**Table S1** Description of CT findings

| CT findings | Subcategories | Description |
|---|---|---|
| Primary lesion | The max diameter | The maximal diameter in the axial plane was recorded for pNENs and the data was categorized as ≥20 mm group and <20 mm group |
| | Location | The location of pNENs was recorded as the uncinate process, head or neck, body and tail |
| | Property | pNENs was divided as purely solid, purely cystic and solid-cystic mixed types according to the hypo-attenuation portion less than 30 HU with no enhancement in both arterial and portal venous phases |
| | Calcification | Calcification in pNENs was recorded with the presence of hyper-attenuation portion more than 80 HU |
| | Shape | The shape of pNENs was classified into 3 types: round shape with clear margin, simple nodular with extra-nodular growth and confluent multinodular |
| | Boundary | If there was a clear line between pNENs' lesion and surrounding tissues, it was recorded as clear boundary. Otherwise, it was recorded as unclear boundary |
| | Vessel involvement | If there was filling defect in pNENs' surrounding vessels (artery observed in arterial phase, venous observed in venous phase), it was recorded as surrounding vessel involvement. Otherwise, it was recorded as surrounding vessel non-involvement |
| | CT ratio | CT ratio was defined as the CT value of pNENs' lesion divided by the non-tumorous pancreatic parenchyma. We recorded CT ratio in unenhanced phase, arterial phase and venous phase, respectively |
| | Relatively enhanced rate | Relatively enhanced ratio was calculated by that increased CT value of pNENs' lesion divided by the increased CT value of aorta in the same plane. We recorded the data in arterial phase and venous phase, respectively |
| Pancreas | Pancreatic duct dilated or cut | The dilation of pancreatic duct was recorded when the diameter of main pancreatic duct measured more than 3 mm. Pancreatic duct cut was defined as a sudden interruption of the main pancreatic duct |
| | Pancreas atrophy | Pancreas atrophy was defined as more than expected loss or of adipose infiltration of pancreas parenchyma |
| Lymph node | Morphology | The maximal diameter of lymph node short axis in the axial plane was recorded and the data was categorized as normal group (<10 mm), enlarged group (≥10 mm) and multinodular confluent group |
| | Enhancement pattern | pNENs' lesion was characterized as heterogeneous enhancement when there was hypo-attenuation area in the solid part and homogeneous enhancement when the solid part appeared as the same attenuation in arterial phase |
| Hepatobiliary system | Fatty liver | Fatty liver was defined as the CT value of liver decreased less than 40 HU |
| | Focal benign lesion | Hepatic focal benign lesions included pure cyst, focal nodular hyperplasia and calcification with typical CT imaging appearance confirmed by hepatic lesion imaging diagnostic expert |
| | Bile duct dilatation | The diameter of bile duct >5 mm was recorded as dilation. The diameter of common hepatic duct and common bile duct >10 mm was recorded as dilation |
| Portal system | Portal vein | The diameter of portal vein was measured at the plane of hepatic hilum |
| | Splenic vein | The diameter of splenic vein was measured at the plane of splenic hilum |
| | Splenic varices | Increased and dilated blood vessels at the splenic hilum were recorded as splenic varices |
| | Splenomegaly | The spleen was beyond 5 costal units on axial plane |

