

## ORIGINAL ARTICLE

# Survival prediction of stomach cancer using expression data and deep learning models with histopathological images

Ting Wei<sup>1,2</sup> | Xin Yuan<sup>1,2</sup> | Ruitian Gao<sup>1,2</sup> | Luke Johnston<sup>2,3</sup> | Jie Zhou<sup>2,3</sup> |  
Yifan Wang<sup>1,2</sup> | Weiming Kong<sup>4</sup> | Yujing Xie<sup>2,3</sup> | Yue Zhang<sup>1,2</sup> | Dakang Xu<sup>5</sup> |  
Zhangsheng Yu<sup>1,2,3,6</sup> 

<sup>1</sup>Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

<sup>2</sup>SJTU-Yale Joint Centre for Biostatistics and Data Sciences, Shanghai Jiao Tong University, Shanghai, China

<sup>3</sup>School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China

<sup>4</sup>Institute of Transactional Medicine, Shanghai Jiao Tong University School of Medicine, Shanghai, China

<sup>5</sup>Faculty of Medical Laboratory Science, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

<sup>6</sup>Clinical Research Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China

## Correspondence

Zhangsheng Yu, Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China.  
Email: [yuzhangsheng@sjtu.edu.cn](mailto:yuzhangsheng@sjtu.edu.cn)

Dakang Xu, Faculty of Medical Laboratory Science, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China.  
Email: [dakang\\_xu@163.com](mailto:dakang_xu@163.com)

## Funding information

National Natural Science Foundation of China, Grant/Award Number: 11671256; Shanghai Jiao Tong University STAR Grant

## Abstract

Accurately predicting patient survival is essential for cancer treatment decision. However, the prognostic prediction model based on histopathological images of stomach cancer patients is still yet to be developed. We propose a deep learning-based model (MultiDeepCox-SC) that predicts overall survival in patients with stomach cancer by integrating histopathological images, clinical data, and gene expression data. The MultiDeepCox-SC not only automatically selects patches with more information for survival prediction, without manual labeling for histopathological images, but also identifies genetic and clinical risk factors associated with survival in stomach cancer. The prognostic accuracy of the MultiDeepCox-SC (C-index = 0.744) surpasses the result only based on histopathological image (C-index = 0.660). The risk score of our model was still an independent predictor of survival outcome after adjustment for potential confounders, including pathologic stage, grade, age, race, and gender on The Cancer Genome Atlas dataset (hazard ratio 1.555,  $p = 3.53e-08$ ) and the external test set (hazard ratio 2.912,  $p = 9.42e-4$ ). Our fully automated online prognostic tool based on histopathological images, clinical data, and gene expression data could be utilized to improve pathologists' efficiency and accuracy (<https://yu.life.sjtu.edu.cn/DeepCoxSC>).

## KEYWORDS

multiomics, pathology, prognostic, stomach cancer, survival analysis

**Abbreviations:** AJCC, American Joint Committee on Cancer; AUC, area under the receiver operating characteristic curve; CNN, convolutional neural network; CPH, Cox proportional hazards; FFPE, formalin-fixed paraffin-embedded; GSCNN, Genomic Survival CNN; HR, hazard ratio; LASSO, least absolute shrinkage and selection operator; OS, overall survival; ROI, region of interest; SIS, sure independence screening; TCGA, The Cancer Genome Atlas; WSISA, Whole Slide Histopathological Images Survival Analysis.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Cancer Science* published by John Wiley & Sons Australia, Ltd on behalf of Japanese Cancer Association.

## 1 | INTRODUCTION

As a clinical gold standard for cancer diagnosis and prognosis, histopathological images guide clinicians to make more precise treatment decisions.<sup>1,2</sup> Pathologists evaluate the morphological characteristics of cancer cells under a microscope to grade.<sup>3</sup> However, manual assessment of large-scale histopathological images is highly time-consuming, subjective, and not repeatable, especially for pathologists working in remote regions.<sup>4</sup> Hence, fully automated prognostic models for survival prediction directly from pathological images have attracted great attention.<sup>5</sup> This computer-assisted tool can be used to improve pathologists' efficiency and accuracy, and ultimately provide better patient treatment.<sup>2,6,7</sup> The main approaches for survival prediction of patients can be divided into three categories: hand-crafted feature-based survival models, deep convolutional network-based survival models, and multimodal fusion.<sup>7</sup>

*Hand-crafted feature-based survival models* typically extract structured image features including cell shape, size, and texture from unstructured pathological images. Yu et al. extracted 9879 quantitative image features of lung cancers and used machine-learning methods to predict cancer prognosis.<sup>8</sup> However, image features had limited abilities in representing image information. In recent years, researchers have attempted to use deep learning to learn survival outcomes directly from histopathological images.

*Deep convolutional network-based survival models* often use deep CNN for survival analysis by replacing the traditional linear risk function of the CPH model with a nonlinear deep fully connected network. According to whether the manually labeled ROIs were needed, a deep convolutional network-based survival model can be categorized into two categories: ROI-based approaches and whole-slide image approaches.

Region of interest-based approaches focused on ROIs of whole-slide images annotated by pathologists. DeepConSurv (deep convolutional network for survival analysis) was a deep CNN for survival prediction with manually labeled discriminative patches from ROIs.<sup>9,10</sup> However, ROI-based methods require ROIs to be labeled by pathologists and were not end-to-end analyses.

Whole-slide image approaches directly used the whole-slide histopathological image to predict patient outcome. The WSISA framework extracted hundreds of patches from whole-slide images and grouped them into clusters, then made patient-level prediction from these clusters in lung cancer and glioblastoma.<sup>11</sup> Kather et al. trained a deep neural network to identify nine tissue subclasses and used CNN to build a predictive score from each of these tissue subclasses in colorectal cancer.<sup>12</sup> Both WSISA and Kather's method used histopathological images and ignored the clinical data and other omics data.

*Multimodal fusion models* integrated histopathological images and other omics data to improve the prognostic accuracy. The GSCNN integrated well-known genomic biomarkers and pathological images into a unified prediction framework to predict patient survival with glioblastoma.<sup>13</sup> However, GSCNN ignored the clinical information of patients. PAGE-Net used histopathology-specific

CNN and genome-specific sparse deep neural networks to extract survival-discriminative features in glioblastoma.<sup>14</sup> S-net integrated computed tomography images and clinicopathological factors to predict survival for patients with stomach cancer.<sup>15</sup>

Stomach cancer is characterized by a complicated and heterogeneous histomorphology,<sup>16</sup> and there has not yet been a model using histopathological images, clinical data, and gene expression data of stomach cancer to predict prognosis. This study develops a fully automated deep CNN for survival prediction using the CPH model directly from histopathological images in stomach cancer (DeepCox-SC). Using stomach cancer data from TCGA datasets and the external test set, we show that the prognostic accuracy of the DeepCox-SC based on histopathological image has similar performance with the clinical benchmark model integrating grade, stage, and clinical data. In addition, the prediction accuracy of the MultiDeepCox-SC integrating histopathological images, clinical data, and high-dimensional gene expression data significantly surpasses the DeepCox-SC model (Wilcoxon signed rank  $p = 0.005$ ). The main contributions of our model are two aspects: (i) improving the survival prediction of stomach cancer patients by integrating histopathological images, clinical data, and gene expression data; and (ii) developing a user-friendly online tool for survival prediction using histopathological images, clinical data, and gene expression data in stomach cancer (<https://yu.life.sjtu.edu.cn/DeepCoxSC>).

