Pancreatic Cancer Detection on CT Scans with Deep Learning: A Nationwide Population-based Study

Po-Ting Chen, MD* • Tinghui Wu, MS* • Pochuan Wang, MS • Dawei Chang, MS • Kao-Lang Liu, MD • Ming-Shiang Wu, MD, PhD • Holger R. Roth, PhD • Po-Chang Lee, MD • Wei-Chih Liao, MD, PhD** • Weichung Wang, PhD**

From the Department of Medical Imaging (P.T.C., K.L.L.) and Division of Gastroenterology and Hepatology, Department of Internal Medicine (M.S.W., W.C.L.), National Taiwan University Hospital, National Taiwan University College of Medicine, Taipei, Taiwan; Institute of Applied Mathematical Sciences (T.W., D.C., W.W.) and Departments of Computer Science and Information Engineering (P.W.) and Internal Medicine, College of Medicine (M.S.W., W.C.L.), National Taiwan University, No. 1, Section 4, Roosevelt Rd, Taipei 10617, Taiwan; Department of Medical Imaging, National Taiwan University Cancer Center, Taipei, Taiwan (K.L.L.); NVIDIA, Bethesda, Md (H.R.R.); and National Health Insurance Administration, Ministry of Health and Welfare, Taipei, Taiwan (P.C.L.). Received March 5, 2022; revision requested April 22; revision received July 15; accepted August 1. **Address correspondence to** W.W. (email: *wwang@ntu.edu.tw*).

* P.T.C. and T.W. contributed equally to this work.

** W.C.L. and W.W. are co-senior authors.

Conflicts of interest are listed at the end of this article.

See also the editorial by Aisen and Rodrigues in this issue.

Background: Approximately 40% of pancreatic tumors smaller than 2 cm are missed at abdominal CT.

Purpose: To develop and to validate a deep learning (DL)–based tool able to detect pancreatic cancer at CT.

Materials and Methods: Retrospectively collected contrast-enhanced CT studies in patients diagnosed with pancreatic cancer between January 2006 and July 2018 were compared with CT studies of individuals with a normal pancreas (control group) obtained between January 2004 and December 2019. An end-to-end tool comprising a segmentation convolutional neural network (CNN) and a classifier ensembling five CNNs was developed and validated in the internal test set and a nationwide real-world validation set. The sensitivities of the computer-aided detection (CAD) tool and radiologist interpretation were compared using the McNemar test.

Results: A total of 546 patients with pancreatic cancer (mean age, 65 years ± 12 [SD], 297 men) and 733 control subjects were randomly divided into training, validation, and test sets. In the internal test set, the DL tool achieved 89.9% (98 of 109; 95% CI: 82.7, 94.9) sensitivity and 95.9% (141 of 147; 95% CI: 91.3, 98.5) specificity (area under the receiver operating characteristic curve [AUC], 0.96; 95% CI: 0.94, 0.99), without a significant difference (P = .11) in sensitivity compared with the original radiologist report (96.1% [98 of 102]; 95% CI: 90.3, 98.9). In a test set of 1473 real-world CT studies (669 malignant, 804 control) from institutions throughout Taiwan, the DL tool distinguished between CT malignant and control studies with 89.7% (600 of 669; 95% CI: 87.1, 91.9) sensitivity and 92.8% specificity (746 of 804; 95% CI: 90.8, 94.5) (AUC, 0.95; 95% CI: 0.94, 0.96), with 74.7% (68 of 91; 95% CI: 64.5, 83.3) sensitivity for malignancies smaller than 2 cm.

Conclusion: The deep learning–based tool enabled accurate detection of pancreatic cancer on CT scans, with reasonable sensitivity for tumors smaller than 2 cm.

© RSNA, 2022

Online supplemental material is available for this article.

Pancreatic cancer (PC) has the lowest 5-year survival rate among cancers and is projected to become the second leading cause of cancer death in the United States by 2030 (1). Because the prognosis worsens precipitously once the tumor grows larger than 2 cm, early detection represents the most effective strategy to improve the dismal prognosis of PC (2). CT is the major imaging modality used to help detect PC, but its sensitivity for small tumors is modest, with approximately 40% of tumors smaller than 2 cm being missed (3). Furthermore, the diagnostic performance of CT is interpreter dependent and may be influenced by disparities in radiologist availability and expertise. Therefore, an effective tool to supplement radiologists in improving the sensitivity for PC detection is needed and constitutes a major unmet medical need.

Recent advances in deep learning (DL) have shown great promise in medical image analysis (4). In our previous single-center proof-of-concept study, we showed that a convolutional neural network (CNN) could accurately distinguish PC from noncancerous pancreas (5). However, in that study, segmentation of the pancreas (ie, identification of the pancreas) for analysis by the CNN was manually performed by radiologists. Segmentation of the pancreas is most challenging (6) because the pancreas borders multiple organs and structures and varies widely in shape and size, especially in patients with PC. However, a clinically applicable computer-aided detection (CAD) tool should enable segmentation and classification (ie, predicting presence or absence of PC), with minimal human annotation or labor.

Abbreviations

AUC = area under the receiver operating characteristic curve, CAD = computer-aided detection, CNN = convolutional neural network, DL = deep learning, LR = likelihood ratio, NHI = National Health Insurance, PC = pancreatic cancer

Summary

An automatic end-to-end deep learning–based detection tool could detect pancreatic cancer on CT scans in a nationwide real-world test data set with 91% accuracy, without requiring manual image labeling or preprocessing.

Key Results

- A deep learning tool for pancreatic cancer detection that was developed using contrast-enhanced CT scans obtained in 546 patients with pancreatic cancer and in 733 healthy control subjects achieved 89.9% sensitivity and 95.9% specificity in the internal test set (109 patients, 147 control subjects), which was similar to the sensitivity of radiologists (96.1%; P = .11).
- In a validation set comprising 1473 individual CT studies (669 patients, 804 control subjects) from institutions throughout Taiwan, the deep learning tool achieved 89.7% sensitivity and 92.8% specificity in distinguishing pancreatic cancer, with 74.7% sensitivity for pancreatic cancers smaller than 2 cm.

To fulfill this unmet clinical need, in the current study, we developed a CAD tool including a segmentation CNN to identify the pancreas on CT images and an ensemble classifier comprising five classification CNNs to predict whether the pancreas harbored PC. To ascertain its generalizability, we tested the CAD tool with the test set retrospectively derived from prospectively collected CT studies from real clinical practice throughout Taiwan. Thus, our aim was to develop and validate a DL-based tool with which to detect pancreatic cancer at CT.

Materials and Methods

This study was approved by the research ethics committee (201710050RINA, 201904116RINC). Informed consent from individual patients was waived because of the retrospective design. Figure 1 summarizes the construction of data sets (details in Appendix E1 [online]). An end-to-end workflow was implemented to analyze the CT images without manual annotation or processing (Fig 2) and consisted of image preprocessing; a segmentation CNN to segment the pancreas, including the tumor (if present); and an ensemble of five classification CNNs to determine whether the pancreas harbored PC. Computer codes are available online *(https://github.com/medalab-dladpcfncp/end_to_end_workflow)*.

