

RESEARCH

Open Access



Denoised recurrence label-based deep learning for prediction of postoperative recurrence risk and sorafenib response in HCC

Yixin Li^{1†}, Ji Xiong^{4†}, Zhiqiu Hu⁶, Qimeng Chang⁶, Ning Ren^{3,5*}, Fan Zhong^{2*}, Qiongzhu Dong^{3*} and Lei Liu^{2*}

Abstract

Background Pathological images of hepatocellular carcinoma (HCC) contain abundant tumor information that can be used to stratify patients. However, the links between histology images and the treatment response have not been fully unveiled.

Methods We trained and evaluated a model by predicting the prognosis of 287 non-treated HCC patients post-operatively, and further explored the model's treatment response predictive ability in 79 sorafenib-treated patients. Based on prognostic relevant pathological signatures (PPS) extracted from CNN-SASM, which was trained by denoised recurrence label (DRL) under different thresholds, the PPS-based prognostic model was formulated. A total of 78 HCC patients from TCGA-LIHC were used for the external validation.

Results We proposed the CNN-SASM based on tumor pathology and extracted PPS. Survival analysis revealed that the PPS-based prognostic model yielded the AUROC of 0.818 and 0.811 for predicting recurrence at 1 and 2 years after surgery, with an external validation reaching 0.713 and 0.707. Furthermore, the predictive ability of the PPS-based prognostic model was superior to clinical risk indicators, and it could stratify patients with significantly different prognoses. Importantly, our model can also stratify sorafenib-treated patients into two groups associated with significantly different survival situations, which could effectively predict survival benefits from sorafenib.

Conclusions Our prognostic model based on pathology deep learning provided a valuable means for predicting HCC patient recurrence condition, and it could also improve patient stratification to sorafenib treatment, which help clinical decision-making in HCC.

Keywords Hepatocellular carcinoma, Prognosis, Sorafenib responses, Tumor pathology, Deep learning, Denoised recurrence labels

[†]Yixin Li and Ji Xiong contributed equally to this work.

*Correspondence:

Ning Ren

ren.ning@zs-hospital.sh.cn

Fan Zhong

zonefan@163.com

Qiongzhu Dong

qzhdong@fudan.edu.cn

Lei Liu

liulei_sibs@163.com

¹ Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

² Intelligent Medicine Institute, Fudan University, No.131 DongAn Road, Xuhui, Shanghai 200032, China

³ Key Laboratory of Whole-Period Monitoring and Precise Intervention of Digestive Cancer, Shanghai Municipal Health Commission, Minhang Hospital, Fudan University, No. 170 XinSong Road, Minhang, Shanghai 201199, China

⁴ Department of Pathology, Huashan Hospital, Fudan University, Shanghai 200040, China

⁵ Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, Fudan University, No. 180 FengLin Road, Xuhui, Shanghai 200032, China

⁶ Department of Hepatobiliary and Pancreatic Surgery, Minhang Hospital, Fudan University, Shanghai 201199, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Hepatocellular carcinoma (HCC) accounts for 75–85% of primary liver cancers with high incidence and mortality, and the 5-year survival rates of HCC patients with hepatectomy are only 21% [1]. Sorafenib is the first FDA-approved first-line targeted therapy for advanced HCC [2] due to its good efficacy and manageable safety [3]. However, only a fraction of patients derive long-term benefits [4, 5]. More recently, the FDA approved many molecular targeted therapies [6, 9] and immune checkpoint inhibitors (ICIs) [10] for the treatment of advanced HCC, but the efficacy of these therapies in HCC is still limited [11]. A standardized risk-stratification system for HCC patients is essential to improve the benefits of adjuvant systemic therapies after curative resection/ablation. Thus, it is urgent to identify biomarkers that predict response to therapy or individuals most likely to develop resistance.

Clinical pathological analysis is crucial and essential for diagnosing and stratifying cancer patients [12, 13]. Pathological images contain numerous phenotypic descriptions and pathological patterns that reflect the comprehensive evaluation of tumor diagnosis, prognosis, and clinical management [14–17]. With the advent of whole-slide imaging (WSI) [18] in digital pathology, deep learning-based pathological analysis exhibits the capacity to dynamically extract and decipher intricate tumor image features [19, 20], thereby presenting an approach for automated diagnostic procedures and prognostic modeling in patient care. Accumulating studies have demonstrated that pathological images can be used for the detection of early-stage [21] and microvascular invasion [22] of HCC based on convolutional neural networks (CNNs). Risk stratification to predict the survival of HCC based on pathological images has been reported recently [23, 24]. However, histopathological images that contain fundamental information on phenotype underlying HCC treatment response and outcomes remain largely unknown. Therefore, it is urgent to translate the histopathological features of HCC into predictive algorithms to inform clinical management and personalized cancer therapy.

Compared to the diagnosis tasks, the performance of deep learning on prognosis tasks is undesirable. Therefore, little research focuses on further investigation of the response to a specific treatment. Besides, it is more challenging to evaluate prognosis based on unreliable labels in the real world because patients' outcomes encompass numerous variables beyond the disease itself. This poses challenges for executing end-to-end prognostic prediction utilizing pathological images. To address this issue, specific research indirectly forecasts outcomes by utilizing intermediate

understandable pathological markers like tumor mutation burden (TMB) [25] and microvascular invasion [26]. In our study, we proposed to denoise the recurrence labels for the first time and use them to train the pathological model, which was more effective in extracting pathological features and predicting the prognosis. In constructing the pathological model, different from the conventional attention and self-attention mechanisms, we considered the mutual correlation between channels or spatial locations of the input data by the cross-self-attention module. It could achieve favorable classification ability by limited parameters.

In this work, we developed a CNN-SASM architecture by integrating CNN with channel and spatial self-attention mechanism modules to extract prognostic relevant pathological signatures (PPS), training by denoised recurrence label (DRL). PPS turned out to be associated with specific clinical histological regions. Furthermore, we constructed a PPS-based prognostic model to explore its crucial role in predicting prognosis of HCC patients compared to existing clinical risk indicators, and then explored the prognostic model's survival benefits predictive capability to sorafenib.

Methods

Patients and follow-up

A total of 372 patients with 744 pathological slides who underwent curative resection for HCC at Zhongshan Hospital, Fudan University (Shanghai, China) were enrolled in the present study. Among them, 292 patients in the non-treatment group did not receive postoperative medication and 80 patients in the sorafenib group were treated with sorafenib postoperatively. Pathological tissue specimens were sampled during the surgery twice per patient. During the follow-up, patients were monitored every 2–3 months after surgery including serum α -fetoprotein (AFP), ultrasonography, and chest X-ray. If tumor recurrence was suspected, computed tomography (CT) scanning or magnetic resonance imaging (MRI) was conducted. The experiment was carried out in line with the principles of the Declaration of Helsinki. The study was reviewed and approved by the medical ethics committee of Zhongshan Hospital, Fudan University, with the approval number B2021-143R. We further excluded patients who followed up less than 2 years. The exclusion criteria of pathological images were (1) loss of tissue structure and (2) an extensive area of out-of-focus. Ultimately, 366 patients with 724 pathological images were included for the subsequent analysis after the screening. The same preprocessing was conducted on the TCGA-LIHC external validation set [27], and 78 patients were incorporated.

Data preprocessing

Limited by the computer's memory and the large size of a single pathological image, we cut all pathological images into 224×224 sizes without overlap in the tissue area. After that, patches with more than 50% background area were excluded, and the remaining patches were stain normalized by Vahadane's approach [28], resulting in an average of 65 patches per image. Data augmentation was also carried out, including random flipping, rotating, adding Gaussian noise, and randomly erasing with a certain probability. Images from TCGA were in WSI format, so we processed them with the Python DeepZooming package, and patches were obtained by undersampling to the same number as the training set from all patches to identify the same distribution. Other preprocessing operations were similar to the internal dataset.

Denoising recurrence label

The real-world prognostic labels are less reliable than diagnostic labels because of the complicated non-disease factors, such as surgical outcomes, postoperative medication types and cycles, postoperative care, and physical or psychological conditions, which make it challenging to predict recurrence directly. To settle this issue, we proposed using DRL, recurrence labels with confidence levels higher than the threshold, instead of raw recurrence labels without filtering to train a pathological model, specifically the CNN-SASM mentioned below. DRL could be further categorized into denoised recurrence (DR) and denoised recurrence-free (DRF). The agreement between patients' recurrence status and overall survival (OS) time determined the confidence of DRL, and we then divided the training set into various subsets according to various confidence levels. For instance, the 1/4 subset consisted of non-recurrence and recurrence patients with OS in the top and bottom quartile, respectively. With the increase in data sizes, DRL's confidence worsened, and we iterated this sampling operation until all training data was included. The subsequent formula defines the subset split criteria:

$$\text{Subset}_t = \text{Patients} \quad \text{which} \begin{cases} OS > OS_t' & \text{and label} = \text{Relapse-Free} \\ OS < OS_t' & \text{and label} = \text{Relapse} \end{cases}$$

where Patients_{OS} represents patients with overall survival of OS, t represents the t th subset of training set, and OS_t and OS_t' are the t -quantiles of OS for relapse-free and relapse patients, respectively. We obtained 6 subsets, including 1/4, 1/3, 1/2, 2/3, 3/4, and 1/1 ratios (Table S1). The external test set of this part was comprised of 14 patients (7 cases in each category) from TCGA following a similar principle.

Model design and construction

Construction of CNN-SASM

The CNN-SASM framework depicted in Fig. 1C chiefly consists of three components: the convolution (Conv) module, the channel self-attention (CSA) and spatial self-attention (SSA) modules, and the classification (Cls) module.

Conv module: Considering the limited data size, we employed ResNet18 pre-trained on ImageNet, except for the last two layers, as the foundational architecture for the convolution module by transfer learning. Assuming the input data $X_{\text{input}} \in \mathbb{R}^{B \times 3 \times H \times W}$, it could produce the output $X_{\text{output}} \in \mathbb{R}^{B \times C \times H' \times W'}$ after one iteration. Here, B refers to the batch size; three corresponds to the RGB color channels; H and W represent the height and width of the input data, while H' and W' correspond to the output feature map. C describes the number of output channels.