CT, computed tomography; pNEN, pancreatic neuroendocrine neoplasm.
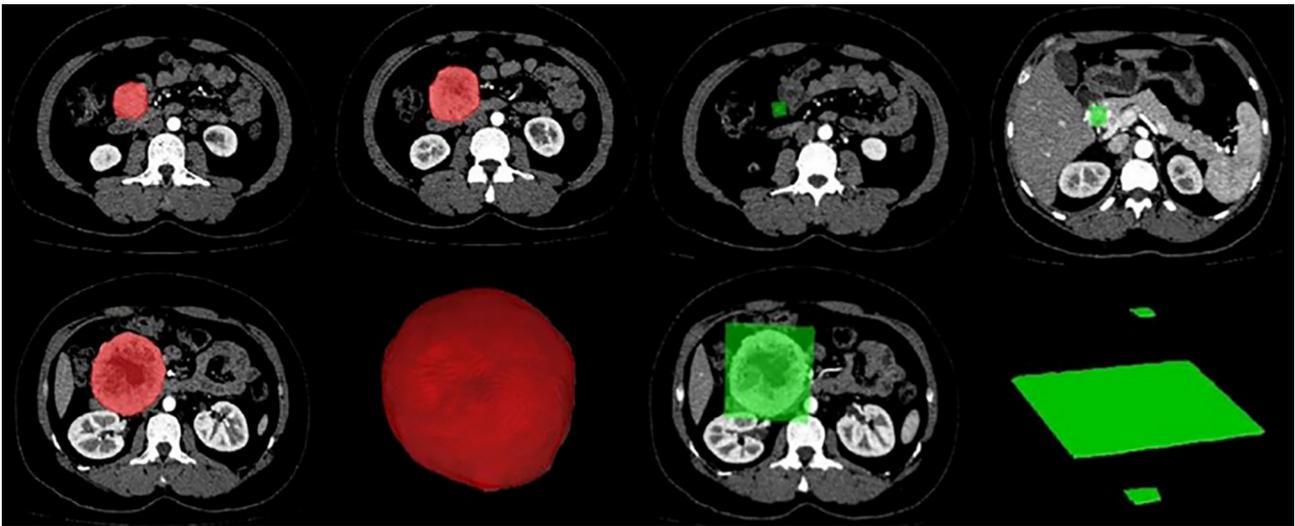
**Figure S1** ROIs for radiomics (red) and DLR (green). ROIs, regions of interest; DLR, deep learning radiomics.
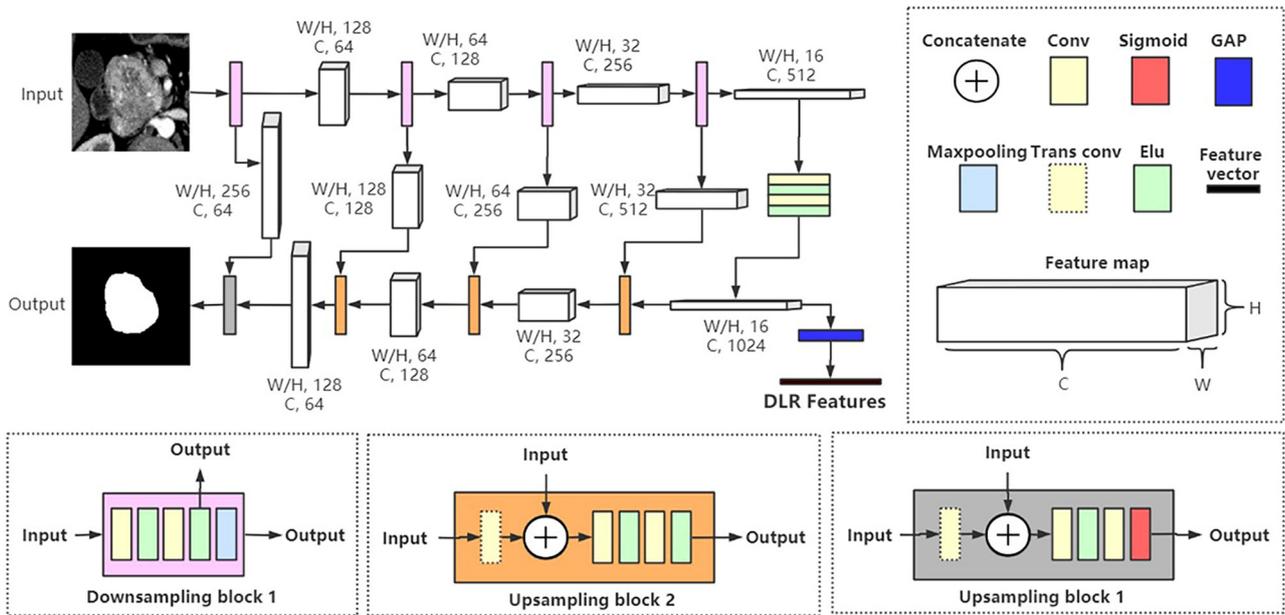


**Figure S2** Network structure of 2D U-net. The W and H and C donate width and height and channel of feature map, respectively. Conv, convolution; GAP, global average pooling; Trans conv, transposed convolution.