Local Data Set

Contrast-enhanced portal venous phase CT studies in 546 patients with PC diagnosed between January 2006 and July 2018 were extracted from the imaging archive of a tertiary referral center for further manual labeling and analysis using convenience sampling. Inclusion criteria were as follows: *(a)* age of at least 18 years; *(b)* findings recorded in cancer registry; and *(c)* pancreatic adenocarcinoma confirmed at histology or cytology. If multiple studies were identified in a patient, then only the study immediately preceding the PC diagnosis was used.

CT studies in 1465 individuals with a normal pancreas obtained between January 2004 and December 2019 were extracted from the imaging archive; these individuals comprised the control group. The control studies were selected from liver or renal donors and randomly selected individuals with a normal or unremarkable pancreas (details in Appendix E1 [online]). The patients with PC were randomly divided (8:2 ratio) into the training and validation set and the local test set, respectively. The control group was randomly divided into two portions, one for constructing the training and validation set and the local test set and the other as part of the nationwide test set. Section thickness and image size were 0.7–1.5 mm and 512 × 512 pixels, respectively. CT scans were obtained in the portal venous phase, 70–80 seconds after intravenous administration of contrast medium.

National Health Insurance Data Set

The National Health Insurance (NHI) test set was provided by the Applications of Artificial Intelligence in Medical Images Database of NHI of Taiwan. The NHI is a single-payer comprehensive compulsory health insurance system covering in- and outpatient care for 99.8% of the Taiwanese population. Among the 23 164 health care facilities throughout Taiwan, 21 463 (93%) have a contract with the NHI (7). All patients with newly confirmed PC between January 2018 and July 2019 throughout Taiwan were identified by searching the registry of the NHI Major Illness/Injury Certificate (details in Appendix E1 [online]). For selection of control subjects with a normal pancreas, we identified all kidney donors and liver donors during the same period and extracted the CT studies performed for predonation evaluation from the NHI database. Demographic information and the imaging protocol used were unavailable. The CT studies of control subjects were reviewed, and the absence of radiologic abnormalities in the pancreas was confirmed by a radiologist (P.T.C.). In total, CT studies of 669 patients with PC and 72 control subjects were extracted from the NHI image database and were further combined with CT studies of 732 control subjects from the tertiary referral center imaging archive to construct the nationwide test data set.

Training of Segmentation and Classification Models

We had previously trained a segmentation CNN (8–10) using local data with radiologist-labeled pancreas and tumor and three US data sets with pancreatic parenchyma and masses labeled (11–13). The tumor and pancreas segmented by the segmentation CNN were combined into a volume of interest to be analyzed by the classification CNNs; therefore, the performance of the classification CNNs would not be negatively affected if the tumor was segmented as pancreas. To provide more robust performance and quantitative information on the confidence of the classification, five independently operating classification CNNs were trained using the same model structure but different subsets of the training and validation set from the tertiary referral center (437 patients with PC, 586 control subjects). CT was considered to show PC if the number of positive-predicting CNNs was equal to or greater than the smallest number yielding a positive likelihood ratio (LR) greater than one in the validation set (details in Appendix E1 [online]).
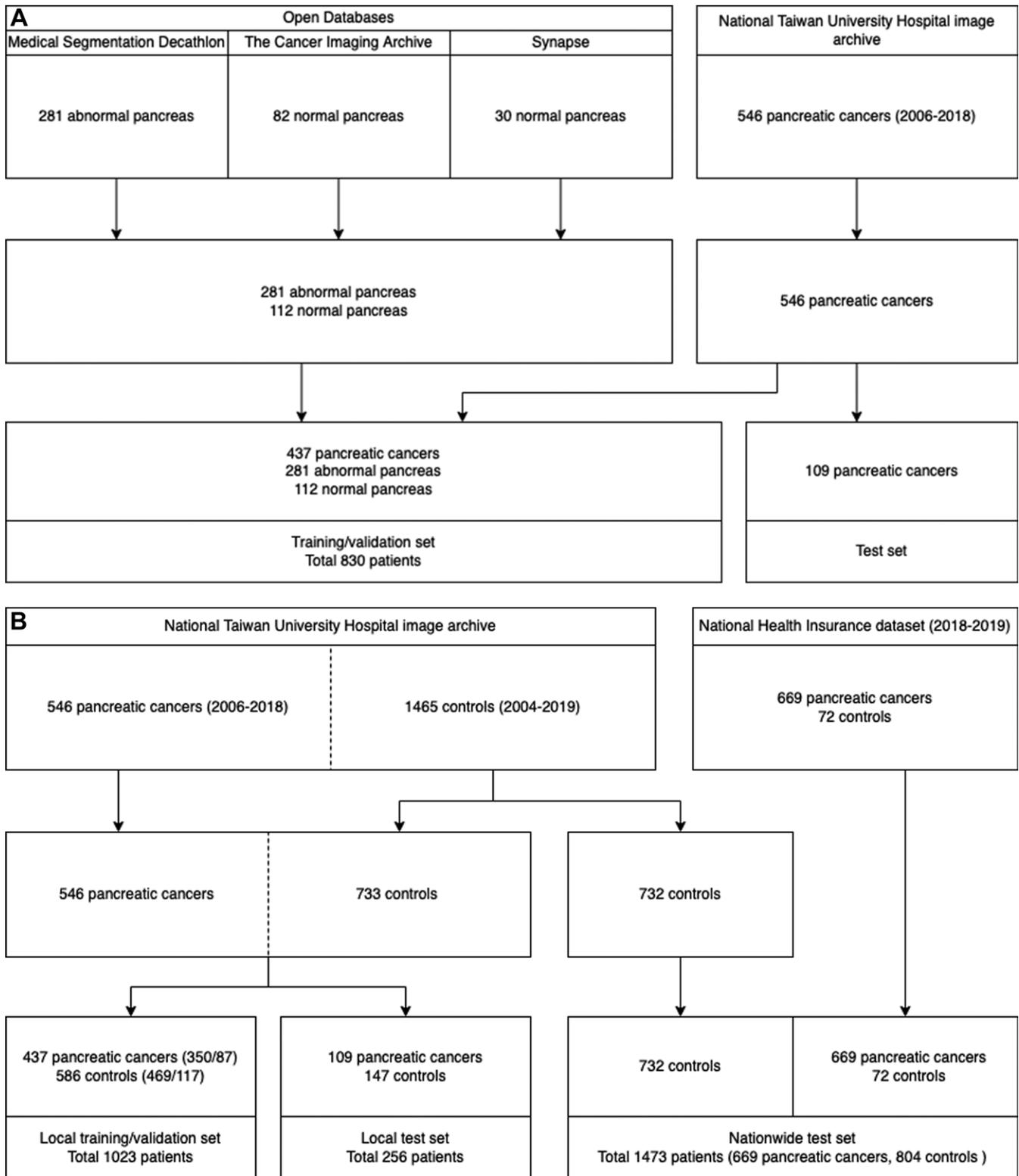
A

Open Databases			National Taiwan University Hospital image archive
Medical Segmentation Decathlon	The Cancer Imaging Archive	Synapse	
281 abnormal pancreas	82 normal pancreas	30 normal pancreas	546 pancreatic cancers (2006-2018)

281 abnormal pancreas
112 normal pancreas

546 pancreatic cancers

437 pancreatic cancers
281 abnormal pancreas
112 normal pancreas

Training/validation set
Total 830 patients

109 pancreatic cancers

Test set

B

National Taiwan University Hospital image archive		National Health Insurance dataset (2018-2019)
546 pancreatic cancers (2006-2018)	1465 controls (2004-2019)	669 pancreatic cancers 72 controls

546 pancreatic cancers

733 controls

732 controls

437 pancreatic cancers (350/87)
586 controls (469/117)

Local training/validation set
Total 1023 patients

109 pancreatic cancers
147 controls

Local test set
Total 256 patients

732 controls

669 pancreatic cancers
72 controls

Nationwide test set
Total 1473 patients (669 pancreatic cancers, 804 controls)

Figure 1: Data sets for the **(A)** segmentation model and **(B)** local and nationwide data sets for classification models.

