nature medicine

Article

https://doi.org/10.1038/s41591-023-02640-w

Large-scale pancreatic cancer detection via non-contrast CT and deep learning

Received: 9 February 2023

Accepted: 12 October 2023

Published online: 20 November 2023

Check for updates

Kai Cao[1,19], Yingda Xia[2,19], Jiawen Yao[3,4,19], Xu Han[5,19], Lukas Lambert[6,19], Tingting Zhang[7,19], Wei Tang[8,19], Gang Jin[9], Hui Jiang[10], Xu Fang[1], Isabella Nogues[11], Xuezhou Li[1], Wenchao Guo[3,4], Yu Wang[3,4], Wei Fang[3,4], Mingyan Qiu[3,4], Yang Hou[12], Tomas Kovarnik[13], Michal Vocka[14], Yimei Lu[8], Yingli Chen[9], Xin Chen[15], Zaiyi Liu[15], Jian Zhou[16], Chuanmiao Xie[16], Rong Zhang[16], Hong Lu[17], Gregory D. Hager[18], Alan L. Yuille[18], Le Lu[2], Chengwei Shao[1] ✉, Yu Shi[12] ✉, Qi Zhang[5] ✉, Tingbo Liang[5] ✉, Ling Zhang[2] ✉ & Jianping Lu[1] ✉

Pancreatic ductal adenocarcinoma (PDAC), the most deadly solid malignancy, is typically detected late and at an inoperable stage. Early or incidental detection is associated with prolonged survival, but screening asymptomatic individuals for PDAC using a single test remains unfeasible due to the low prevalence and potential harms of false positives. Non-contrast computed tomography (CT), routinely performed for clinical indications, offers the potential for large-scale screening, however, identification of PDAC using non-contrast CT has long been considered impossible. Here, we develop a deep learning approach, pancreatic cancer detection with artificial intelligence (PANDA), that can detect and classify pancreatic lesions with high accuracy via non-contrast CT. PANDA is trained on a dataset of 3,208 patients from a single center. PANDA achieves an area under the receiver operating characteristic curve (AUC) of 0.986–0.996 for lesion detection in a multicenter validation involving 6,239 patients across 10 centers, outperforms the mean radiologist performance by 34.1% in sensitivity and 6.3% in specificity for PDAC identification, and achieves a sensitivity of 92.9% and specificity of 99.9% for lesion detection in a real-world multi-scenario validation consisting of 20,530 consecutive patients. Notably, PANDA utilized with non-contrast CT shows non-inferiority to radiology reports (using contrast-enhanced CT) in the differentiation of common pancreatic lesion subtypes. PANDA could potentially serve as a new tool for large-scale pancreatic cancer screening.

Pancreatic ductal adenocarcinoma (PDAC) is the deadliest solid malignancy worldwide, and causes approximately 466,000 deaths per year[1]. Despite the poor prognosis of PDAC, its early or incidental detection has been shown to substantially improve patient survival[2–7]. Recent studies indicate that high-risk individuals with screen-detected PDAC have a median overall survival of 9.8 years, substantially longer than the 1.5 years for those diagnosed outside of surveillance (for example, via standard clinical diagnostic techniques)[6]. As such,

A full list of affiliations appears at the end of the paper. ✉e-mail: cwshao@sina.com; 18940259980@163.com; qi.zhang@zju.edu.cn; liangtingbo@zju.edu.cn; ling.z@alibaba-inc.com; cjr.lujianping@vip.163.com