**Table S2** Clinical information and CT findings in recurrence and recurrence-free pNENs (56 patients from Hospital I)

| Variables | Recurrence-free (n=46) | Recurrence (n=10) | Statistics* | P |
|---|---|---|---|---|
| Clinical information | | | | |
| Age | 14.25 (3.00) | 14.75 (4.50) | m 140.000 | 0.054 |
| Sex | | | x 0.487 | 0.730 |
| F | 24 | 4 | | |
| M | 22 | 6 | | |
| Symptom | | | f | 0.032 |
| N | 23 | 9 | | |
| Y | 23 | 1 | | |
| Primary lesion | | | | |
| The max diameter | | | f | 0.032 |
| <20 mm | 23 | 1 | | |
| ≥20 mm | 23 | 9 | | |
| Location | | | f | 0.850 |
| Uncinate process | 16 | 5 | | |
| Head and neck | 14 | 2 | | |
| Body | 4 | 1 | | |
| Tail | 12 | 2 | | |
| Property | | | f | 0.054 |
| Cystic | 0 | 1 | | |
| Mixed | 18 | 6 | | |
| Solid | 28 | 3 | | |
| Calcification | | | f | 0.390 |
| Y | 38 | 7 | | |
| N | 8 | 3 | | |
| Shape | | | f | 0.022 |
| Round | 29 | 2 | | |
| Local lobulated | 11 | 4 | | |
| Confluent multinodular | 6 | 4 | | |
| Boundary | | | f | 0.140 |
| Clear | 32 | 4 | | |
| Unclear | 14 | 6 | | |
| Vessel involvement | | | | 1.000 |
| N | 42 | 9 | | |
| Y | 4 | 1 | | |
| CT ratio | | | | |
| Unenhanced | 1.16 (0.38) | 0.85 (0.45) | m 136.000 | 0.044 |
| Arterial phase | 1.21 (0.61) | 0.90 (0.88) | m 172.000 | 0.260 |
| Venous phase | 1.16 (0.38) | 0.85 (0.45) | m 136.000 | 0.044 |
| Relatively enhanced rate | | | | |
| Arterial phase | 0.43 (0.39) | 0.28 (0.34) | | 0.052 |
| Venous phase | 0.64 (0.37) | 0.48 (0.45) | | 0.120 |
| Pancreas | | | | |
| Pancreatic duct dilated or cut | | | f | 0.028 |
| N | 39 | 5 | | |
| Y | 7 | 5 | | |
| Pancreas atrophy | | | f | 0.680 |
| N | 36 | 7 | | |
| Y | 10 | 3 | | |
| Lymph node | | | | |
| Morphology | | | f | 0.023 |
| Normal | 41 | 6 | | |
| Enlarged | 5 | 3 | | |
| Confluent multinodular | 0 | 1 | | |
| Enhancement pattern | | | f | 0.029 |
| Homogeneous | 46 | 8 | | |
| Heterogeneous | 0 | 2 | | |
| Hepatobiliary system | | | | |
| Fatty liver | | | f | 1.000 |
| N | 43 | 10 | | |
| Y | 3 | 0 | | |
| Focal benign lesion | | | f | 0.490 |
| N | 26 | 4 | | |
| Y | 20 | 6 | | |
| Bile duct dilatation | | | f | 1.000 |
| N | 40 | 9 | | |
| Y | 6 | 1 | | |
| Portal system | | | | |
| Portal vein | 14.25 (3.00) | 14.75 (4.50) | m 189.500 | 0.380 |
| Splenic vein | 8.83±2.15 | 7.60±2.95 | t −1.244 | 0.240 |
| Splenomegaly | | | f | 1.000 |
| N | 29 | 6 | | |
| Y | 17 | 4 | | |
| Splenic varices | | | f | 0.560 |
| N | 43 | 9 | | |
| Y | 3 | 1 | | |

*, $t$ represents Student's $t$-test, m represents Mann Whitney U test, x represent Pearson chi-square test, f represents fisher exact probability test. CT, computed tomography; pNEN, pancreatic neuroendocrine neoplasm.