CSA and SSA modules: We introduced cross multi-head self-attention mechanisms in both channel and spatial dimensions to capture interactions across channels and spatial locations, followed by the softmax layer to allocate more weight to vital channels and spatial locations. In the CSA computation, the number of channels corresponded to the number of tokens in sequence models. Specifically, the feature map $Fea_i \in \mathbb{R}^{H' \times W'}$ was reshaped into a vector $Vec_i \in \mathbb{R}^{H' \times W'}$ in the i th channel, which acted as the embedding representation for that particular channel. The vectors q_m , k_m , and v_m were derived from the linear projection of Vec_i and could be decomposed into multiple single attention head vectors q_s^n , k_s^n , and v_s^n , where m stands for multi-heads, s denotes the single-attention head, and n represents the n th attention head. The self-attention scores, termed attention-score (AS), and corresponding weighted vectors were independently calculated for each single-attention head. Then, the final output was formed by concatenating the weighted results from all attention heads.

$$q_m = w_q \times Vec_i + b_q = \text{Concat}(q_s^n)$$

$$k_m = w_k \times Vec_i + b_k = \text{Concat}(k_s^n)$$

$$v_m = w_v \times Vec_i + b_v = \text{Concat}(v_s^n)$$

$$AS^n = \text{Dropout}\left(\text{SoftMax}\left(\frac{q_s^n \times (k_s^n)^T}{\sqrt{d_s}}\right)\right)$$

$$CSA(X_{\text{Conv}}) = \text{Dropout}(W \times \text{Concat}(AS^n \times v_s^n) + b)$$

$$X_{\text{CSA}} = \text{SoftMax}(\text{LayerNorm}(X_{\text{Conv}} + CSA(X_{\text{Conv}})))$$

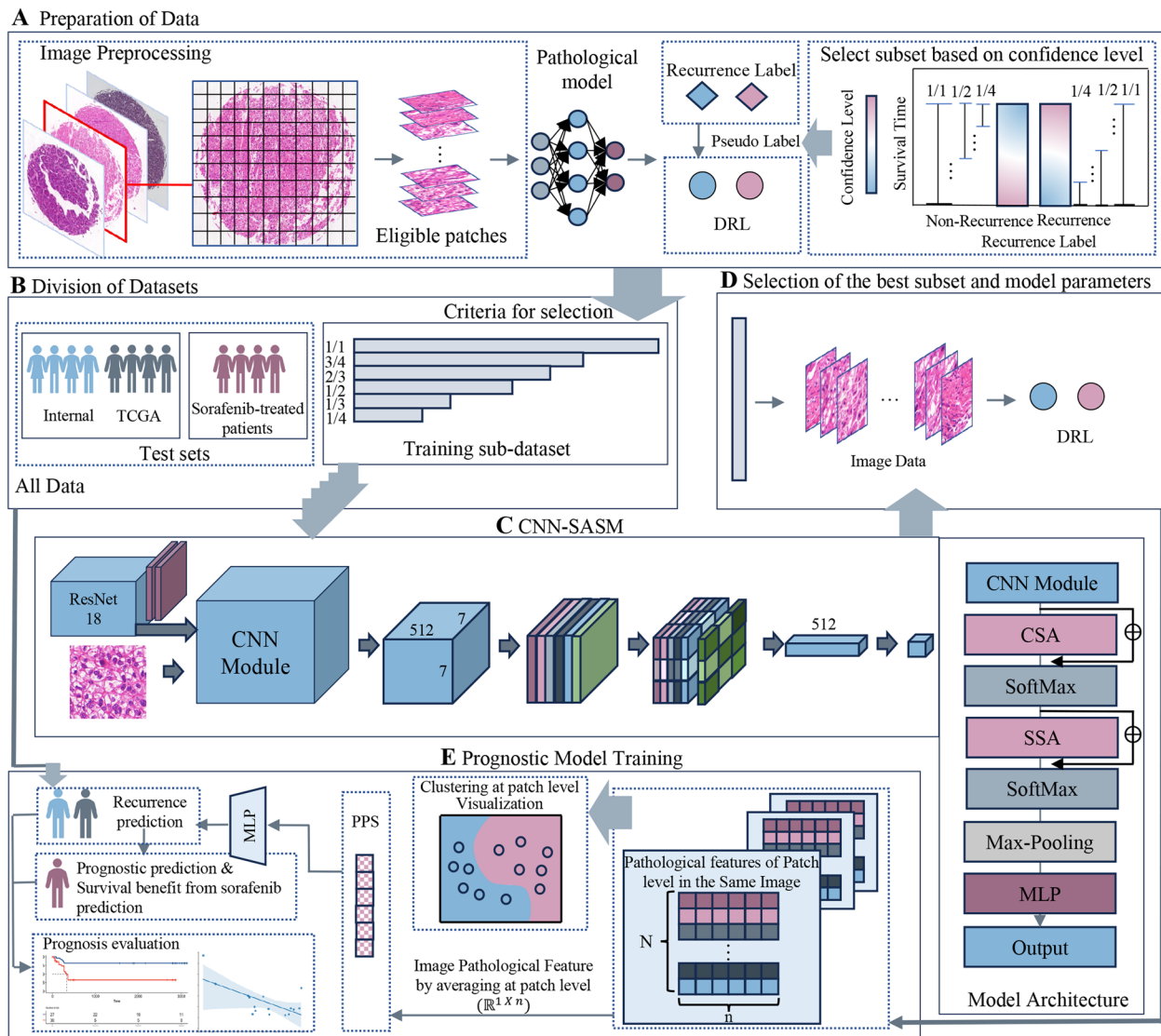


Fig. 1 Workflow diagram of the study. **A** Process of the pre-processing of pathological images and the criteria for subset selection based on confidence level. Patches were deemed non-qualified if they had an excessive background or lacked cellular structures. The pathological model was trained using DRL for labeling, with recurrence status as a pseudo-label, and the confidence was determined by the consistency between the patient's recurrence status and overall survival. **B** Dataset structure. All subsets were selected from the training set, and the test set contained three sections: non-treatment internal test set, non-treatment TCGA, and sorafenib-treated patients in the test set, respectively. **C** The framework of the CNN-SASM detailed the data's forward propagation and the model's architecture. **D** Selection of the best subset with the corresponding optimum parameters CNN-SASM. **E** Training process of the prognostic model and the corresponding prognostic tasks

which $n = [n_1, n_2, n_3, \dots, n_{num_heads}]$. The softmax layer was incorporated, followed by self-attention, allowing the model to distinguish the significant features across the sequence dimension and normalize the trainable parameters. The design of SSA mirrored CSA but differed in that it treated spatial locations as tokens and recognized channel features as embeddings.

$$X_{SSA} = \text{SoftMax}(\text{LayerNorm}(\text{Output}_{CSA} + \text{SSA}(\text{Input})))$$

CIs module: This module consisted of a max-pooling layer and a multi-layer perceptron (MLP). Their primary function was to extract feature vectors and then execute the final classification. Max-pooling was performed on each channel dimension so that the feature dimension changed from $\mathbb{R}^{B \times C \times H \times W}$ to $\mathbb{R}^{B \times C}$. The MLP was the combination of two fully connected layers incorporating activation functions. Finally, the embedding features of patches were encoded into 128-dimensional patch-level

feature vectors and binary classification outputs after Cls, and the former was PPS_{patch} .

$$PPS_{patch} = \text{Maxpooling}(\text{Output}_{SSA})$$

$$DRS_{patch} = \text{Mlp}(PPS_{patch})$$

During the training phase, we employed the cross-entropy loss function with a learning rate of 0.0001, the batch size was set to 128, and the training encompassed 50 epochs. Early termination occurs if the loss decreases for 5 consecutive epochs. The goal of CNN-SASM convergence was it could distinguish DR and DRF.

The aggregation of PPS_{image} : After being well-trained on the subset, CNN-SASM could be applied to inference, and we could extract the PPS_{patch} directly from it. The image-level PPS was derived from averaging the corresponding PPS_{patch} belong to the same image.

$$PPS_{image}^t = \frac{1}{N} \sum_{i=1}^N PPS_{patch}^i$$

where N denotes the patch number of the t th image.

We took the CNN architecture ResNet18 and DenseNet121 as the pathological baseline models, along with the current SOTA model, Swin Transformer, based on self-attention mechanisms. All model parameters were initialized with pre-trained weights via transfer learning.

Construction of PPS-based prognostic model

The prognostic model was designed to evaluate the recurrence risk of postoperative HCC patients based on PPS extracted from CNN-SASM. We applied a MLP similar to the Cls in CNN-SASM to generate the prediction. Specifically, the input was the $PPS_{image} \in \mathbb{R}^{1 \times 128}$, and the output was binary classification nodes with a dimension $\mathbb{R}^{1 \times 2}$, where the first value represented the probability of non-recurrence risk for the image, and the second value denoted the probability of recurrence risk, with the sum of the two equal to one. The recurrence risk probabilities would be averaged if the patient had more than one pathological image. At this stage, the learning rate was fixed at 0.001, with all other parameters maintained consistently with the CNN-SASM training process.

Software and statistical analysis

This study utilized Python 3.9.12 and Pytorch 2.0.1. It evaluated the performance of CNN-SASM and the prognostic model primarily using AUROC. Patch-level clustering visualization was achieved through PCA and t -SNE. The significance tests were conducted by independent t -tests if two groups of data both conformed to the normal distribution or else we applied Wilcoxon

rank-sum tests, which were both used in our study. The correlation test used the Pearson correlation coefficient. Significance scoring in survival analysis was conducted using the log-rank test, and prognostic ability was quantified through the concordance index (C-index) and hazard ratio (HR). All significance thresholds were set to 5%.

Results

Study design and patient characteristics

The overall study design is shown in Fig. 1. We first trained and validated the pathology model CNN-SASM by a series of pathological image subsets derived from the training set, then selected the optimal CNN-SASM in both performance and generalizability from all subset models as the PPS extractor. The rationale for splitting training subsets is that we believe complicated factors beyond the disease lead to the recurrence label in the real world, so we need to denoise the recurrence label and make it more associated with HCC itself; the new label was named DRL. We finally constructed a simple MLP using PPS to predict HCC patients' postoperative recurrence at different points of time and survival benefits from sorafenib treatment.

A total of 366 patients met the inclusion criterion from 372 recruited patients and were divided into training and test sets in a ratio of 7:3. The detailed clinicopathological characteristics are listed in Table 1. Among these, 79 patients were treated with sorafenib postoperatively (Additional file 1: Fig. S1). The detailed dataset structure can be found in Additional file 1: Fig. S2. A total of 78 patients from the TCGA-LIHC project were applied to test the generalizability of the prognostic model.