https://doi.org/10.1038/s41591-023-02640-w

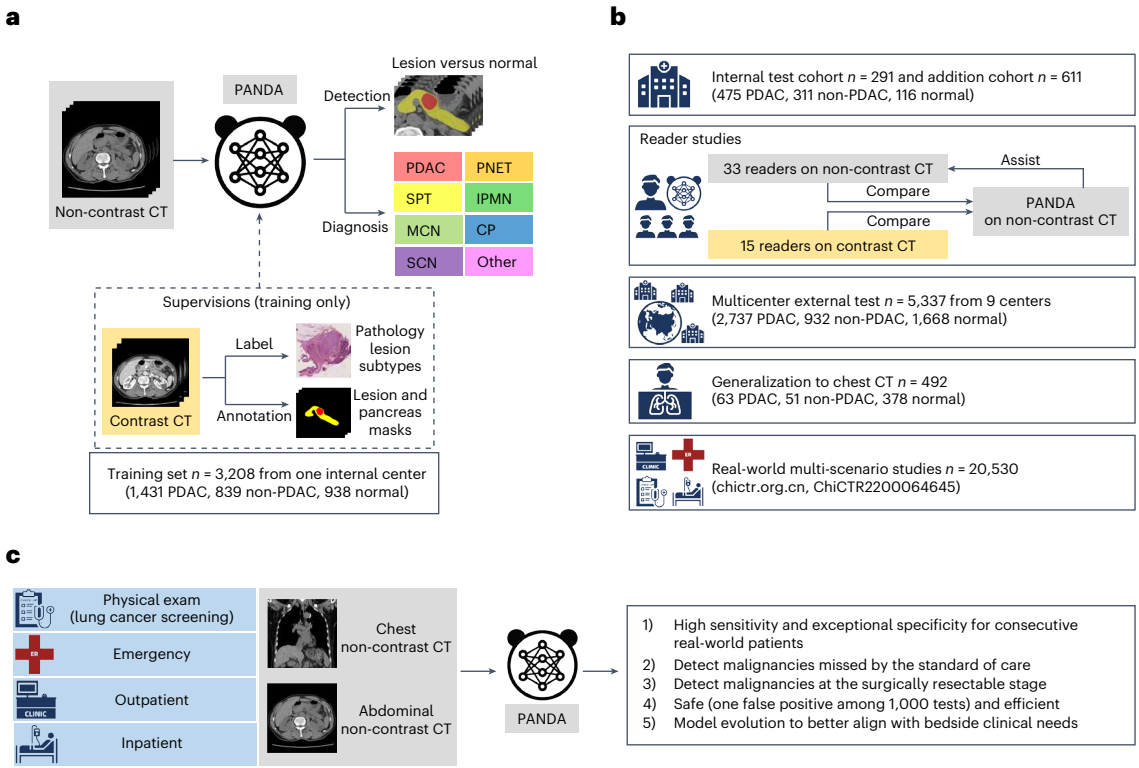


Fig. 1 | Overview of PANDA's development, evaluation and clinical translation. **a**, Model development. PANDA takes non-contrast CT as input and outputs the probability and the segmentation mask of possible pancreatic lesions, including PDAC and seven non-PDAC subtypes; PANDA was trained with pathology-confirmed patient-level labels and lesion masks annotated on contrast CT images. CP, chronic pancreatitis. **b**, Model evaluation. We evaluate the performance of PANDA on the internal test cohort, two reader studies (on non-contrast and contrast CT, respectively), external test cohorts consisting of nine centers, a chest CT cohort, and real-world multi-scenario studies (the clinical trial includes two real-world studies; chictr.org.cn, ChiCTR2200064645). **c**, Model clinical translation. The real-world clinical evaluation answers five critical questions to close the clinical translational gap for PANDA.

screening of PDAC holds the greatest promise to reduce PDAC-related mortality[8]. However, due to the relatively low prevalence of PDAC, effective screening in the general population requires high sensitivity and exceptionally high specificity to mitigate the risk of over-diagnosis. Current screening techniques are limited in this regard, and thus cannot be implemented in the general population as urgently needed[9,10].