## 2 | MATERIALS AND METHODS

### 2.1 | Patients cohorts

The H&E stained histopathological images, clinical data, and gene expression data of stomach cancer were obtained from the publicly available TCGA datasets (<https://portal.gdc.cancer.gov/projects/TCGA-STAD>). The clinical information of each sample is presented in Table S1 and the TCGA manifest file for download data is presented in Table S2. Histopathological images from FFPE tissues were included because FFPE slides were the gold standard for diagnostic medicine. Initially, 442 whole-slide histopathology images in .svs format from 416 unique patients (36 patients included two whole slide images) were downloaded. Among these patients, 41 patients with no 20× magnification (0.5  $\mu\text{m}$  per pixel) whole-slide images and 18 patients with missing survival data were excluded. Finally, a total of 382 whole-slide images (20× magnification) from 357 unique patients were analyzed. Figure S1 shows the flowchart of the stomach cancer dataset.

The external test set including 30 gastric cancer patients was previously collected at Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine. This study was approved by the Ethics Committee of Ruijin Hospital (ID: 2021-194). We used these data for the secondary analysis.

The staging is based on the 7th edition of the AJCC TNM staging system for stomach cancer.<sup>17</sup> The TCGA gastric cancer samples included different versions of the AJCC TNM staging systems (AJCC

4th–7th editions). Clinical staging with different versions of staging systems was adjusted to the 7th edition.<sup>18,19</sup> The grade by pathologist depends on what the cancer cells look like under a microscope according to the WHO classification.<sup>20</sup>

## 2.2 | Whole-slide images preprocessing

### 2.2.1 | Step 1: Cropping whole-slide images into patches

Whole-slide images are typically gigapixels in size and so should not be directly used as the input of CNN. An increase in the size of the input image increased the parameter to be estimated and computational power. Taking these into account, whole-slide images with 20× magnification were cropped into small patches with 1536×1536 pixels. The margin of whole-slide images contained white background, so some patches contained white background when we processed the whole slide images into patches. Patches containing white background space of more than 30% were filtered out.

### 2.2.2 | Step 2: Color normalization

Many factors, such as staining and the slide digitization process,<sup>21</sup> resulted in drastically different appearances between two histopathological images (Figure S2). This difference significantly affected the performance of the CNN model.<sup>22</sup> Patches were normalized to a gold standard H&E histopathological image using the Macenko method.<sup>23</sup>

### 2.2.3 | Step 3: Nuclei segmentation

After patch filtration, an average of 2937 patches for each case remained. Thousands of patches of each case cannot be directly used for the CNN model. Cellular features such as nuclear size and texture played an important role in diagnosis and grading. We segmented cell nuclei of each patch with hierarchical multi-level thresholding<sup>24</sup> and then selected the patch with the largest frequency of cell nuclei for each case. The selected patches (1536×1536 pixels) consisted of nine patches of 512×512 pixels and we used CellProfiler to extract the 1100 features of patches (512×512 pixels). The results of the univariate Cox regression showed that the *p* value of the feature (Median\_PrimaryObject\_AreaShape\_Zernike\_7\_7) was smallest and the absolute value of the coefficient was largest. We selected the patch of 512×512 pixels with the smallest “Median\_PrimaryObject\_AreaShape\_Zernike\_7\_7” among nine patches, because the coefficient of this feature was negative. The selected patches of 512×512 pixels contained more histopathological information and were used to train the DeepCox-SC model to predict patient risk.

## 2.3 | DeepCox-SC model architecture

Figure 1 shows the DeepCox-SC model architecture. We combined the InceptionResNetV2 convolutional network with CPH model to predict the survival risk of stomach cancer from selected patches. The DeepCox-SC model takes selected patches as the input and outputs the predicted patient's risk of death (DeepCox-SC risk score). The higher DeepCox-SC risk score means the higher risk of death.

One standard survival model is the CPH model that calculates the effects of structured covariates, such as age, gender, and stage, on the risk of death. The DeepCox-SC model combines CNN with the CPH model to predict the risk of death from histopathological images (unstructured data). The DeepCox-SC risk score is the estimated risk function ( $\hat{h}_\theta(x)$  in Equation 1) based on histopathological images. Convolutional layers of InceptionResNetV2 pretrained on ImageNet first extract the spatial features of histopathological images using convolutional kernels and pooling operations.<sup>25</sup> These features are then connected to fully connected layers. The output of the network was a single node, which estimates the risk function  $\hat{h}_\theta(x)$ . The weights of the network are trained with the time-to-event (death) outcomes to optimize the Cox likelihood function. The Cox partial likelihood was defined as:

$$L(\theta) = \prod_{i=1}^n \left[ \frac{\exp(\hat{h}_\theta(x_i))}{\sum_{j \in \mathcal{R}(X_i)} \exp(\hat{h}_\theta(x_j))} \right]^{\delta_i} \quad (1)$$

When we minimized negative log partial likelihood, the loss function was:

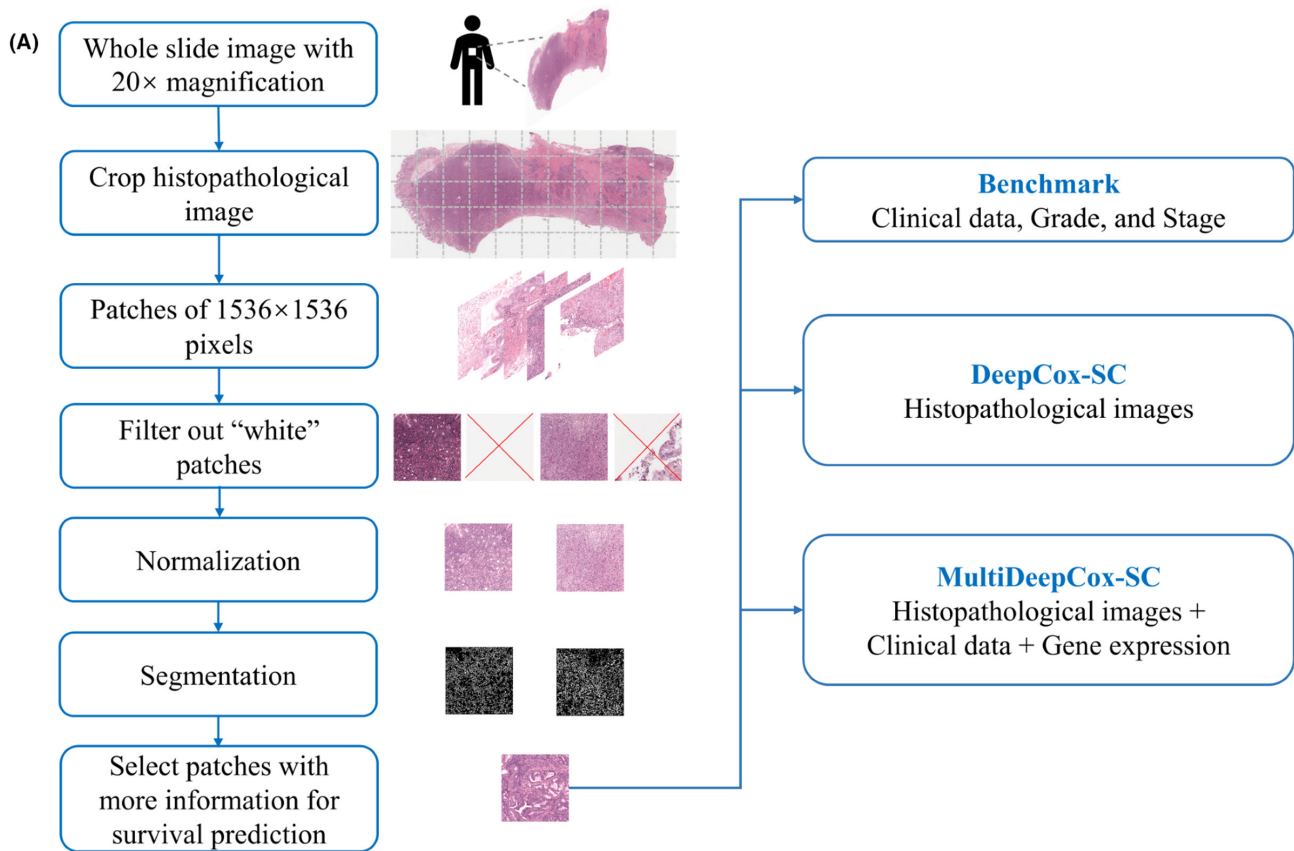
$$l(\theta) = - \sum_{i=1}^n \delta_i \left( \hat{h}_\theta(x_i) - \log \sum_{j \in \mathcal{R}(X_i)} \exp(\hat{h}_\theta(x_j)) \right) \quad (2)$$