CNN-SASM accurately captures PPS of patches

We trained CNN-SASM on each subset to select the optimal subset with the best model parameters. CNN-SASM's classification performance in DRL declined in both patch and TMA core levels as the samples of subsets increased (Fig. 2B), and a similar trend was also observed for accuracy (Additional file 1: Fig. S4B), proving effective in denoising recurrence labels. To trade off the classification performance and the sample size, we carried out generalizability tests in a TCGA subset with strictly DRL, containing 910 patches. The result suggested that the generalizability increased steadily until the 1/2 subset, then fluctuated, which might be due to the instability of the added samples. Therefore, we selected the CNN-SASM trained on the 1/2 subset as the PPS extractor, which achieved an AUROC of 0.912 and an AUPR of 0.921 at the patch level on the corresponding test set, with an accuracy of 0.830 computing from the confusion matrix (Fig. 2D), indicating CNN-SASM captured prognostic relevant signatures from pathological images.

Table 1 Clinical characteristics of the subjects

Variable	Training set		Test set	
	Number (n = 293)	Frequency (%)	Number (n = 73)	Frequency (%)
Age				
≤ 50	139	47.4	38	52.1
> 50	154	52.5	35	47.9
Gender				
Male	248	84.6	62	84.9
Female	45	15.4	11	15.1
HBsAg				
+	268	91.5	67	91.8
–	25	8.5	6	8.2
AFP				
+	201	68.6	49	67.1
–	92	31.4	24	32.9
Tumor diameter				
≤ 5 cm	202	68.9	54	74.0
> 5 cm	91	31.1	19	26.0
BCLC stage				
I	124	42.3	29	39.7
II	169	57.7	44	60.3
Vascular invasion				
+	108	36.9	31	42.5
–	185	63.1	42	57.5

To visualize the representation ability of PPS qualitatively, we further displayed the PCA and *t*-SNE results in the internal test set and TCGA subset (Fig. 2G, H). DRF and DR patches could be separated well according to PPS despite slight mixing existing and that might be caused by noised labels. In this way, clustering results on the rest subsets could also be conducted (Additional file 1: Figs. S6 and S7), from clear boundaries to blending, they followed the same trend as AUROC and accuracy.

ResNet18 [29], DenseNet121 [30], and Swin-Transformer [31] represent classical model architectures within the domain of deep learning, frequently employed for image-related tasks. We found that our proposed pathological model, CNN-SASM, was superior to ResNet18, DenseNet121, and even Swin-Transformer in classifying DRF and DR on 1/2 subset and average performance with just little more trainable parameters than ResNet18 (Fig. 2A).

To further explore the interpretability of the CSA and SSA modules, we employed Grad-CAM to visualize their respective attention regions (Additional file 1: Fig. S5). As expected, the CSA module assigned varying attention weights to different channels, with channel 10 and 257 receiving the highest attention,

indicating their importance for the model, while channel 39 and 154 had minimal impact on the prediction results. Similarly, the SSA module generated distinct attention weights across spatial locations of the image. The activation map of the DR image predicted as recurrence exhibited more prominent activation, whereas the DRF image showed the opposite pattern. By assigning higher weights to the most important channels and spatial regions, the CSA and SSA modules enable CNN-SASM to achieve substantial prediction performance.

Additionally, ablation experiments were performed to ascertain the efficacy of the CSA, SSA modules, and the softmax layer in CNN-SASM quantitatively. Eliminating either the CSA or SSA module resulted in a decline in model performance. Among them, the CSA module played a more significant role in CNN-SASM. Specifically, after removing the CSA module, the model's AUROC on 1/2 subset dropped from 0.951 to 0.932 (Table 2). The softmax layer could function analogously as the conventional attention layer within limited samples (Table 2, Additional file 1: Fig. S4A).

Correlation between the PPS and known pathologic characteristics

In order to have a better understanding of PPS in morphological and biological interpretation, we visualized the attention map of the convolutional layer using the Grad-CAM method. CNN-SASM concentrated on the cell nucleus for both DR and DRF patches. It was correlated with clinical benchmarks for tumor assessment, demonstrating the significance and reliability of CNN-SASM in distinguishing DR and DRF patches (Fig. 3A). Regarding patch morphology, tumor cell nuclei on DRF patches were smaller with noticeable glandular structures. In contrast, nuclei on DR patches were larger with visible nucleoli. The biological significance of these morphological differences in prognosis needs to be further investigated.

Additionally, we displayed heatmaps of different denoised recurrence risk levels using the output of CNN-SASM to explore the connections between PPS and biological tissues. Compared to the non-recurrence WSI, there was a larger area of high risk in the recrudescence WSI, and almost all high-risk regions under the extreme threshold were distributed along the tumor edges or junctions. In contrast, low-risk regions resembled lymphocyte-rich regions (Fig. 3B, C). These suggested that the interface between tumor tissue, adjacent tissues, and lymphoid tissue regions were potential prognostic indicators. Further magnified the tissues, the high-risk tissues contained dispersed lymphocytes among collagenous fibrous tissue and remnants of

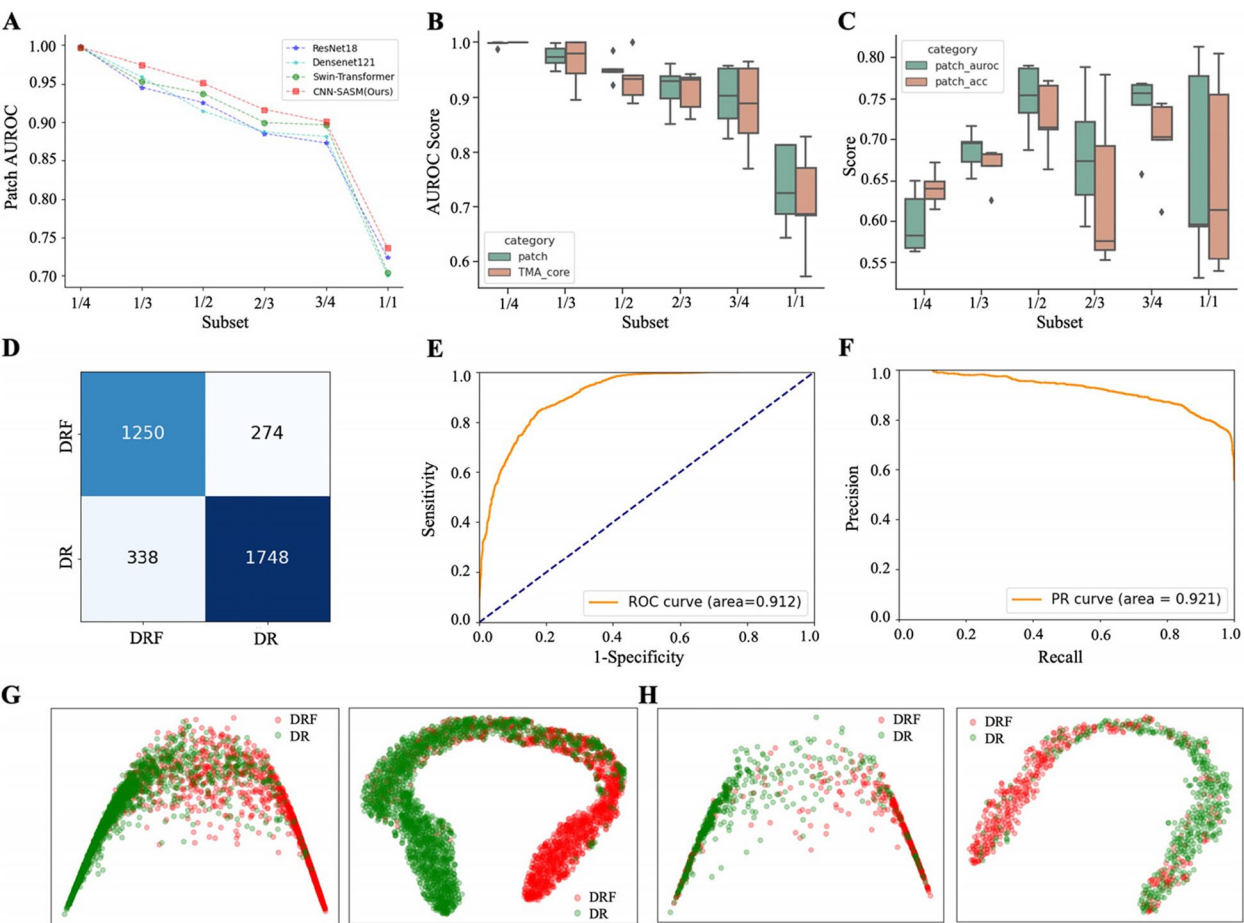


Fig. 2 Selection and performance of CNN-SASM. **A** Five-fold cross-validation results of CNN-SASM compared with the baseline models on different subsets. **B–C** The AUROC scores of CNN-SASM for all subsets, along with the model's generalization performance on the TCGA dataset. **D–F** The confusion matrix, ROC curve, and PR curve for the internal test set at the patch level. **G–H** The PCA and *t*-SNE clustering visualizations at the patch level for the internal test set and the TCGA dataset. ROC, receiver-operating characteristic

Table 2 The performance of various models

Model	1/4	1/3	1/2	2/3	3/4	1/1	Average	Parameters
ResNet18	0.999	0.946	0.926	0.885	0.873	0.724	0.892	44.60 M
DenseNet121	0.998	0.959	0.915	0.887	0.881	0.700	0.890	26.53 M
Swin-Transformer	0.997	0.953	0.938	0.900	0.897	0.704	0.898	334.82 M
CNN-SASM ^a	0.997	0.975	0.951	0.916	0.901	0.737	0.913	46.94 M
CNN-SelfAttn ^a	0.998	0.974	0.938	0.922	0.897	0.733	0.909	46.94 M
CNN-SelfAttn-Attn ^a	0.997	0.970	0.941	0.913	0.902	0.743	0.911	47.06 M
CNN-CSA ^a	0.998	0.972	0.949	0.909	0.898	0.743	0.912	42.92 M
CNN-SSA ^a	0.997	0.965	0.932	0.887	0.888	0.718	0.898	46.90 M

^a represents our proposed models in this study. All three combined CNN with self-attention mechanisms. CNN-SASM is the pathological feature extractor. CNN-SelfAttn was constructed by removing the softmax layer of CNN-SASM. CNN-SelfAttn-Attn was developed by substituting the softmax layer of CNN-SASM to the conventional attention layer. Both of them were applied to explore the role of the softmax layer. CNN-CSA is the CNN-SASM without the SSA module, and CNN-SSA is the CNN-SASM without the CSA module

small bile duct structures. The low-risk tissues primarily featured densely arranged cancer cells in cord-like structures with abundant cytoplasm and numerous lymphocytes infiltrating the interstitial space, potentially accounting for the favorable prognosis associated with this region.