Visual Assessment of Tumor Location Identified by the Segmentation Model

Consensus review by two experienced abdominal radiologists (P.T.C., K.L.L.) was conducted to assess whether the segmentation model correctly predicted the location of the tumor in the CT study classified as PC by the ensemble classifier. To explore whether the secondary signs of PC might have contributed to classification, the dilated pancreatic duct was masked by randomly selected neighboring pixels of the pancreas parenchyma.
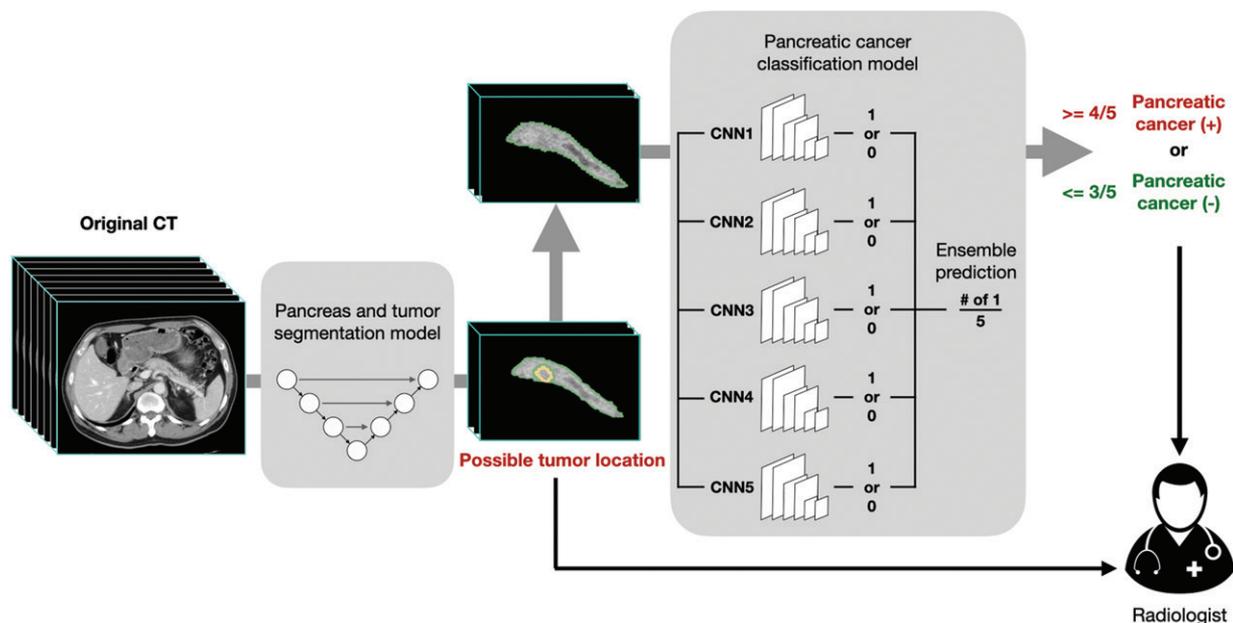
Figure 2: Workflow of the deep learning–based computer-aided detection tool. The segmentation masks passed from the segmentation convolutional neural network (CNN) to the classification CNNs included the pancreas and tumor (if present) combined, without separate delineation or identification between the pancreas and tumor. Solid arrows indicate output of the computer-aided detection tool.

Table 1: Characteristics of Pancreatic Cancer and Control Groups in Local Data Sets

Variable	Training and Validation Set		Test Set		All	
	PC Group	Control Group	PC Group	Control Group	PC Group	Control Group
No. of individuals	437	536	109	147	546	683
Age (y)	65 ± 12	54 ± 16	64 ± 12	55 ± 16	65 ± 12	54 ± 16
Sex						
Female	201 (46)	235 (44)	48 (44)	74 (50)	249 (46)	309 (49)
Male	236 (54)	301 (56)	61 (56)	73 (50)	297 (54)	374 (51)
Stage						
I	23 (5)	NA	4 (4)	NA	27 (5)	NA
II	175 (400)	NA	46 (42)	NA	221 (40)	NA
III	66 (15)	NA	13 (12)	NA	79 (14)	NA
IV	173 (40)	NA	46 (42)	NA	219 (40)	NA
Tumor size (cm)	2.9 (2.1–4.4)	NA	2.6 (2.1–3.9)	NA	2.9 (2.1–4.3)	NA

Note.—Continuous variables are presented as either means ± SDs or as medians with IQRs in parentheses. Categorical variables are presented as number of individuals, with the percentage in parentheses. Stage and tumor size are only available in patients with pancreatic cancer. NA = not applicable.

Statistical Analyses

The performance of the segmentation CNN was evaluated with Dice score per patient. The performance of each classification CNN and the ensemble classifier was assessed in various test sets to ascertain the respective sensitivity, specificity, and accuracy and the exact 95% CI for each. The sensitivity of the CAD tool was further assessed with stratification by tumor size and stage and was compared with sensitivity of radiologist interpretation as assessed in the original radiology report (details in Appendix E1 [online]). The receiver operating characteristic curves of each classification CNN and the ensemble classifier were constructed by plotting sensitivity against 1-specificity, as the threshold varied from 0 to 1. The area under the receiver operating characteristic curve (AUC) and its asymptotic 95% CI were calculated to evaluate the performance of the classifiers. LR was calculated as the ratio between the probability of the specific test result in individuals with the disease and the probability of the specific test result in those without the disease (14). The Fisher exact test and the Mann-Whitney U test were used to compare categorical and continuous variables between groups, respectively. The exact McNemar test was used to compare sensitivity for PC between the CAD tool and radiologist interpretation for CT studies in which radiologist reports were available (15). Trend analysis for proportions (ie, P trend) were implemented in STATA 17 (Stata, College Station, Tex) (16). Statistical analysis was conducted (T.W., D.C.), and $P < .05$ was considered to indicate a significant difference.

Results

Data Set Characteristics

In the local data set, the mean age of patients with PC (54% male, 46% female) was 65 years, and the mean age of control subjects (51% male, 49% female) was 54 years (Table 1). The median age of patients with PC in Taiwan was 67 years in male patients and 69 years in female patients, with a slight male predominance (54%) (17). There were 106 healthy donors in the local training and validation set and 22 in the local test set. The nationwide test set included 210 healthy donors, 72 from the NHI and 138 from the tertiary referral center. In the local test set, the segmentation model yielded median Dice scores of 0.70 (IQR, 0.63–0.75) and 0.54 (IQR, 0.28–0.73) in segmenting the pancreas and the tumor, respectively, and 0.76 (IQR, 0.71–0.80) for the pancreas and tumor combined. The performance of the segmentation model could not be assessed in the NHI data set because segmentation of the pancreas and tumor by radiologists as the ground truth was not feasible due to NHI regulations.

