

Deep Learning for Prediction of Hepatocellular Carcinoma Recurrence After Resection or Liver Transplantation: a Discovery and Validation Study

Zhikun Liu

Hangzhou First People's Hospital

Yuanpeng Liu

Zhejiang University

Wenhui Zhang

Hangzhou First People's Hospital

Yuan Hong

Zhejiang University

Jinwen Meng

Hangzhou First People's Hospital

Jianguo Wang

Hangzhou First People's Hospital

Shusen Zheng

First Hospital of Zhejiang Province: Zhejiang University School of Medicine First Affiliated Hospital

Xiao Xu (✉ zjxu@zju.edu.cn)

Hangzhou First People's Hospital <https://orcid.org/0000-0002-2761-2811>

Research Article

Keywords: Deep learning, MobileNetV2, HCC, Prognosis, LT

Posted Date: November 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1053962/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Hepatology International on March 29th, 2022. See the published version at <https://doi.org/10.1007/s12072-022-10321-y>.

Abstract

Background: There is a growing need for new improved classifiers of prognosis in hepatocellular carcinoma (HCC) patients to stratify them effectively.

Methods: A deep learning model was developed on a total of 1118 patients from 4 independent cohorts. A nucleus map set (n = 120) was used to train U-net to capture the nuclear architecture. The training set (n = 552) included HCC patients that had been treated by resection. The liver transplantation (LT) set (n = 144) contained patients with HCC that had been treated by LT. The train set and its nuclear architectural information extracted by U-net were used to train the MobileNet V2-based classifier (MobileNetV2_HCC_Class). The classifier was then independently tested on the LT set and externally validated on the TCGA set (n = 302). The primary outcome was recurrence-free survival (RFS).

Results: The MobileNetV2_HCC_Class was a strong predictor of RFS in both LT set and TCGA set. The classifier provided a hazard ratio of 3.44 (95% CI 2.01–5.87, $p < 0.001$) for high risk versus low risk in the LT set, and 2.55 (95% CI: 1.64–3.99, $p < 0.001$) when known prognostic factors, remarkable in univariable analyses on the same cohort, were adjusted. The MobileNetV2_HCC_Class maintained a relatively higher discriminatory power [time-dependent area under curve (AUC)] than other factors after LT or resection in the independent validation set. A pathological review demonstrated that tumoral areas with the highest recurrence predictability featured the following features: the presence of stroma, a high degree of cytological atypia, nuclear hyperchromasia, and a lack of immune cell infiltration.

Conclusion: A prognostic classifier for clinical purposes had been proposed based on the use of deep learning on histological slides from HCC patients. This classifier assists in refining the prognostic prediction of HCC patients and identifies patients who have been benefited from more intensive management.

Introduction

Hepatocellular carcinoma (HCC) is the world's seventh-leading solid malignancy and the third primary cause of cancer-related death [1]. The prevalence of HCC is relatively high in the Asia-Pacific countries [2]. Hepatectomy and liver transplantation (LT) are still the main treatment approaches of HCC. Despite significant progress in diagnostic and management techniques for HCC, the recurrence rate is still high (up to 70–80%) following hepatectomy and 20–40% following LT [3–5]. A refinement of prognostic models, especially those based on the accessible data, could easily allow for detection of early warning signs during follow-up and prolonging adjuvant therapeutic decisions [6]. The prognosis of a disease is closely associated with its pathological features. Histological analysis of tumor tissues can certainly provide information for patient stratification and treatment allocation. Histological slides contain a vast amount of information that can be quantitatively assessed by deep learning algorithms. In recent times, convolutional neural networks (CNNs) have been extensively employed in the fields of speech recognition, traffic sign management, and face recognition [8, 9, 10]. CNNs have successfully passed numerous

image interpretation tests and also retrieved extra information from histopathological images. Recently, a pioneering survey revealed that simulating routine pathology workflows, i.e., by using subset algorithms based on deep learning, achieved a better diagnostic performance than an expert group consisting of 11 pathologists in monitoring lymph node metastasis in breast cancers [7]. Coudray et al. have shown in their work that CNNs successfully identified the prime histological subtypes of non-small cell lung cancer and predicted the gene mutation status of genes such as *STK11* and *EGFR* [8]. CNNs are also found to be useful for predicting the aggressiveness of colorectal cancer [9]. More recently, there is growing evidence that suggests the computational processing applied on histological slides better refines prediction for patient prognosis, thus leading to an improvement in treatment allocation. A model based on deep learning developed by Saillard et al. could correctly predict the survival in HCC patients. They used pre-trained CNNs to extract features from images and then the network selected 25 tiles with maximum and minimum scores to predict patient survival [10]. Skrede et al. successfully developed a marker to predict the prognosis of colorectal cancer in large cohorts using MobileNet V2, one of the CNNs [11], building the model using Multiple Instance Learning (MIL). Interestingly, the local spatial arrangement of nuclei in the histopathology images proved to be vital information of high prognostic value in cases of oropharyngeal cancer [12].

In the present study, four independent cohorts of postoperative HCC patients were investigated for developing and validating a MobileNet V2-based model to improve prognosis and prediction. The scientific and innovative features of our method are inspired by two studies: 1) the study by Skrede et al. trained MobileNet V2 using the MIL, which allowed training on large tile collections labeled with the associated whole-slide images [11], and 2) the study by Ji Meng-Yao et al. in which nuclear architectural information was used in building a model, which proved to be effective for cancer grading and prediction of patient outcomes [13]. To capture localized nuclear architectural information, local nuclei measurements were constructed by U-net in the independent cohort. Here we conclude that the models have higher accuracy in survival prediction relative to conventional methods. Our research aimed at using MobileNet V2 to analyze histopathology images and propose an automatic prognostic classifier exclusive to HCC patients that have undergone liver resection. In addition, we validated the prognostic power of MobileNet V2 across different cohorts.

Methods

Patients and samples

Four different cohorts were enrolled in this study. Stained tumor tissue sections obtained from patients with adequate quality and titles were used. The first cohort was used to train U-net to capture the localized nuclear architectural information (Nucleus map set, n = 120). The second cohort was from HCC patients who had received surgical resection treatment from the First hospital of Zhejiang University in 2010-2016 and have a so-called distinct outcome (Train set, n = 552). Patients exhibiting obvious good or bad outcomes (good: 274, poor: 278) were used as training cohorts. Patients with a four-year follow-up history after resection and no recorded recurrence were grouped under the good outcome cohort. At the

same time, the bad outcome group included patients who relapsed within a period of 1.6 years (exclusive) after surgery. The third cohort was from HCC patients who received liver transplantation from the First hospital of Zhejiang University between 2015 and 2019 (LT set, $n = 144$). The nucleus map set, train set, and LT set were collated from these three different batches of HCC patients after obtaining approval from the ethics committee of the institution. A fourth dataset, namely the TCGA set, with complete follow-up data ($n = 302$) from the TCGA database, was included for external validation.

The nucleus map set, train set, and LT set were scanned and digitized using a P250FLASH2 (3DHISTECH3) at 20x magnification. Nucleus map sets were used for training U-net. The train set was used to train the MobileNet V2, while the LT set was used to externally validate the model in HCC treated by LT. The histology slides, clinical follow-up data, and histological annotation were retrieved from the TCGA database (<https://cancergenome.nih.gov/>).

