patient's clinical characteristics. To verify the predictive performance of the proposed model, we drew a Kaplan–Meier survival curve of the patients in the test set, which is shown in Figure 3. Since prognostic analysis is time dependent, we calculated specificity and sensitivity by time-varying ROC using Hegerty's method [39]. We used the value corresponding to the year with the largest Youden index as the cutoff value for the risk grouping.



**Figure 3.** The Kaplan-Meier curve of the test set by selecting the cutoff value based on the time-varying ROC. **(a, b)** Our Cox model with OS as event by clinical features. **(c, d)** Our GATSurvMIL model with OS as event by fusion features. **(e–h)** The DFS model corresponding to the above figures. s are fused with clinical features after GAT and Attention modules to output risk values.

Figure 3 (a, b) shows the OS Cox model established by clinical features. Figure 3 (c, d) shows GATSurvMIL fused with imaging features. The high-risk and low-risk groups of WCH could not be significantly distinguished by clinical characteristics ( $P = 0.19$ ), but GATSurvMIL could significantly

distinguish the high- and low-risk groups of WCH ( $P = 0.034$ ) and TCGA ( $P = 0.0016$ ). Figure 3(e–h) shows the results of the corresponding DFS model, in which patients were indistinguishable by clinical features ( $P = 0.34$ ), and even after 20 months, the high-risk group appeared safer, but the GATSurvMIL fusion feature could preliminarily separate patients ( $P = 0.079$ ). We performed the same analysis for the models built by LMN+ and LMN-, and the results are shown in Figure 4.



**Figure 4.** The Kaplan-Meier curves of the LMN- and LMN+ model. **(a, b)** LMN- model with DFS as the event by clinical features. **(c, d)** GATSurvMIL model with DFS as the event by fusion features. **(e–h)** The LMN+ model corresponding to the above figures.

---

## 4. Discussion

Breast cancer has become the most common tumor in the world. HER2-positive tumors are aggressive and prone to recurrence and metastasis, and the prognosis of patients is poor, making these tumors a major threat to public health. Unfortunately, no effective model or product is currently available to predict the prognosis of HER2-positive patients. Currently, the C-index for WSI-based breast cancer prognosis prediction is generally 0.5 to 0.7, and the performance of image-based models must be improved.

Based on the performance of GATSurvMIL on DFS, the proposed model can effectively integrate the advantages of images and clinical features to obtain better performance when clinical features cannot accurately represent patient risk. We modeled patients separately by lymph node status, where clinical features barely describe a patient's prognostic risk, but the proposed model still combines the strengths of imaging and clinical features. We believe that when the number of patients is small, the image features still have a strong ability to represent patients' risk because the information in pathological images is rich. However, clinical features cannot accurately represent the prognosis of patients due to differences in patient distribution. When the number of patients is large, clinical features have a more significant discriminative ability, and the performance of the proposed fused feature model can also be improved more than only WSIs. Therefore, there is still much room for improvement in the prognostic analysis of pathological images. The proposed model is also based on MIL and thus does not require pixel-level manual annotation and can be directly applied to most current datasets. The introduced GAT and attention network provide better interpretability with improved model performance. Currently, the prediction of patient prognosis through pathological images remains in the experimental stage. With an increasing number of patients and follow-up data, more advanced methods and models should appear in the future. Compared to using one feature alone to predict patient outcomes, multimodal models can predict patient status from a more comprehensive perspective. In contrast, multimodal data often require patients to have experienced multiple diagnostic methods and sufficient follow-up time, which is rare. WSI, MRI, and clinical features are often different, and it is also difficult to find a suitable feature fusion method.

## 5. Conclusions

This paper proposes an MIL-based multimodal model to predict the prognosis of HER2-positive breast cancer patients. The proposed model can be directly applied to most current hospital data, which can help provide personalized treatment of patients and auxiliary decision-making for doctors. However, this study currently has some limitations: 1) since TCGA provides fewer clinical features, we only included 5 clinical variables. 2) Because survival analysis requires long-term tracking records, the amount of data is small. 3) The "black box" problem of neural networks is still unavoidable, even though we have added GAT and attention modules. We continue to work on improving the prediction of patient prognosis and plan to study more modal models and algorithms in the future by combining features such as gene expression and radiomics. Finally, we look forward to more central datasets and longer follow-up records to improve model performance.

