10.1148/rycan.2021210010

Appendix E1 Supplemental Methods

Preparation of Local Data Set

Patients with histologically or cytologically confirmed pancreatic ductal adenocarcinoma (PDAC) were identified from Cancer Registry of National Taiwan University, a tertiary referral center in Taiwan. CT images of identified patients with PDAC and individuals with normal pancreas were extracted from the imaging archive for further review to construct the local datasets. In patients who underwent more than one CT examination, the one that immediately preceded the date of diagnosis was used.

Formal radiologist reports of the CT images of patients with PDAC in the local test sets were retrieved from electronic health records to ascertain radiologist performance. The radiologist reports were jointly reviewed by two radiologists (P.T.C. and K.L.L.) who had 5 years and 20 years of experience, respectively, after removal of the identities of the patient and interpreting radiologist, without reference to the images. A radiologist report was considered to have classified the patient as having PDAC if the report mentioned a definite or suspicious pancreatic tumor; otherwise, the PDAC was considered as being missed by the interpreting radiologist.

Preparation of External Data Set

The external test set consisted of 12,769 CT images (5715 PDACs and 7054 controls) obtained in 264 patients (182 with PDAC and 82 controls) from two open source data sets. The Medical Segmentation Decathlon Dataset from Memorial Sloan Kettering Cancer Center contained CT images of 281 patients who underwent surgical resection of pancreatic masses with PDAC, intraductal papillary mucinous neoplasm (IPMN), and pancreatic neuroendocrine tumor (PNET). Because the Medical Segmentation Decathlon Dataset did not provide the exact diagnosis of each case, two radiologists (P.T.C. and K.L.L.) reviewed the CT images to identify and exclude cases in which IPMN or PNET was possible, leaving 182 patients with PDAC for the study. The Cancer Imaging Archive (TCIA) data set contained CT images of 82 individuals with normal pancreas from National Institutes of Health (NIH) Clinical Center. The indication for scan was screening before kidney donation in 17 individuals. The remaining 65 individuals were selected by a radiologist from patients with neither major abdominal pathologies nor pancreatic lesions.

Details of Segmentation and Image Preprocessing

The pancreases on CT images were segmented and labeled as regions of interest (ROIs) for further analysis. In patients with PDACs, the tumor was also segmented and labeled as ROI. The images were manually segmented using 3D Slicer (version 4.8.1) by one of two experienced abdominal radiologists (P.T.C. and K.L.L.), with reference to the findings of other examinations or surgery when needed. Because of potential interobserver differences regarding the exact extent of the pancreases and PDACs, all labeled images were jointly reviewed by both radiologists for consensus.

To avoid potential bias resulting from differences in the spacing between CT images across patients, all the images and segmentation labels were resampled to the spacing of 1 × 1 × 5 mm using linear interpolation and nearest neighbor interpolation, respectively. The tumor part was cropped into patches of 20 × 20 pixels sequentially using a row-wise moving window with a stride of 1 pixel from upper left to lower right, and the patches in which the tumor occupied more than 50% of the total area were marked as cancerous patches for subsequent analysis. If more than 200 cancerous patches were generated from a patient, patches were selected from the first patch (top left) toward the final patch (bottom right) at a fixed interval such that 200 patches were selected. If fewer than 10 cancerous patches could be generated from a patient with PDAC with the 50% area threshold due to small tumor size, the patches in which cancerous pancreas occupied more than 5% of the total area were also marked as cancerous. Noncancerous patches were generated from the pancreas part (excluding tumor) using the same approach, except that the stride of the moving window was 5 pixels. Only 0.70% of the patches were generated with the 5% threshold.