Tile cropping and color normalization

Due to the limitations of graphic card memory, it was almost impossible to process whole-sliced pathological images, which are usually at a resolution of 100000 X 100000, on GPU or main memory during the training phase. To circumvent this problem, the current best practice is to cut large images into hundreds of smaller images, which are called tiles or patches [11; 14; 15]. In our study, these tiles were 512×512 pixels (px) and $0.25 \mu\text{m}$ per px. They were cropped from the nucleus map set, train set, and LT set. They were finally resized to a resolution of 224×224 px. The tiles were normalized as described previously [16].

Extending features with a segmentation heat map of nuclear architectural information using U-net

Before feeding these data into the model, we used a trained image segmentation model to get the heat map of nuclei segmentation for each tile. The segmentation model was a U-net neural network trained with a nucleus map set. Let I denote an image slice, p indicate the U-net output, and y denote the ground-truth label in the image slice where $\epsilon = 0.00000001$ is a smoothing term to make the denominator non-zero. The loss function is Dice loss (1) and the final Dice Score on the TCGA test set can reach up to 82%. The segmentation result is not desired to be too perfect, since information other than nuclei, such as cytoplasm and shape of the whole cell, also contributes to the heat map.

$$\{L\}_{\text{dice}} = 1 - 2 \times \frac{\sum_{i \in I} \{p\}_i \{y\}_i + \epsilon}{\sum_{i \in I} \{p\}_i + \sum_{i \in I} \{y\}_i + \epsilon} \quad (1)$$

Realization of MIL in the MobileNet V2

The main guiding methodology in our work is MIL, which is a kind of weak supervised learning method to deal with a lack of annotations. All the tiles could be fed to train the learning model. However, such an approach has serious drawbacks during classification. In many cases, the content of one small tile conflicts with the label of the original pathological image, especially in HCC cases with great heterogeneity. To solve this problem, MobileNet V2 was developed using MIL for training only on tile

collections that carried a label for the associated whole-slide image. In this way, we could use MIL to take information on features from every tile. Instead of annotating each tile with its ancestor's label and dumping it into the network directly, we packed all the tiles into a bag with a label identical to the original pathological image. Each bag, which represents one pathological image, was then passed through a trained neural network to calculate the scores of each tile in the bag, and an aggregational function was used to produce a weight-average score for the whole bag. By setting a threshold, the pathological image was classified into a certain class.

Each 224×224 tile was color-normalized using the method described by Vahadane [17]. After nuclei segmentation, the color-normalized RGB tiles were then concatenated with their heat map in channel level to produce a 4-channel tile. Then, these bags of 4-channel tiles were dumped into a feature extractor, which is a MobileNet V2 model, and the score of each tile was calculated. A generalized mean was used as the aggregation function since it could keep the extremes while taking into account the average. The aggregation function reads as (2), where p is a hyperparameter.

$$\text{S} = \left(\sum \text{s}_i^{\text{p}} \right)^{\frac{1}{\text{p}}} \quad (2)$$

The output of the aggregation function, which represents the score of the pathological image, was activated by a sigmoid function and compared with a given threshold t , where t is also a hyperparameter. Based on this, the image was finally classified into a certain class.

Training strategy

During the training process, we deployed a decay learning rate, which was initiated with 0.0001 and halves every 10 epochs. Due to the limitation of GPU memory, the training batch size could only be set to 1. Besides, threshold t was 0.4457, and aggregation function p was 3. Cross-entropy with L2 regulation (3)-(4) was selected as the loss function, and regularized factor α was 0.02.

$$\text{p}_i = \begin{cases} \frac{\text{t} - \text{s}_i}{\text{t} + 1} * 0.5, & \text{s}_i \leq \text{t} \\ \frac{\text{s}_i - \text{t}}{1 - \text{t}} * 0.5, & \text{s}_i > \text{t} \end{cases} \quad (3)$$

$$\text{L} = \left(\text{p}_i \right), \left(\text{y}_i \right), \left(\text{W} \right) = -\frac{1}{\text{N}} \sum \left[\text{y}_i * \log \left(\text{p}_i \right) + \left(1 - \text{y}_i \right) * \log \left(1 - \text{p}_i \right) \right] + \alpha \sum \text{w}_i^2 \quad (4)$$

Analysis on tiles with high predictive value

To deepen our understanding of features related to tumor aggressiveness, tiles having high and low-risk scores were retrieved for further in-depth analysis. Altogether four histological features of tumoral liver tissues have been systematically documented.

Statistical analysis

Log-rank tests were performed to compare the stratification of patients into subgroups in terms of survival distribution. Time-dependent AUC (area under curve) was adopted as an index to assess the proposed model's performance and the baseline clinical, biological, and pathological features. Statistical analyses were carried out with R (version 3.6.0) using ggplot2, survival, and Survminer packages. The training and deployment of CNNs were conducted with Python using a standard desktop workstation (Nvidia Tesla P40 GPUs each with 24GM memory). P-value < 0.05 indicated statistical significance.

Results

Patient Characteristics and Model development

The nucleus map set was used to train U-net to capture localized nuclear architectural information (n = 120). The other three sets that were used for training and validation were: 1) 552 patients from the train set that joined in the development of the model, 2) 144 patients from the LT set, and 3) the TCGA set that was used to externally validate the model; 552 patients from the train set exhibited obvious outcomes (good: 274, bad: 278) and were enrolled for obtaining definite facts. The patients' demographics are presented in Table 1.

Table 1
Baseline characteristics in the nucleus map set, train set, LT set, and the TCGA set.

Variables	Nucleus map set (n=120)	Train set (n=552)	LT set (n=144)	TCGA set (n=302)
Age (year)	59 (49-65)	55(47-63)	52 (45-58)	60 (51-68)
Gender (male)	104(86.7%)	478(86.6%)	130 (90.3%)	208 (68.9%)
AFP (ng/ml)	34.7 (6.6- 708.0)	76.5(7.2-888.0)	49.3 (7.7-1418.3) ^a	11.0 (4.0-231.5) ^b
Grade				
G1	15 (12.5%)	32 (5.8%)	3 (2.1%)	43 (14.2%)
G2	54 (45.0%)	243 (44.0%)	42 (29.2%)	142 (47.0%)
G3	42 (35.0%)	217 (39.3%)	35 (24.3%)	103 (34.1%)
G4	9 (7.5%)	54 (9.8%)	0 (0.0%)	10 (3.3%)
Missing	0 (0.0%)	6 (1.1%)	64 (44.4%)	4 (1.3%)
Total tumor size				
<5cm	79 (65.8%)	334 (60.5%)	35 (24.3%)	
≥5cm	41 (34.2%)	210 (38.0%)	109 (75.7%)	
Missing	0 (0.0%)	8 (1.4%)	0 (0.0%)	302 (100.0%)
Tumor number				
Single	108 (90.0%)	479 (86.8%)	52 (36.1%)	
Multiple	12 (10.0%)	67 (12.1%)	92 (63.9%)	
Missing	0 (0.0%)	6 (1.1%)	0 (0.0%)	302 (100.0%)
Variables	Nucleus map set (n=120)	Train set (n=552)	LT set (n=144)	TCGA set (n=302)
Stage_AJCC				
Stage II	25 (20.8%)	145 (26.3%)	34 (23.6%)	66 (21.9%)
Stage III	5 (4.2%)	53 (9.6%)	63 (43.8%)	69 (22.8%)
Stage IV	0 (0.0%)	12 (2.3%)	21 (14.6%)	3 (1.0%)
Missing	0 (0.0%)	7 (1.3%)	1 (0.7%)	20 (6.6%)
Data are median (IQR) or n (%). a, NA = 1. b, NA = 71.				

