Supplementary Materials

Summary of contents

Supplementary Figure 5. Prediction profiles of pathomics-driven ensemble model

Supplementary Figure 6. Comparison of prediction performance between the pathomics-driven ensemble model and single kind of pathomics features models in the training cohort and internal validation cohort

Supplementary Figure 7. Overall survival Kaplan-Meier curve analysis of prediction populations

Supplementary Figure 8. Progression-free survival Kaplan-Meier curve analysis of actual populations

Supplementary Figure 9. Overall survival Kaplan-Meier curve analysis of actual populations

Supplementary Figure 10. Forest plot for the Univariate cox regression analysis of progression-free survival

Supplementary Figure 11. Kaplan-Meier survival analysis according to the prediction populations stratified by clinicopathological risk factors

Supplementary Figure 12 SHAP heatmap plot

Supplementary Figure 13. Heatmaps of the important features correlation.

Supplementary Figure 14. SHAP decision plot

Section 4. Supplementary Reference

Section 1. Supplementary Methods

1.1 Study design and participants

This study retrospectively selected patients with histologically confirmed gastric cancer who received anti-PD-1 immunotherapy at Sun Yat-sen University Cancer Center (SYUCC) in Guangzhou, China, in order to predict the response to anti-PD-1 immunotherapy and assess its clinical outcomes. Patients were included if they met the following criteria: (1) histologically confirmed gastric adenocarcinoma; (2) treatment with anti-PD-1 or anti-PD-L1 inhibitors for at least two cycles, either as monotherapy or combined with chemotherapy; (3) available complete information about clinicopathological characteristics and follow-up data; (4) available high-quality H&E-stained slides that were obtained before the start of immune checkpoint inhibition. and (5) no history of other malignancies. Patients were excluded if the clinical response could not be evaluated. The identical inclusion and exclusion criteria were applied to patients with gastric cancer undergoing immunotherapy at Nanfang Hospital of Southern Medical University (SMU) and Guangdong Provincial Hospital of Chinese Medicine (GPHCM), both located in Guangzhou, China, to establish external validation cohorts. Pipeline of pathomics feature computations.

1.2.1 Image acquisition and processing

The H&E-stained slides of all patients included in this study were prepared using formalin-fixed paraffin-embedded samples. The director of the Department of Pathology selected sections that were most representative of the depth of invasion in

each case. These selected slides were then scanned using the Aperio ScanScope Scanner system (Leica Biosystems) with the ×40 objective, and the resulting images were digitized as svs. format files. The Aperio ImageScope software (version 12.4.6) was used to manage these files.

To ensure adequate image quality, all whole slide images (WSIs) were reviewed. Whenever possible, WSIs at 40× magnification (0.25 μm/pixel) were processed and analyzed. In some cases, slides were scanned at 20× magnification and the corresponding images were used. Tissue segmentation was performed using the publicly available CLAM repository[1]. The resulting regions of interest (ROIs) were then carefully examined and refined by two expert pathologists using ImageScope software (Supplementary Figure 2).

For a subset of the WSIs, there were regions with pen marks, folds, and blurred artifacts from TCGA. To address this, we utilized the Openslide software to down-sample the whole-slide images by a factor of 32. We then applied appropriate color filters to remove these regions with pen marks, folding, and blurring artifacts (https://github.com/histolab/histolab).

1.2.2 Pathomics feature extraction

In order to develop a pathomics-based model, three types of quantitative pathomics features were extracted, including nucleus features, single-cell spatial distribution features, and deep microenvironment features. These categories provide comprehensive information on individual cell morphology, cellular spatial distribution, and the overall microenvironment of the tumor (Supplementary Figure 2).

First, features related to tumor nuclei were extracted. After segmenting tumor nuclei using a HoverNet model[2], three categories of nucleus features were extracted, including nuclear intensity, morphology, and texture features. This was done using the 'MeasureObjectIntensity', 'MeasureObjectSizeShape', and 'Measure Texture' modules in the CellProfiler platform[3]. The extracted features were then aggregated using statistical measures such as mean, median, standard deviation, 25th quantiles, and 75th quantiles for each region of interest (ROI) in the slide. In total, 525 pNUC features were generated for each patient.

Second, deep microenvironment features were extracted. Image patches of size 256x256 were extracted from all identified tissue regions without overlap. These patches were then encoded into 1024-dimensional feature vectors using a ResNet50 model pretrained on ImageNet. This encoding was performed by applying spatial average pooling after the 3rd residual block. To expedite this process, multiple GPUs were utilized for parallel computation with a batch size of 256 per GPU.

Finally, single-cell spatial distribution features were extracted. For each whole slide image (WSI), a ROI image of size 8192x8192 pixels was cropped at 40x magnification using Openslide. A HoverNet model[2] pretrained on the Pannuke dataset[4] was employed to segment and classify cells in the ROI, including tumor cells, lymphocytes, stromal cells, dead cells, and non-neoplastic epithelial cells. The number of tumor cells, lymphocytes, and stromal cells per unit square was computed on a 16x16 μm^2 grid to generate an RGB image. In this image, the red, green, and blue channels represent the density maps of tumor cells, lymphocytes, and stromal cells, respectively. The same ResNet50 model used for tumor microenvironment feature extraction was applied to capture different cell types and their spatial organization patterns in the RGB image.