Details of Extraction of Radiomic Features

Radiomic features were extracted from the noncancerous pancreas in noncancerous patches and the tumor in cancerous patches using PyRadiomics (version 2.2.0). Because the tumor or pancreas was processed into patches, we excluded shape-based features and features that were confounded by the volume of the analyzed object, including first order: energy, first order: total energy, and first order: root mean squared. The remaining 88 nonfiltered radiomic features were extracted, including 15 first order features, 22 gray-level co-occurrence matrix (GLCM) features, 16 gray-level run-length matrix (GLRLM) features, 16 gray-level size zone matrix (GLSZM) features, 5 neighboring gray-tone difference matrix (NGTDM) features, and 14 gray-level dependence matrix (GLDM) features. The bin width for computing texture features was fixed as 25.

Cutoff Selection

For differentiation between cancerous and noncancerous patches (patch-based analysis), features of all patches in the training set were input into XGBoost (version 1.0.2). The trained model was applied on all patches in the respective validation set and a receiver operating characteristic (ROC) curve was generated according to the predictive scores, and the score with the highest Youden index (ie, sensitivity + specificity − 1) was selected as cutoff to balance the trade-off between the need for high sensitivity and specificity. For differentiation between patients with PDAC and controls (patient-based analysis), the tumors in patients with PDAC and the pancreas of controls were treated as ROIs of the PDAC cases and controls, respectively. The score for classifying a patient as with or without PDAC was defined as the proportion of patches in the ROI that were predicted as cancerous in the patch-based analysis. An ROC curve was generated with the scores of all individuals in the validation set, and the score with the highest Youden index was chosen as the cutoff.

Details of Training Configuration of XGBoost Models

All the XGBoost models in this study were trained with scikit-learn (version 0.22.2.post1) API with the following parameters: learning_rate = 0.1, n_estimators = 1000, max_depth = 5, min_child_weight = 1, gamma = 0, subsample = 0.8, colsample_bytree = 0.8, nthread = 4,

objective = ‘binary:logistic’, scale_pos_weight = 1, seed = 27. The training process was terminated when the area under the ROC (AUC) on the validation set did not increase for 30 iterations. The model that achieved the highest AUC in the validation set was selected as the model for subsequent analysis.

Assessment of Potential Sampling Error in Random Dataset Splitting

Thirty additional rounds of random splitting were conducted to access potential sampling error in randomly splitting the data set into the training and validation set and the test set, and the features selected to construct the respective local and generalized models in each split were compared with those selected with the original split as presented in the results section. The results showed no significant sampling error/bias with our results, as all the features selected with the original split were selected in multiple additional random splits (Table E7). Specifically, in the additional 30 random splits, the 14 features selected in the local model obtained with the original data split were selected in seven to 30 of the 30 additional splits, with nine features being selected in half or more of the 30 splits. The 20 features in the generalized model selected with the original data split were selected in five to 30 of the 30 additional splits, with 16 features being selected in half or more of the 30 splits. Furthermore, of the 11 features selected by both models with the original data split, five were selected in all 30 random splits, and one was selected in all but one split.

Table E1. Performance of XGBoost Models including Different Number of Radiomic Features in Local Training and Validation Set

	Patch Based				Patient Based			
Number of Features	Sensitivity	Specificity	Accuracy	AUC	Sensitivity	Specificity	Accuracy	AUC
88	32586/34164 (95.4) [95.2, 95.6]	92684/100955 (91.8) [91.6, 92.0]	125270/135119 (92.7) [92.6, 92.8]	0.98 [0.98, 0.98]	325/349 (93.1) [89.9, 95.5]	365/383 (95.3) [92.7, 97.2]	690/732 (94.3) [92.3, 95.8]	0.97 [0.95, 0.98]
4	30906/34164 (90.5) [90.1, 90.8]	87951/100955 (87.1) [86.9, 87.3]	118857/135119 (88.0) [87.8, 88.1]	0.96 [0.95, 0.96)	284/349 (81.4) [76.9, 85.3]	356/383 (93.0) [89.9, 95.3]	640/732 (87.4) [84.8, 89.7]	0.88 [0.85, 0.91]
7	31710/34164 (92.8) [92.5, 93.1]	91527/100955 (90.7) [90.5, 90.8]	123237/135119 (91.2) [91.1, 91.4]	0.97 [0.97, 0.98]	306/349 (87.7) [83.8, 90.9]	354/383 (92.4) [89.3, 94.9]	660/732 (90.2) [87.8, 92.2]	0.93 [0.90, 0.95]
14	32492/34164 (95.1) [94.9, 95.3]	92576/100955 (91.7) [91.5, 91.9]	125068/135119 (92.6) [92.4, 92.7]	0.98 [0.98, 0.98]	333/349 (95.4) [92.7, 97.4]	360/383 (94.0) [91.1, 96.2]	693/732 (94.7) [92.8, 96.2]	0.97 [0.96, 0.99]