Firstly, we used an image segmentation model to get the heat map of nuclei segmentation for each tile. This segmentation model was a U-net neural network trained using the nucleus map set. The loss function was Dice and the final Dice Score for the nucleus map set could reach 82%. The segmentation result was not desired to be too precise, since information other than nuclei, such as cytoplasm and shape of the whole cell, was also accountable in the heatmap. A total of 57415 tiles (small image patches with 224×224 pixels) were extracted from the train set (good: 28534, poor: 28881). The pre-trained U-net was used to get a heat map of nuclei segmentation for each tile before finally training our model. We concatenated the heat map of nuclei segmentation and the color-normalized RGB tiles at channel level and produced a 4-channel tile. The bags containing 4-channel tiles were then dumped onto a feature extractor of the MobileNet V2 model. We used a generalized mean with a sign as the aggregation function since it could keep the extremes while simultaneously taking the average into account. The output of the aggregation function, which represents the score of the pathological image was activated using a sigmoid function and then compared with a given threshold of 0.4457, where 0.4457 is a hyperparameter. Finally, the images are classified into certain a class based on their scores. The pipeline for MobileNet V2 HCC classification (MobileNet V2_HCC_Class) is shown in Figure 1.

The model generalized to LT for the HCC dataset

The output of our neural networks could categorize patients into low-risk and high-risk subgroups. In the LT set, 144 patients with complete follow-up data were included, of which 65 patients relapsed during follow-up. The available variables for analysis are age at diagnosis, gender, serum alpha-fetoprotein (AFP), Child-Pugh score, the model for end-stage liver disease (MELD), tumor size, tumor number, grade, and tumor stage according to the American Joint Committee on Cancer (Stage AJCC). Univariable analyses indicated that the variables AFP, tumor size, grade, tumor number, and Stage AJCC were associated with a shorter RFS (Table S1). Tiles from the tissue array of these patients were retrieved and processed under the proposed model. The MobileNetV2_HCC_Class was a strong predictor of RFS in the whole LT set and was even capable of stratification of other common prognostic features (Stage AJCC, AFP, tumor number, and tumor size) (Figure 2).

Multivariate analyses showed that the MobileNetV2_HCC_Class had an independent prognostic value (HR = 3.44 (2.01–5.87), $p < 0.001$) after adjusting known prognostic markers remarkable in the univariable analyses, such as Stage AJCC, AFP, tumor number, and tumor size (Figure 3A). The time-dependent AUC curves are depicted in Figure 3B. During the entire course of the 3-year follow-up, the MobileNetV2_HCC_Class maintained relatively higher AUC values than the other factors in the first two years after LT.

The model generalized to the TCGA dataset

The robustness of our model was evaluated on an independent series from the TCGA. 302 patients satisfied the inclusion criteria, and 165 of them with recurrence were recorded. The slides were gathered from various centers. The available variables that were entered for analysis are age at diagnosis, age, gender, AFP, vascular invasion, stroma tumor ratio (STR), tumor-infiltrating lymphocyte (TIL), grade, and

Stage AJCC. The clinical, biological, and pathological feature most related to a shorter survival should be the AJCC stage in univariable analyses (Table S2). Tiles from WSIs of the 302 patients were retrieved and processed under the proposed model. In the TCGA set, MobileNetV2_HCC_Class predicted the RFS while also following the stratification of other significant prognostic features like Stage AJCC, AFP, grade, or vascular invasion (Figure 4).

The classifier seemed strong during multivariable analysis (HR = 2.55 (1.64–3.99), $p < 0.001$), upon adjusting the known prognostic markers remarkable in univariable analyses, such as Stage AJCC, AFP, grade, and vascular invasion (Figure 5A). The results show that the model can capture complicated non-redundant patterns in which baseline variables influence HCC patients' survival. The time-dependent AUC curves are depicted in Figure 5B. During the entire course of the 6-year follow-up, the MobileNetV2_HCC_Class maintained relatively higher AUC values than other factors after HCC resection.

Histological analysis of tiles

The MobileNetV2_HCC_Class could extract tiles with the highest predictability from thousands of tiles. The prime histological features related to recurrence could be surveyed by retrieving 400 tiles with the highest predictability (high recurrence risk: 200, low recurrence risk: 200) among 302 patients of the TCGA with MobileNetV2_HCC_Class. Four such histological features were found from tumoral areas. The presence of stroma, high degree of cytological atypia, and nuclear hyperchromasia were related to high risk ($p = 0.0003$, $p = 0.0010$, $p = 0.0012$, respectively), while immune cell infiltration was associated with low risk ($p = 0.0019$) (Figure 6, Table S3). The above findings show that the proposed deep learning model detects established histological patterns related to recurrence among HCC patients.

Discussion

Based on the latest developments in deep learning, this study proposed the use of MobileNetV2_HCC_Class for automatic prognosis prediction in HCC patients. This would enable automated analyses of standard histological sections stained with H&E. These algorithms exhibit a higher accuracy in prognosis prediction relative to classical clinical, biological, and pathological features.

Deep learning-driven methods in medical image processing have proven to be extremely useful in standardizing cancer diagnosis as well as in improving patient stratification [7; 18]. Recently, a pioneering survey reported that deep learning-based models can detect and categorize lung cancer cases with accuracy similar to that of pathologists [8]. Previous studies suggest that deep learning is highly efficient in developing markers, which utilize basic morphology for the prediction of outcomes in cancer patients [19; 20]. A deep learning-based model developed by Coudray et al. could even predict six of the most frequent genetic alterations directly from the slides [8]. In cases of gastrointestinal cancer, a deep learning-based model could directly estimate microsatellite instability based on just histological images [15]. Kather et al. also reported that a CNN could extract the tumor components and predict patient survival directly from histology images [9]. Saillard et al. predicted the survival of HCC patients by extracting features from images using a pre-trained CNN, following which the network selected 25 tiles

having maximum and minimum scores to predict survival [10]. In our study, a different method was used to develop the MobileNetV2_HCC_Class to improve the prediction of prognosis in HCC patients treated by surgical resection and LT. The innovative features of our method were: 1) random tiles were used for each patient, like Skrede [11], 2) the MobileNet V2 was trained using MIL, which allowed for training on tile collections labeled with the associated whole-slide image, and 3) the use of nuclear architectural information in building of the model, which proved to be efficient for cancer grading and prediction of patient outcomes [13]. Genetic instability was demonstrated through diversifying nuclear shape and texture, which had a major effect on metastasis and proliferation that might lead to cancer recurrence. The MobileNetV2_HCC_Class proved to be a strong predictor of RFS in HCC patients treated with resection or LT and generalized in the TCGA set across different centers.