## Acknowledgments

This work was supported by Natural Science Foundation of Sichuan, China (2023NSFSC1393); Scientific Research Starting Project of SWPU (2021QHZ001); Sichuan Nanchong Science and Technology Bureau (SXQHJH046); West China Hospital of Sichuan University-University of Electronic Science and Technology of China Medical-Industrial Integration Interdisciplinary Talent Training Fund (ZYGX2022YGRH015); 1.3.5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (ZYJC21035) and Fund of Sichuan Provincial Department of Science and Technology (2023YFH0095).

## Conflict of interest

The authors declare there is no conflict of interest.

## References

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2022, *CA Cancer J. Clin.*, **72** (2022), 7–33. <https://doi.org/10.3322/caac.21708>
2. E. A. Perez, E. H. Romond, V. J. Suman, J. Jeong, G. Sledge, C. E. Geyer Jr, et al., Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2–Positive breast cancer: Planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831, *JCO*, **32** (2014), 3744–3752. <https://doi.org/10.1200/JCO.2014.55.5730>
3. C. L. Arteaga, M. X. Sliwkowski, C. K. Osborne, E. A. Perez, F. Puglisi, L. Gianni, Treatment of HER2-positive breast cancer: current status and future perspectives, *Nat. Rev. Clin. Oncol.*, **9** (2012), 16–32. <https://doi.org/10.1038/nrclinonc.2011.177>
4. J. N. Wang, B. H. Xu, Targeted therapeutic options and future perspectives for HER2-positive breast cancer, *Sig. Transduct. Target Ther.*, **4** (2019), 34. <https://doi.org/10.1038/s41392-019-0069-2>
5. D. Cameron, M. J. Piccart-Gebhart, R. D. Gelber, M. Procter, A. Goldhirsch, E. de Azambuja, et al., 11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: Final analysis of the HERceptin Adjuvant (HERA) trial, *Lancet*, **389** (2017), 1195–1205. [https://doi.org/10.1016/S0140-6736\(16\)32616-2](https://doi.org/10.1016/S0140-6736(16)32616-2)
6. Director's challenge consortium for the molecular classification of lung adenocarcinoma, Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study, *Nat. Med.*, **14** (2008), 822–827. <https://doi.org/10.1038/nm.1790>
7. M. Y. Park, T. Hastie, L1-regularization path algorithm for generalized linear models, *J Royal Statistical Soc B*, **69** (2007), 659–677. <https://doi.org/10.1111/j.1467-9868.2007.00607.x>
8. E. Bair, R. Tibshirani, Semi-Supervised methods to predict patient survival from gene expression data, *PLoS Biol.*, **2** (2004), 512–522. <https://doi.org/10.1371/journal.pbio.0020108>

9. A. Warth, T. Muley, M. Meister, A. Stenzinger, M. Thomas, P. Schirmacher, et al., The novel histologic international association for the study of lung cancer/American thoracic society/European respiratory society classification system of lung adenocarcinoma is a Stage-Independent predictor of survival, *JCO*, **30** (2012), 1438–1446. <https://doi.org/10.1200/JCO.2011.37.2185>
10. B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, et al., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, *JAMA*, **318** (2017), 2199. <https://doi.org/10.1001/jama.2017.14585>
11. Y. Yuan, H. Failmezger, O. M. Rueda, H. R. Ali, S. Gräf, S. Chin, et al., Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling, *Sci. Transl. Med.*, **4** (2012). <https://doi.org/10.1126/scitranslmed.3004330>
12. J. Xu, Y. Cao, Y. Sun, J. Tang, Absolute exponential stability of recurrent neural networks with generalized activation function, *IEEE Trans. Neural Networks*, **19** (2008), 1075–1089. <https://doi.org/10.1109/TNN.2007.2000060>
13. J. Tang, X. Liu, H. Cheng, K. M. Robinette, Gender recognition using 3-D human body shapes, *IEEE Trans. Syst. Man Cybern. C*, **41** (2011), 898–908. <https://doi.org/10.1109/TSMCC.2011.2104950>
14. X. Liu, J. Liu, X. Xu, L. Chun, J. Tang, Y. Deng, A robust detail preserving anisotropic diffusion for speckle reduction in ultrasound images, *BMC Genom.*, **12** (2011), S14. <https://doi.org/10.1186/1471-2164-12-S5-S14>
15. J. Tang, S. Millington, S. T. Acton, J. Crandall, S. Hurwitz, Ankle cartilage surface segmentation using directional gradient vector flow snakes 2004. *IEEE Int. Conf. Inf. Process.*, **4** (2004), 2745–2748. <https://doi.org/10.1109/ICIP.2004.1421672>
16. J. Tang, S. Acton, An image retrieval algorithm using multiple query images, *ISSPA 2003*, **1** (2003), 193–196. <https://doi.org/10.1109/ISSPA.2003.1224673>
17. E. H. Cain, A. Saha, M. R. Harowicz, J. R. Marks, P. K. Marcom, M. A. Mazurowski, Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set, *Breast Cancer Res. Treat.*, **173** (2019), 455–463. <https://doi.org/10.1007/s10549-018-4990-9>
18. H. Wang, F. Xing, H. Su, A. Stromberg, L. Yang, Novel image markers for non-small cell lung cancer classification and survival prediction, *BMC Bioinform.*, **15** (2014), 310. <https://doi.org/10.1186/1471-2105-15-310>
19. Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, Q. Sun Jr, Deep learning for image-based cancer detection and diagnosis-A survey, *Pattern Recognition*, **83** (2018), 134–149. <https://doi.org/10.1016/j.patcog.2018.05.014>
20. J. Yang, J. Ju, L. Guo, B. Ji, S. Shi, Z. Yang, et al., Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning, *Comput. Struct. Biotel.*, **20** (2022), 333–342. <https://doi.org/10.1016/j.csbj.2021.12.028>