$\delta_i$  was an indicator of whether the survival time was censored ( $\delta_i = 0$ ) or observed ( $\delta_i = 1$ ).  $\theta$  was the weight of the CNN.  $\mathcal{R}(X_i)$  denoted the set of individuals who were at risk for failure time of individual *i*.

## 2.4 | DeepCox-SC model training

### 2.4.1 | Ten-fold cross-validation

The model was trained using 10-fold cross-validation. Stratified random sampling was used to split the data, making the censoring rate and time distribution the same across each fold. Some patients had two whole-slide images, so if a patient was assigned to the training set, then all whole-slide images of this patient were assigned to the training set. This measure ensured histopathological images of one sample were not in both training and validation sets, and effectively avoided information disclosure. The final



(B)

1 **SHANGHAI JIAO TONG UNIVERSITY** Introduction DeepCox-SC Analysis

**What is DeepCox-SC?**

A fully automated **Deep** convolutional network for survival prediction using the **Cox** Proportional-Hazard model directly from histopathological images in **Stomach Cancer** (DeepCox-SC). The DeepCox-SC model automatically selected patches with the largest frequency of nuclei as the input, which was showed to be more informative for survival prediction. DeepCox-SC multi-modal fusion model further integrated DeepCox-SC risk score, demographics, and gene expression data to improve survival prediction. The DeepCox-SC model could be utilized as a computer-assisted tool to improve pathologists' efficiency and accuracy.

2 **SHANGHAI JIAO TONG UNIVERSITY** Introduction DeepCox-SC Analysis

**Submit Your Histopathological Image**

Uri of Whole Slide Image (WSI):

Email:

Age (optional):

Gene Expression (optional):

3 **Results**

- Your input histopathological image: [http://10.119.4.162/stomach\\_cancer/TCGA-BR-A452-DX1.svs](http://10.119.4.162/stomach_cancer/TCGA-BR-A452-DX1.svs).
- The size of the image is: 118048 X 89209 pixels
- Time used: 0.0000s

The patch with more information for survival prediction	DeepCox-SC risk score
	DeepCox-SC risk score of your input histopathological image is 3.08. This patient belongs to High-Risk group.

Grey line is the median of risk scores of stomach cancer patients in TCGA dataset  
 Red line is the risk score of your input histopathological image

**Online tool**
<https://yu.life.sjtu.edu.cn/DeepCoxSC/>

**FIGURE 1** Overview of DeepCox-SC workflow. (A) DeepCox-SC preprocessing. (B) The DeepCox-SC online tool (<https://yu.life.sjtu.edu.cn/DeepCoxSC>) takes whole-slide image as an input, and outputs the predicted risk score. Online tools include introduction page, analysis page, and results page. The user needs to input the histopathological image and email (age and gene expression are optional). An email containing the results page is sent to user when the job is finished

DeepCox-SC risk score is calculated by averaging the scores of all histopathological images of a patient that has more than one whole-slide image.

### 2.4.2 | Data augmentation

Data augmentation was adopted to prevent overfitting from insufficient training data and improve generalization performance. We applied random rotations by 40° and random horizontal and vertical flips to the training data. The pixel value of training and validation data was rescaled to 0–1.

### 2.4.3 | Hyperparameters

The Adam optimization algorithm was used to minimize the loss function (Equation 1). The initial learning rate was set to 0.002 and the decay was 0.05. Models were trained for 100 epochs using the mini-batch size of 32. To avoid overfitting, we adopted early stopping based on the performance of validation data. The training process would break when the validation loss did not reduce after 10 epochs.

### 2.4.4 | Hardware and software

Models were trained using TensorFlow (version 2.2.0) framework on a cluster (Pi 2.0 High Performance Computing of Shanghai Jiao Tong University) equipped with NVIDIA Tesla V100 32GB Graphics Card. HistomicsTK (<https://digitalslidearchive.github.io/HistomicsTK/>) was used for fundamental histopathological image analysis.

## 2.5 | Variable selection

### 2.5.1 | Genes

The number of genes is much larger than the sample size, so we conducted two-step variable selection (Figure S3). First, we reduced the high-dimensional gene expression data to moderate ( $n/\log(n)$ , where  $n$  is the number of patients) using the SIS variable selection method in all samples<sup>26</sup>. Second, the CPH model with the LASSO variable selection was conducted in each fold only using the training data. The Bayesian information criterion was used for LASSO penalty selection. The samples in each fold were the same as DeepCox-SC. The intersection of genes non-zero coefficients in

ten folds were finally added in the MultiDeepCox-SC multimodal fusion model. There were 10 genes (*CHAF1A*, *REPIN1*, *SERPINE1*, *HTRA3*, *PWP2*, *GPR173*, *NCLN*, *NT5E*, *MYL4*, and *YWHABP2*) associated with survival.

### 2.5.2 | Features of image patches

We used CellProfiler to extract features from each patch.<sup>27</sup> Using the “UnmixColor” module, each histopathological image was separated into a hematoxylin staining grayscale image and an eosin staining grayscale image. For each patch, 1100 features were obtained using “Measure Texture”, “Measure Object Intensity”, “Measure Object Size Shape”, “Measure Granularity”, and “Measure Object Intensity Distribution” modules. The mean and SD values of each feature were calculated for each patch. The variable selection of the features of image patches was similar to gene expression data by using SIS for preliminary screening and then LASSO in each fold.

## 2.6 | MultiDeepCox-SC model

The MultiDeepCox-SC integrates the DeepCox-SC risk score, age, and expression of 10 genes through the CPH model. The DeepCox-SC model was first trained using histopathological images, and then the output of this model (DeepCox-SC risk score) was combined with clinical data (age) and gene expression (10 genes) using the Cox model.

## 2.7 | Statistical analysis

Prediction accuracy was assessed using Harrell's C-index and the AUC of time-dependent receiver operating characteristic on the validation set. C-index (ranges from 0–1) was used to measure the concordance between the predicted risk and actual survival on validation sets. The higher C-index means more accurate prediction. The mean and SD of C-index and AUC values for 10 validation sets were calculated. The Wilcoxon signed rank test was used to compare the improvements of performance metrics (C-index and AUC value) among compared models. The univariate and multivariate Cox regression analyses were undertaken and the likelihood ratio test was used to calculate the  $p$  value of the multivariate Cox model. The Kaplan–Meier survival curves were plotted, and the survival differences between groups were compared by the log-rank test. All statistical analyses were performed in R, version 3.6.1. The  $p$  value < 0.05 was regarded as a significant level.