It is well established that molecular and/or genetic features can predict the survival of patients with HCC [21; 22]. Chaudhary et al. had adopted deep-learning methods for RNA sequencing and methylation data from the TCGA database and successfully predicted HCC survival in multiple patient cohorts. High-throughput gene expression profiling/sequencing techniques are restrictive due to high cost and poor reproducibility clinically. The proposed method uses merely the histological slides that are routinely available at surgical treatment centers. Additionally, the processing and computing time in this approach is brief enough to avoid delay in therapy. Therefore, this method facilitates the easy application of the risk stratification system clinically.

To sum up, we successfully built a prognostic model for clinical use based on deep learning approaches applied on histological slides from patients. The model was widely assessed among independent patient populations receiving different types of treatments and gave consistently excellent results across the classical clinical, biological, and pathological features. The proposed CNN-based approach can potentially improve patient prognosis evaluation and help guide clinicians in their decision-making process about the use of adjuvant therapy on their patients.

Declarations

Author contributions

LZ, XX ZS and XX designed the study. ZW, HY, WJw, WJg had full access to all the data. LZ and LY analyzed the data, wrote the manuscript, and were responsible for the decision to submit the manuscript. All authors provided substantial comments on drafts and approved the final report.

Funding

This research gained support from the Key Program of the National Natural Science Foundation of China (No. 81930016)) and the National Natural Science Foundation of China (No. 81802889). Research results are partially based on data from the TCGA database.

Availability of data and material

The data that support the findings of this study are available from the corresponding author on reasonable request.

Conflict of interest

Zhikun Liu, Yuanpeng Liu, Wenhui Zhang, Yuan Hong, Jinwen Meng, Jianguo Wang, Shusen Zheng and Xiao Xu have nothing to disclose.

Ethics approval

Ethics approval was obtained by the Ethics Committee of Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine.

Consent to participate

All patients gave written informed consent.

Consent for publication

All authors approved the manuscript for publication.

Code availability

Not applicable.

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68(6):394-424.
2. Omata, M., Cheng, A.L., Kokudo, N., Kudo, M., Lee, J.M., Jia, J., et al., 2017. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. *Hepatol Int* 11(4):317-370.
3. Kudo, M., 2017. Systemic Therapy for Hepatocellular Carcinoma: 2017 Update. *Oncology* 93 Suppl 1:135-146.
4. Llovet, J.M., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwartz, M., Sherman, M., et al., 2016. Hepatocellular carcinoma. *Nat Rev Dis Primers* 2:16018.
5. Llovet, J.M., Montal, R., Villanueva, A., 2019. Randomized trials and endpoints in advanced HCC: Role of PFS as a surrogate of survival. *J Hepatol* 70(6):1262-1277.
6. Fujiwara, N., Friedman, S.L., Goossens, N., Hoshida, Y., 2018. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. *J Hepatol* 68(3):526-549.

7. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., et al., 2017. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *Jama* 318(22):2199-2210.
8. Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., et al., 2018. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 24(10):1559-1567.
9. Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., et al., 2019. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med* 16(1):e1002730.
10. Saillard, C., Schmauch, B., Laifa, O., Moarii, M., Toldo, S., Zaslavskiy, M., et al., 2020. Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides. *Hepatology*.
11. Skrede, O.J., De Raedt, S., Kleppe, A., Hveem, T.S., Liestøl, K., Maddison, J., et al., 2020. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 395(10221):350-360.
12. Lu, C., Koyuncu, C., Corredor, G., Prasanna, P., Leo, P., Wang, X., et al., 2020. Feature-driven local cell graph (FLock): New computational pathology-based descriptors for prognosis of lung cancer and HPV status of oropharyngeal cancers. *68*:101903.
13. Ji, M.Y., Yuan, L., Jiang, X.D., Zeng, Z., Zhan, N., Huang, P.X., et al., 2019. Nuclear shape, architecture and orientation features from H&E images are able to predict recurrence in node-negative gastric adenocarcinoma. *J Transl Med* 17(1):92.
14. Corredor, G., Wang, X., Zhou, Y., Lu, C., Fu, P., Syrigos, K., et al., 2019. Spatial Architecture and Arrangement of Tumor-Infiltrating Lymphocytes for Predicting Likelihood of Recurrence in Early-Stage Non-Small Cell Lung Cancer. *Clin Cancer Res* 25(5):1526-1534.
15. Kather, J.N., Pearson, A.T., Halama, N., Jager, D., Krause, J., Loosen, S.H., et al., 2019. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*.
16. Tam, A., Barker, J., Rubin, D., 2016. A method for normalizing pathology images to improve feature extraction for quantitative pathology. *Med Phys* 43(1):528.
17. Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., et al., 2016. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Trans Med Imaging* 35(8):1962-1971.
18. Niazi, M.K.K., Parwani, A.V., Gurcan, M.N., 2019. Digital pathology and artificial intelligence. *Lancet Oncol* 20(5):e253-e261.
19. Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega, J.E., et al., 2018. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A* 115(13):E2970-e2979.
20. Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P.E., Verrill, C., et al., 2018. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 8(1):3395.

21. Zucman-Rossi, J., Villanueva, A., Nault, J.C., Llovet, J.M., 2015. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology* 149(5):1226-1239.e1224.
22. Pan, Y., Chen, H., Yu, J., 2020. Biomarkers in Hepatocellular Carcinoma: Current Status and Future Perspectives. *Biomedicines* 8(12).