21. K. Yu, C. Zhang, G. J. Berry, R. B. Altman, C. Ré, D. L. Rubin, et al., Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features, *Nat. Commun.*, **7** (2016), 12474. <https://doi.org/10.1038/ncomms12474>
22. X. Liu, Z. Guo, J. Cao, J. Tang, MDC-net: A new convolutional neural network for nucleus segmentation in histopathology images with distance maps and contour information, *Comput. Biol. Med.*, **135** (2021), 104543. <https://doi.org/10.1016/j.compbiomed.2021.104543>
23. R. Yan, F. Ren, Z. Wang, L. Wang, T. Zhang, Y. Liu, Breast cancer histopathological image classification using a hybrid deep neural network, *Methods*, **173** (2020), 52–60. <https://doi.org/10.1016/j.ymeth.2019.06.014>
24. X. Zhu, J. Yao, F. Zhu, J. Huang, Wsisa: Making survival prediction from whole slide histopathological images, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 7234–7242. <https://doi.org/10.1109/CVPR.2017.725>
25. R. Li, Graph CNN for survival analysis on whole slide pathological images, in *Medical Image Computing and Computer Assisted Intervention*, Springer International Publishing, (2018), 174–182. [https://doi.org/10.1007/978-3-030-00934-2\\_20](https://doi.org/10.1007/978-3-030-00934-2_20)
26. J. Yao, X. Zhu, J. Jonnagaddala, N. Hawkins, J. Huang, Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks, *Med. Image Anal.*, **65** (2022), 101789. <https://doi.org/10.1016/j.media.2020.101789>
27. M. Y. Lu, D. F. K. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, F. Mahmood, Data-efficient and weakly supervised computational pathology on whole-slide images, *Nat. Biomed. Eng.*, **5** (2021), 555–570. <https://doi.org/10.1038/s41551-020-00682-w>
28. F. Wu, P. Liu, B. Fu, Y. Ye, DeepGCNMIL: Multi-head attention guided multi-instance learning approach for whole-slide images survival analysis using graph convolutional networks, *ICMLC 2022*, (2022), 67–73. <https://doi.org/10.1145/3529836.3529942>
29. G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. Silva, K. J. Busam, et al., Clinical-grade computational pathology using weakly supervised deep learning on whole slide images, *Nat. Med.*, **25** (2019), 1301–1309. <https://doi.org/10.1038/s41591-019-0508-1>
30. R. J. Chen, M. Y. Lu, J. Wang, D. F. K. Williamson, S. J. Rodig, N. I. Lindeman, et al., Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis., *IEEE Trans. Med. Imaging*, **4** (2022), 757–770. <https://doi.org/10.1109/TMI.2020.3021387>
31. C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, et al., Mutational landscape and significance across 12 major cancer types, *Nature*, **502** (2013), 333–339. <https://doi.org/10.1038/nature12634>
32. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv: 1409.1556.
33. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>

34. T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, preprint, arXiv: 1609.02907.
35. W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, *Adv. Neural Inform. Proc. Syst.*, (2017), 30. <https://doi.org/10.5555/3294771.3294869>
36. P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, preprint, arXiv: 1710.10903.
37. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, preprint, arXiv: 1706.03762.
38. J. Devlin, M. W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, preprint, arXiv: 1810.04805.
39. P. J. Heagerty, T. Lumley, M. S. Pepe, Time-dependent ROC curves for censored survival data and a diagnostic marker, *Biometrics*, **56** (2000), 337–344. <https://doi.org/10.1111/j.0006-341X.2000.00337.x>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)