Section 2. Supplementary Tables

Supplementary Table 1. Comparison of predictive performance among ensemble models combining three different base models in multiple cohorts

Ensemble model	Training cohort AUC (95% CI)	Internal validation cohort AUC (95% CI)	External validation cohort 1 AUC (95% CI)	External validation cohort 2 AUC (95% CI)
KNN +LASSO+ DT+ RF	0.985(0.971-0.999)	0.921(0.839-0.999)	0.914(0.837-0.990)	0.927 (0.802-0.999)
SVM+LASSO+ DT+ RF	0.943(0.906-0.981)	0.879(0.774-0.983)	0.8226(0.7861-0.859)	0.9115(0.787-0.999)
LR+LASSO+ DT+ RF	0.927(0.876-0.977)	0.788(0.641-0.934)	0.6738(0.6388-0.709)	0.6052(0.5781-0.7323)

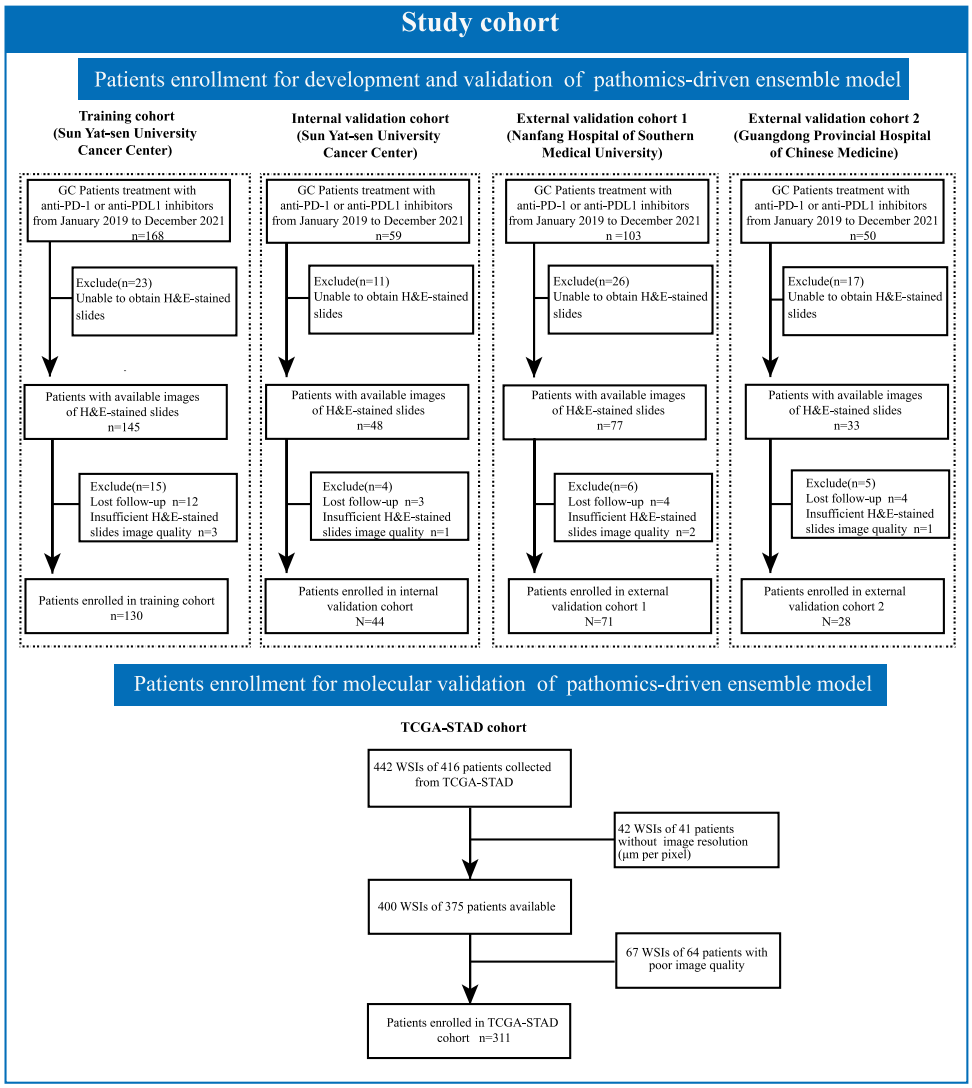
Data are mean (95% CI). AUC, area under the curve; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; KNN, k-nearest neighbors; LR, logistic regression; DT, decision trees; KNN, k-nearest neighbors; RF, random forests.

Supplementary Table 2. Delong test for ROC improvements of pathomics-driven ensemble model compared to individual prediction models and CPS in multiple cohorts