Figures

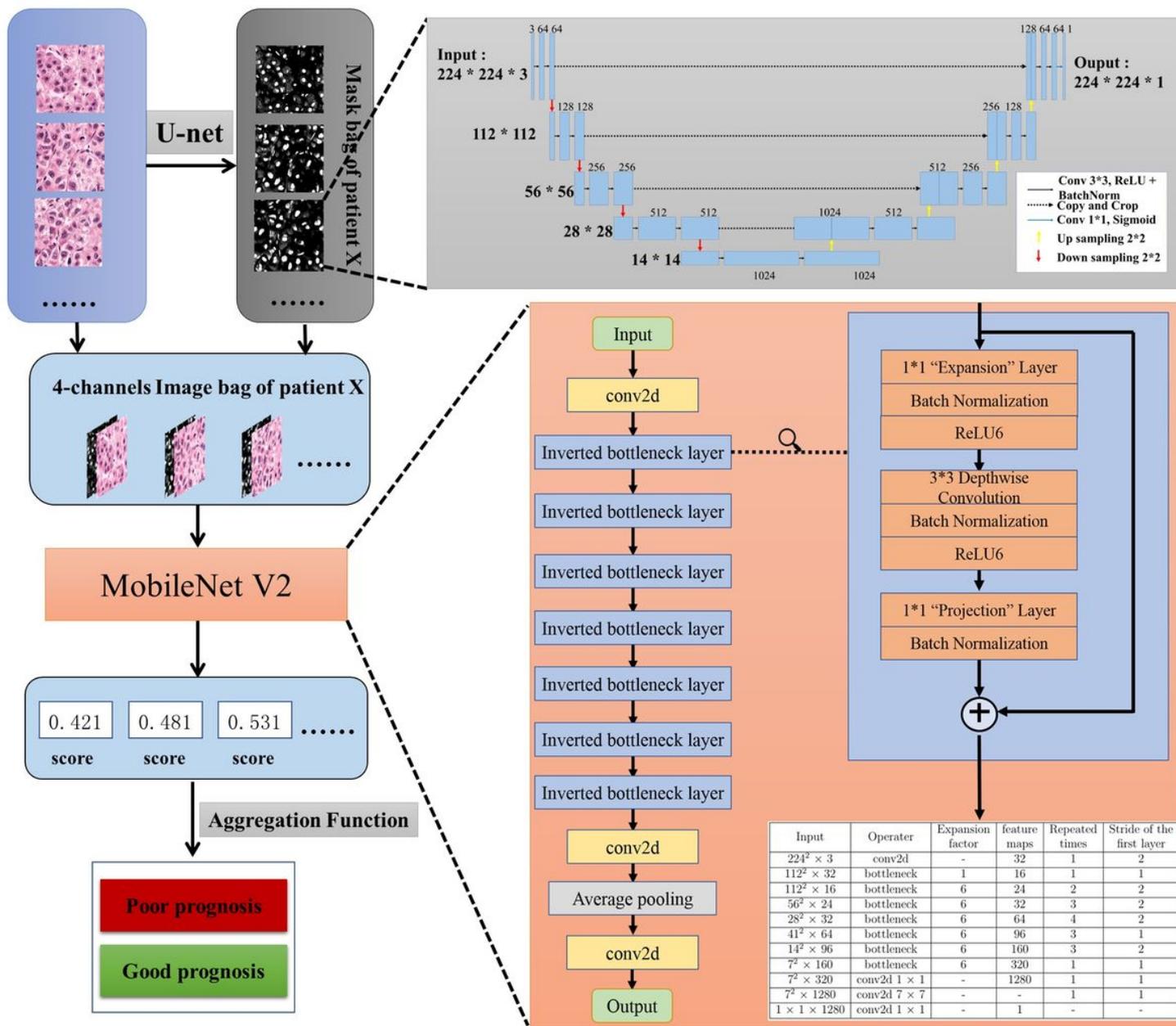


Figure 1

The pipeline for MobileNetV2_HCC_Class. From the small image patches in 224×224 pixels of the train set, the heat map of nuclei segmentation for each tile was obtained using a pre-trained U-net. The heatmap of nuclei segmentation and the color-normalized RGB tiles were concatenated at channel level,

and a 4-channel tile was obtained. Bags containing 4-channel tiles were then dumped into a feature extractor of the MobileNetV2 model. A generalized mean with a sign was used as the aggregation function.

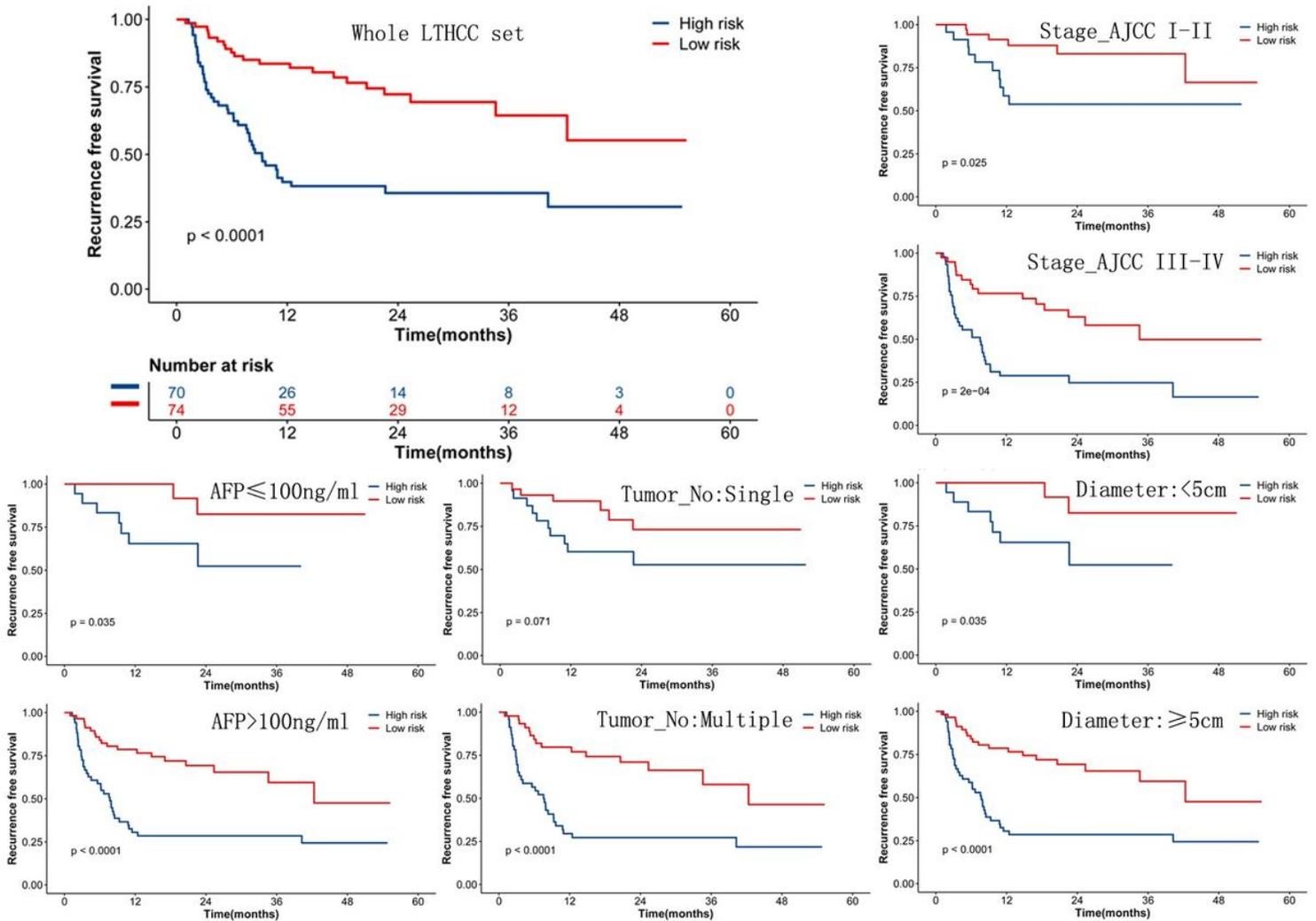


Figure 2

Prognostic value of MobileNetV2_HCC_Class in LT set and the stratification of common prognostic variables. MobileNetV2_HCC_Class categorized patients into low-risk and high-risk subgroups. The prognostic value for MobileNetV2_HCC_Class was conservative, even following the stratification of common clinical and pathological variables. AFP: alpha-fetoprotein; Tumor No: tumor number.; Diameter: total tumor diameter.