Non-contrast computed tomography (CT) is widely used in physical examination centers and hospitals in low-resource regions. Compared with contrast-enhanced CT (the primary imaging modality for diagnosing PDAC[9]), non-contrast CT exposes patients to lower radiation doses and eliminates the risk of adverse reactions to the contrast agents. In addition to abdominal non-contrast CT routinely used in emergency departments and community hospitals, chest non-contrast CT also can fully or partially scan the pancreas region and is the most frequently performed CT exam (that is, it accounts for nearly 40% of all performed CT exams)[11] in multiple clinical scenarios, such as for lung cancer screening. Although identifying PDAC from non-contrast CT is challenging even for experienced radiologists, recent studies have shown that artificial intelligence (AI) can match or surpass human experts on various medical image analysis tasks[12–17]; moreover, AI is capable of synthesizing contrast-enhanced medical images from regular images[18–21]. AI-based opportunistic screening[22] via non-contrast CT has the potential to advance early detection of PDAC in the vast population of asymptomatic patients under several clinical domains, with minimal additional cost and exposure to radiation.

In this study we show that our proposed AI approach, PANDA (pancreatic cancer detection with AI, Fig. 1), is capable of detecting and diagnosing PDAC and non-PDAC lesions on non-contrast CT with high accuracy and can be readily utilized for opportunistic screening in large-scale asymptomatic patient populations. This will result in safe and effective detection of early-stage malignancies missed by standard of care diagnostic techniques, and in some cases will enable timely treatment with intent to cure. Our study first evaluates PANDA internally on abdominal non-contrast CT scans and compares its performance with results from two reader studies involving 48 radiologists on non-contrast and contrast CT imaging, respectively. We then validate PANDA on a large external multicenter test cohort (n = 5,337) to assess its generalizability to various settings. To increase the range of applicable patient populations, we study the feasibility of applying PANDA on chest CT. Finally, to validate the critical issues related to realistic clinical translation, we explore the integration of PANDA into large-scale real-world multi-scenarios of routine clinical processes, involving 20,530 consecutive patients from four settings (that is, physical exam, emergency, outpatient, and inpatient) with available abdominal or chest non-contrast CT scans.

Results

The PANDA network

We present a deep learning model, PANDA, to detect and diagnose PDAC and seven subtypes of non-PDAC lesions (Methods), that is, pancreatic neuroendocrine tumor (PNET), solid pseudopapillary tumor (SPT), intraductal papillary mucinous neoplasm (IPMN), mucinous cystic neoplasm (MCN), serous cystic neoplasm (SCN), chronic pancreatitis, and 'other' (cf. Supplementary Table 1), from abdominal and chest non-contrast CT scans. Our model can detect the presence or absence of a pancreatic lesion, segment the lesion, and classify the lesion subtypes (Fig. 1a).

PANDA was trained on a training set of abdominal non-contrast CT scans of 3,208 patients from a high-volume pancreatic cancer institution, Shanghai Institution of Pancreatic Diseases (SIPD), directly

affiliated with a tertiary hospital (a major comprehensive academic medical center in Shanghai, China). The patient characteristics are listed in Extended Data Table 1. The ground truth labels were confirmed either by surgical pathology for lesions or by a 2 year follow-up for normal controls. PANDA was also supervised by pixel-wise annotations, including both the pancreas and the lesion, transferred by image registration from annotations on paired contrast-enhanced CT scans in which tumors were more visible. Dataset and annotation details are given in the Methods section.

PANDA consists of a cascade of three network stages that increase in model complexity and the difficulty level of the tasks performed (Extended Data Fig. 1; Methods). The first stage (Stage 1) involves pancreas localization, using an nnU-Net model[23]. The second stage (Stage 2) carries out lesion detection, and we build convolutional neural networks (CNNs) together with a classification head to distinguish the subtle texture change of lesions in non-contrast CT. We tune the Stage 2 model to achieve a specificity of 99% for lesion detection on cross-validation of the training set to reduce false-positive predictions. The third stage (Stage 3) involves the differential diagnosis of pancreatic lesions if any abnormalities are detected in the second stage, integrated with an auxiliary memory transformer branch[24,25] to automatically encode the feature prototypes of the pancreatic lesions, such as local texture, position and pancreas shape, for more accurate fine-grained classification.