Differentiation between PC and Control Groups and Associated Positive Likelihood Ratios

Figure 3A–3C shows ROC curves of classification CNNs in the training and validation set, local test set, and nationwide test set, respectively. The AUCs of the five classification CNNs ranged from 0.95 to 0.97 in the local test set and from 0.94 to 0.95 in the nationwide test set. The final classification (with PC vs without PC) was determined as 4 in the training phase because the corresponding positive LR began to exceed 1 at this threshold in the validation set (Table 2). In the local test set of 256 individuals, the ensemble classifier achieved 89.9% sensitivity (98 of 109; 95% CI: 82.7, 94.9), 95.9% specificity (141 of 147; 95% CI: 91.3, 98.5), and 93.4% accuracy (239 of 256; 95% CI: 89.6, 96.1) (AUC, 0.96; 95% CI: 0.94, 0.99) in distinguishing CT images in the PC group from those in the control group (Table 3, Fig 3B). In the nationwide test set of 1473 individuals, the ensemble classifier achieved 89.7% sensitivity (600 of 669, 95% CI: 87.1, 91.9), 92.8% specificity (746 of 804, 95% CI: 90.8, 94.5), and 91.4% accuracy (1346 of 1473; 95% CI: 89.8, 92.8) (AUC, 0.95; 95% CI: 0.94, 0.96] (Table 3, Fig 3C). There is no statistical difference ($P = .20$) between specificity in the 72 control patients in the NHI database (97.2%; 95% CI: 90.3, 99.7) and specificity in the 732 control patients in the local data set (92.3%; 95% CI: 90.2, 94.2).

In both the local test set and the nationwide test set, the number of classification CNNs predicting PC was positively associated with the positive LR ($P_{trend} = .04$ and $P_{trend} = .03$, respectively) (Table 2). The positive LRs for prediction as PC by four and five CNNs were 2.22 (95% CI: 1.46, 3.37) and 25.01 (95% CI: 17.10, 36.56), respectively, in the nationwide test set and 8.09 (95% CI: 1.85, 35.42) and 29.0 (95% CI: 10.98, 76.60), respectively, in the local test set.
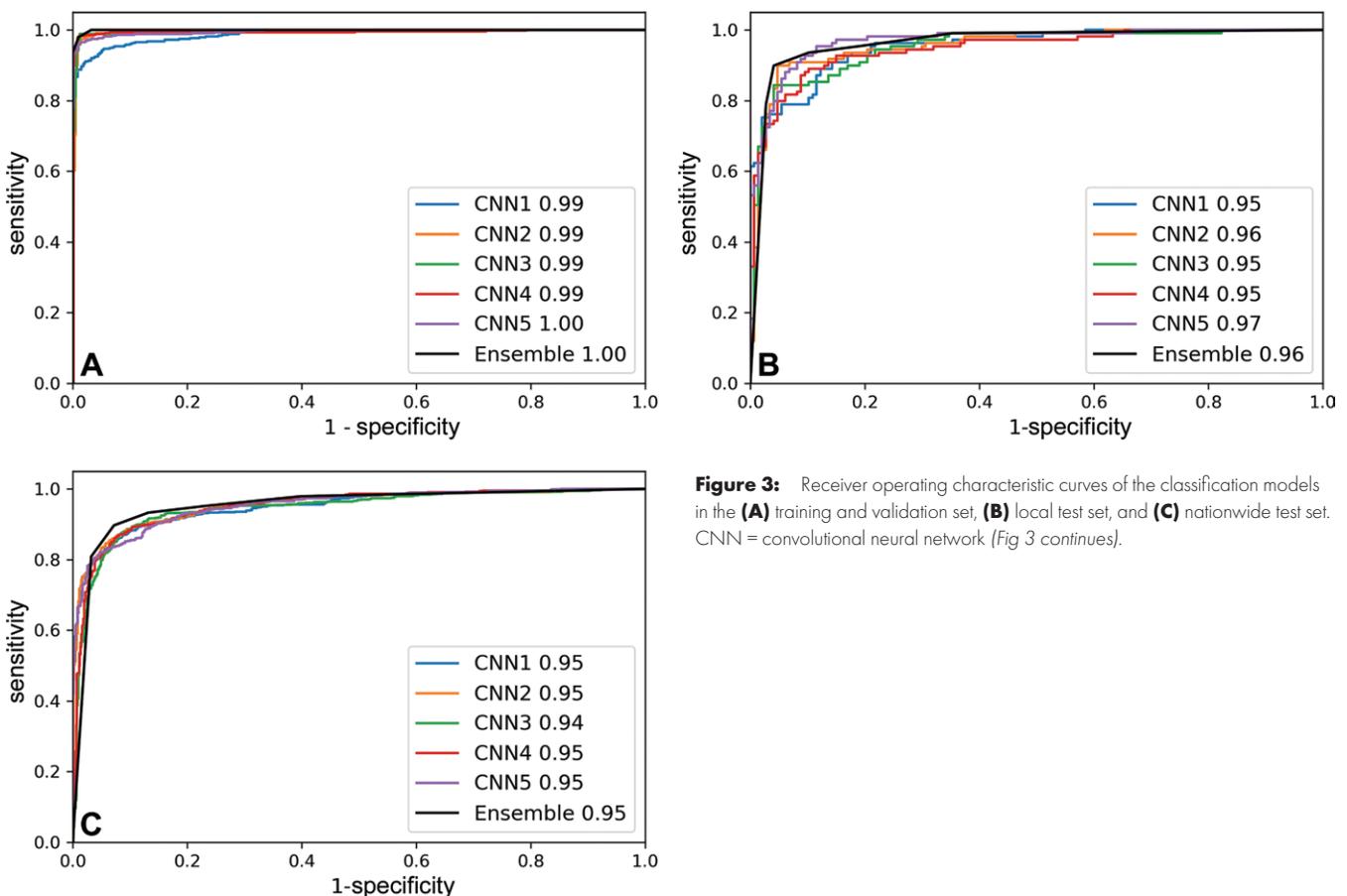