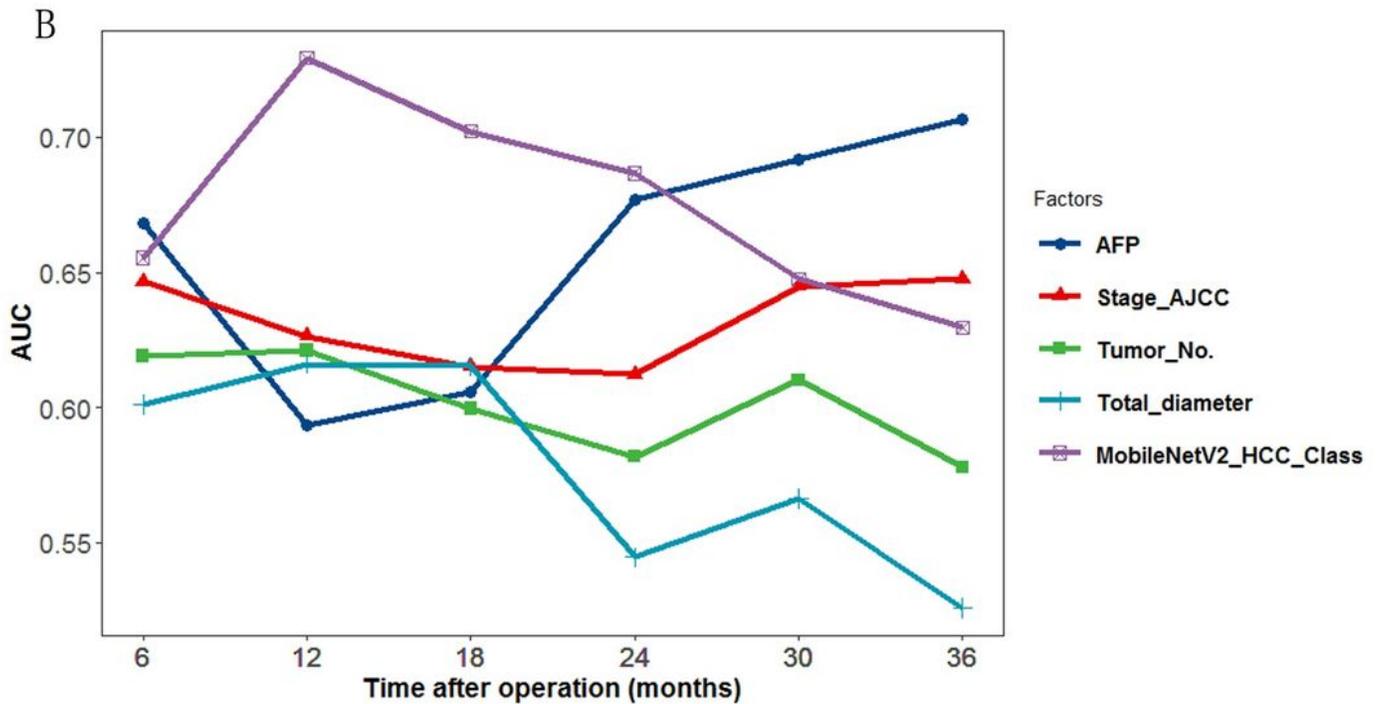
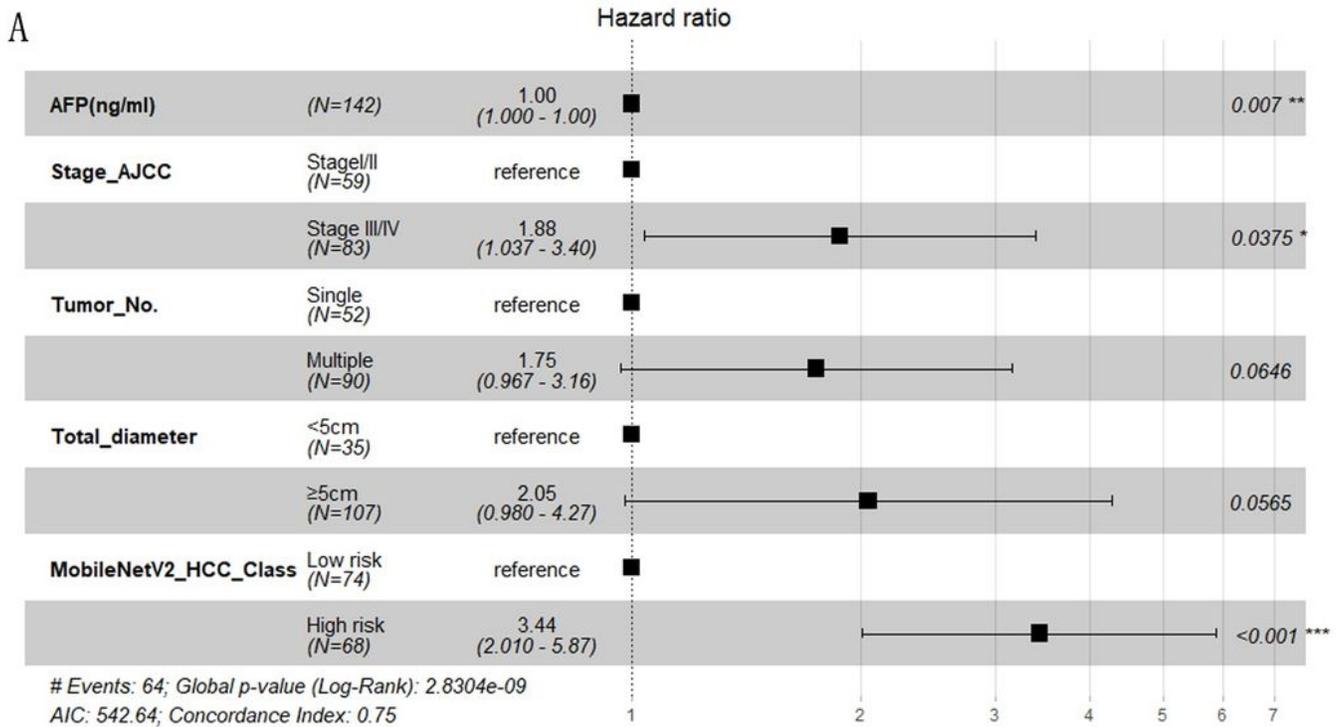


Figure 3

The performance of different risk factors for tumor recurrence after LT. A. Multivariate analyses. B. The time-dependent area in the receiver operating characteristic curve (AUC) value for different criteria based on tumor recurrence.