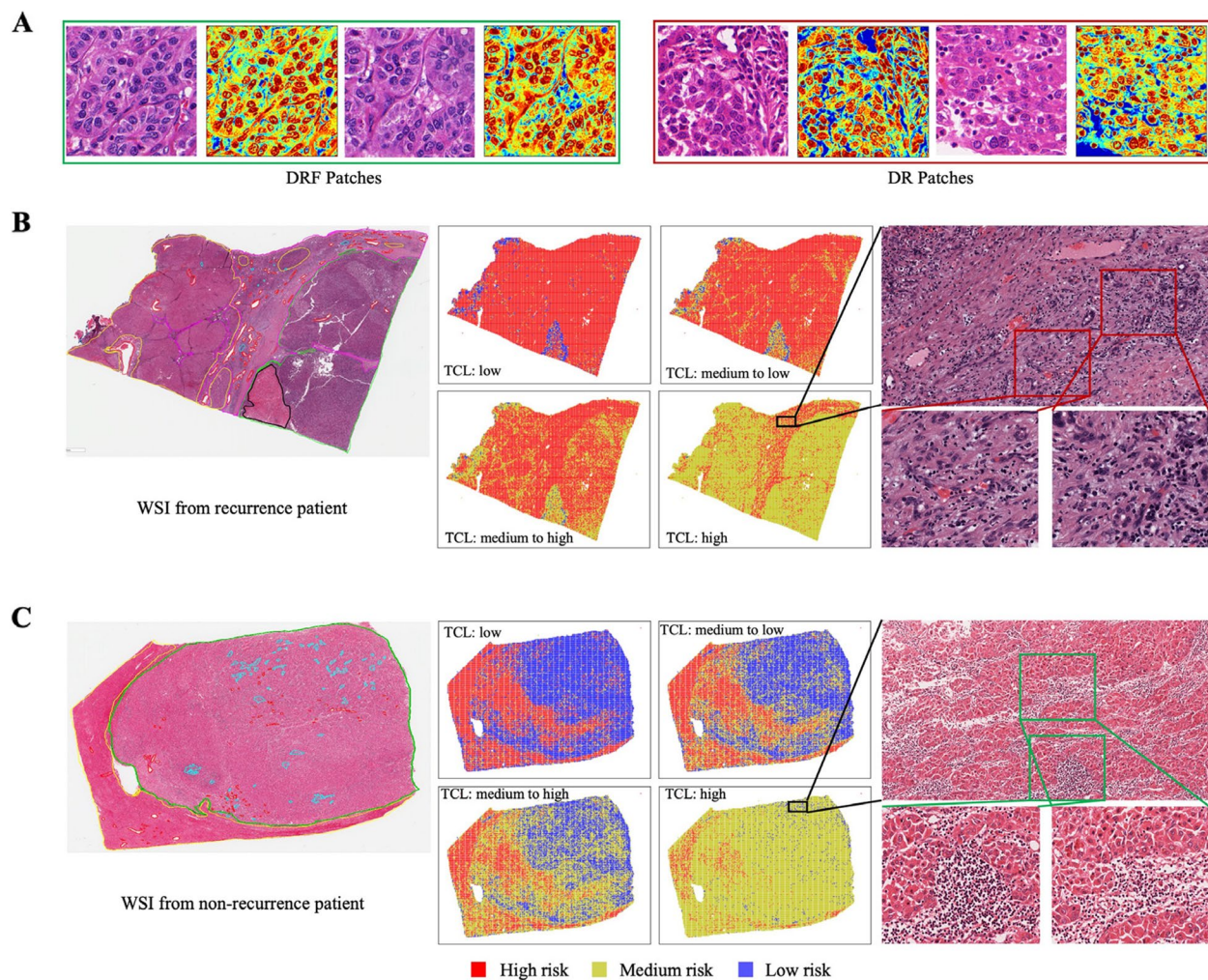


Fig. 3 Visualization of CNN-SASM prediction results and the correlation with pathologic characteristics. **A** The malignant and benign patches with overlaid heatmaps, indicated areas of high attention as identified by Grad-CAM. **B–C** WSIs from TCGA patients with and without postoperative recurrence, including visual differentiation between high- and low-risk patches. In **B**, annotations are color-coded: green for tumor tissue, pink for the collagenous stroma, yellow for adjacent liver tissue, red for blood vessels, and blue for lymphoid tissue. **C** uses a similar color scheme: green for tumor tissue, yellow for adjacent liver tissue, blue for lymphocytes, and red for blood vessels (four diagrams with a gradation of thresholds define high and low risk, becoming increasingly strict from left to right and top to bottom). TCL, threshold cutoff level for high-risk and low-risk tissues

The PPS-based prognostic model predicts outcomes of HCC patients

Based on qualitative and quantitative evaluations, PPS had been identified to be reliable and could represent prognostic relevant pathological signatures. We then developed a prognostic model to predict the postoperative recurrence of HCC patients using PPS at different time points. In the internal test set, the prognostic model achieved a favorable performance AUROC of 0.818, 0.811, and 0.709 at 1-year, 2-year, and final recurrence prediction tasks, respectively (Fig. 4A). We then divided patients of the test set into two groups, which were predicted non-recurrence and recurrence, by comparing

their predicted risk scores with the medium value of predicted recurrence risk score in the training set to investigate the association between patients' PFS and predicted results. We found that, except for the classification ability, the correlation analysis illustrated that the predicted results were also associated with patients' PFS. Specifically, patients predicted recurrence had significantly worse PFS than those predicted non-recurrence ($P < 0.0001$ for all three prognostic models) (Fig. 4B). The regression analysis results also showed that patients with higher recurrence risk scores had worse PFS (Additional file 1: Fig. S8B). We then conducted the survival analysis of patients in the test set, with the additional endpoint

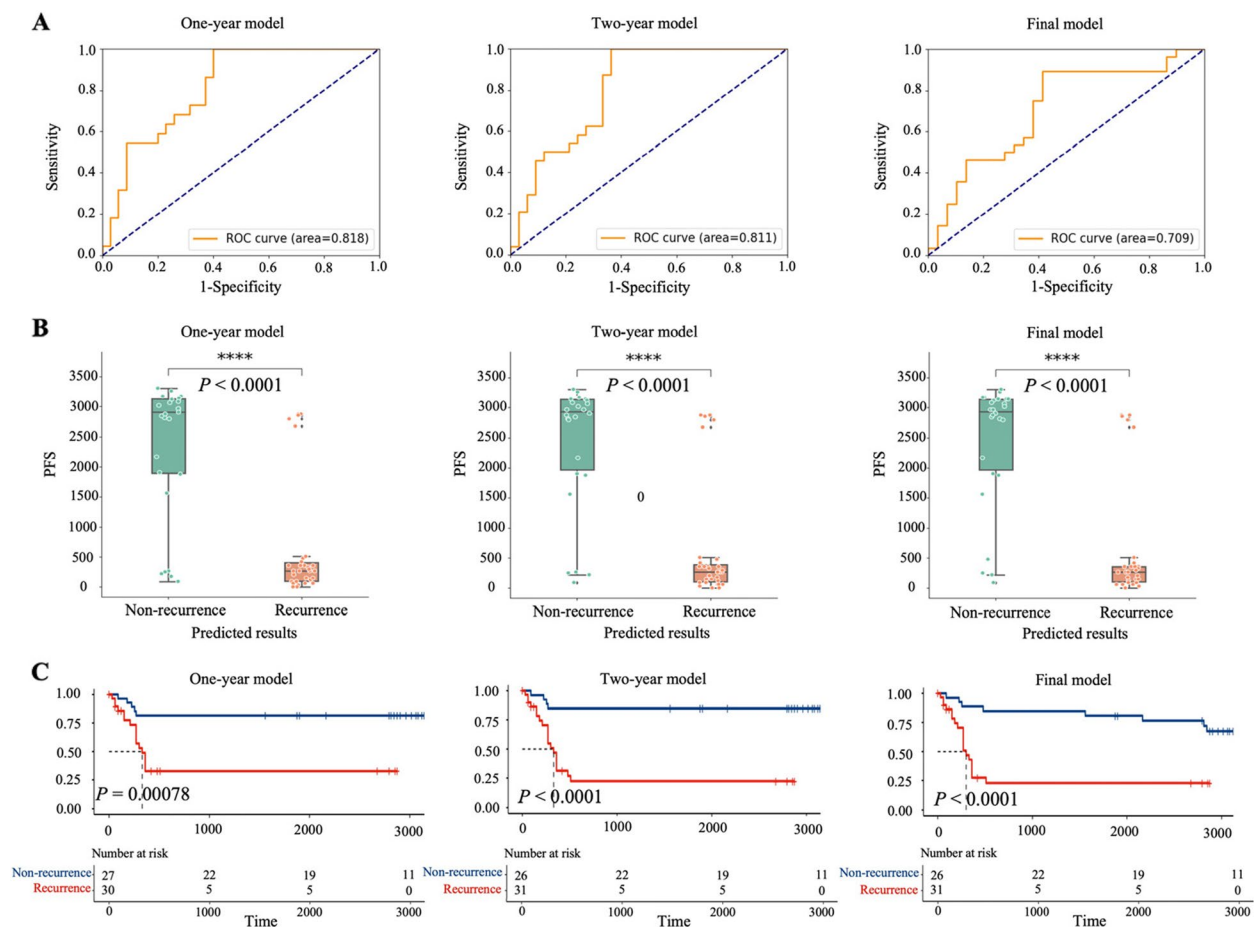


Fig. 4 Performance of the prognostic models in the internal test set. **A** The ROC curves of 1-year, 2-year, and final prognostic models. **B** The significance tests between patients' PFS and the prediction results of 1-year, 2-year, and final prognostic models. The binary classification results were derived by comparing the predicted risk scores of individuals in the test set with the medium value of predicted risk scores in the training set. **C** The survival curves of patients in the non-treatment test set. The endpoints were defined as the recurrence status at 1-year, 2-year, and final time points, and the log-rank test was used to assess the significance level between the two groups (from left to right panels are 1-year, 2-year, and final prognostic models, respectively)

events defined as the recurrence status at different time points. Differences in PFS at 1-year, 2-year, and final time points were observed significantly between patients with predicted non-recurrence and recurrence groups (Fig. 4C), indicating the PPS-based prognostic model could discern high recurrence risk patients from all HCC patients by pathological images.