Figure 3: Receiver operating characteristic curves of the classification models in the **(A)** training and validation set, **(B)** local test set, and **(C)** nationwide test set. CNN = convolutional neural network *(Fig 3 continues).*
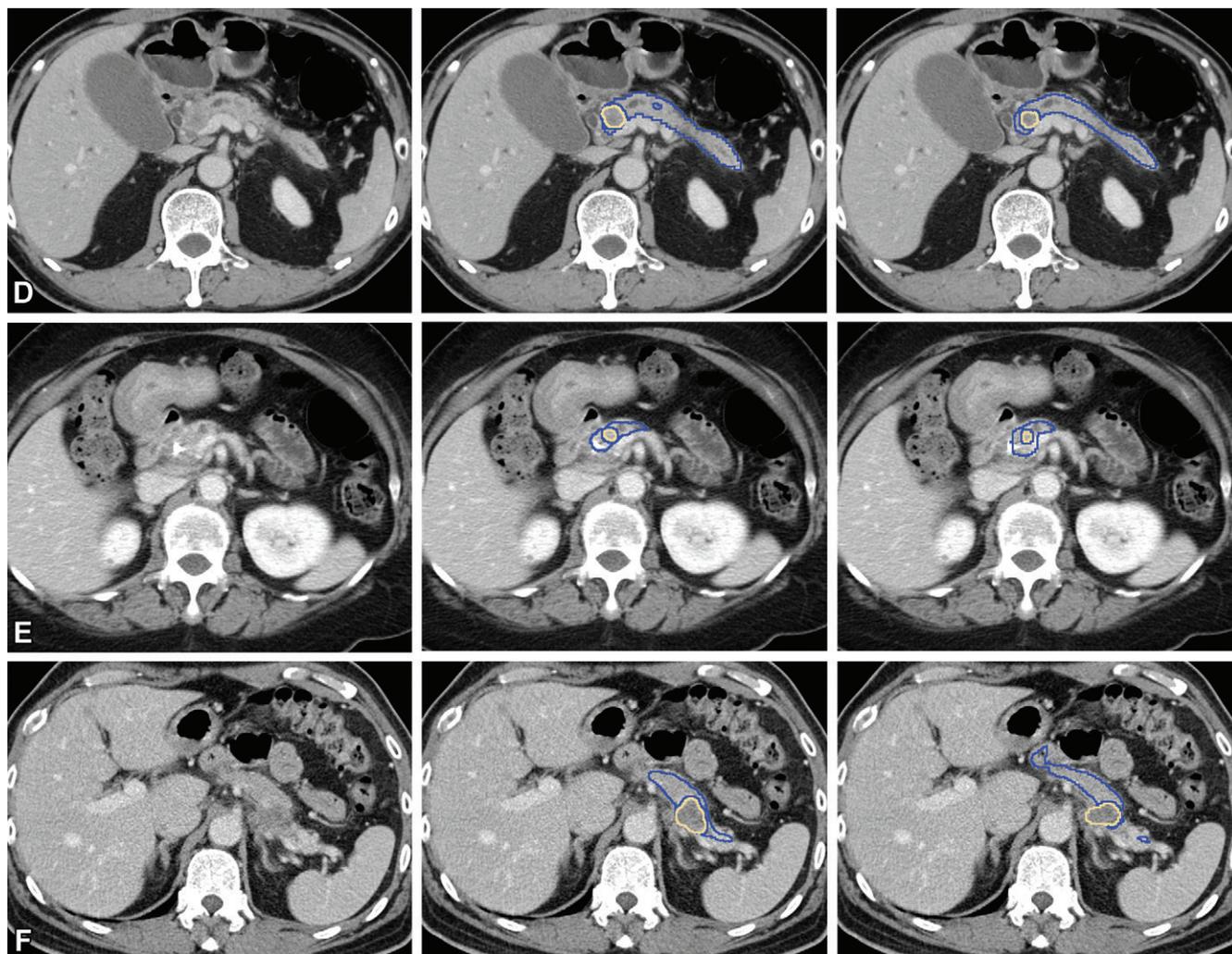
Figure 3 *(continued):* Representative CT scans (left column) with tumor at the pancreas **(D)** head, **(E)** body, and **(F)** tail show correspondence in tumor location between manual segmentation by radiologists (middle column) and predictions by the segmentation model (right column). Blue outline indicates the pancreas; yellow outline indicates tumor.

Table 2: Positive Likelihood Ratio according to Number of Classification Convolutional Neural Networks Predicting as Having Pancreatic Cancer

No. of CNNs	Local Training and Validation set			Local Test Set			Nationwide Test Set–National Health Insurance Data set		
	PC Group	Control Group	Positive LR	PC Group	Control Group	Positive LR	PC Group	Control Group	Positive LR
5	409	0	Infinity	86	4	29.00 (10.98, 76.60)	541	26	25.01 (17.10, 36.56)
4	19	6	4.25 (1.71, 10.54)	12	2	8.09 (1.85, 35.42)	59	32	2.22 (1.46, 3.37)
3	9	13	0.93 (0.40, 2.15)	4	9	0.60 (0.19, 1.90)	24	48	0.60 (0.37, 0.97)
2	0	14	0	2	13	0.21 (0.05, 0.90)	13	82	0.19 (0.11, 0.34)
1	0	75	0	4	24	0.22 (0.08, 0.63)	18	132	0.16 (0.10, 0.27)
0	0	478	0	1	95	0.01 (0.00, 0.10)	14	484	0.03 (0.02, 0.06)

Note.—The threshold of the number of positive-predictive convolutional neural networks (CNNs) for predicting pancreatic cancer by the ensemble classifier was 4 based on the finding that positive likelihood ratio (LR) exceeded 1 in the local validation set at this threshold. Data in parentheses are 95% CIs. PC = pancreatic cancer.

Table 3: Performance of Computer-aided Detection Tool and Radiologists Based on the Original Report in Differentiating Between CT Studies with and Those without Pancreatic Cancer

Test Set	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	CAD vs Radiologist			
					Sensitivity of CAD (%)	Sensitivity of Radiologist (%)	Difference	P Value
Local training and validation set (cancer group: 437 patients; control group: 586 patients)	97.9 (96.1, 99.1) [428/437]	99.0 (97.8, 99.6) [580/586]	98.5 (97.6, 99.2) [1008/1023]	1.00 (1.00, 1.00)	97.8 (95.9, 99.0) [406/415]*	96.1 (93.8, 97.8) [399/415]*	0.017 (−0.006, 0.040)	.17
Local test set (cancer group: 109 patients; control group: 147 patients)	89.9 (82.7, 94.9) [98/109]	95.9 (91.3, 98.5) [141/147]	93.4 (89.6, 96.1) [239/256]	0.96 (0.94, 0.99)	90.2 (82.7, 96.2) [92/102]†	96.1 (90.3, 98.9) [98/102]†	−0.059 (−0.128, 0.010)	.11
Nationwide test set (cancer group: 669 patients; control group: 804 patients)	89.7 (87.1, 91.9) [600/669]	92.8 (90.8, 94.5) [746/804]	91.4 (89.8, 92.8) [1346/1473]	0.95 (0.94, 0.96)	…	…	…	…

Note.—The ensemble classifier deemed CT findings positive if four or more of the five classification CNNs deemed it positive. Data in parentheses are 95% CIs. Data in brackets are numerators and denominators used to calculate percentages. AUC = area under the receiver operating characteristic curve, CAD = computer-aided detection.

* Excluding 22 patients without radiologist report.

† Excluding seven patients without radiologist report.

Comparison between Model Segmentation and Radiologist Interpretation with Respect to Tumor Location

In the local test set, 98 PC cases were correctly classified by the ensemble classifier, and the segmentation model-predicted tumor location was correct in 86 (87.8%) (Fig 3D–3F) and immediately adjacent to the radiologist-determined tumor in three (3.1%). In the remaining nine (9.2%) PC cases with a median tumor size of 2.3 cm, the location of the segmentation model-predicted tumor was inaccurate. In the nine patients with PC in which segmentation CNN-predicted tumor location was inaccurate, the segmentation model-predicted pancreas encompassed the tumor in six patients and included approximately half of the tumor in one patient. In the remaining two cases, the segmentation CNN classified the tumor as neither pancreas nor tumor but correctly segmented the nontumorous portion of the pancreas, which displayed secondary signs of PC (18), including upstream dilation of the pancreatic duct with abrupt duct cutoff in one case and upstream dilation of the pancreatic duct with parenchymal atrophy in the other (Fig 4A, 4B). To understand whether the secondary signs of PC in those two patients might have contributed to their classification as PC, we repeated the analysis after masking the dilated duct in the former PC case (Fig 4A), and the case was classified as noncancerous by the ensemble classifier (Fig 5). In the other case (Fig 4B), masking of secondary signs was not feasible due to lack of normal-appearing pancreas parenchyma. Among the six control patients in whom PC was incorrectly predicted by the ensemble classifier in the local test set, no apparent tumor was identified by the segmentation CNN in four cases. In the remaining two control patients, collateral veins secondary to idiopathic portal vein thrombosis were segmented as tumor in one patient by the segmentation CNN (Fig 4C); in the other, the pancreas parenchyma beside a biliary stent for palliation of obstructive jaundice from hepatocellular carcinoma was segmented as tumor (Fig 4D).