### 3 | RESULTS

#### 3.1 | Patient characteristics

Table 1 presents the clinical characteristics of 357 stomach cancer patients from the TCGA dataset and 30 stomach cancer patients in Ruijin Hospital (the external test set). The censoring rate is 60%

TABLE 1 Clinical characteristics of stomach cancer patients

Characteristics	TCGA cohort (357 samples)	Independent test set (30 samples)
Age (years)		
NA	3	
Mean $\pm$ SD	64.76 $\pm$ 10.40	65.40 $\pm$ 12.48
Gender		
Male	234	24
Female	123	6
7th TNM stage (AJCC)		
I	38	8
II	116	6
III	175	13
IV	21	1
NA	7	2
Grade		
I	7	4
II	125	11
III	216	14
X	9	1
Race		
White	251	/
Asian	81	/
Black or African American	12	/
Native Hawaiian or other Pacific islander	1	/
NA	37	/
Status		
Dead	142	14
Alive	215	16
Censoring rate	0.6	0.53
Survival time (days): mean $\pm$ SD		
Dead	430.26 $\pm$ 358.31	791.29 $\pm$ 635.80
Alive	743.15 $\pm$ 589.33	2257.7 $\pm$ 61.48

Abbreviations: AJCC, American Joint Committee on Cancer; NA, missing data; TCGA, The Cancer Genome Atlas.

(40% patients died) for the TCGA dataset and 53% (47% patients died) for the external test set.

#### 3.2 | Univariate and multivariate Cox regression

The output of the model was the predicted risk score (DeepCox-SC risk score) reflecting the risk of death. We applied a univariate and multivariate Cox regression model to evaluate the predictive power of the DeepCox-SC risk score. The DeepCox-SC risk score and clinical data (age, gender, race, stage, and grade) were coded as categorical variables (Figure 2A) and continuous variables (Figure 2B) for univariate analysis. We divided the patients into high- and low-risk groups according to the median value of DeepCox-SC risk scores.

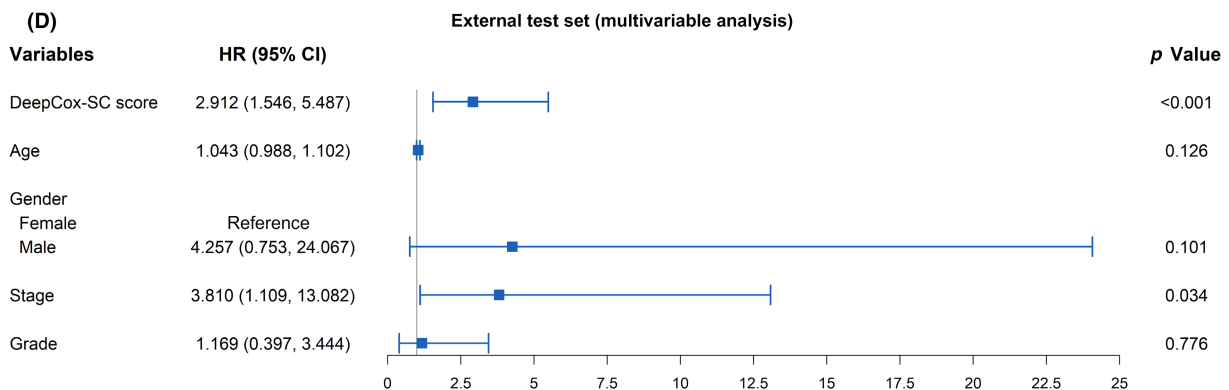
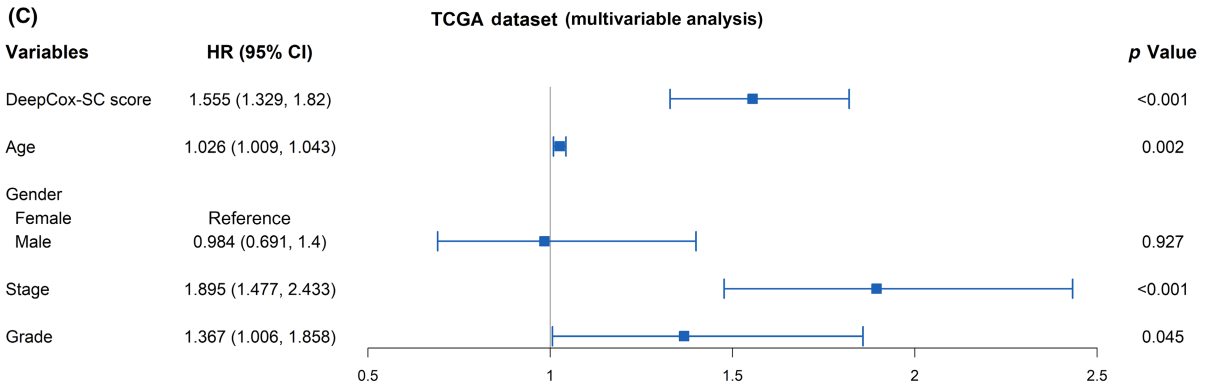
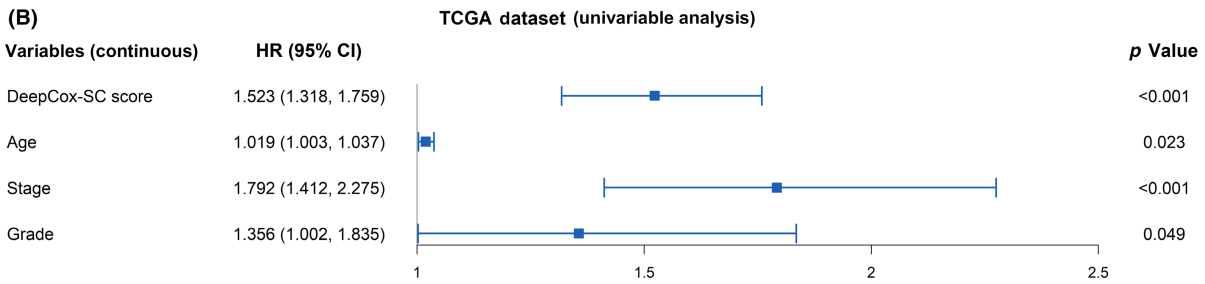
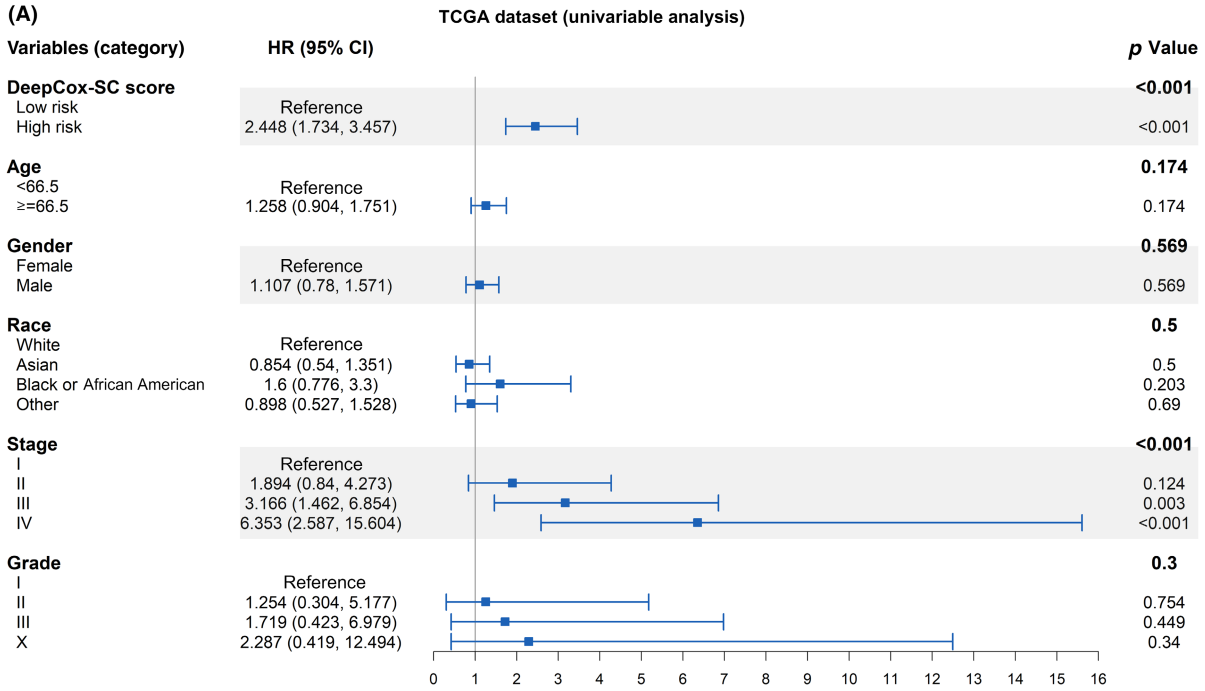
The stage and grade were prognostic factors in the clinical setting, and higher stage and grade indicated a higher risk. This was also demonstrated by HRs of pathologic grade and stage, where the higher stage and grade had the higher HR (Figure 2A). The HRs for stage II, III, and IV were 1.894, 3.166, and 6.353 (stage I as reference), respectively. The HR for grade II and grade III were 1.254 and 1.719 (grade I as reference), respectively. We coded stage and grade as continuous variables to consider this correlation for multivariate analysis (Figure 2B). The grade was a significant predictor only when it was coded as a continuous variable. The stages coding as either categorical or continuous variables were both significant predictors of OS. However, the *p* value was smaller for the continuous variables (likelihood ratio test, 2e-05 vs. 6e-07).