We validated the generalizability of the PPS-based prognostic model in the external dataset. To ensure the reliability of outcomes, we excluded patients with less than 2 years of follow-up data and applied 1-year and 2-year prognostic models to validate. The AUROC at 1 and 2 years after surgery were 0.713 and 0.707 (Fig. 5A), and the corresponding AUPR were 0.785 and 0.776 (Additional file 1: Fig. S9), suggesting the models had good generalizability and binary classification ability on recurrence. Furthermore, we also divided the external

patients into predicted recurrence and non-recurrence groups according to the medium value of risk scores in the training set. The correlation analyses between patients' PFS and the predicted results were both significant (Fig. 5B). Patients who predicted recurrence tended to have a worse PFS compared to those who predicted non-recurrence. In addition, similar results were also observed that there were significant survival differences between the two groups on the external test set, and the P values of 1-year and 2-year models were 0.003 and 0.0014, respectively (Fig. 5C). In conclusion, these results showed that the PPS-based prognostic models could predict the outcomes of HCC patients at 1 and 2 years after surgery in both the internal and external test sets, exhibiting the powerful predictable ability of tumor pathological signatures.

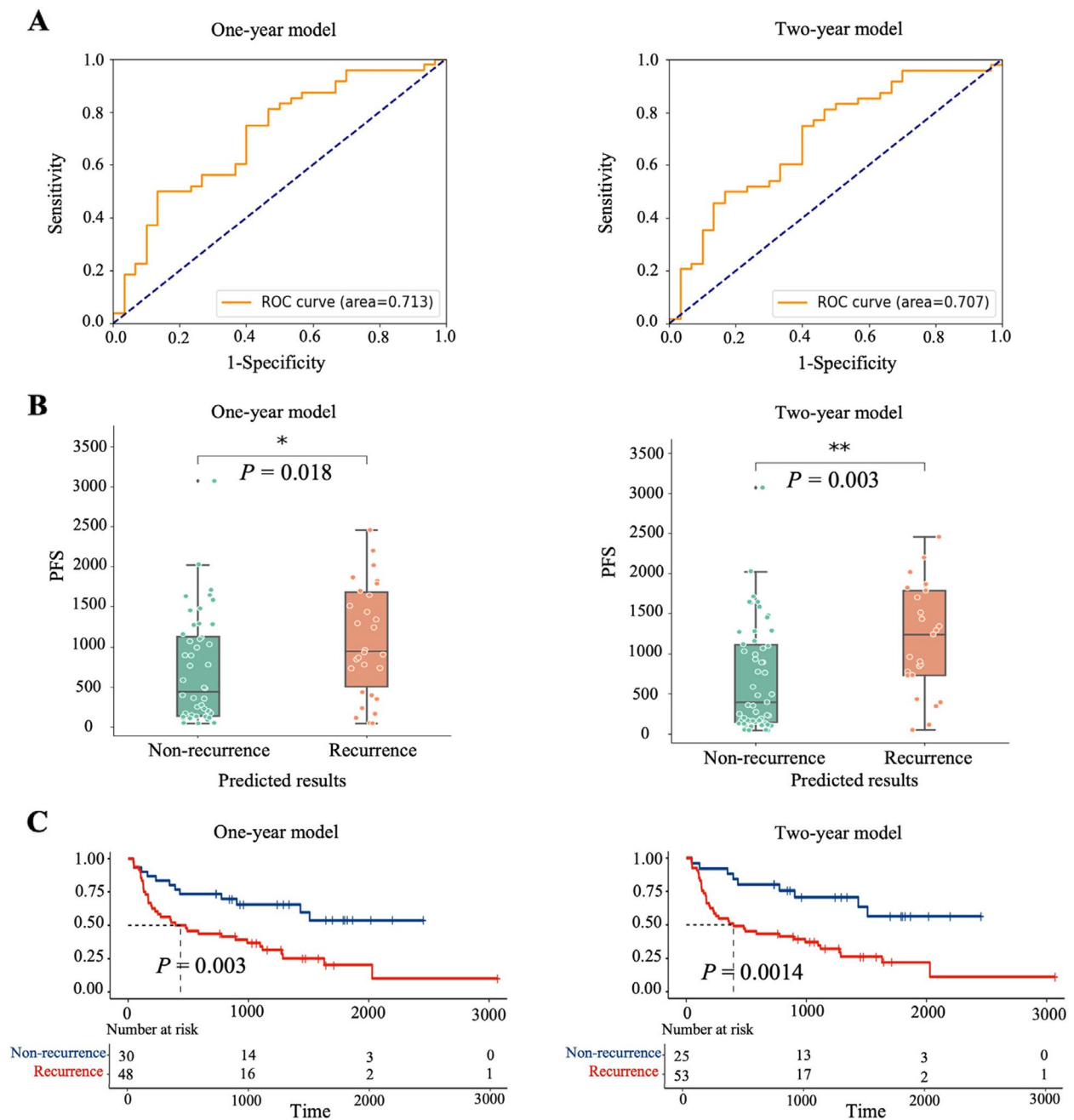


Fig. 5 Performance of the prognostic model in the external test set TCGA-LIHC. **A** The ROC curves of the 1-year and 2-year prognostic model. **B** The significance tests between patients' PFS and the prediction results of 1-year and 2-year prognostic models. Patients were divided into predicted recurrence and non-recurrence groups by comparing their risk score to the medium value of the training set. **C** The survival curves of patients at 1-year and 2-year time points in the non-treatment TCGA set, and the log-rank test was used to assess the significance level of the difference between the two groups (from left to right panels are 1-year and 2-year prognostic models, respectively)

To prove it is more effective to extract PPS by DRL rather than real-world recurrence labels, we compared the performance of our method and the end-to-end method under the same configuration (Table 3). Our method exceeded the end-to-end one in almost all

evaluation metrics at all time points, with an AUROC increase of up to 7.1%, 3.7%, and 5.5% at different time points, suggesting denoising the recurrence label will make the model behave better in prognostic tasks.

Table 3 Comparison of our method and end-to-end strategy

Time interval	Method	AUROC	AUPR	Accuracy	Precision	Recall	F1
One-year	PPS-based	0.818	0.697	0.702	0.632	0.545	0.585
	End-to-end	0.721	0.566	0.649	0.550	0.500	0.524
Two-year	PPS-based	0.811	0.708	0.684	0.636	0.583	0.609
	End-to-end	0.774	0.671	0.684	0.65	0.542	0.591
Final	PPS-based	0.709	0.685	0.684	0.656	0.750	0.700
	End-to-end	0.654	0.639	0.632	0.652	0.536	0.588

The PPS-based recurrence risk outperforms clinical risk indicators

We compared the discrimination performance of the PPS-based prognostic model with traditional clinical risk indicators in terms of AUROC and C-index for prognosis prediction. Our PPS-based recurrence risk at 1 year outperformed all clinical indicators, including HBsAg, ALT, AFP, tumor size, BCLC stage, and vascular invasion, even the comprehensive logistic regression model, which incorporated all clinical risk indicators and the integrated model (Fig. 6A, Additional file 1: Fig. S10). The result above suggested that PPS contained powerful prognostic signatures beyond existing clinical characteristics. On the TCGA set, 1-year and 2-year recurrence risks based on PPS also had the highest AUROC among other clinical indicators, showing good generalizability (Fig. 6A).

Previous studies had demonstrated that the vascular invasion [32, 33] and tumor size-related signatures [34] may contribute to the prognostic outcomes in HCC patients, which was also observed in our study. The vascular invasion and tumor size had the highest predictive powers among all clinical indicators in both the internal test set and TCGA set. We then attempted to perform the survival analysis within these two variable subgroups, and all the *P* values were significant, demonstrating recurrence risk based on PPS could further divide patients into two groups (Fig. 6B). Cox univariate analysis PPS-based recurrence risks had higher HR than clinical indicators, reaching 7.00 (95% CI 2.07–23.70) and 4.51 (95% CI 1.61–12.62) in the internal and external set at 1 year and 4.74 (95% CI 1.61–13.93) in the external set at 2 years (Fig. 6C). In conclusion, PPS-based recurrence risk was the independent predictive indicator in prognostic tasks and could become the powerful tool in estimating patients' survival.

The PPS-based prognostic model predicts response to sorafenib treatment

Sorafenib is recognized as the first-line target drug in HCC patients, but it is still challenging to forecast its

effectiveness [35]. We then investigated the predictive ability of the PPS-based prognostic model for the sorafenib treatment response in HCC. To examine whether there were differences in the clinical backgrounds of patients in the sorafenib group, we performed significance testing between the recurrence and non-recurrence groups. The results indicated that the *P* values for all clinical features were greater than 0.05, suggesting no clinical indicator significantly influences the patients' treatment response (Table 4). The significance test of clinical features between the predicted recurrence and non-recurrence in the testing set was also conducted to ensure alignment of the clinical backgrounds between the two groups (Additional file 2: Table S2). In the test set, sorafenib-treated patients were divided into predicted recurrence and non-recurrence groups by comparing their recurrence risk scores to the medium value of the training set at the 1-year time point. Patients in the 1-year recurrence subgroup were significantly associated with poor PFS, $P=0.003$ (Fig. 7A), suggesting the PPS-based prognostic model could also distinguish high recurrence risk HCC patients after being treated with sorafenib.

Aside from the prognostic model, we also investigated the predictive ability of CNN-SASM, which was trained by DRL. CNN-SASM's output was defined as the denoised recurrence score (DRS) representing the theoretical recurrence risk. Then, we computed the correlation coefficient between the PFS of sorafenib-treated patients and the DRS and found that there was a significant negative correlation ($r = -0.66$, $P=0.0058$). Patients exhibited worse PFS showed the significant increasing of DRS (Fig. 7B). These results indicated that PPS-based prognostic model's output could not only predict outcomes of HCC patients but also strongly associated with the outcomes of sorafenib-treated HCC patients, offering more specific treatment management strategy in clinical.