Sensitivity according to Tumor Size and Stage and Comparison with Radiologist Interpretation

In the local test set excluding seven patients without formal radiologist reports, there was no significant difference between the sensitivity of the CAD tool and that of the original radiologist report (CAD tool, 90.2%; radiologist report, 96.1%; P = .11) (Table 3). Four PCs were missed by radiologists, of which two (1.3 and 3.3 cm in size) were correctly classified by the ensemble classifier. Stratified analysis according to primary tumor size and stage also showed similar sensitivity between the CAD tool and attending radiologists (Table 4). For tumors smaller than 2 cm, the sensitivities for detecting PC with the CAD were 87.5% (21 of 24, 95% CI: 67.6, 97.3) in the local test set and 74.7% (68 of 91, 95% CI: 64.5, 83.3) in the nationwide test set. Radiologist reports and information on cancer stage were not available in the NHI data set. Comparison between the CAD tool and radiologists with respect to specificity was not feasible, given that control patients were selected based on the statement of a normal or unremarkable pancreas in the radiologist report, because otherwise it was inappropriate to exclude the possibility of PC.

Discussion

Enhancing the sensitivity of CT for small or early pancreatic cancer (PC) is a major unmet clinical need. This study developed and validated an end-to-end deep learning (DL)–based computer-aided detection (CAD) tool to differentiate between CT studies with and those without PC without requiring manual image labeling or preprocessing. In a test set retrospectively derived from prospectively collected real-world images from institutions throughout Taiwan, the tool achieved 89.7% sensitivity (74.7% for PCs <2 cm) and 92.8% specificity, demonstrating high robustness and generalizability. No significant difference in sensitivity was noted between the CAD tool (90.2%) and attending radiologists (96.1%) in a tertiary academic institution with a large volume of patients with PC (P = .11); therefore, the CAD tool might have higher sensitivity compared with radiologists who are less experienced with PC.

Generalizability to new data sets is a prerequisite for clinical applications, but the performance of CNNs often deteriorates significantly in external validation (19–22), and few studies of medical image analysis by artificial intelligence undergo external validation (23). A recent study also investigated end-to-end diagnosis of PC by DL and reported an accuracy of 87.6% (24), but no external validation was attempted.

Our current study had several methodologic advances compared with the previous study (5). To obviate the need for manually segmenting the pancreas, we trained a segmentation CNN based on a coarse-to-fine neural architecture search, which is most suitable for pancreas segmentation (8,9). The segmentation CNN was trained with images from multiple institutions and races or ethnicities and yielded Dice scores
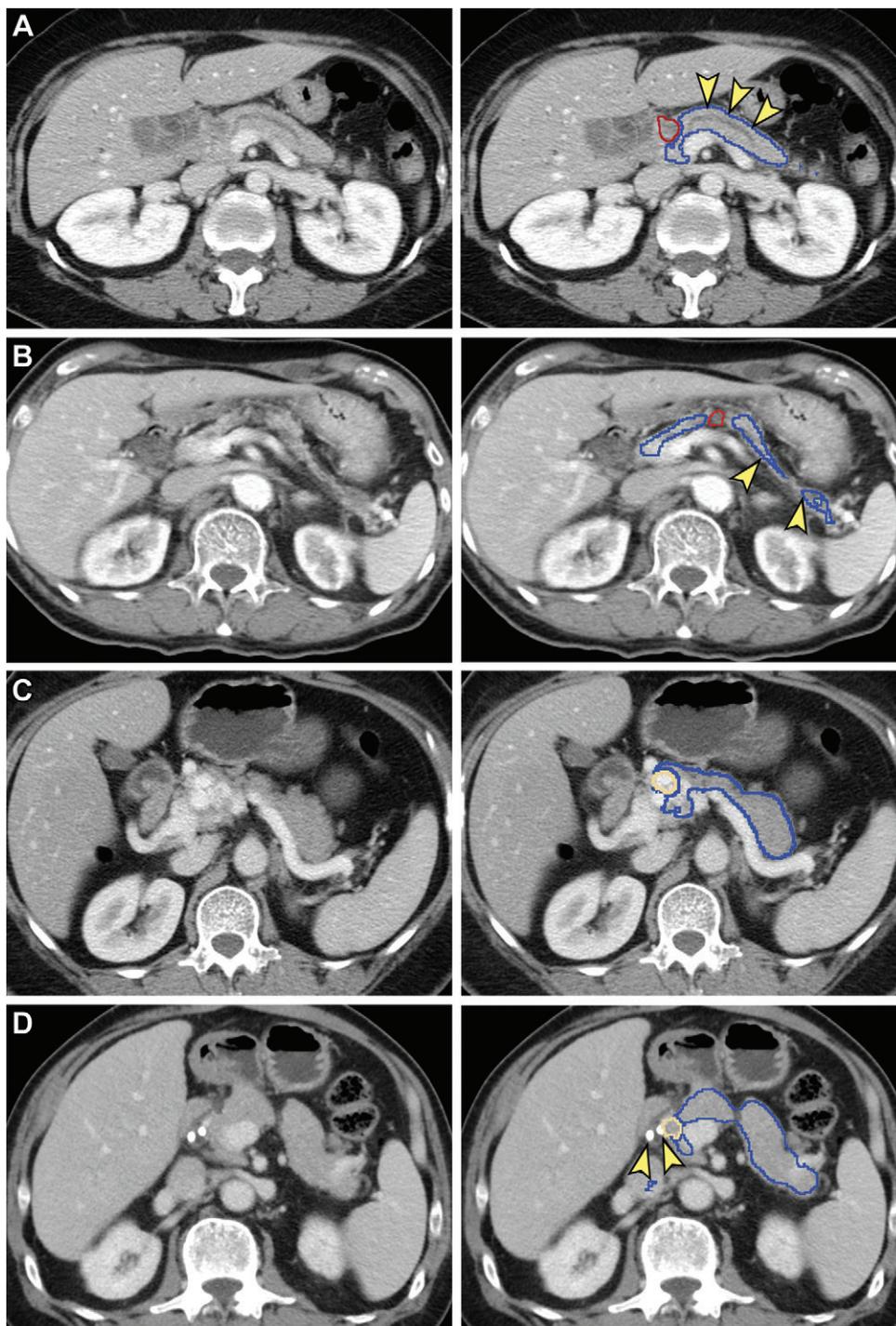

Figure 4: False-negative **(A, B)** and false-positive **(C, D)** tumor segmentation by the segmentation model. Blue and yellow outlines indicate normal pancreas and tumor segmented with the segmentation model, respectively. Images in the left column are original unannotated CT scans. **(A, B)** Tumors (red outline) were not segmented by the segmentation convolutional neural network. The upstream pancreas shows secondary signs of pancreatic cancer, including dilation of the pancreatic duct with abrupt cutoff (arrowhead in **A**) and parenchymal atrophy with dilation of the pancreatic duct (arrowhead in **B**). **(C)** Collateral veins secondary to idiopathic portal vein thrombosis were incorrectly segmented as tumor by the segmentation model. **(D)** Pancreatic parenchyma adjacent to biliary stents (arrowhead) placed for relieving obstructive jaundice from hepatocellular carcinoma was incorrectly segmented as tumor by the segmentation model.