Note.—Values are numbers with percentages in parentheses and 95% CIs in brackets. AUC = area under the receiver operating characteristic curve.

Table E2. Performance of XGBoost Models including Different Number of Radiomic Features in Combined Training and Validation Set

	Patch Based				Patient Based			
Number of features	Sensitivity	Specificity	Accuracy	AUC	Sensitivity	Specificity	Accuracy	AUC
88	38220/41088 (93.0) [92.8, 93.3]	121518/137955 (88.1) [87.9, 88.3]	159738/179043 (89.2) [89.1, 89.4]	0.97 [0.97, 0.97]	447/495 (90.3) [87.3, 92.8]	434/449 (96.7) [94.5, 98.1]	881/944 (93.3) [91.5, 94.8]	0.96 [0.95, 0.98]
3	33842/41088 (82.4) [82.0, 82.7]	118173/137955 (85.7) [85.5, 85.8]	152015/179043 (84.9) [84.7, 85.1]	0.91 [0.90, 0.91]	300/495 (60.6) [56.1, 64.9]	421/449 (93.8) [91.1, 95.8]	721/944 (76.4) [73.5, 79.1]	0.78 [0.75, 0.81]
11	37280/41088 (90.7) [90.4, 91.0]	122402/137955 (88.7) [88.6, 88.9]	159682/179043 (89.2) [89.0, 89.3]	0.96 [0.96, 0.96]	441/495 (89.1) [86.0, 91.7]	429/449 (95.5) [93.2, 97.3]	870/944 (92.2) [90.3, 93.8]	0.95 [0.93, 0.96]
20	37891/41088 (92.2) [92.0, 92.5]	124140/137955 (90.0) [89.8, 90.1]	162031/179043 (90.5) [90.4, 90.6]	0.97 [0.97, 0.97]	457/495 (92.3) [89.6, 94.5]	431/449 (96.0) [93.7, 97.6]	888/944 (94.1) [92.4, 95.5]	0.96 [0.95, 0.98]

Note.—Values are numbers with percentages in parentheses and 95% CIs in brackets. AUC = area under the receiver operating characteristic curve.

Table E3. Eleven Radiomic Features included in Both the Local and Generalized Models for Patch-based Analysis

Feature	Description
First order: 90th percentile	The 90th percentile of the pixel intensity
First order: mean	The mean of the pixel intensity within ROI
First order: median	The median of the pixel intensity within ROI
First order: skewness	Asymmetry of the distribution of the pixel intensity
GLCM: correlation	Linear dependency of the marginal distribution in GLCM
GLCM: cluster shade	Asymmetry about the mean in GLCM
GLDM: dependence nonuniformity	Variability of similarity of pixel and its neighbors
GLDM: gray level nonuniformity	Variability of gray-level intensity
GLDM: large dependence low gray level emphasis	Concentration of large dependency with lower gray-level
GLRLM: run entropy	Uncertainty of the lengths of pixels with same gray-levels
NGTDM: busyness	A measure of changes from a pixel to its neighbors

Note.—GLCM = gray-level co-occurrence matrix, GLDM = gray-level dependence matrix, GLRLM = gray-level run-length matrix, NGTDM = neighboring gray-tone difference matrix, ROI = region of interest.