To evaluate the independent predictive power of the DeepCox-SC risk score, we performed a multivariate analysis including the DeepCox-SC risk score, age, gender, stage, and grade on the TCGA dataset (Figure 2C) and the external test set (Figure 2D). After adjustment for age, sex, stage, and grade, the DeepCox-SC risk score was still a significant predictor of OS in the multivariate model on the TCGA dataset (HR 1.555; 95% CI, 1.329–1.820; *p* = 3.53e-08) and the external test set (HR 2.912; 95% CI, 1.546–5.487; *p* = 9.42e-4). The higher DeepCox-SC risk score means the higher risk of death. This was also demonstrated by HR of DeepCox-SC risk score (Figure 2A,C,D).

#### 3.3 | Prediction accuracy of DeepCox-SC and MultiDeepCox-SC

We assessed the prediction accuracy by comparing our model (DeepCox-SC and MultiDeepCox-SC) with other methods. According to the input data, models can be divided into three categories: image-based, image and age-based, and image, age, and gene-based. Ten-fold cross-validation was used to compare the C-index, 1-year

FIGURE 2 Forest plot for univariate and multivariate Cox regression analysis of patients with stomach cancer. (A) Univariate Cox regression on The Cancer Genome Atlas (TCGA) dataset and variables are coded by category. (B) Univariate Cox regression on the TCGA dataset and variables are coded continuously. (C) Multivariate Cox regression on the TCGA dataset. (D) Multivariate Cox regression on the external test set. CI, confidence interval; HR, hazard ratio



AUC, 2-year AUC, and 3-year AUC of the different models on the TCGA dataset. We further assessed the model performance on the external test set.

### 3.3.1 | Image-based models

We assess the performance by comparing DeepCox-SC with other two models: randomly selecting the patch-based model and structured features-based model (Figure 3A, green violin plots). The randomly selecting patch-based model used randomly selected patches (512×512 pixels) from cropped images as the input of the deep neural Cox model, and model training was the same as the DeepCox-SC.<sup>11,13</sup> The DeepCox-SC model selected patches according to the cellularity and CellProfiler software (see Materials and Methods). The structured features-based model used CellProfiler to extract 1100 features from each patch, and then used SIS and LASSO for variable selection. The structured features-based model was fitted by the traditional CPH model (see Materials and Methods).<sup>27</sup>

For TCGA stomach cancer patients, the DeepCox-SC model predicted overall survival with C-index, 1-year AUC, 2-year AUC, and 3-year AUC of  $0.660 \pm 0.057$ ,  $0.701 \pm 0.146$ ,  $0.766 \pm 0.081$ , and  $0.699 \pm 0.123$ , respectively (mean±SD) (Figures 3A and S4, green violin plots). The DeepCox-SC risk score had the better performance than the randomly selecting patch-based model (C-index  $0.609 \pm 0.078$ ). These selected patches improved the performance by over 8.37% on average compared with the randomly selected patches (Wilcoxon signed rank  $p = 0.049$ ). This indicated that selected patches (according to the cellularity and CellProfiler software) were more predictive for OS than the randomly selected patches. The performance for the structured features-based model was worse (C-index  $0.601 \pm 0.065$ ). The C-index of the DeepCox-SC model was 0.657 on the external test set (Figure 3B). Figure 3C shows representative histopathological images of DeepCox-SC selected patches and the randomly selected patches of the DeepCox-SC high-risk group and DeepCox-SC low-risk group.

### 3.3.2 | Image and age-based model

According to the results of univariate and multivariate Cox regression models, age was a significant predictor of OS. As shown in the yellow violin plots of Figure 3A, we further integrated the DeepCox-SC risk score and clinical data (age). For comparison, we built the benchmark model based on manual pathologic grade, stage, and age using a linear Cox model. Integrating the DeepCox-SC risk score and age achieved a C-index of 0.667 on the TCGA dataset (Figure 3A, yellow violin plots) and 0.702 on the external test set (Figure 3B). The C-index of the three benchmark models were  $0.566 \pm 0.080$ ,  $0.634 \pm 0.095$ , and  $0.645 \pm 0.092$  for “grade + age” model, “stage + age” model, and “grade + stage + age” model, respectively. The C-index of “DeepCox-SC risk score + age” model (0.667) is higher than the clinical benchmark model integrating grade, stage, and age

(0.649), although the p-value between these two models was not significant (Wilcoxon signed rank  $p = 0.188$ ).

### 3.3.3 | Image, age, and gene-based model

The MultiDeepCox-SC integrated histopathological images, clinical data (age), and gene expression data (Figure 3A, blue violin plots). The preprocessing of high-dimensional gene expression data used SIS and LASSO for variable selection (see Materials and Methods). Finally, we obtained 10 genes (*CHAF1A*, *REPIN1*, *SERPINE1*, *HTRA3*, *PWP2*, *GPR173*, *NCLN*, *NT5E*, *MYL4*, and *YWHABP2*) associated with survival.