Additionally, we then attempted to predict patients' survival benefits from sorafenib with PPS-based prognostic model. Patients with specific physical conditions would be treated, and such clinical practice would bring

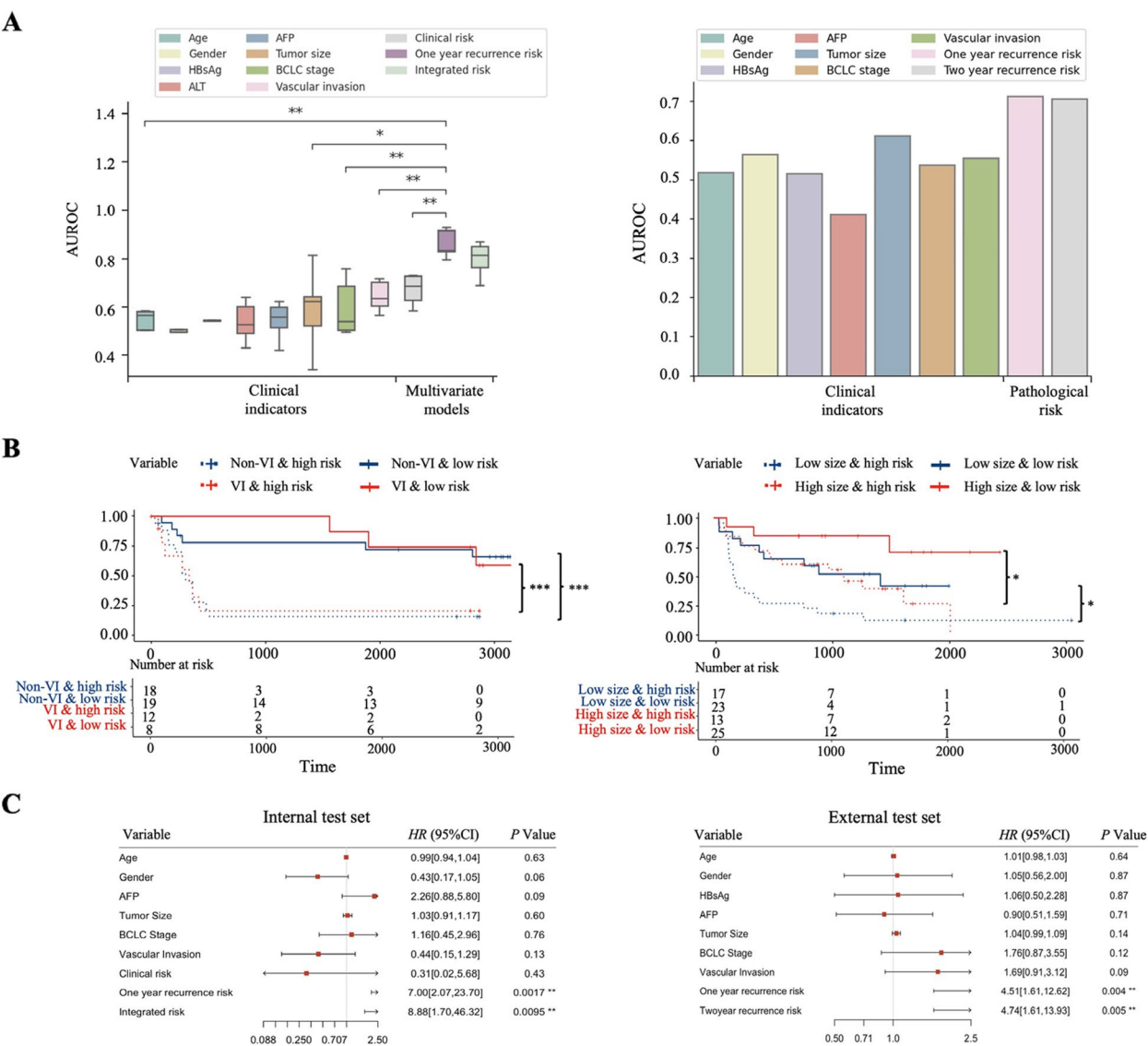


Fig. 6 The comparison between the risk score of the PPS-based model and clinical risk factors. **A** The AUROC values in clinical univariable and multivariate scores in the internal test set and TCGA set. The clinical risk was computed by the logistic regression (LR) model using all clinical indicators, and the integrated risk was also derived from the LR model by clinical indicators and the PPS-based prognostic model risk score. All comparisons of the left panel were conducted based on a fivefold cross-validation strategy. The lines within the boxes represent the medium AUROC value, the bounds of the boxes represent the interquartile range (IQR), and the whiskers represent the 95% confidence intervals. **B** The survival difference of the vascular invasion (VI) and tumor size subgroups in the internal and external test sets. **C** The HR values of the risk score of the PPS-based prognostic model and clinical indicators in the internal and external test sets

bias, so the baseline physical situations of non-treated and sorafenib-treated patients varied hugely, $P=0.017$ (Fig. 7C). However, for patients predicted with recurrence at the 1-year time point, the PFS between non-treated and sorafenib-treated patients had no significant difference (Fig. 7D), which meant patients in predicted recurrence subgroup would decrease the PFS gap between two treatments because of the benefit from sorafenib, while the PFS of patients predicted with non-recurrence remained the significant difference between

two treatments (Fig. 7E), demonstrating patients under this restrictions got few benefits from sorafenib. In this way, we can screen patients who tend to respond well to sorafenib after surgery and improve the medication efficacy in clinical.

Discussion

Accumulating studies have demonstrated that digital pathological images could contribute to predicting patient prognosis based on deep learning [36, 37]. In

Table 4 The baseline characteristic comparison of clinical indicators between recurrence and non-recurrence patients in the sorafenib group

Clinical indicator	Subgroup	Non-recurrence	Recurrence	P value
Age	≥ 50	14	18	0.658
	< 50	17	30	
Gender	Male	29	45	1
	Female	2	3	
HBsAg	+	3	28	0.229
	−	10	38	
AFP	+	25	43	0.431
	−	6	5	
Tumor size	≥ 5	19	32	0.805
	< 5	12	16	
BCLC stage	1	3	13	0.674
	2	28	45	
Vascular invasion	+	20	38	0.239
	−	11	10	

The chi-square test was used if the number of individuals in each subgroup exceeded five or Fisher's exact test was applied

this work, we developed a pathology-based deep learning model that can adaptively extract as well as decode the prognostic-related features and facilitate the prediction of HCC patients' recurrence condition. The predictive ability of PPS-based prognostic model was superior to all conventional clinicopathological factors and could stratify patients with significantly different prognoses in HCC. Additionally, we explored the PPS-based prognostic model's sorafenib response predictive ability and demonstrated that it could identify patients who were more likely to benefit from sorafenib, which would offer help in clinical medication management.

It is challenging to extract new prognostic features from pathological images and train a deep neural network in image analysis. The conventional attention layers in CNN have channel and spatial location dimensions [38]; they can directly compute weights from sequences and labels by gradient renewal but fail to consider connections between sequences. The self-attention mechanism [39] can address this issue by computing attention scores of sequences. Herein, we successfully applied the architecture of CNN-SASM incorporated the integration of multi-head self-attention mechanisms

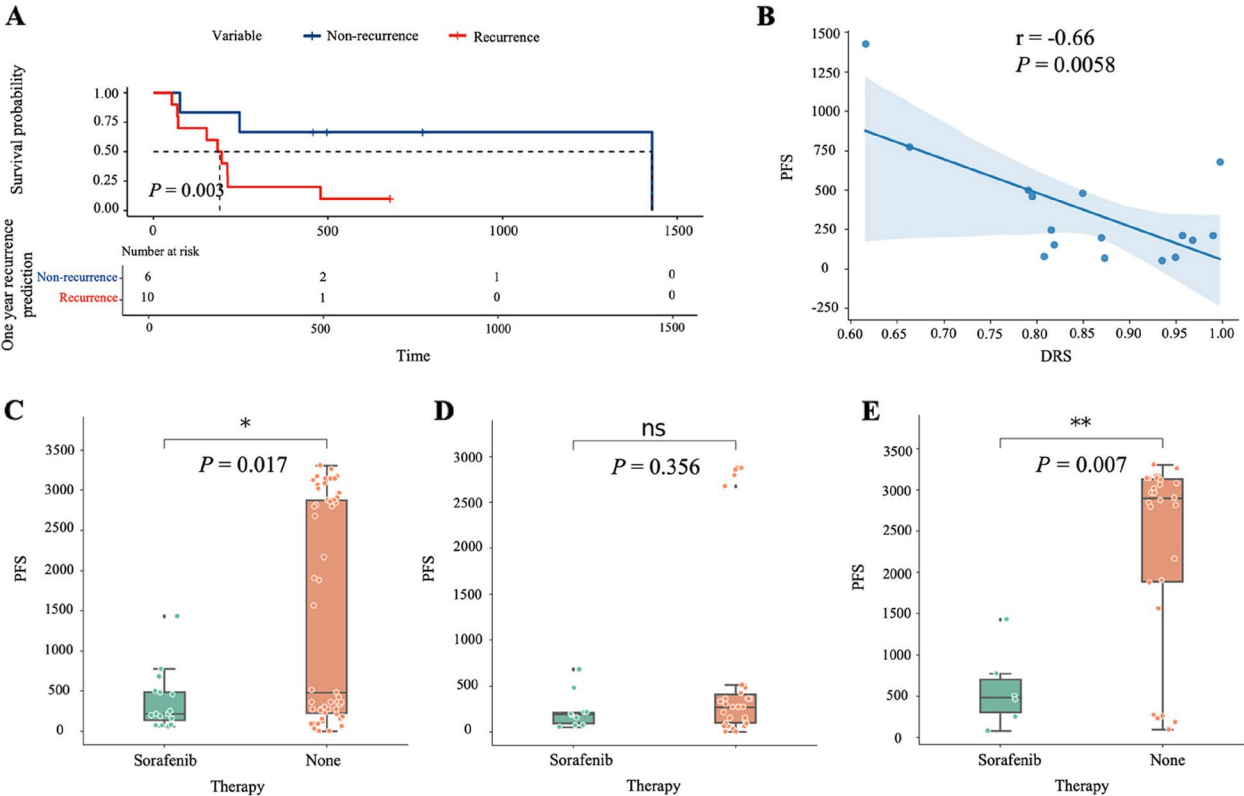


Fig. 7 Performance of the PPS-based prognostic model in predicting the survival and benefits in patients treated with sorafenib. **A** The survival curve for sorafenib-treated patients in the test set. **B** The correlation test between PFS and DRS. A series of significant tests between PFS and therapy strategies in all **(C)**, predicted recurrence **(D)**, and predicted non-recurrence **(E)** patients in the test set, respectively

across two distinct dimensions: channel and spatial location. This design contemplates the interconnections between sequences within two dimensions and can offer the flexibility to be independently integrated into any framework. This design could obtain substantial and stable classification performance even though we employed the simple architecture ResNet18. In our study, the softmax layer in CNN-SASM functioned as a conventional attention layer, aiming to bestow more weights on particular channels or spatial locations without extra parameters. As the sample size grew and noise increased, the softmax struggled to discern vital channels and spatial locations alone. Thus, CNN-SASM can improve the classification performance with limited parameter scale, making it suitable as a foundational framework for small dataset.