Table E4. Median of Normalized Features between Cancerous and Noncancerous Patches from Local and External Sets

	Local Test Sets			External Test Sets		
Features	Noncancerous	Cancerous	Difference	Noncancerous	Cancerous	Difference
Features reflecting image grayscale intensity						
First order: median	0.4413	–1.0577	–1.4990	0.1994	–0.2892	–0.4886
First order: 90th percentile	0.3150	–0.9709	–1.2858	–0.0211	–0.2914	–0.2703
First order: mean	0.4056	–1.0090	–1.4146	0.1148	–0.1760	–0.2908
Features reflecting heterogeneity						
NGTDM: busyness	–0.0339	0.0129	0.0468	–0.4194	0.4250	0.8443
GLDM: gray-level nonuniformity	–0.2659	0.5425	0.8084	–0.2373	0.4760	0.7133
GLDM: dependence nonuniformity	–0.2703	0.7547	1.0250	–0.1182	0.6184	0.7366

Note.—GLDM = gray-level dependence matrix, NGTDM = neighboring gray-tone difference matrix.

Table E5. PDAC Cases Missed by Analyses Based on Local Model, Generalized Model, or Radiologist Report in Local Test Set 1

Dataset and Patient	Local Model	Generalized Model	Radiologist	Tumor Size (cm)
1–1	+	+	-	1.7
1–2	+	+	-	2.3
1–3	+	+	-	2.4
1–4	+	+	-	4.7
1–5	-	+	+	1.1
1–6	-	-	+	2.4
1–7	-	-	+	2.5
1–8	+	-	+	2.7

Note.— + = detected, - = missed.

Table E6. PDAC Cases Missed by Analyses Based on Local model, Generalized Model, or Radiologist Report in Local Test Set 2

Dataset and Patient	Local Model	Generalized Model	Radiologist	Tumor Size (cm)
2–1	+	+	-	1.3
2–2	+	+	-	1.6
2–3	+	+	-	1.8
2–4	+	+	-	1.9
2–5	+	+	-	2.2
2–6	+	+	-	2.7
2–7	+	+	-	2.8
2–8	+	+	-	3.6
2–9	+	+	-	3.6
2–10	+	+	-	4.1
2–11	-	+	+	1.1
2–12	-	+	+	1.2
2–13	-	-	+	1.8
2–14	-	-	+	1.8
2–15	-	-	+	2.0
2–16	-	+	+	2.1
2–17	-	-	+	2.2
2–18	-	-	+	2.3
2–19	+	-	+	2.8
2–20	-	-	?	3.1

Note.— + = detected, - = missed, ? = radiologist report could not be identified.

Table E7. The Frequency of Each Feature Being Selected in the Local and Generalized Models among 30 Additional Random Data Set Splits

Feature	Local Model Frequency (*n* = 30)	Generalized Model Frequency (*n* = 30)
Feature selected by both local and generalized model with original splitting		
First order: 90th percentile	30	30
First order: median	30	30
GLCM: cluster shade	30	30
GLDM: gray-level nonuniformity	30	30
NGTDM: busyness	30	30
First order: skewness	29	30
GLDM: large dependence low gray-level emphasis	25	8
GLDM: dependence nonuniformity	21	28
First order: mean	14	7
GLCM: correlation	13	24
GLRLM: run entropy	7	20
Feature selected by only local model with original splitting		
First order: interquartile range	19	
GLRLM: run length nonuniformity normalized	9	
GLCM: sum entropy	7	
Feature selected by only generalized model with original splitting		
GLCM: IMC2		23
First order: 10th percentile		22
GLCM: autocorrelation		22
First order: minimum		20
GLCM: joint average		19
GLRLM: long run low gray-level emphasis		18
GLDM: large dependence high gray-level emphasis		15
GLSZM: zone entropy		13
GLSZM: zone variance		5

Note.—GLCM = gray-level co-occurrence matrix, GLDM = gray-level dependence matrix, GLRLM = gray-level run-length matrix, GLSZM = gray-level size-zone matrix, IMC2 = information measure of correlation 2, NGTDM = neighboring gray-tone difference matrix.