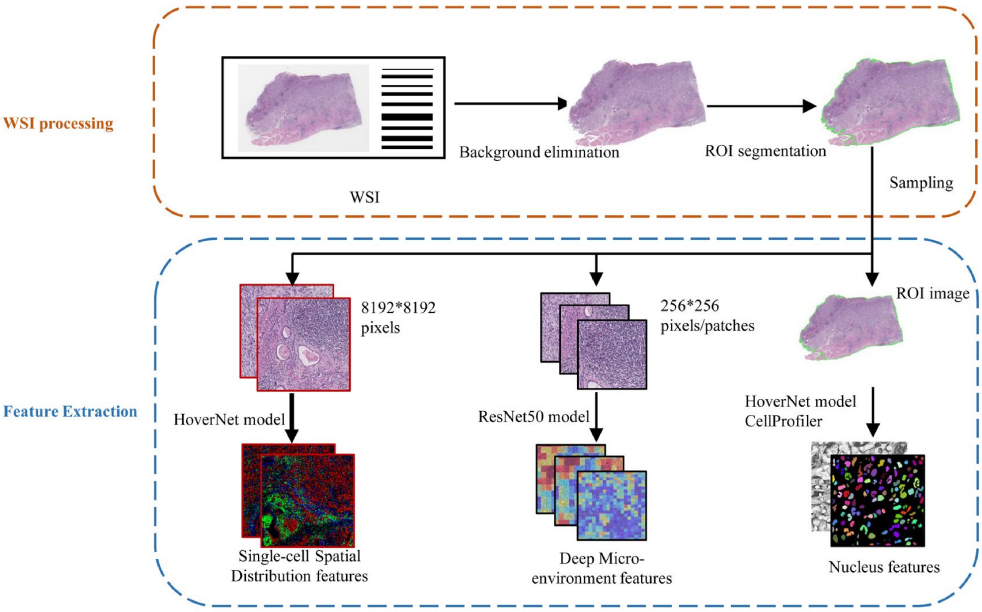
Model	Training cohort	p value	Internal validation cohort	p value	External validation cohort 1	p value	External validation cohort2	p value
Pathomics-driven ensemble model	/	/	/	/	/	/	/	/
LASSO	3.074	0.002	3.130	0.001	3.803	0.000	1.902	0.057
Logistic Regression	8.998	<0.001	6.210	<0.001	9.060	<0.001	2.952	0.003
SVM	3.004	0.003	3.896	<0.001	4.822	<0.001	2.866	0.004
Decision Tree	3.060	0.002	2.638	0.008	2.396	0.017	3.391	<0.001
Random Forest	4.830	<0.001	2.927	0.003	1.739	0.082	2.011	0.044
KNN	5.559	<0.001	2.851	0.004	4.209	<0.001	4.380	<0.001
CPS	8.24	<0.001	4.1568	<0.001	6.0171	<0.001	3.94	<0.001

Note: Data were metric value. LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; KNN, k-nearest neighbors; CPS, combined positive score of PDL1 expression. p<0·05 indicated significant difference between models in the test.

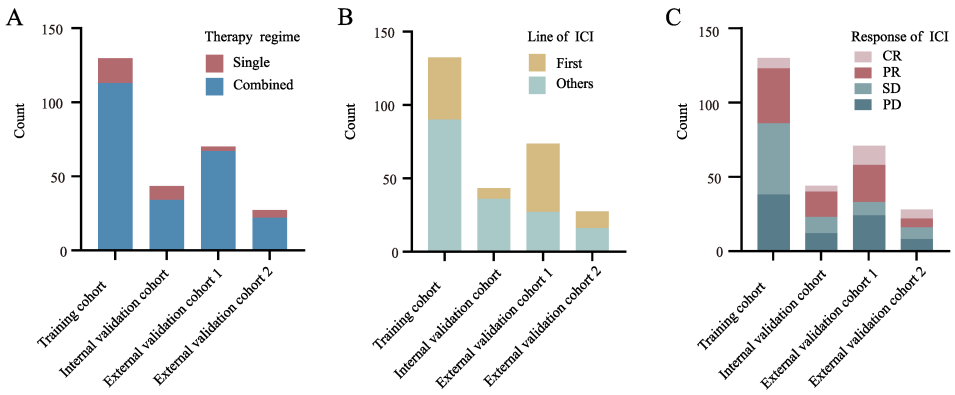
Section 3. Supplementary Figures



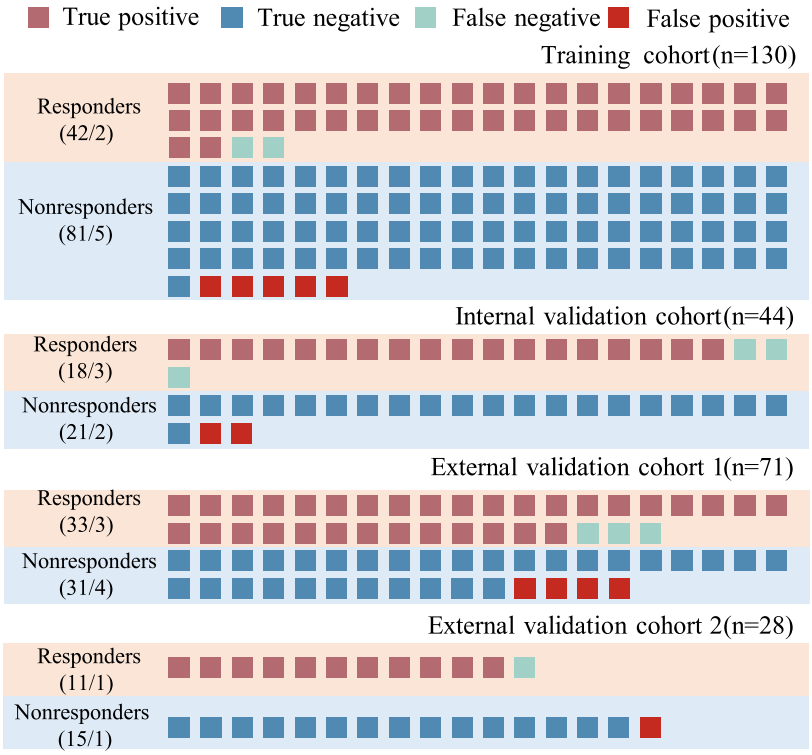
Supplementary Figure 1. Patient enrollment for development and validation of pathomics-driven ensemble model. H&E, hematoxylin-eosin; WSI, whole slide image; TCGA-STAD, the cancer genome atlas-stomach adenocarcinoma.