We mainly evaluate the performance of PANDA on three tasks (Methods). The first task is lesion detection: that is, lesion versus normal, which also includes detection rates stratified by lesion type and by cancer stage. The second task is primary diagnosis: PDAC versus non-PDAC versus normal, which also includes evaluation of one versus others, for example, PDAC identification (PDAC versus non-PDAC + normal). The third task is differential diagnosis: that is, classification of PDAC and seven non-PDAC lesion subtypes.

Internal evaluation

Our independent internal test cohort consisted of 291 patients (108 patients with PDAC, 67 patients with non-PDAC, and 116 normal controls) from the SIPD (Extended Data Table 1; Methods). These patient labels were confirmed on surgical pathology or a 2 year follow-up. For lesion detection, PANDA achieved an area under the receiver operating characteristic curve (AUC) of 0.996 (95% confidence interval (CI) 0.991–1.00, Fig. 2a), a sensitivity of 94.9% (95% CI 91.4–97.8%) and a specificity of 100% (95% CI 100–100%); for the PDAC subgroup the sensitivity for detection was 97.2% (95% CI 93.5–100%) overall, 97.1% (95% CI 91.4–100%; $n = 35$; Fig. 2c) for stage I, and 96.2% (95% CI 90.4–100%; $n = 52$; Fig. 2c) for stage II. For small PDACs (diameter <2 cm, T1 stage), the sensitivity for detection was 85.7% (95% CI 64.3–100%; $n = 14$; Fig. 2c). For PDAC identification, the AUC was 0.987 (95% CI 0.975–0.996, Fig. 2b), the sensitivity was 92.6% (95% CI 87.3–97.0%) and the specificity was 97.3% (95% CI 94.6–99.5%, Fig. 2b).

For the internal differential diagnosis cohort ($n = 786$; Extended Data Table 1; Methods), PANDA achieved an accuracy of 79.6% (95% CI 76.8–82.6%) and a balanced accuracy (averaged class-level accuracy) of 60.7% (95% CI 55.7–65.4%). The accuracy is non-inferior ($P = 0.0018$ at a pre-specified 5% margin) to the second-reader radiology reports (Fig. 2f, Supplementary Fig. 1 and Supplementary Table 4), which is a secondary analysis of a primary standard of care clinical radiology report that includes access to the contrast-enhanced CT, clinical information and patient history, and represents the standard of care of pancreatic lesion management practice in the internal center. The results for IPMN subtype classification (main or mixed-duct versus branch-duct IPMN) and the full pipeline (detection + diagnosis) in the internal cohorts are given in Supplementary Table 13 and Supplementary Fig. 7a, respectively.

Ablation studies were conducted to analyze the performance of PANDA's Stage 2 and Stage 3 modules on the internal training cohort ($n = 3,208$) (Extended Data Fig. 2; Methods). Stage 2 and Stage 3 had significantly better performance than their related baseline methods ($P = 0.00022$ and $P = 0.0002$, respectively). We also analyzed the effect of training data size on the performance of PANDA. More training data led to better performance for all tasks, and the margins of improvement increased as the tasks became more challenging (Extended Data Fig. 3). PANDA is an interpretable AI model that directly outputs the segmentation mask of the pancreas and the detected lesion (see Supplementary Table 5 for segmentation accuracy). Additional analyses of interpretability via the visualization of the Stage 2 activation maps and Stage 3 attention maps are provided in Extended Data Fig. 4 and the Methods section.

Reader studies

We conducted two reader studies (Methods and Extended Data Table 2). The aim of the first study was to compare PANDA with non-contrast CT readers consisting of pancreatic imaging specialists, general radiologists and radiology residents, and validate whether PANDA could assist them in making more accurate decisions. The second reader study was designed to compare PANDA, using only non-contrast CT, with a clinical expert upper-bound set-up, that is, a pancreatic imaging specialist reading a contrast-enhanced CT.