comparable to the best-performing pancreas segmentation DL models reported (25–27). Second, we adopted three-dimensional volumetric analyses in segmentation and classification, whereas the previous study used two-dimensional analyses,
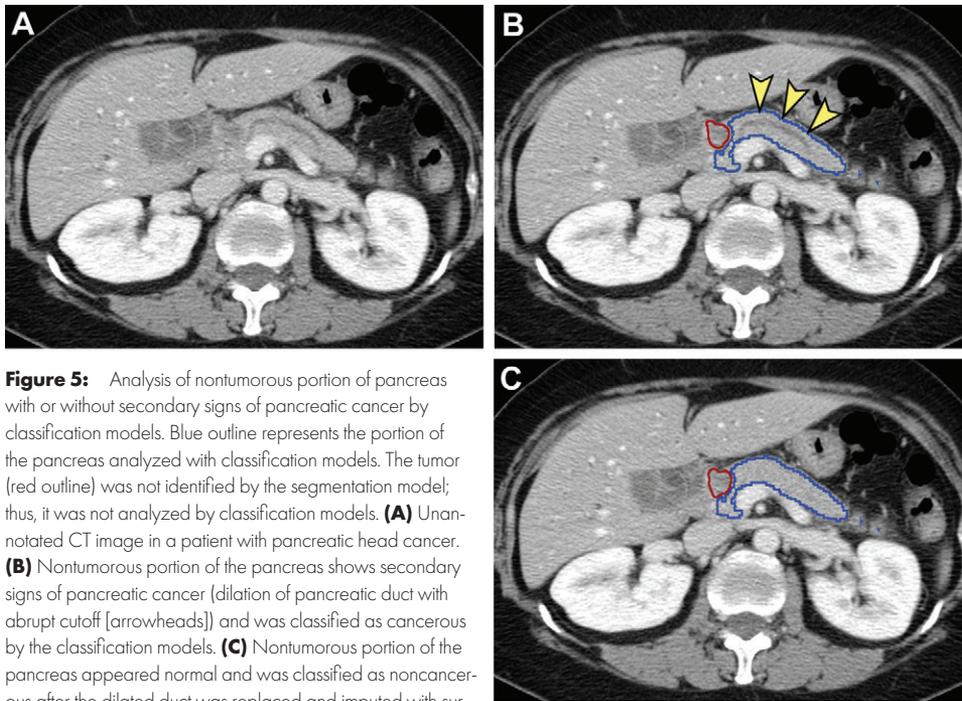
Figure 5: Analysis of nontumorous portion of pancreas with or without secondary signs of pancreatic cancer by classification models. Blue outline represents the portion of the pancreas analyzed with classification models. The tumor (red outline) was not identified by the segmentation model; thus, it was not analyzed by classification models. **(A)** Unannotated CT image in a patient with pancreatic head cancer. **(B)** Nontumorous portion of the pancreas shows secondary signs of pancreatic cancer (dilation of pancreatic duct with abrupt cutoff [arrowheads]) and was classified as cancerous by the classification models. **(C)** Nontumorous portion of the pancreas appeared normal and was classified as noncancerous after the dilated duct was replaced and imputed with surrounding normal-appearing pancreas parenchyma.

losing information between neighboring CT sections and ignoring the spatial correlations among sections. Integrating the segmentation and classification CNNs enabled automatic analysis of a CT study in approximately 30 seconds, supporting the feasibility for clinical deployment.

This CAD tool could provide a multitude of information to assist clinicians. Besides determining whether the images showed PC, the tool could indicate the region of suspicion to expedite radiologist interpretation. In approximately 90% of PCs correctly classified by the CAD tool, the segmentation CNN correctly identified the tumor location. As the segmentation CNN might falsely segment normal pancreas parenchyma as tumor when encountering uncommon conditions or artifacts, the area determined as tumor by the segmentation CNN should be interpreted as the probable location of the tumor only in cases classified as PC by the classification CNNs rather than being overinterpreted as another set of classification. Furthermore, the CAD tool could provide the positive LR, a measure of the confidence of the classification (PC vs non-PC), which could be multiplied with the pretest odds determined based on clinical parameters and experience to derive the posttest odds and probability of PC, thereby better informing the subsequent diagnostic-therapeutic process than a simple binary classification.

The comparable sensitivity between the CAD tool and experienced radiologists at a tertiary referral center supports the idea that the CAD tool might be useful for reducing the miss rate attributed to disparities in expertise. The radiologists' high sensitivity for PCs smaller than 2 cm observed in this tertiary referral center might not be generalizable to other institutions. Previous research showed that approximately 40% of PCs smaller than 2 cm were missed on CT scans (28), whereas our CAD tool achieved 87.5% and 74.7% sensitivity for PCs smaller than 2 cm in the local and nationwide test sets, respectively.

The interesting finding that the classification CNNs correctly classified two cases of PC by analyzing only the nontumorous portion of the pancreas underscored the diagnostic value of secondary signs of PC, including pancreatic duct dilatation, upstream pancreatic parenchymal atrophy, and abrupt cutoff of the pancreatic duct (18). Besides features in the tumor, secondary signs in the nontumorous portion of the pancreas are important clues to occult PCs and thus should be taken advantage of in developing CAD tools for PC. Our explorative analysis showed that the nontumorous portion of the pancreas with secondary signs of PC and the portion without such signs were classified as with PC and without PC, respectively. Our results suggest that the classification CNNs might have learned the secondary signs of PC, in line with the notion that DL can spontaneously capture distinguishing imaging features through learning from examples (29,30).

This study had limitations. First, radiologist reports were unavailable in the NHI data set; thus, comparison between the CAD tool and radiologists was not feasible. Second, the difference in sensitivity between the CAD tool and radiologists at a tertiary referral center might not be generalizable to other institutions. Third, although the nationwide NHI data set included variations in imaging parameters and quality occurring in real clinical practice and thus represented a most rigorous test set, the Taiwanese population in the test set was predominantly Asian and relatively homogenous in race and ethnicity. Fourth, control subjects from the NHI database were healthy donors and were likely younger. However, no significant difference in specificity was noted between control subjects from the NHI and those from local databases; thus, this limitation should not have favorably biased the results. Last, the control group did not include patients with pancreatic abnormalities other than PC, many of which require tissue sampling for confirmatory diagnosis. We seek to include other pancreatic abnormalities and prospectively assess the potential usefulness of the CAD tool in clinical settings in a future study.