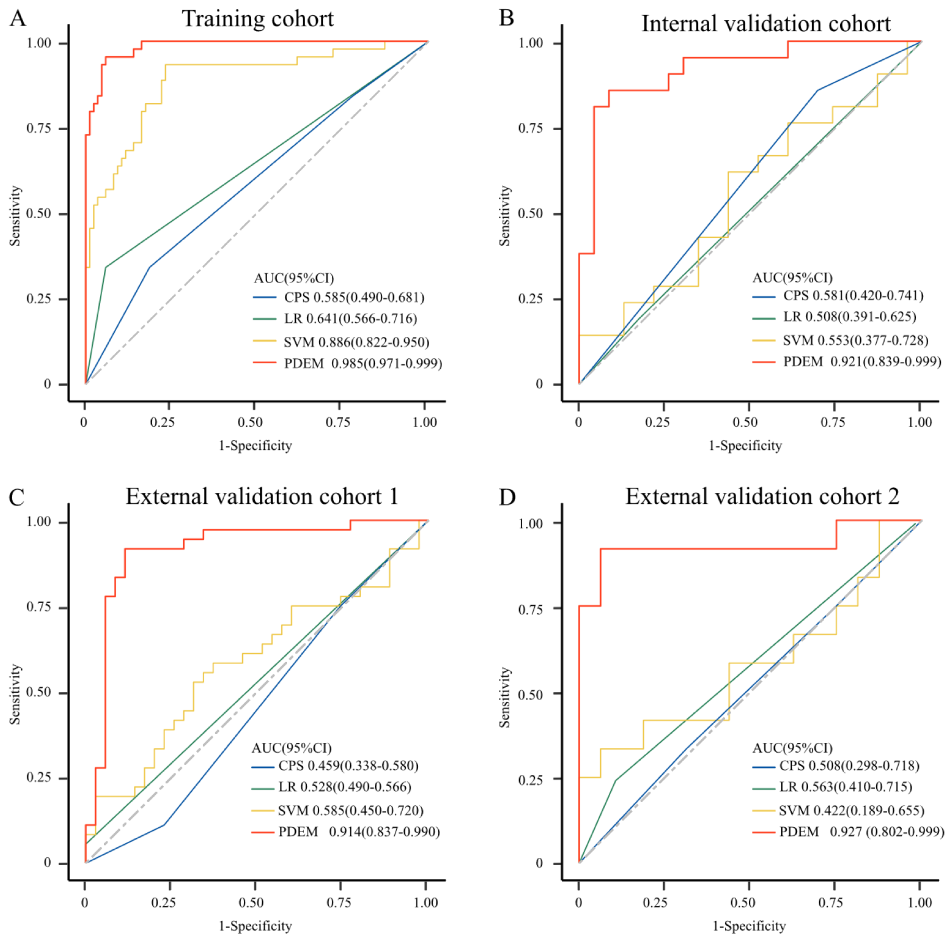
Supplementary Figure 2. Pipeline of WSI processing and pathomics feature extraction. WSI, whole slide image; ROI, region of interest.



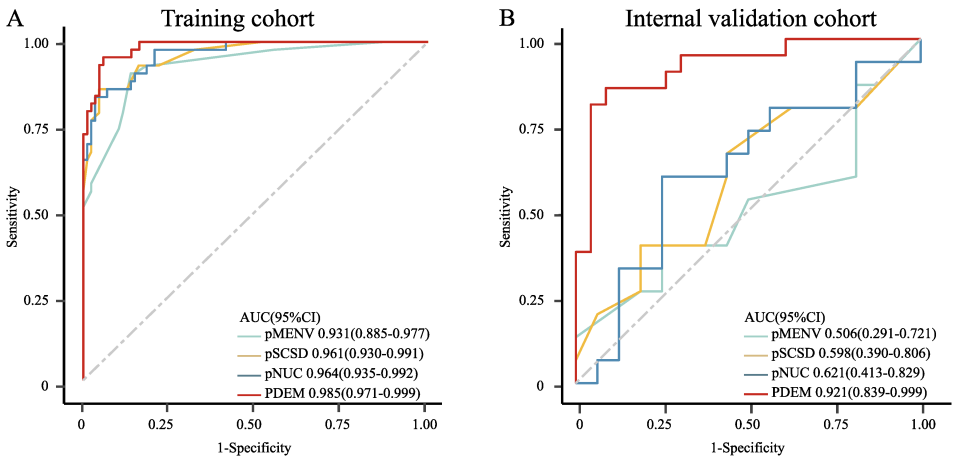
Supplementary Figure 3. Histograms of treatment strategy(A), line of therapy(B) and response(C) in different cohorts of patients treated with immune checkpoint inhibitor (ICI) therapy.