In the first reader study, 33 readers from 12 institutions interpreted 291 non-contrast CT scans in the internal test cohort. Alongside the CT images, readers were provided with each patient's age and sex, and rated each case as PDAC, non-PDAC or normal (Supplementary Fig. 2). For lesion detection, the performance values of all 33 readers fell below PANDA's receiver operating characteristic (ROC) curve (Fig. 3a). PANDA significantly outperformed the average reader performance by 14.7% (95% CI 10.8–18.8%, $P = 0.0002$) in sensitivity and 6.8% (95% CI 5.6–8.1%, $P = 0.0002$) in specificity for lesion detection (Supplementary Table 6a), and by a significant margin of 34.1% (95% CI 29.3–38.9%, $P = 0.0002$) in sensitivity and 6.3% (95% CI 4.1–8.4%, $P = 0.0002$) in specificity for PDAC identification (Supplementary Table 6b). Notably, for PDAC identification the sensitivity was as low as 16.7–35.2% for some radiology residents who were not specialized in pancreatic imaging.

After at least a 1 month washout period, readers were additionally provided with the AI lesion segmentation and primary diagnosis probabilities (Supplementary Fig. 3) and re-rated each patient. With AI assistance, for lesion detection the mean reader performance was significantly improved by 8.5% in sensitivity (95% CI 6.5–10.3%, $P = 0.0002$) and 5.3% in specificity (95% CI 4.3–6.3%, $P = 0.0002$; Supplementary Table 7a). For PDAC identification, the mean reader performance was significantly improved by 20.5% (95% CI 17.8–23.4%, $P = 0.0002$) in sensitivity and by 3.1% (95% CI 2.1–4.1%, $P = 0.0002$) in specificity (Supplementary Table 7b). Overall, the largest improvement was observed in readers not specialized in pancreatic imaging. The residents' performance with AI could even approach that of pancreatic radiology specialists (evaluated using balanced accuracy in Fig. 3d,e and Supplementary Tables 7 and 9). Detailed confusion matrices are given in Supplementary Figs. 2 and 4.

In the second reader study, another 15 pancreatic imaging specialists from the internal center (SIPD) interpreted multi-phase contrast-enhanced CT scans of the same 291 patients. Each reader was provided with the non-contrast, arterial, and venous phase of CT images along with the age and sex information and carried out the same rating (Supplementary Fig. 5). PANDA (on non-contrast CT imaging) did better than the mean performance of the specialists (using contrast-enhanced CT scans) by 2.9% (95% CI 0.1–5.8%, $P = 0.0002$ for non-inferiority) in sensitivity and by 2.1% (95% CI 1.4–3.0%, $P = 0.0002$ for difference) in specificity, for lesion detection (Supplementary Tables 10a and 11a); and by a margin of 13.0% (95% CI 8.5–17.8%, $P = 0.0002$ for difference) in sensitivity and 0.5% (95% CI −0.7 to 1.9%, $P = 0.0002$ for non-inferiority) in specificity, for PDAC identification (Supplementary Tables 10b and 11b).