Recently, some novel pathological features have been reported in HCC, such as tumor necrosis [40], inflammatory score [41], and immune-related biomarkers [42]. Deep learning models based on pathological features face challenges in predicting real-world recurrence due to complicated influencing factors. For the first time, we proposed utilizing DRL under different thresholds to train pathological models, and it can extract biologically interpretable pathological signatures maximally, PPS, by ignoring irrelevant noised labels, which can be visualized on heatmaps. Compared with the end-to-end strategy, training the deep learning model by predictable labels would be more effective than the noising labels (Table 3). It may provide a new paradigm for complicated tasks involved in intricate influencing factors.

Considering the comprehensive performance of the pathological model, we traded off the classification ability and generalizability in the multiple subsets training process to determine the best-performing CNN-SASM. Despite displaying good generalization on the 3/4 subset, CNN-SASM was in a fluctuating phase. We attributed the fluctuation phase to the quality variability of added data and the rare sample size of the 14 TCGA cases used for the generalization test. Ultimately, we selected the 1/2 subset CNN-SASM to guarantee the robustness of the model. The risk area visualization on WSI shows that CNN-SASM effectively discriminated against DR and DRF. The high-risk areas were primarily located in the tumor's peripheral tissues, indicating the potential relevance of the surrounding tissue to the patient's prognosis. This observation aligns with Zhu et al.'s findings [43].

A growing number of studies has substantiated that the morphological characteristics observable in pathological images are indicative of the prognostic outcomes for patients in various cancers, including breast [44], gastric [45], cervical cancers [46], and

HCC. However, either the methodologies and models employed are more complicated, or their efficacy in prognostic classification necessitates further enhancement [24, 47]. In our study, the approach of denoising recurrence labels facilitated the employment of less complex network architectures for training. Moreover, our AUROC for prognosis classification achieved a value exceeding 0.8 and further exhibited robust generalization using TCGA HCC data. Notably, the profound signature markers, identified as PPS in our study, exhibited superior performance relative to clinical risk indicators across all evaluation metrics, encompassing C-index, AUROC, and HR. The proposed model based on PPS was even better than the integrated multimodal model which combined PPS and clinical risk factors, suggesting PPS might already incorporate clinical characteristics from pathological images. The prognostic model based on PPS can greatly help doctors focus on the postoperative adverse reaction population so that preventive measures can be taken in advance to improve patients' prognostic survival.

Although substantial progress has been made in the imaging-based deep learning model for cancer detection and diagnosis, there is limited success in the prediction of the response to specific treatments of cancer patients. Among the limited studies, the predominant focus has been on utilizing more accessible data modalities, including CT or MRI [48], or in other carcinomas, such as lung [49] and ovarian [50] cancers. To our knowledge, this study is the initial application to predict patient response to targeted therapy in HCC based on histopathological images. According to our results, PPS-based prognostic model could further predict sorafenib treatment response, manifested in further dividing sorafenib-treated patients into high and low risks associated with postoperative survival situations. In addition, DRS, the output of the CNN-SASM, was also significantly related to sorafenib-treated patients' PFS. Our results indicate a significant predictive effect of the PPS-based prognostic model for prognosis and the benefit of target therapy in HCC. Patients who predicted recurrence by the PPS-based prognostic model were more likely to benefit from sorafenib than non-recurrence patients. In this way, it would be useful to provide supplemental information for better selection of therapeutic strategies and avoid unnecessary medication.

This study has several limitations. First, we only employed ResNet18 architecture to develop CNN-SASM although CNN-SASM yielded improved results and showed better generalization. In addition, the results would be more convincing if prospective multicenter and large-sample research is conducted.

Conclusions

Overall, our study proposed the CNN-SASM, a pathological model integrating CNN with channel and spatial self-attention mechanisms to predict recurrence condition in HCC patients and sorafenib survival benefits. Furthermore, the proposed concept that uses DRL as training labels rather than real-world recurrence labels to develop pathological models is broadly applicable to prognostic tasks.

Abbreviations

AFP	Serum a-fetoprotein
AS	Attention-score
AUROC	Area under the receiver operating characteristic curve
AUPR	Area under the precision versus recall curve
C-index	Concordance index
CNNs	Convolutional neural networks
CSA	Channel self-attention
CT	Computed tomography
DR	Denosed recurrence
DRF	Denosed recurrence-free
DRL	Denosed recurrence label
HCC	Hepatocellular carcinoma
HR	Hazard ratio
ICIS	Immune checkpoint inhibitors
MLP	Multi-layer perceptron
MRI	Magnetic resonance imaging
OS	Overall survival
PFS	Progression-free survival
PPS	Prognostic relevant pathological signatures
SSA	Spatial self-attention
TCL	Threshold cutoff level for high-risk and low-risk tissues
WSI	Whole slide imaging

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-025-03977-4>.

Additional file 1: Figures S1–S10. Fig. S1. Architecture and organization of the dataset. Fig. S2. The detailed data structure diagram. Fig. S3. Data flow diagrams and organization of concepts in our study. Fig. S4. Supplementary data for CNN-SASM model selection. Fig. S5. Grad-CAM of CSA and SSA modules. Fig. S6. PCA clustering of the CNN-SASM at patch and image levels on the test set. Fig. S7. *t*-SNE clustering visualization for models from all subsets. Fig. S8. Supplementary data for the performance of the PPS-based prognostic model on the internal test set. Fig. S9. Supplementary data for the performance of the PPS-based prognostic model on TCGA. Fig. S10. The comparison between the risk score of the PPS-based model and clinical risk factors at C-index values.

Additional file 2: Tables S1–S2. Table S1. Statistics on sub-datasets. Table S2. The baseline characteristic comparison of clinical indicators between predicted recurrence and non-recurrence patients in the sorafenib test set.

Acknowledgements

We acknowledge the support of Medical Science Data Center in Shanghai Medical College of Fudan University.

Authors' contributions

YL, JX; Conceptualization: LL, QD, YL; Data curation: YL; Formal analysis: YL; Funding acquisition: QD, NR, ZH; Investigation: YL, FZ; Methodology: YL, FZ; Project administration: QD, YL, FZ; Resources: QD, NR, ZH, QC; Supervision: LL, QD, FZ; Validation: YL, JX, FZ; Visualization: YL, JX; Writing – original draft: YL; Writing – review & editing: YL, QD, FZ. All authors read and approved the final manuscript.

Funding

This work was supported by the following: the National Key Research and Development Program of China (2021YFC2500403), the Shanghai Shen Kang Hospital Development Center New Frontier Technology Joint Project (SHDC12021109), Medical Research Specialized Program of Beijing Huatong Guokang Foundation for Industry-University-Research Innovation Fund of Chinese Universities, National Ministry of Education (2023HT060). The authors thank the Medical Science Data Center in Shanghai Medical College of Fudan University.

Data availability

The datasets used during the current study are available from the corresponding author on reasonable request. The data used for external validation are available in the TCGA repository, <https://portal.gdc.cancer.gov>, and the TCGA-LIHC dataset is described in the original publication [27]. The underlying code for this study is available in <https://github.com/liyixin12139/Prediction-Sorafenib-Response-by-DRL>.

Declarations

Ethics approval and consent to participate

The study protocol was conducted in accordance with ethical guidelines (Declaration of Helsinki) and approved by the Institutional Ethics Committee of Zhongshan Hospital (approval number: B2021-143R).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 September 2024 Accepted: 27 February 2025

Published online: 18 March 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209–49. <https://doi.org/10.3322/caac.21660>.
2. Abou-Alfa GK, Blanc JF, Miles S, Ganten TM, Trojan J, Cebon JS, et al. Phase II study of first-line trebananib plus sorafenib in patients with advanced hepatocellular carcinoma (HCC). *J Clin Oncol*. 2014;32:286–286. https://doi.org/10.1200/jco.2014.32.3_suppl.286.
3. Cheng AL, Kang YK, Chen ZD, Tsao CJ, Qin SK, Kim JS, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol*. 2009;10:25–34. [https://doi.org/10.1016/S1470-2045\(08\)70285-7](https://doi.org/10.1016/S1470-2045(08)70285-7).
4. Hsueh KC, Lee CC, Huang PT, Liang CY, Yang SF. Survival benefit of experience of liver resection for advanced recurrent hepatocellular carcinoma treated with sorafenib: a propensity score matching analysis. *Curr Oncol*. 2023;30:3206–16. <https://doi.org/10.3390/curroncol30030243>.
5. Wu LQ, Pang XF, Cao JY, Wang ZS. Prognosis of advanced hepatocellular carcinoma (HCC) in patients undergoing surgery combined with peri- or postoperative treatment with sorafenib. *J Clin Oncol*. 2014;32:377–377. https://doi.org/10.1200/jco.2014.32.3_suppl.377.
6. FDA, FDA expands approved use of Stivarga to treat liver cancer. 2017. <https://www.fda.gov/news-events/press-announcements/fda-expands-approved-use-stivarga-treat-liver-cancer>.
7. FDA, FDA approves lenvatinib for unresectable hepatocellular carcinoma. 2018. <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-lenvatinib-unresectable-肝细胞癌>.
8. FDA, FDA approves cabozantinib for hepatocellular carcinoma. 2019. <https://www.fda.gov/drugs/fda-approves-cabozantinib-hepatocellular-carcinoma>.