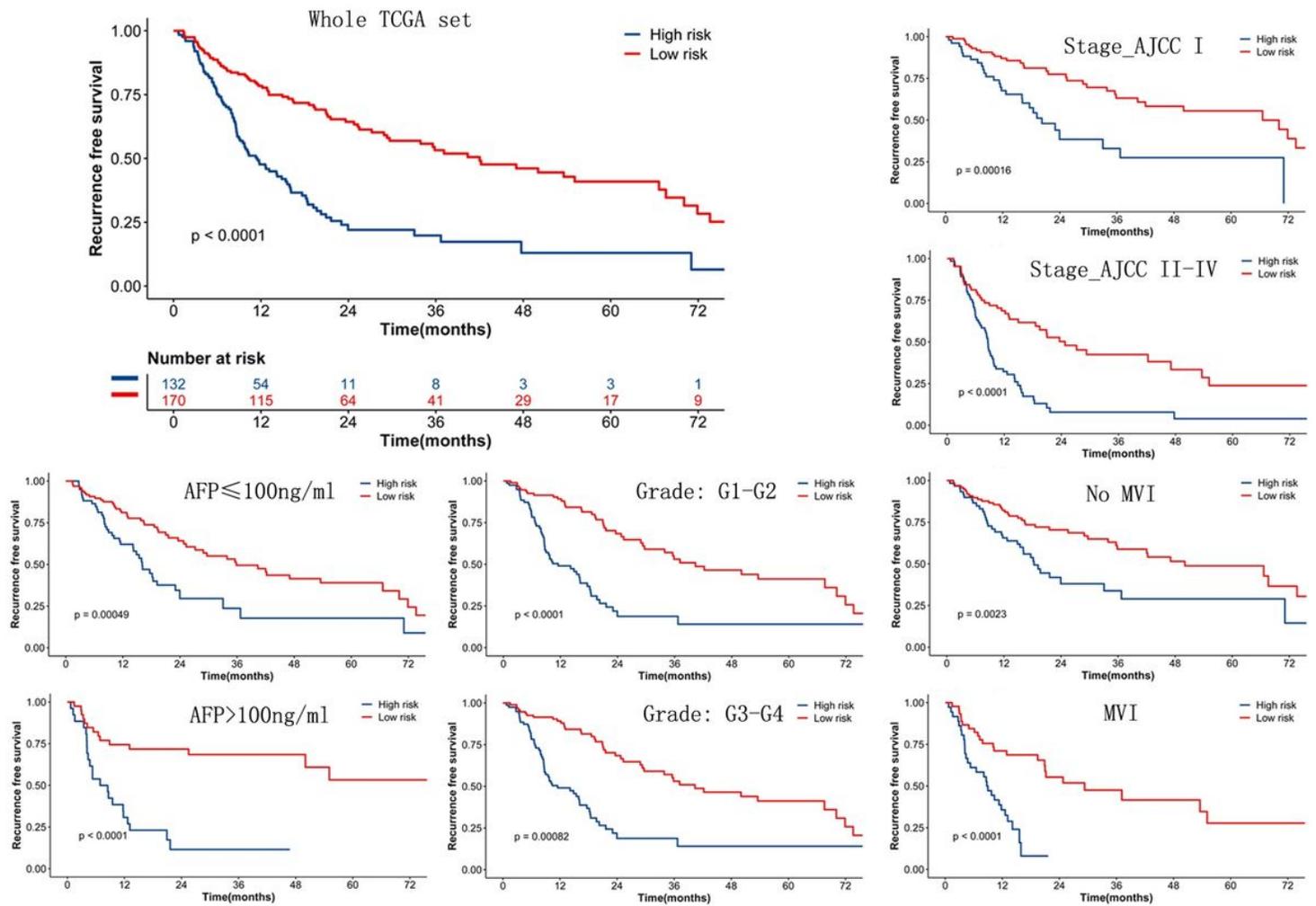


Figure 4

Prognosis of MobileNetV2_HCC_Class in TCGA set and the stratification of common baseline variables. MobileNetV2_HCC_Class predicts RFS while also following the stratification of other common baseline variables.

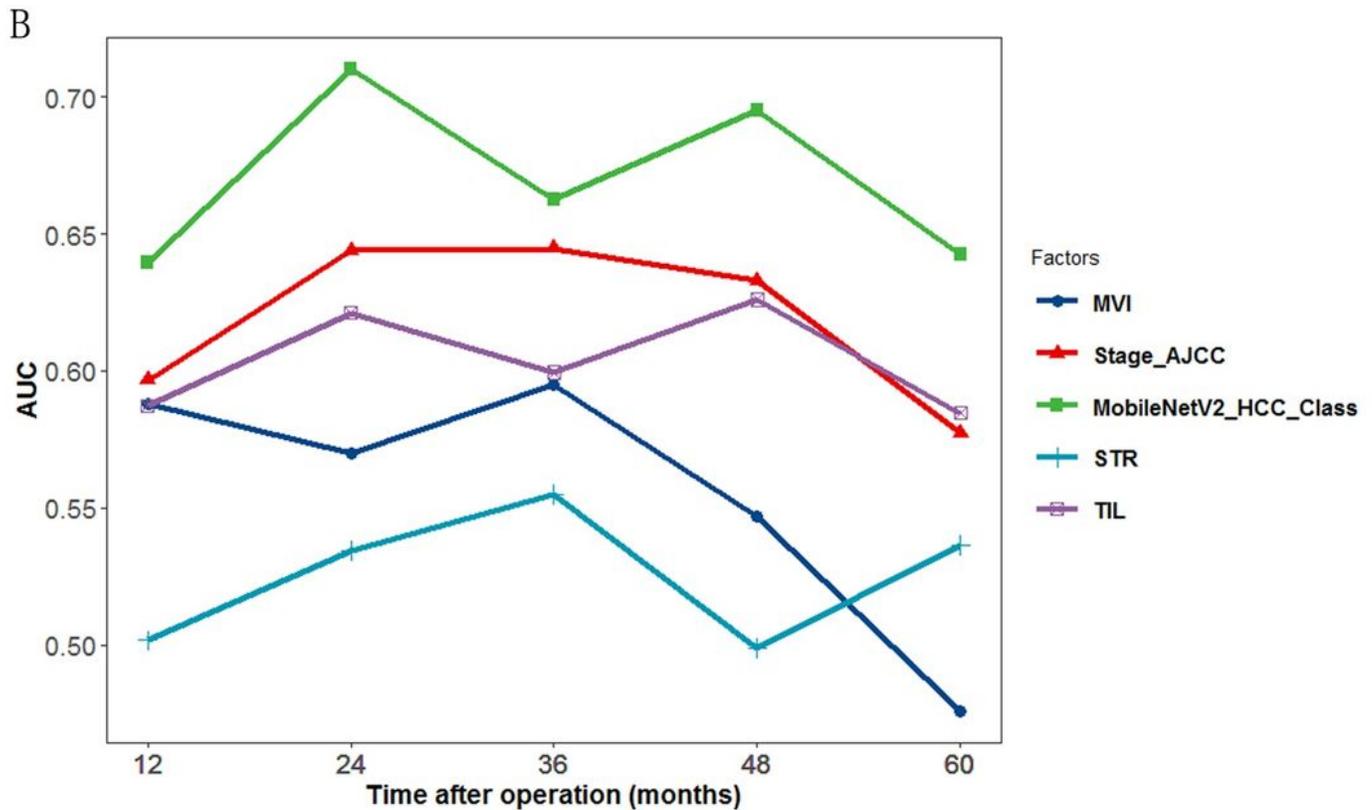
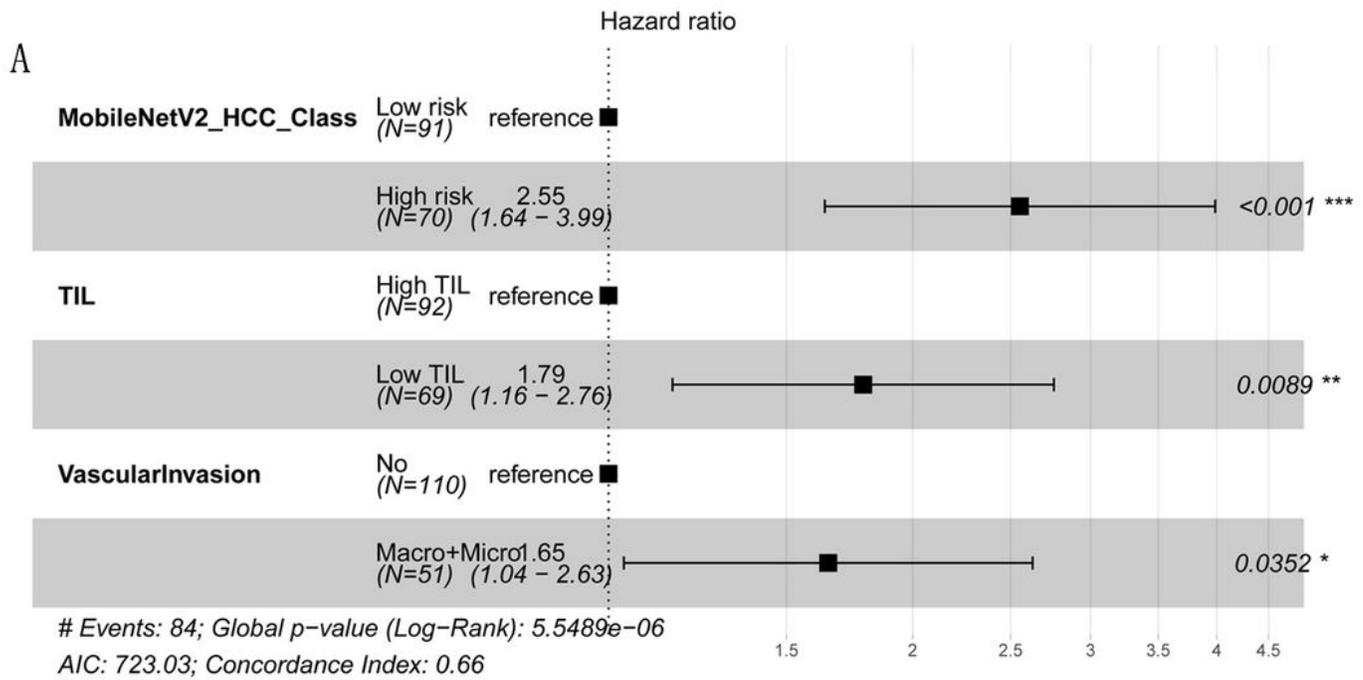