**Table S3** DeLong test results (P value) of ROC comparisons for all models based on Hospital I image datasets

| Model | DLR-A | DLR-V | DLR-A&V | Radiomics-A | Radiomics-V | Radiomics A&V | CT findings |
|---|---|---|---|---|---|---|---|
| DLR-A | – | 0.0632 | 0.1519 | 0.5952 | 0.2808 | 0.4309 | 0.1191 |
| DLR-V | – | – | 0.1618 | 0.1719 | 0.3590 | 0.2756 | 0.7364 |
| DLR-A&V | – | – | – | 0.8552 | 0.6310 | 0.9041 | 0.2474 |
| Radiomics-A | – | – | – | – | 0.6500 | 0.7994 | 0.1046 |
| Radiomics-V | – | – | – | – | – | 0.5058 | 0.2855 |
| Radiomics-A&V | – | – | – | – | – | – | 0.1966 |
| CT findings | – | – | – | – | – | – | – |

ROC, receiver operating characteristic; DLR, deep learning radiomics; A, arterial; V, venous; A&V, arterial & venous; CT, computed tomography.
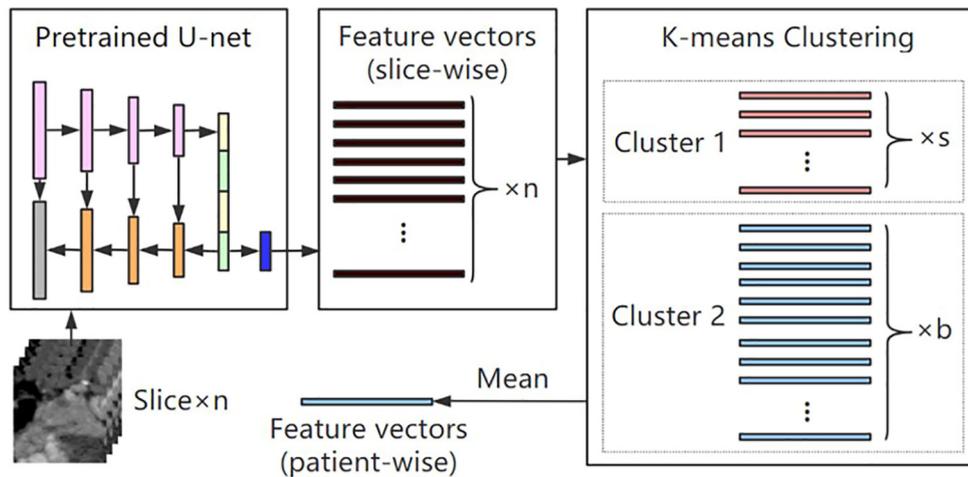