https://doi.org/10.1038/s41591-023-02640-w

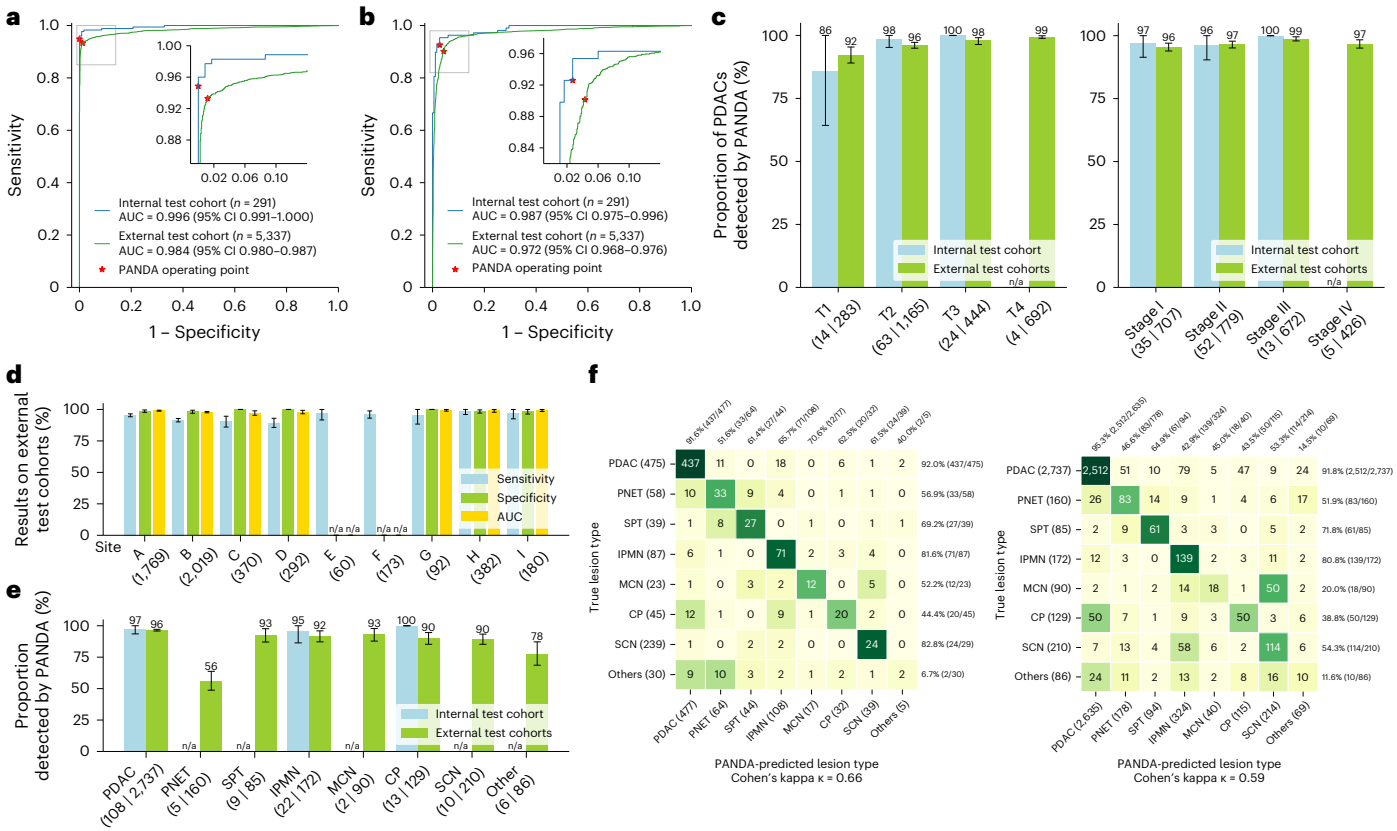


Fig. 2 | Internal and external validation. a,**b** Receiver operating characteristic curves of lesion detection (**a**) and PDAC identification (**b**) for the internal and external test cohorts. **c**, Proportion of PDACs detected by PANDA in terms of American Joint Committee on Cancer (AJCC) T stage (left) and TNM (tumor, nodes, metastasis) stage (right) in the internal test cohort ($n = 105$) and external test cohort ($n = 2,584$). **d**, Sensitivity, specificity and AUC of lesion detection in the external center cohorts (sites A–I, $n = 5,337$). **e**, Proportion of different lesion subtypes detected by PANDA in the internal test cohort ($n = 175$) and external test cohort ($n = 3,669$). **f**, Confusion matrices of differential diagnosis in the internal differential diagnosis cohort (left) and external test cohorts (right). **c**–**e**, Error bars indicate 95% CI. The center shows the computed mean of the metric specified by its respective axis labels. The results of subgroups with too few samples to be studied reliably (≤10) are omitted and marked as not applicable (n/a).