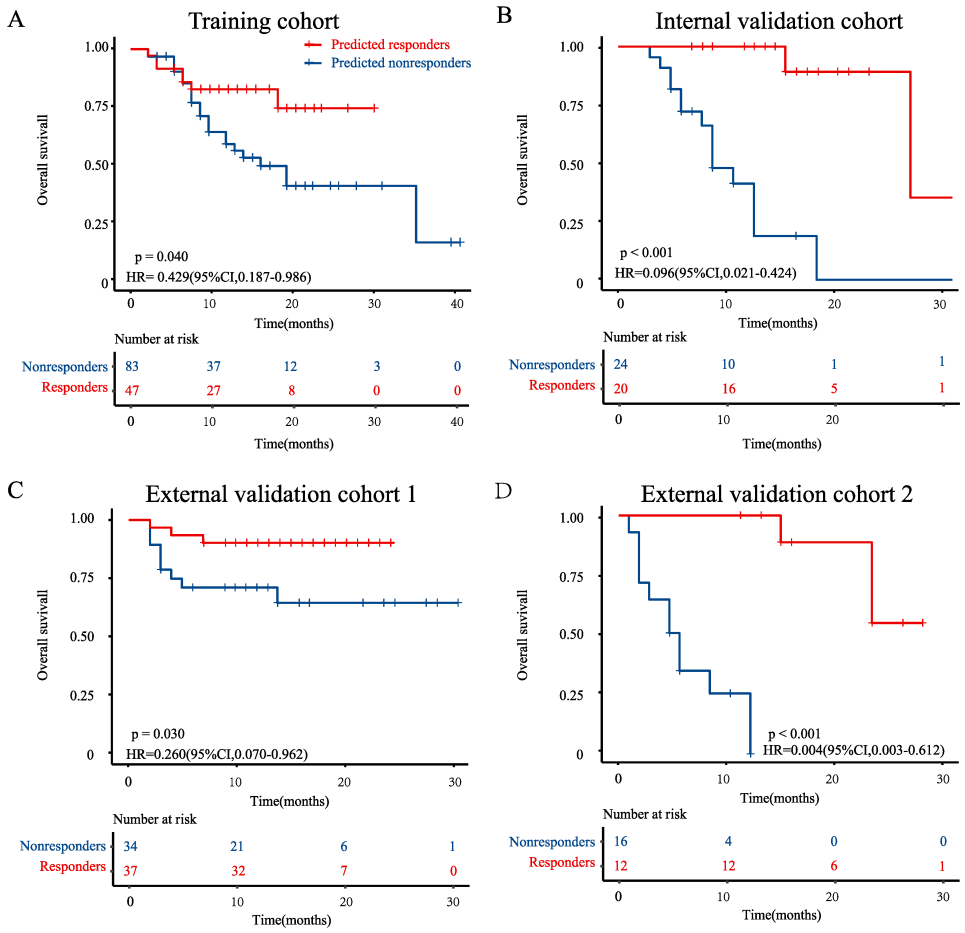
Supplementary Figure 4. Prediction profiles of pathomics-driven ensemble model.All individual participants were identified as responders or nonresponders by the ensemble model, and recognized as true postive (red), false negative(cyan), true negative (blue) and false positive(carmine) according to their true lables of immunotherapy response in all datasets.



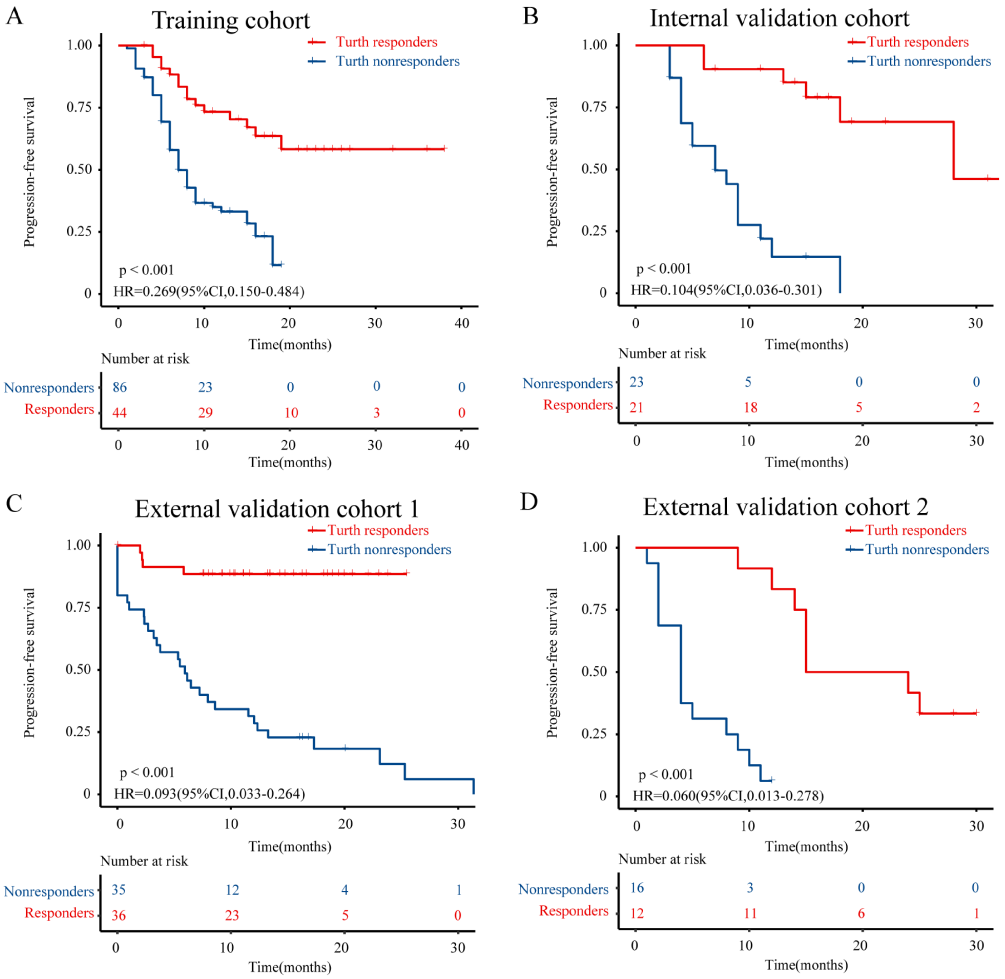
Supplementary Figure 5. Comparison of prediction performance between the pathomics-driven ensemble model, other models, and CPS in multiple cohorts. For comparison methods, the ensemble model was compared with other models (logistic regression, SVM) and CPS. Receiver operating characteristic curves of predictive performance for immunotherapy effect in patients with gastric cancer among logistic regression, SVM, CPS and the pathomics-based ensemble model in the training cohort (A), internal validation cohort (B), external validatipression; LR, logistic regression; SVM, support vector machine; PDEM, the Pathomics-Driven Ensemble Model.