**Figure S3** Flow chart of deep learning radiomics feature extraction.
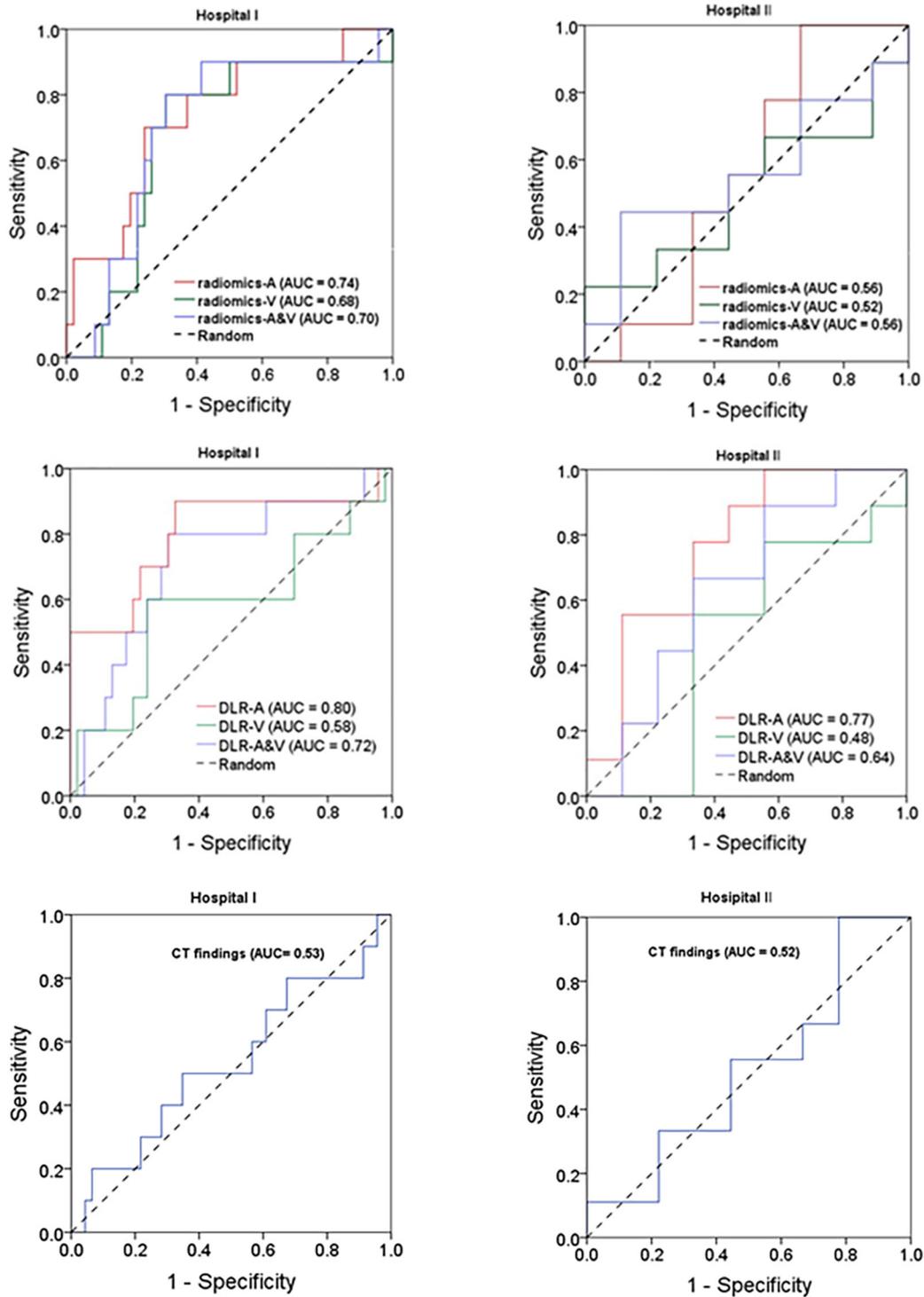
**Figure S4** ROCs of different phases with radiomics, deep learning radiomics (DLR) and CT findings in the internal and external groups. ROC, receiver operating characteristic.
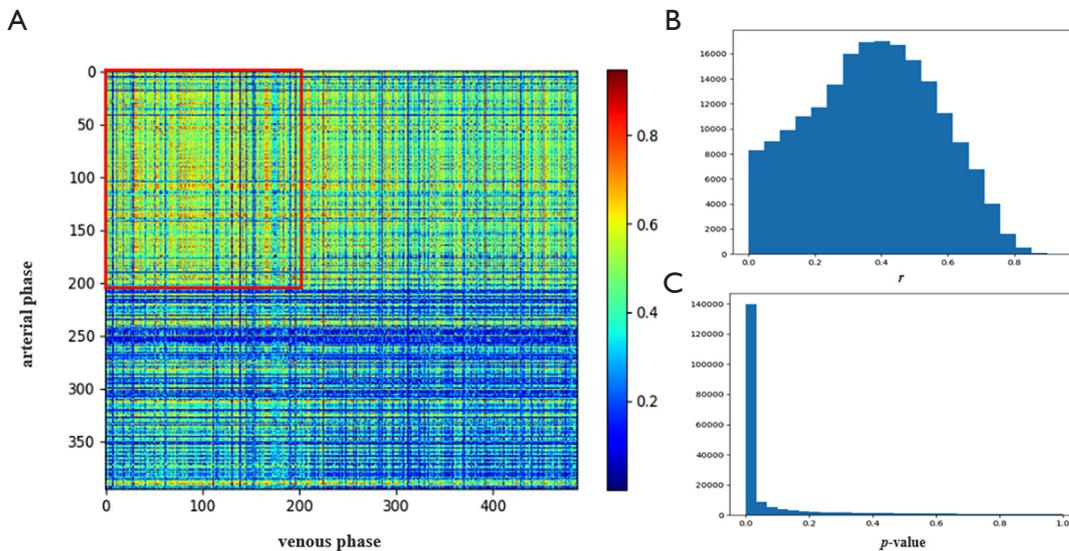
**Figure S5** Collinearity analysis results of DLR features in arterial and venous phases. (A) Heatmap of absolute value of correlation coefficient r, 206 features overlap between arterial and venous phases (which correlation coefficients are shown in the red box). (B) Histogram of Pearson's correlation coefficient r distribution. (C) Histogram of P value distribution. DLR, deep learning radiomics.