9. FDA, FDA approves ramucirumab for hepatocellular carcinoma. 2019. <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-ramucirumab-hepatocellular-carcinoma>.
10. FDA, FDA grants accelerated approval to pembrolizumab for hepatocellular carcinoma. 2018. <https://www.fda.gov/drugs/fda-grants-accelerated-approval-pembrolizumab-肝细胞-癌>.
11. Luo X-Y, Wu K-M, He X-X. Advances in drug development for hepatocellular carcinoma: clinical trials and potential therapeutic targets. *J Exp Clin Cancer Res*. 2021;40:172. <https://doi.org/10.1186/s13046-021-01968-w>.
12. Rastogi A. Changing role of histopathology in the diagnosis and management of hepatocellular carcinoma. *World J Gastroenterol*. 2018;24:4000–13. <https://doi.org/10.3748/wjg.v24.i35.4000>.
13. Zhan Q, Shen B-Y, Deng X-X, Zhu Z-C, Chen H, Peng C-H, et al. Clinical and pathological analysis of 27 patients with combined hepatocellular-cholangiocarcinoma in an Asian center. *J Hepatobiliary Pancreat Sci*. 2012;19:361–9. <https://doi.org/10.1007/s00534-011-0417-2>.
14. Nakamura Y, Miyaaki H, Miuma S, Akazawa Y, Fukushima M, Sasaki R, et al. Automated fibrosis phenotyping of liver tissue from non-tumor lesions of patients with and without hepatocellular carcinoma after liver transplantation for non-alcoholic fatty liver disease. *Hepatol Int*. 2022;16:555–61. <https://doi.org/10.1007/s12072-022-10340-9>.
15. Zucman-Rossi J, Jeannot E, Van Nhieu JT, Scoazec JY, Guettier C, Rebouissou S, et al. Genotype-phenotype correlation in hepatocellular adenoma: new classification and relationship with HCC. *Hepatology*. 2006;43:515–24. <https://doi.org/10.1002/hep.21068>.
16. Calderaro J, Couchy G, Imbeaud S, Amaddeo G, Letouze E, Blanc J-F, et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. *J Hepatol*. 2017;67:727–38. <https://doi.org/10.1016/j.jhep.2017.05.014>.
17. Qin LX, Tang ZY. The prognostic molecular markers in hepatocellular carcinoma. *World J Gastroenterol*. 2002;8:385–92. <https://doi.org/10.3748/wjg.v8.i3.385>.
18. Kumar N, Gupta R, Gupta S. Whole slide imaging (WSI) in pathology: current perspectives and future directions. *J Digit Imaging*. 2020;33:1034–40. <https://doi.org/10.1007/s10278-020-00351-z>.
19. Bidgoli AA, Rahnamayan S, Dehkharghanian T, Riasatian A, Tizhoosh HR. Evolutionary computation in action: hyperdimensional deep embedding spaces of gigapixel pathology images. *IEEE Trans Evol Comput*. 2023;27:52–66. <https://doi.org/10.1109/TEVC.2022.3178299>.
20. Lee Y, Park JH, Oh S, Shin K, Sun J, Jung M, et al. Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. *Nat Biomed Eng*. 2022;1–15. <https://doi.org/10.1038/s41551-022-00923-0>.
21. Lin Y-S, Huang P-H, Chen Y-Y. Deep learning-based hepatocellular carcinoma histopathology image classification: accuracy versus training dataset size. *IEEE Access*. 2021;9:33144–57. <https://doi.org/10.1109/ACCESS.2021.3060765>.
22. Chen Q, Xiao H, Gu Y, Weng Z, Wei L, Li B, et al. Deep learning for evaluation of microvascular invasion in hepatocellular carcinoma from tumor areas of histology images. *Hepatol Int*. 2022;16:590–602. <https://doi.org/10.1007/s12072-022-10323-w>.
23. Shi J-Y, Wang X, Ding G-Y, Dong Z, Han J, Guan Z, et al. Exploring prognostic indicators in the pathological images of hepatocellular carcinoma based on deep learning. *Gut*. 2021;70:951–61. <https://doi.org/10.1136/gutjnl-2020-320930>.
24. Saillard C, Schmauch B, Laifa O, Moarii M, Toldo S, Zaslavskiy M, et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. *Hepatology*. 2020;72:2000–13. <https://doi.org/10.1002/hep.31207>.
25. Liu X, Liu Z, Yan Y, Wang K, Wang A, Ye X, et al. Development of prognostic biomarkers by TMB-guided WSI analysis: a two-step approach. *IEEE J Biomed Health*. 2023;27:1780–9. <https://doi.org/10.1109/JBHI.2023.3249354>.
26. Wang K, Xiang Y, Yan J, Zhu Y, Chen H, Yu H, et al. A deep learning model with incorporation of microvascular invasion area as a factor in predicting prognosis of hepatocellular carcinoma after R0 hepatectomy. *Hepatol Int*. 2022;16:1188–98. <https://doi.org/10.1007/s12072-022-10393-w>.
27. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*. 2017;169:1327–+. <https://doi.org/10.1016/j.cell.2017.05.046>.
28. Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, et al. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Trans Med Imag*. 2016;35:1962–71. <https://doi.org/10.1109/TMI.2016.2529665>.
29. He KM, Zhang XY, Ren SO, Sun J. Deep Residual Learning for Image Recognition. *Proc IEEE Conf Comput Vis Pattern Recog CVPR*. 2016:770–778. <https://doi.org/10.1109/CVPR.2016.90>.
30. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. *Proc. IEEE Conf. Comput Vis Pattern Recog*. 2017;2261–9. <https://doi.org/10.1109/CVPR.2017.243>.
31. Liu Z, Lin YT, Cao Y, Hu H, Wei YX, Zhang Z, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *Proc. IEEE Int Conf Comput Vis ICCV*. 2021:9992–10002. <https://doi.org/10.1109/ICCV48922.2021.00986>.
32. Mokdad AA, Singal AG, Marrero JA, Zhu H, Yopp AC. Vascular invasion and metastasis is predictive of outcome in Barcelona clinic liver cancer stage C hepatocellular carcinoma. *J Natl Compr Cancer Netw*. 2017;15:197–204. <https://doi.org/10.6004/jnccn.2017.0020>.
33. Zhao N, Ni CS, Zhang DF, Che N, Li YL, Wang X. Identification of a vascular invasion-related signature based on lncRNA pairs for predicting prognosis in hepatocellular carcinoma. *BMC Gastroenterol*. 2024;24:33. <https://doi.org/10.1186/s12876-023-03118-2>.
34. Chen ZL, Zheng H, Zeng W, Liu MY, Chen Y. Prognostic analysis on different tumor sizes for 14634 hepatocellular carcinoma patients. *Eur J Cancer Care*. 2023;2023: 1106975. <https://doi.org/10.1155/2023/1106975>.
35. Samawi HH, Sim HW, Chan KK, Alghamdi MA, Lee-Ying RM, Knox JJ, et al. Prognosis of patients with hepatocellular carcinoma treated with sorafenib: a comparison of five models in a large Canadian database. *Cancer Med*. 2018;7:2816–25. <https://doi.org/10.1002/cam4.1493>.
36. Yamashita R, Long J, Saleem A, Rubin DL, Shen J. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. *Sci Rep*. 2021;11:2047. <https://doi.org/10.1038/s41598-021-81506-y>.
37. McCaffrey C, Jahangir C, Murphy C, Burke C, Gallagher WM, Rahman A. Artificial intelligence in digital histopathology for predicting patient prognosis and treatment efficacy in breast cancer. *Expert Rev Mol Diagn*. 2024;24:363–77. <https://doi.org/10.1080/14737159.2024.2346545>.
38. Woo SH, Park J, Lee JY, Kweon IS. CBAM: Convolutional Block Attention Module. *Proc. Eur Conf Comput Vis*. 2018;3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
39. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30. <https://dl.acm.org/doi/10.5555/3295222.3295349>.
40. Yen YH, Kuo FY, Eng HL, Liu YW, Yong CC, Li WF, et al. Tumor necrosis as a predictor of early tumor recurrence after resection in patients with hepatoma. *PLoS One*. 2023;18: e0292144. <https://doi.org/10.1371/journal.pone.0292144>.
41. Wu BA, Wu YM, Guo XL, Liu YF, Yue YP, Zhao WJ, et al. Prognostic significance of preoperative integrated liver inflammatory score in patients with hepatocellular carcinoma. *Med Sci Monit*. 2022;28:e937005–1. <https://doi.org/10.12659/MSM.937005>.
42. Wan HF, Lu S, Xu L, Yuan KF, Xiao Y, Xie KL, et al. Immune-related biomarkers improve performance of risk prediction models for survival in patients with hepatocellular carcinoma. *Front Oncol*. 2022;12: 925362. <https://doi.org/10.3389/fonc.2022.925362>.
43. Zhu HW, Lin YP, Lu DY, Wang SS, Liu YJ, Dong LQ, et al. Proteomics of adjacent-to-tumor samples uncovers clinically relevant biological events in hepatocellular carcinoma. *Natl Sci Rev*. 2023;10:nwad167. <https://doi.org/10.1093/nsr/nwad167>.
44. Yang JL, Ju J, Guo L, Ji BB, Shi SF, Yang ZX, et al. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Comput Struct Biotechnol J*. 2022;20:333–42. <https://doi.org/10.1016/j.csbj.2021.12.028>.
45. Huang BL, Tian S, Zhan N, Ma JJ, Huang ZW, Zhang CK, et al. Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: a retrospective multicentre study. *EBio-Medicine*. 2021;73: 103631. <https://doi.org/10.1016/j.ebiomed.2021.103631>.
46. Ye ZX, Zhang YX, Liang YB, Lang JD, Zhang XL, Zang GL, et al. Cervical cancer metastasis and recurrence risk prediction based on deep

- convolutional neural network. *Curr Bioinf.* 2022;17:164–73. <https://doi.org/10.2174/1574893616666210708143556>.
47. Qu W-F, Tian M-X, Lu H-W, Zhou Y-F, Liu W-R, Tang Z, et al. Development of a deep pathomics score for predicting hepatocellular carcinoma recurrence after liver transplantation. *Hepatol Int.* 2023;17:927–41. <https://doi.org/10.1007/s12072-023-10511-2>.
 48. Li W, Partridge SC, Newitt DC, Steingrimsdottir J, Marques HS, Bolan PJ, et al. Breast multiparametric MRI for prediction of neoadjuvant chemotherapy response in breast cancer: the BMMR2 challenge. *Radiol-Imag Cancer.* 2024;6:e230033. <https://doi.org/10.1148/rycan.230033>.
 49. Zhang YB, Yang ZJ, Chen RQ, Zhu YL, Liu L, Dong JY, et al. Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer. *NPJ Digit Med.* 2024;7:15. <https://doi.org/10.1038/s41746-024-01003-0>.
 50. Yang ZJ, Zhang YB, Zhuo LL, Sun KD, Meng FL, Zhou M, et al. Prediction of prognosis and treatment response in ovarian cancer patients from histopathology images using graph deep learning: a multicenter retrospective study. *Eur J Cancer.* 2024;199:113532. <https://doi.org/10.1016/j.ejca.2024.113532>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.