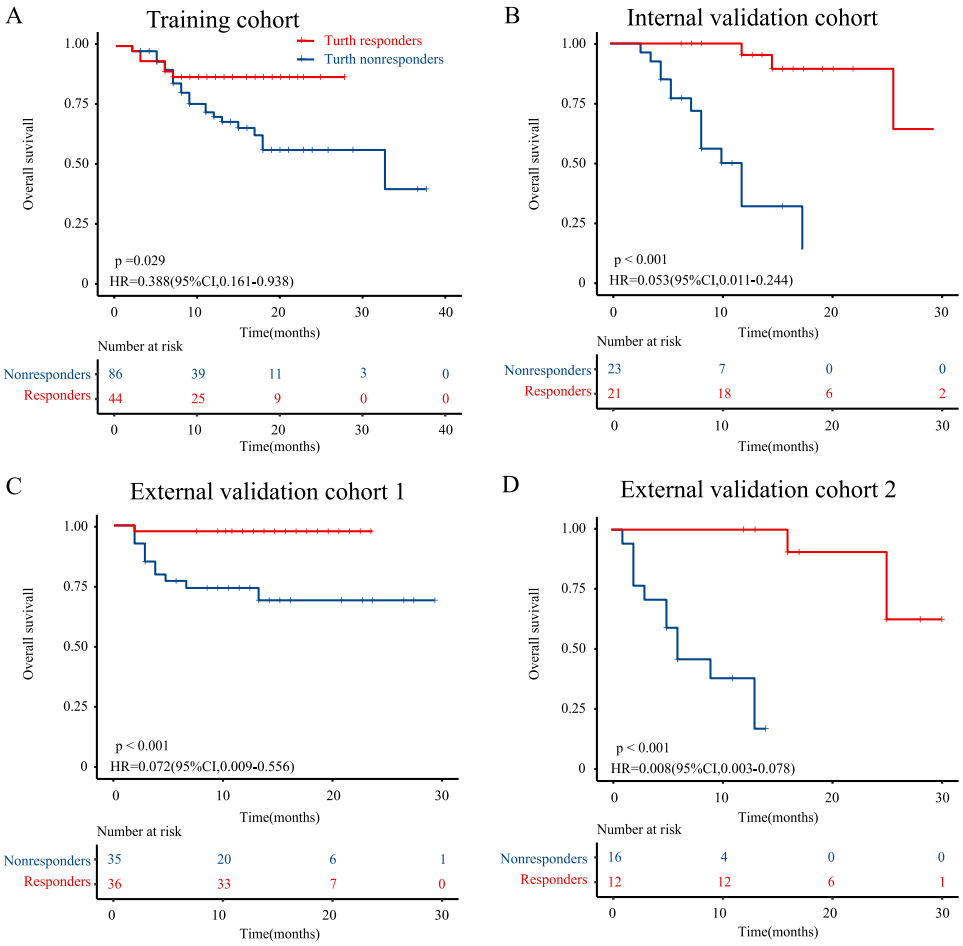
Supplementary Figure 6. Comparison of prediction performance between the pathomics-driven ensemble model and single kind of pathomics features models in the training cohort (A) and internal validation cohort (B). AUC, area under curve; 95%CI, 95%Confidence Interval; pNUC, pathomics nucleus features; pMENV, pathomics deep microenvironment features; pSCSD, pathomics single-cell spatial distribution features; PDEM, the Pathomics-Driven Ensemble Model.



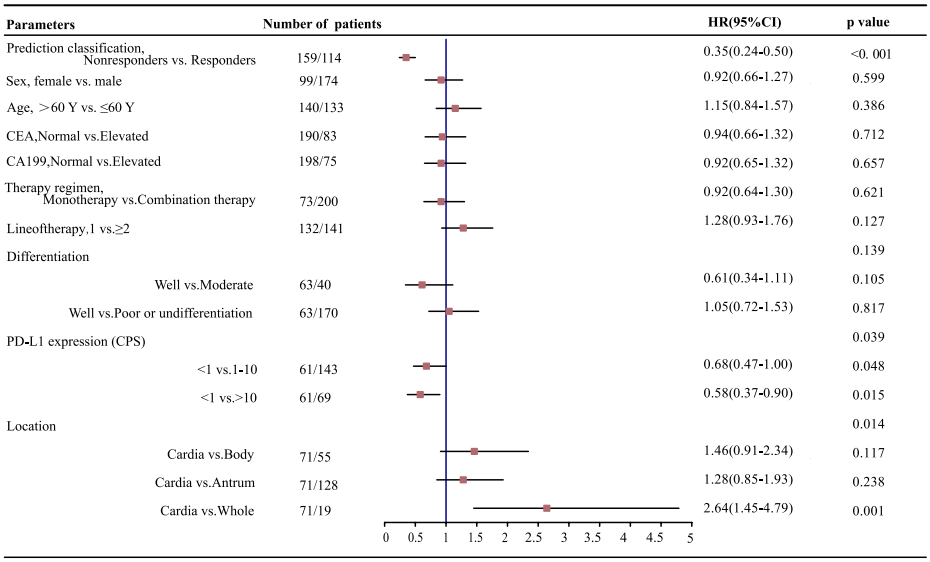
Supplementary Figure 7. Overall survival Kaplan-Meier curve analysis of prediction populations. Patients identified as 'predicted responders' by pathomics-driven ensemble model presented favorable overall survival than that of patients identified as 'predicted nonresponders ' in the training cohort (A), internal validation cohort (B), external validation cohort 1 (C), and external validation cohort 2(D).