## Supplementary Materials and Methods

### Section 1 Training U-net for DLR

We used a 2D U-net to extract DLR features (*Figure S2*). The encoder of the U-net contained 4 downsampling modules, and the decoder contained 4 upsampling modules constructed based on transposed convolution. Skip connections were set between the upsampling and downsampling modules to provide more high-resolution information for the decoder. The initial learning rate was set as $1\times10^{-5}$, the optimizer was Adam, and we used cross-entropy as loss function. Dice similarity coefficient (DSC) was calculated on the validation set for evaluating the performance of segmentation, and the calculation formula of DSC was as follow, where A and B are the ground truth (GT) and predicted segmentation mask of the image, respectively.

$$DSC(A,B) = \frac{2\,|\,A\cap B\,|}{|\,A\,|+|\,B\,|} \qquad [1]$$

### Section 2 DLR features extraction

In the feature extraction process (*Figure S3*), we first took the smallest externalized cube of the region of interest (ROI) roughly annotated by the radiologists in 3D space as processed ROI, then for each patient we inputted each slice of CT image in processed ROI and extracted the feature map [after exponential linear unit (ELU) activation] of the last convolution layer before the decoder. Then a global average pooling (GAP) was performed to convert the feature map with size of $16\times16\times1,024$ into a feature vector with size of $1\times1,024$.

The input of segmentation network was a 2D slice of the tumor on CT image, and the recurrence annotation was patient-wise, so it was necessary to aggregate all slice-wise feature vectors of the same patient into a patient-wise feature vector. The feature vectors extracted from the multi-layer images of the same sample was $n\times1,024$, and n was the number of tumor slices. All feature vectors were clustered into 2 clusters based on K-means algorithm, and the maximum cluster was preserved. Then we took the mean value in the maximum cluster along feature dimension to get the final vector with a size of $1\times1,024$.

### Section 3 Model integration

For model integration, we used models in each fold of cross-validation on internal group to predict the recurrence risk of each patient in external group, and the average of the multi-model predicted recurrence risk was used to calculate the evaluation metric. The whole process of model integration can be expressed as following equation,

$$Y_i = \left\{ F(x_{i,p}) \,|\, x_{i,p} \in X_i \right\} \qquad [2]$$

$$Z_i = \frac{1}{N}\sum_n g_n(K(Y_i)) \qquad [3]$$

where X and x represent the CT image (in processed ROI) and its slice, respectively. And i is the patient index, p is the slice index. F denotes the segmentation feature extraction process (whose output is a feature vector), and Y is the feature vector set of all slices of tumor X. In the latter formula, K is the feature aggregation operation (K-means clustering), and g denotes the recurrence prediction model (whose input is a feature vector). N is the number of classification models, and n is the cross-validation model index. Z is the final predicted recurrence risk of patient in external group.