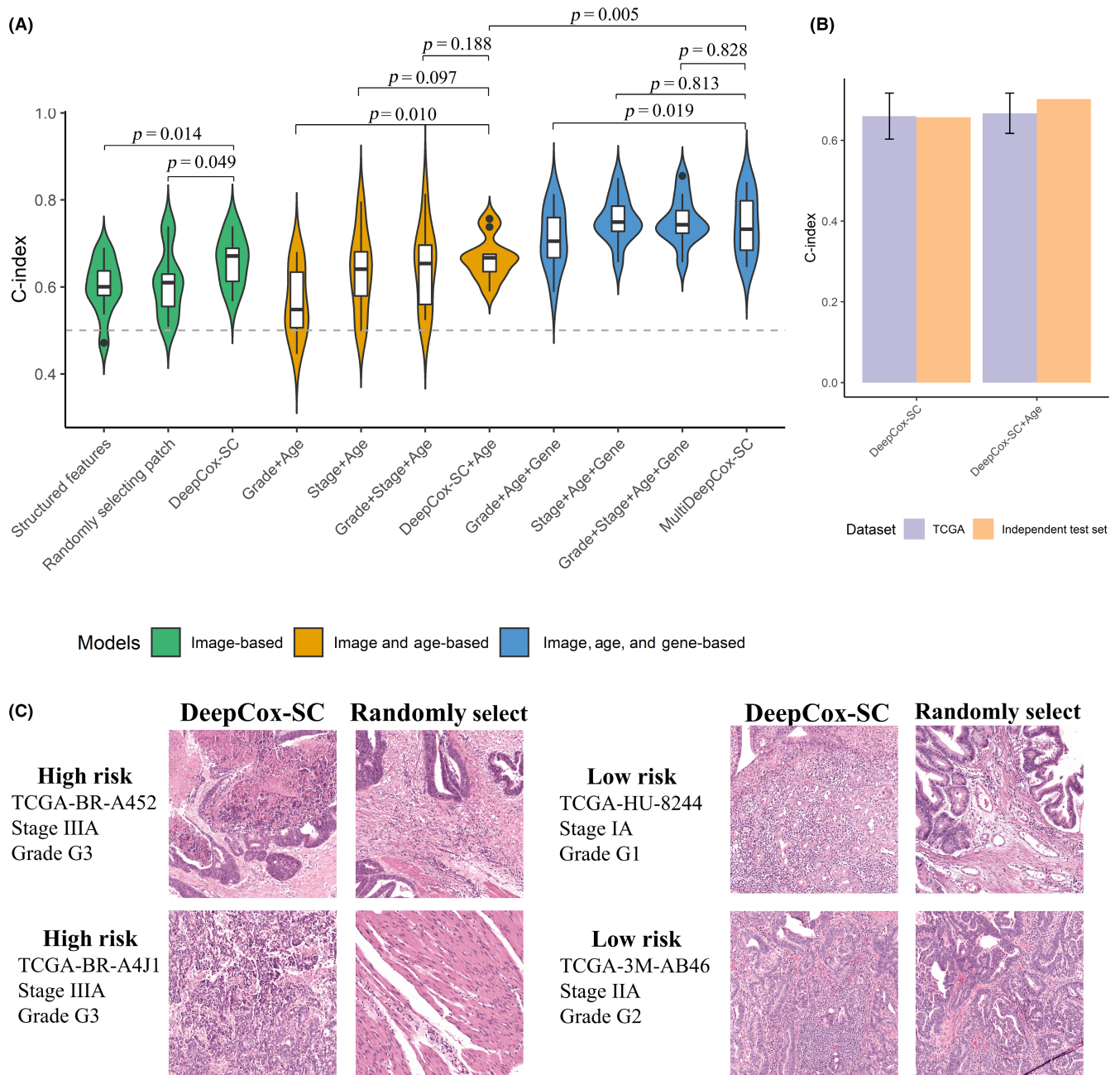
As showed in the blue violin plots of Figure 3A, the MultiDeepCox-SC integrating histopathological image, age, and gene expression data improved the C-index by over 13% on average (C-index  $0.744 \pm 0.070$ , Wilcoxon signed rank  $p = 0.005$ ) compared with the DeepCox-SC model (C-index  $0.660 \pm 0.057$ ). The C-index of the MultiDeepCox-SC was almost the same as the benchmark model integrating stage, grade, age, and gene expression (C-index  $0.751 \pm 0.055$ , Wilcoxon signed rank  $p = 0.652$ ). The 1-year AUC and 2-year AUC of the MultiDeepCox-SC multimodal fusion model were  $0.800 \pm 0.091$  and  $0.833 \pm 0.055$ , respectively.

Six of 10 genes used in the multimodal fusion model are relevant to survival. Histone chaperone *CHAF1A* promotes cell aggressive and inhibits apoptosis in many cancers.<sup>28</sup> Replication initiator *REPIN1* is associated with clinical outcome of survival.<sup>29</sup> *SERPINE1* is a cancer-promoting factor in stomach cancer.<sup>30</sup> *HTRA3* contributes to tumor metastasis.<sup>31</sup> *PWP2* is prognosis-related gene in stomach cancer.<sup>32</sup> *NT5E* is a correlative factor of patient survival in many kinds of cancers.<sup>33</sup> The other three genes have abnormal expression in tumor. *NCLN*, *MYL4*, and *GPR173* are overexpressed in human cancer.<sup>34–36</sup> *YWHABP2* is a tyrosine hydroxylase pseudogene, and tyrosine hydroxylase is essential for animal development and survival.<sup>37</sup>

## 3.4 | Subgroup analyses

There were 113, 163, 125, and 216 patients in stage II, stage III, grade II, and grade III, respectively. More than 100 patients were in the same group; however, the survival risk of these patients was different. It was meaningful to stratify these patients into subgroups with significant survival differences. According to the DeepCox-SC risk score median, 113 patients of stage II were stratified into two subgroups (high- and low-risk groups), showing statistically significant survival differences (Figure 4A). The Kaplan–Meier survival curves of the two subgroups were well separated, and the log-rank  $p$  values of the survival difference were  $5.66e-4$ ,  $1.97e-4$ ,  $8.79e-4$ , and  $4.48e-3$  for grade II, grade III, stage II, and stage III, respectively (Figure 4). There were small a number of patients in grade I (7 patients), stage I (38 patients), and stage IV (21 patients). The survival differences between two subgroups were 0.19, 0.14, and 0.21 for grade I, stage I, and stage IV, respectively (Figure S5).





**FIGURE 3** Comparison of C-indexes for different models in stomach cancer. (A) Comparison of the C-index of the different models using 10-fold cross-validation. Green violin plots represent image-based models: DeepCox-SC model, randomly selecting the patch from whole-slide image (randomly selecting patch-based model), and extracting the features of the image (structured features-based model). Our DeepCox-SC model outperformed the other two models. Yellow violin plots represent models integrating histopathological images and clinical data (age). Blue violin plots represent models integrating histopathological images, clinical data (age), and gene expression data. (B) C-indexes of different models on the external test set. (C) Representative histopathological images (patches). TCGA, The Cancer Genome Atlas