Figure 5

Performance of different risk factors in tumor recurrence after resection. A. Multivariate analyses. B. The time-dependent area in the receiver operating characteristic curve (AUC) value for different criteria based on tumor recurrence.

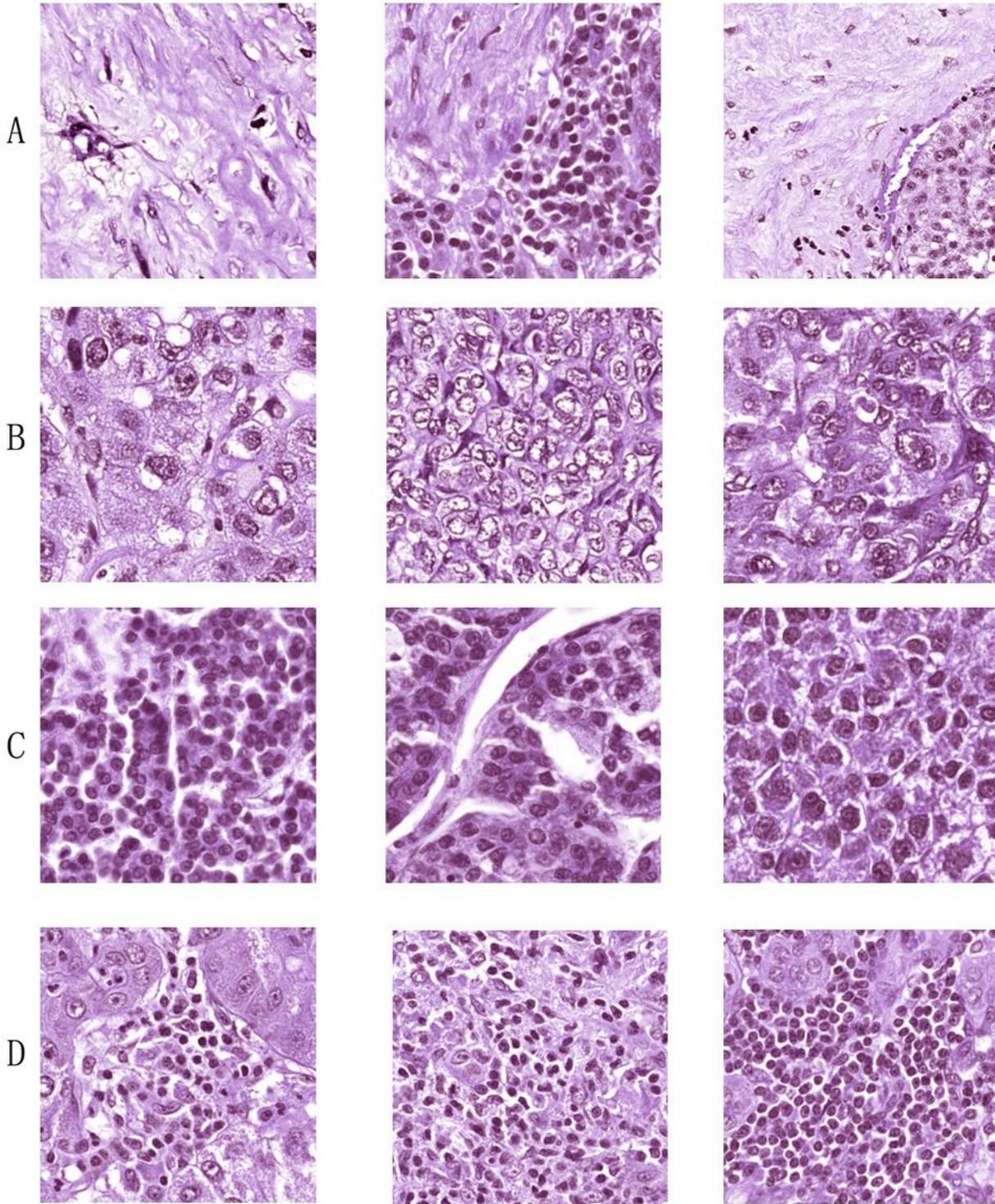


Figure 6

Typical tiles were found to have low or high risks using MobileNetV2_HCC_Class. Four hundred tiles with the highest predictability were investigated. The features used to predict high recurrence risk were stroma (A), cellular atypia (B), and nuclear hyperchromasia (C). The feature indicating low recurrence risk was the existence of immune cells (D).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial1.docx](#)