Generalization to external multicenter test cohorts

To assess the generalizability of PANDA to different patient populations and imaging protocols we validated our model on external multicenter ($n = 9$) test cohorts, which consisted of preoperative non-contrast abdominal CT scans of 5,337 patients (2,737 with PDAC, 932 with non-PDAC and 1,668 normal controls) from China, Taiwan ROC and the Czech Republic (Extended Data Table 1; Methods). The patient labels were confirmed by surgical or biopsy pathology diagnosis reports or a 2 year follow-up visit diagnosis. PANDA achieved an AUC of 0.984 (95% CI 0.980–0.987, Fig. 2a), sensitivity of 93.3% (95% CI 92.5–94.1%) and specificity of 98.8% (95% CI 98.3–99.4%) for lesion detection. For the PDAC patient subgroup, the detection rate was 96.5% (95% CI 95.8–97.2%) overall, 95.6% (95% CI 93.9–97.0%; Fig. 2c) for stage I, and 96.5% (95% CI 95.3–97.8%; Fig. 2c) for stage II. For small PDAC lesions (diameter <2 cm, T1 stage), the sensitivity for detection was 92.2% (95% CI 89.0–95.4%; $n = 283$; Fig. 2c). The lesion detection results for each center are shown in Fig. 2d and the performance stratified by lesion subtype is given in Fig. 2e. For PDAC identification, the sensitivity was 90.1% (95% CI 89.0–91.2%) and the specificity was 95.7% (95% CI 94.9–96.5%; Fig. 2b).

For differential diagnosis (Fig. 2f, $n = 3,669$) our model achieves an accuracy of 81.4% (95% CI 80.2–82.6%) and a balanced accuracy of 52.6% (95% CI 50.0–55.1%). The confusion matrices, accuracy and balanced accuracy of each external center with pathology-confirmed lesion types are shown in Supplementary Fig. 6 and Supplementary Table 12. The results for IPMN subtype classification and the full pipeline are given in Supplementary Table 13 and Supplementary Fig. 7b, respectively.

Feasibility study of lesion detection on chest computed tomography

PANDA's ability can be coupled with established clinical indications such as chest CT for lung cancer screening. We validated the feasibility of pancreatic lesion detection using PANDA on chest CT (Fig. 4). From SIPD we collected non-contrast chest CT scans of 492 patients, consisting of 63 with PDAC, 51 with non-PDAC, and 378 normal controls, as a test cohort independent of the training data. The patient labels were confirmed by surgical pathology or a 2 year follow-up visit diagnosis (Methods).

Without tuning on any chest CT scans, PANDA achieved an AUC of 0.979 (95% CI 0.962–0.993), a sensitivity of 86.0% (95% CI 79.4–91.9%) and a specificity of 98.9% (95% CI 97.8–100%) for lesion detection (Fig. 4c), and a sensitivity of 92.1% (95% CI 85.7–98.4%) for the PDAC subgroup. Depending on detailed chest CT protocols, certain pancreatic lesions could not be completely scanned. We analyzed the lesion scanning completeness in chest CT by referring to the lesion location in contrast-enhanced abdominal CT scans (Fig. 4a), and found that 67% of patients with PDAC and 43% of patients with non-PDAC were not fully scanned (Fig. 4b). For those patients whose pancreatic lesions were not captured in the CT scan's field of view (and thus were not directly observable), 75% of PDAC cases in these patients were detected by PANDA, that is, the patients were classified as having a lesion (Fig. 4d)