### 3.5 | Web server for survival prediction

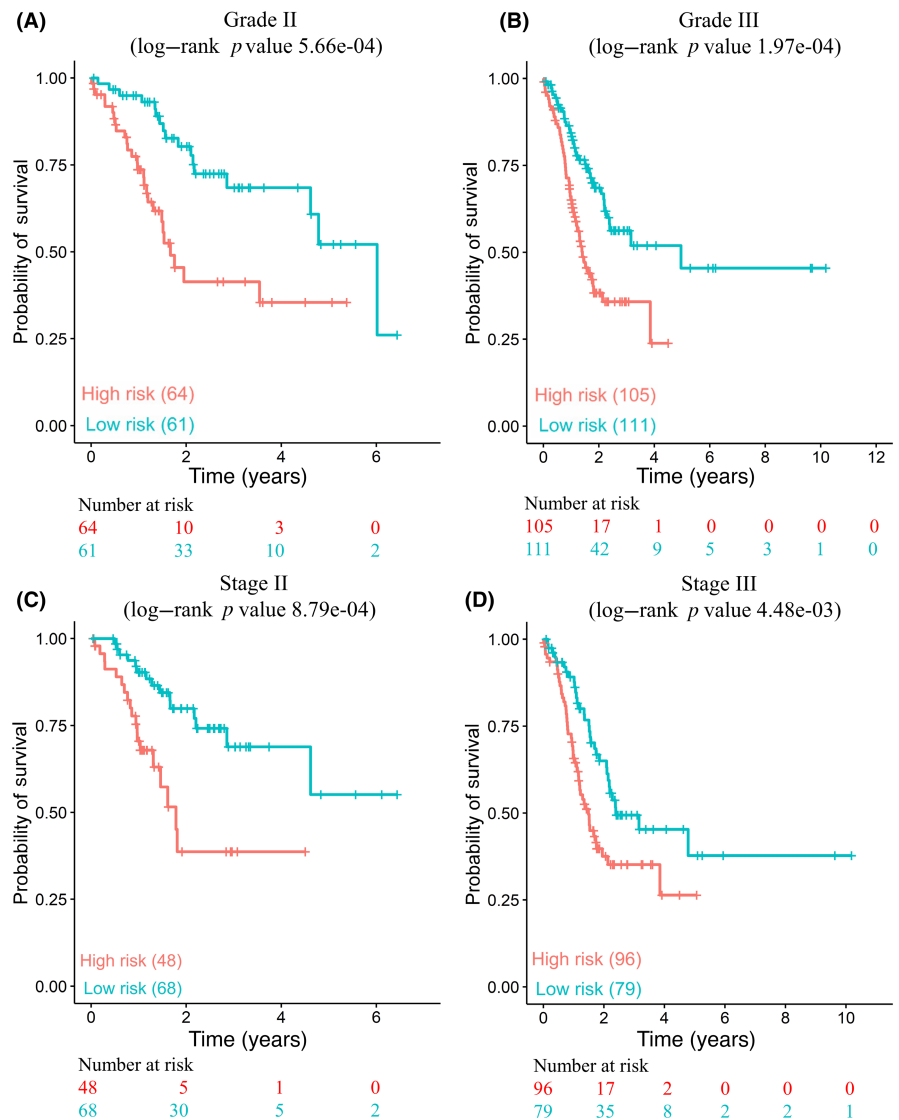
As shown in Figure 1B, we also developed a user-friendly online tool for the DeepCox-SC and MultiDeepCox-SC model (<https://yu.life.sjtu.edu.cn/DeepCoxSC>). This online tool needs the URL of the histopathological image and email as the input. The age and gene expression data of the patient are optional. The results including the predicted risk score, subgroups (high- and low-risk groups), and the patch with more information for survival prediction is automatically

displayed. An email containing the result link is sent to the user when the job is complete.

## 4 | DISCUSSION

This study devised a fully automated prognostic model for predicting survival outcome for patient with stomach cancer based on histopathological images, clinical data, and gene expression

**FIGURE 4** Survival differences between high- and low-risk groups of patients with stomach cancer. (A) Kaplan–Meier plots for the DeepCox-SC risk scores in patients with grade II disease. (B) Kaplan–Meier plots for patients with grade III disease. (C) Kaplan–Meier plots for patients with stage II disease. (D) Kaplan–Meier plots for patients with stage III disease



data. The prognostic accuracy of the proposed model surpassed the current clinical benchmark model based on pathologic grade, stage, and clinical data in stomach cancer. Our model could be utilized as a computer-assisted tool to improve pathologists' efficiency and accuracy and ultimately allow clinicians to select appropriate therapies.

We systematically examined the performance of the DeepCox-SC model in stomach cancer, one of the leading causes of cancer-related death worldwide.<sup>38</sup> The DeepCox-SC risk score remained an independent predictor after adjustment for all other variables, including pathologic grade, stage, age, race, and gender on the TCGA dataset (HR 1.555,  $p = 3.53\text{e-}08$ ) and the external dataset (HR 2.912,  $p = 9.42\text{e-}4$ ). The DeepCox-SC risk score might have the potential to complement the future staging system.

We used 10-fold cross-validation, a common solution effectively improving model robustness, to assess model performance. The DeepCox-SC model showed predictive power, achieving a mean C-index, 1-year AUC, 2-year AUC, and 3-year AUC of 0.660, 0.701, 0.766, and 0.699, respectively, on the TCGA dataset. The model performance was further validated on the external test set (C-index

0.657). According to the results of the multivariate Cox regression model, age was a significant predictor of survival. The C-index of the "DeepCox-SC risk score + age" model (0.667) is higher than the clinical benchmark model integrating grade, stage, and age (0.649), although the  $p$  value between these two models was not significant (Wilcoxon signed rank  $p = 0.188$ ).

Many studies have investigated possible gene biomarkers to determine prognosis and customize treatment in stomach cancer.<sup>39</sup> The MultiDeepCox-SC multimodal fusion model including DeepCox-SC risk score, age, and gene expression achieved a mean C-index, 1-year AUC, 2-year AUC, and 3-year AUC of 0.744, 0.800, 0.833, and 0.778, respectively. The MultiDeepCox-SC multimodal fusion model significantly outperformed the DeepCox-SC model (C-index 0.660, Wilcoxon signed rank  $p = 0.010$ ). The deep learning could discover additional information relevant to prognosis and consider these large numbers of features together.<sup>2,40</sup> In addition, the multimodal fusion model incorporating more information (histopathological image, clinical data, and gene expression data) significantly improves the model performance and could be an area of future research.<sup>13</sup>

Although we developed a fully automated assistance method to predict patient survival, it has some limitations. First, we selected patches using cellularity and CellProfiler software for training, but this is only a small fraction of a whole-slide image. We should incorporate more patches to better account for intratumoral heterogeneity. Second, we should input both high and low magnification images. Because cell shape is well captured in high-power field microscopic images, structural information made of many cells is better captured in lower-power field images.<sup>41,42</sup> Finally, although the stomach cancer patients in the TCGA dataset came from 22 centers, the robust performance of DeepCox-SC needs to be further tested in a large cohort.

In conclusion, we developed a fully automated deep CNN for survival prediction from histopathological images, clinical data, and gene expression data in stomach cancer. The DeepCox-SC model and MultiDeepCox-SC multimodal fusion model showed prognostic accuracy on the TCGA dataset and the external dataset. Our online tool (<https://yu.life.sjtu.edu.cn/DeepCoxSC>) could be used as an assisted tool to improve pathologists' efficiency and accuracy.

#### AUTHOR CONTRIBUTIONS

Conceptualization, ZY and TW; methodology, ZY, TW, RG, LJ, YW, JZ, and YX; software, TW; validation, ZY and TW; formal analysis, TW; investigation, TW; data curation, DX, XY, and TW; writing—original draft preparation, TW; writing—review and editing, ZY, RG, LJ, and TW; Online tool, TW and WK; supervision, ZY; project administration, ZY; funding acquisition, ZY and YZ. All authors have read and agreed to the published version of the manuscript.

#### ACKNOWLEDGMENTS

The work was funded by the National Natural Science Foundation of China 11671256 and Shanghai Jiao Tong University STAR Grant.

#### DATA AVAILABILITY STATEMENT

The histopathological images, clinical data, and expression data of stomach cancer were obtained from the TCGA dataset (<https://portal.gdc.cancer.gov/projects/TCGA-STAD>). The data are publicly available. The TCGA manifest file for download data is presented in Table S2.