In conclusion, this study developed an end-to-end deep learning–based computer-aided detection (CAD) tool that could accurately and robustly detect pancreatic cancers (PCs) on contrast-enhanced CT scans. The CAD tool may be a useful supplement for radiologists to enhance detection of PC. Our results also suggest that the classification convolutional neural networks might have learned the secondary signs of PC, which warrants further investigation. While the results of this study provide strong

Table 4: Sensitivity of Computer-sided Detection Tool and Radiologists Stratified by Tumor Stage and Size

Stage and Size	CAD Tool	Local Test Set: CAD vs Radiologist				Nationwide Test Set: CAD Tool
		Sensitivity of CAD (%)	Sensitivity of Radiologist (%)	Difference	P Value	
Stage						
I	50.0 (6.8, 93.2) [2/4]	50.0 (6.8, 93.2) [2/4]	75.0 (19.4, 99.4) [3/4]	−0.250 (−0.898, 0.398)	>.99	NA*
II	89.1 (76.4, 96.4) [41/46]	88.6 (75.4, 96.3) [39/44]†	97.7 (88.0, 99.9) [43/44]†	−0.091 (−0.195, 0.013)	.13	NA*
III	92.3 (64.0, 99.8) [12/13]	91.7 (61.5, 98.8) [11/12]‡	91.7) (61.5, 98.8) [11/12]‡	0.000 (−0.221, 0.221)	>.99	NA*
IV	93.5 (82.1, 98.6) [43/46]	95.2 (83.8, 99.4) [40/42]§	97.6 (87.4, 99.9) [41/42]§	−0.024 (−0.104, 0.056)	>.99	NA*
Size (cm)						
<2	87.5 (67.6, 97.3) [21/24]	86.4 (65.1, 97.1) [19/22]‖	95.5 (77.2, 99.9) [21/22]‖	−0.091 (−0.258, 0.076)	.63	74.7 (64.5, 83.3) [68/91]
2–4	89.8 (79.2, 96.2) [53/59]	90.9 (80.0, 97.9) [50/55]#	96.4) (87.5, 99.6) [53/55]#	−0.055 (−0.146, 0.036)	.38	91.4 (88.0, 94.0) [339/371]
>4	92.3 (74.9, 99.1) [24/26]	92.0 (74.0, 99.0) [23/25]**	96.0 (79.6, 99.9) [24/25]**	−0.040 (−0.171, 0.091)	>.99	93.2 (88.9, 96.3) [193/207]

Note.—Data in parentheses are 95% CIs, and data in brackets are numerators and denominators used to calculate percentages. The cancer stage in the local test set was extracted from the cancer registry and determined in accordance with current standards according to the American Joint Committee on Cancer (seventh edition) based on histologic findings in patients undergoing surgical resection and on complete staging work-up in patients not undergoing surgery. Tumor size was determined according to the histologic report if the patient underwent surgery. Otherwise, tumor size was based on the CT report. In the nationwide data set, information on cancer stage was not available, and tumor size was measured on CT images and was recorded by a radiologist. CAD = computer-aided detection, NA = not applicable.

* Information on tumor stage was not available in the nationwide test set.

† Excluding two patients without radiologist report.

‡ Excluding one patient without radiologist report.

§ Excluding four patients without radiologist report.

‖ Excluding two patients without radiologist report.

Excluding four patients without radiologist report.

** Excluding one patient without radiologist report.

support for the generalizability of the CAD tool in the Taiwanese and perhaps Asian populations, the performance of the CAD tool in other populations needs to be evaluated further.

References

1. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 2014;74(11):2913–2921.
2. Chari ST. Detecting early pancreatic cancer: problems and prospects. Semin Oncol 2007;34(4):284–294.
3. Kang JD, Clarke SE, Costa AF. Factors associated with missed and misinterpreted cases of pancreatic ductal adenocarcinoma. Eur Radiol 2021;31(4):2422–2432.
4. Chen PT, Chang D, Wu T, Wu MS, Wang W, Liao WC. Applications of artificial intelligence in pancreatic and biliary diseases. J Gastroenterol Hepatol 2021;36(2):286–294.
5. Liu KL, Wu T, Chen PT, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. Lancet Digit Health 2020;2(6):e303–e313.
6. Yao X, Song Y, Liu Z. Advances on pancreas segmentation: a review. Multimed Tools Appl 2020;79(9):6799–6821.
7. National Health Insurance Administration Ministry of Health and Welfare. https://www.nhi.gov.tw/English/Content_List.aspx?n=8FC0974BBFEFA56D&topn=ED4A30E51A609E49. Published 2016. Accessed May 21, 2021.
8. Wang P, Shen C, Roth HR, et al. Automated Pancreas Segmentation Using Multi-institutional Collaborative Deep Learning. In: Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning. Cham, Switzerland: Springer International Publishing, 2020; 192–200.
9. Shen C, Wang P, Roth HR, et al. Multi-task federate learning for heterogeneous pancreas segmentation. In: Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning. Cham, Switzerland: Springer International Publishing, 2021; 101–110.
10. Yu Q, Yang D, Roth H, et al. C2FNAS: Coarse-to-Fine Neural Architecture Search for 3D Medical Image Segmentation. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, June 13–19, 2020. Piscataway, NJ: IEEE, 2020; 4125–4134.
11. Simpson AL, Antonelli M, Bakas S, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv [preprint] https://arxiv.org/abs/1902.09063. Posted February 25, 2019. Accessed February 27, 2022.
12. Pancreas-CT - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT. Accessed June 14, 2019.
13. Gibson E, Giganti F, Hu Y, et al. Multi-organ Abdominal CT Reference Standard Segmentations. . Published February 22, 2018. Accessed February 27, 2022.
14. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;329(7458):168–169.

15. Zhou XH. Statistical methods in diagnostic medicine. Hoboken, NJ: Wiley, 2011.

16. Royston P. PTREND: Stata module for trend analysis for proportions. Boston College Department of Economics. https://EconPapers.repec.org/RePEc:boc:bocode:s426101. Published 2014. Accessed May 21, 2022.

17. Health Promotion Administration. https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=269&pid=14913. Published 2016. Accessed May 21, 2022.

18. Tamada T, Ito K, Kanomata N, et al. Pancreatic adenocarcinomas without secondary signs on multiphasic multidetector CT: association with clinical and histopathologic features. Eur Radiol 2016;26(3):646–655.

19. Bai W, Sinclair M, Tarroni G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J Cardiovasc Magn Reson 2018;20(1):65.

20. Barish M, Bolourani S, Lau LF, Shah S, Zanos TP. External validation demonstrates limited clinical utility of the interpretable mortality prediction model for patients with COVID-19. Nat Mach Intell 2020;3(1):25–27.

21. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016;316(22):2402–2410.

22. Beede E, Baylor E, Hersch F, et al. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery, 2020; 1–12.

23. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. Korean J Radiol 2019;20(3):405–410.

24. Si K, Xue Y, Yu X, et al. Fully end-to-end deep-learning-based diagnosis of pancreatic tumors. Theranostics 2021;11(4):1982–1990.

25. Turečková A, Tureček T, Komínková Oplatková Z, Rodríguez-Sánchez A. Improving CT Image Tumor Segmentation Through Deep Supervision and Attentional Gates. Front Robot AI 2020;7:106.

26. Zhang Y, Wu J, Liu Y, et al. A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set. Med Image Anal 2021;68:101884.

27. Antonelli M, Reinke A, Bakas S, et al. The Medical Segmentation Decathlon. arXiv [preprint] https://arxiv.org/abs/2106.05735. Posted June 10, 2021. Accessed February 27, 2022.

28. Dewitt J, Devereaux BM, Lehman GA, Sherman S, Imperiale TF. Comparison of endoscopic ultrasound and computed tomography for the preoperative evaluation of pancreatic cancer: a systematic review. Clin Gastroenterol Hepatol 2006;4(6):717–725; quiz 664.

29. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

30. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019;25(1):24–29.