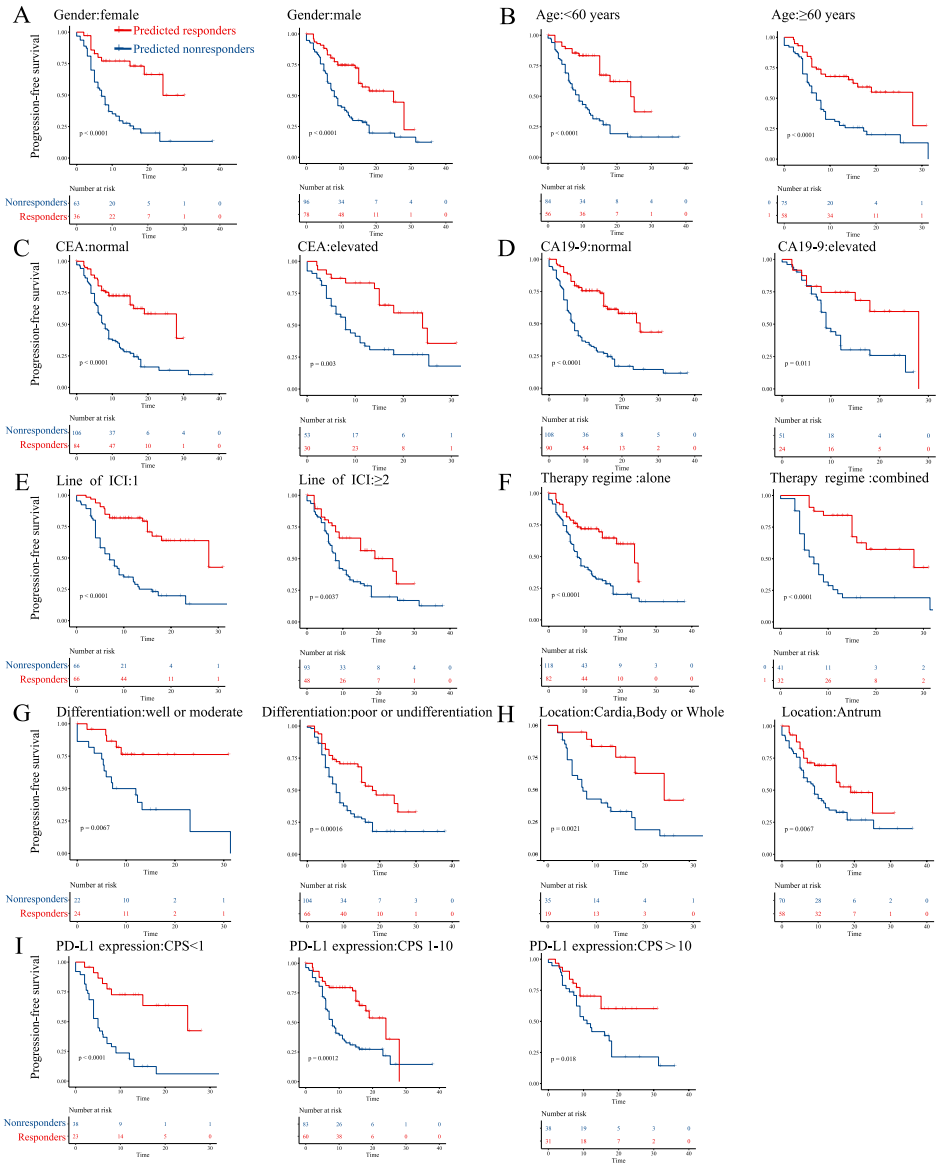
Supplementary Figure 8. Progression-free survival Kaplan-Meier curve analysis of actual populations. Patients with 'Turth responders ' presented favorable progression-free survival than that of patients with ' Turth nonresponders ' in the training cohort (A), internal validation cohort (B), external validation cohort 1 (C), and external validation cohort 2(D).



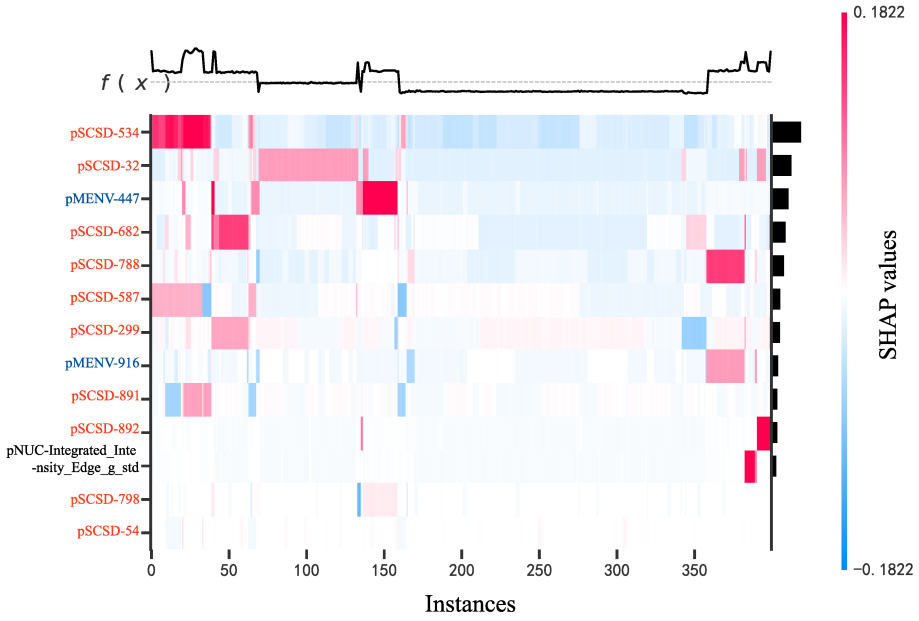
Supplementary Figure 9. Overall survival Kaplan-Meier curve analysis of actual populations. Patients with 'Turth responders ' presented favorable overall survival than that of patients with ' Turth nonresponders ' in the training cohort (A), internal validation cohort (B), external validation cohort 1 (C), and external validation cohort 2(D).