#### DISCLOSURE

The authors have no conflict of interest.

#### ETHICAL APPROVAL

Approval of the research protocol by an institutional reviewer board: The external test set including 30 gastric cancer patients was previously collected in Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine. Ethics approval for this study was approved by the Ethics Committee of Ruijin Hospital (ID: 2021-194).

#### INFORMED CONSENT

This is a secondary data analysis. Informed consent is not applicable.

#### REGISTRY AND THE REGISTRATION NO. OF THE STUDY/TRIAL

N/A.

#### ANIMAL STUDIES

N/A.

#### WEB SERVER

The data that the user uploaded to the web server would be automatically deleted just 24 h after the analysis is completed. These data would not be used for other purposes.

#### ORCID

Zhangsheng Yu  <https://orcid.org/0000-0002-8189-5330>

#### REFERENCES

- Histopathology is ripe for automation. *Nat Biomed Eng.* 2017;1:925.
- Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol.* 2019;20:e253-e261.
- Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. *Med Image Anal.* 2016;33:170-175.
- Wang S, Yang DM, Rong R, et al. Artificial intelligence in lung cancer pathology image analysis. *Cancers (Basel).* 2019;11:1-16.
- Song Z, Zou S, Zhou W, et al. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. *Nat Commun.* 2020;11:4294.
- Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* 2019;25:1301-1309.
- Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol.* 2019;16:703-715.
- Yu KH, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun.* 2016;7:1-10.
- 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;18:1-12.
- Zhu X, Yao J, Huang J. Deep convolutional neural network for survival analysis with pathological images. *IEEE Int Conf Bioinform Biomed.* 2016;2017:544-547.
- Zhu X, Yao J, Zhu F, et al. WSISA: Making survival prediction from whole slide histopathological images. *Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017* 2017; 2017-Janua: 6855-6863.
- Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med.* 2019;16:1-22.
- Mobadersany P, Yousefi S, Amgad M, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A.* 2018;115:E2970-E2979.
- Hao J, Kosaraju SC, Tsaku NZ, et al. PAGE-net: interpretable and integrative deep learning for survival analysis using histopathological images and genomic data. *Pacific Symp Biocomput.* 2020;25:355-366.
- Jiang Y, Jin C, Yu H, et al. Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer. *Ann Surg.* 2021;274:e1153-e1161.

16. Calderaro J, Kather JN. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. *Gut*. 2021;70:1183-1193.
17. Sano T, Kodera Y. Japanese gastric cancer treatment guidelines 2010 (ver. 3). *Gastric Cancer*. 2011;14:113-123.
18. Qiu MZ, Wang ZQ, Zhang DS, et al. Comparison of 6th and 7th AJCC TNM staging classification for carcinoma of the stomach in China. *Ann Surg Oncol*. 2011;18:1869-1876.
19. Fang WL, Huang KH, Chen JH, et al. Comparison of the survival difference between AJCC 6th and 7th editions for gastric cancer patients. *World J Surg*. 2011;35:2723-2729.
20. Nagtegaal ID, Odze RD, Klimstra D, et al. The 2019 WHO classification of tumours of the digestive system. *Histopathology*. 2020;76:182-188.
21. Veta M, Pluim JPW, Van Diest PJ, et al. Breast cancer histopathology image analysis: a review. *IEEE Trans Biomed Eng*. 2014;61:1400-1411.
22. Sun C, Xu A, Liu D, Xiong Z, Zhao F, Ding W. Deep learning-based classification of liver cancer histopathology images using only global labels. *IEEE J Biomed Heal Informatics*. 2020;24:1643-1651.
23. Macenko M, Niethammer M, Marron JS, et al. A method for normalizing histology slides for quantitative analysis. *IEEE Int Symp Biomed Imaging*. 2009;1107-1110.
24. Ahmady Phoulady H, Goldgof DB, Hall LO, et al. Nucleus segmentation in histology images with hierarchical multilevel thresholding. *Med Imaging 2016 Digit Pathol*. 2016;9791:979111.
25. Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning. *31st AAAI Conf Artif Intell AAAI 2017*; 2017: 4278-4284.
26. Bickel P, Buhlmann P, Yao Q, et al. Discussion on 'sure independence screening for ultra-high dimensional feature space' by Fan, J and Lv, J. *J R Stat Soc Ser B Methodol*. 2008;70:849-911.
27. Soliman K. CellProfiler: novel automated image segmentation procedure for super-resolution microscopy. *Biol Proced Online*. 2015;17:1-7.
28. Barbieri E, De Preter K, Capasso M, et al. Histone chaperone CHAF1A inhibits differentiation and promotes aggressive neuroblastoma. *Cancer Res*. 2014;74:765-774.
29. Qi T, Qu J, Tu C, et al. Super-enhancer associated five-gene risk score model predicts overall survival in multiple myeloma patients. *Front Cell Dev Biol*. 2020;8:1-12.
30. Yang J-D, Ma L, Zhu Z. SERPINE1 as a cancer-promoting gene in gastric adenocarcinoma: facilitates tumour cell proliferation, migration, and invasion by regulating EMT. *J Chemother*. 2019;31:408-418.
31. Zhao J, Feng M, Liu D, et al. Antagonism between HTRA3 and TGFβ1 contributes to metastasis in non-small cell lung cancer. *Cancer Res*. 2019;79:2853-2864.
32. Zhou W, Li J, Lu X, et al. Derivation and validation of a prognostic model for cancer dependency genes based on CRISPR-Cas9 in gastric adenocarcinoma. *Front Oncol*. 2021;11:617289.
33. Wang J, Matosevic S. NT5E/CD73 as correlative factor of patient survival and natural killer cell infiltration in glioblastoma. *J Clin Med*. 2019;8:1526.
34. Moghimi M, Bakhtiari R, Mehrabadi JF, et al. Interaction of human oral cancer and the expression of virulence genes of dental pathogenic bacteria. *Microb Pathog*. 2020;149:104464.
35. Eichenmüller M, Bauer R, Von Schweinitz D, et al. Hedgehog-independent overexpression of transforming growth factor-beta1 in rhabdomyosarcoma of Patched1 mutant mice. *Int J Oncol*. 2007;31:405-412.
36. Orentas RJ, Sindiri S, Duris C, et al. Paired expression analysis of tumor cell surface antigens. *Front Oncol*. 2017;7:173.
37. Gomez-Flores R, Gutierrez-Leal I, Caballero-Hernández D, et al. Association of tyrosine hydroxylase expression in brain and tumor with increased tumor growth in sympathectomized mice. *BMC Res Notes*. 2021;14:94.
38. Van Cutsem E, Sagaert X, Topal B, et al. Gastric cancer. *Lancet*. 2016;388:2654-2664.
39. Machlowska J, Baj J, Sitarz M, Maciejewski R, Sitarz R. Gastric cancer: epidemiology, risk factors, classification, genomic characteristics and treatment strategies. *Int J Mol Sci*. 2020;21:4012.
40. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18:500-510.
41. Kosaraju SC, Hao J, Koh HM, Kang M. Deep-Hipo: multi-scale receptive field deep learning for histopathological image analysis. *Methods*. 2020;179:3-13.
42. Skrede OJ, De Raedt S, Kleppe A, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet*. 2020;395:350-360.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Wei T, Yuan X, Gao R, et al. Survival prediction of stomach cancer using expression data and deep learning models with histopathological images. *Cancer Sci*. 2023;114:690-701. doi: [10.1111/cas.15592](https://doi.org/10.1111/cas.15592)