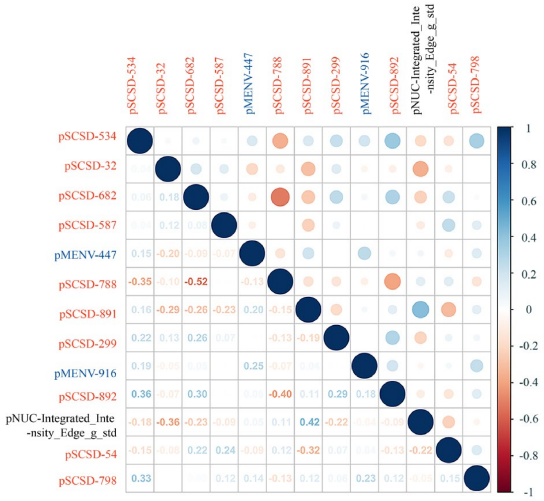
Supplementary Figure 10. Forest plot for the Univariate cox regression analysis of progression-free survival. CPS,combined positive score of PD-L1 expression.



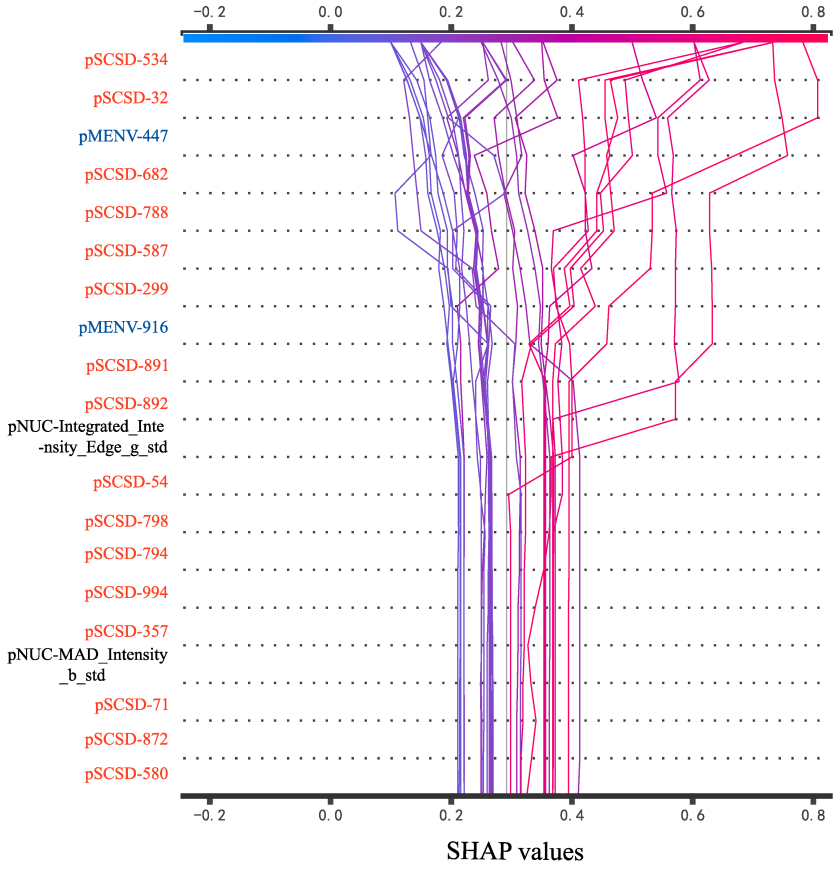
Supplementary Figure 11. Kaplan-Meier survival analysis according to the prediction populations stratified by clinicopathological risk factors. P-values were calculated by log-rank test.



Supplementary Figure 12. SHAP heatmap plot. Passing a matrix of SHAP values to the heatmap plot function creates a plot with the instances on the x-axis, the model inputs on the y-axis, and the SHAP values encoded on a color scale. The output of the model is shown above the heatmap matrix, and the global importance of each model input shown as a bar plot on the right-hand side of the plot. SHAP, SHapley Additive exPlanations;pNUC, pathomics nucleus features; pMENV, pathomics deep microenvironment features; pSCSD, pathomics single-cell spatial distribution features.



Supplementary Figure 13. Heatmaps of the important features correlation. pNUC, pathomics nucleus features; pMENV, pathomics deep microenvironment features; pSCSD, pathomics single-cell spatial distribution features.



Supplementary Figure 14. SHAP decision plot.The x-axis represents the model's output. All SHAP values are relative to the model's expected value like a linear model's effects are relative to the intercept. The y-axis lists the model's features. By default, the features are ordered by descending importance. The importance is calculated over the observations plotted. Each observation's prediction is represented by a colored line. At the top of the plot, each line strikes the x-axis at its corresponding observation's predicted value. This value determines the color of the line on a spectrum. * Moving from the bottom of the plot to the top, SHAP values for each feature are added to the model's base value. This shows how each feature contributes to the overall prediction. SHAP, SHapley Additive exPlanations;pNUC,

pathomics nucleus features; pMENV, pathomics deep microenvironment features; pSCSD, pathomics single-cell spatial distribution features.

Section 4. Supplementary Reference

1. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* **5**, 555-570 (2021).
2. Graham S*, et al.* Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis* **58**, 101563 (2019).
3. Weisbart E*, et al.* CellProfiler plugins – An easy image analysis platform integration for containers and Python tools. *Journal of Microscopy*, jmi.13223 (2023).
4. Gamper J*, et al.* PanNuke Dataset Extension, Insights and Baselines.). arXiv (2020).