https://doi.org/10.1038/s41591-023-02640-w

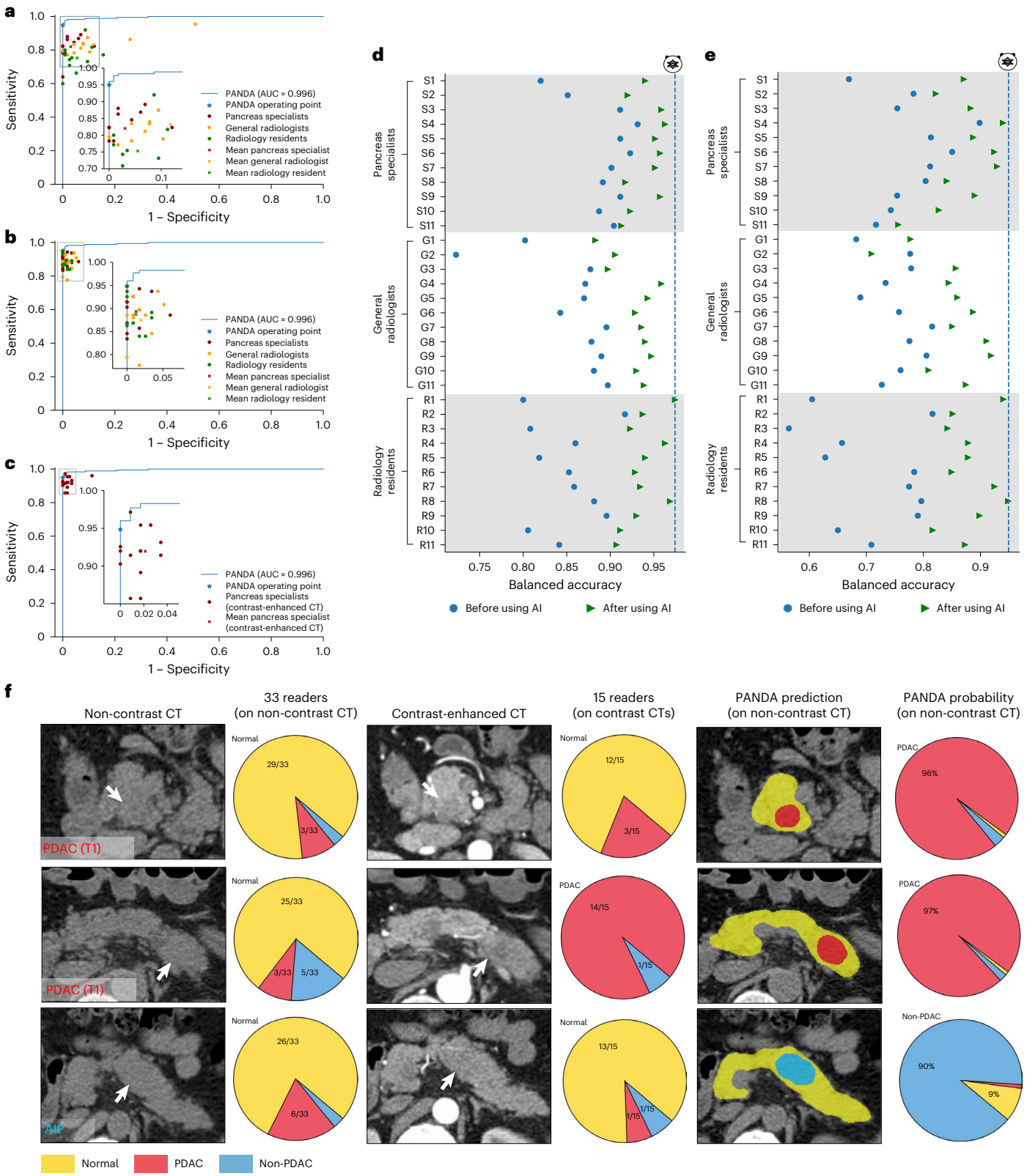


Fig. 3 | Reader studies. a, Comparison between PANDA and 33 readers with different levels of expertise on non-contrast CT for lesion detection. **b**, Lesion detection performance of the same set of readers with the assistance of PANDA on non-contrast CT. **c**, Comparison between PANDA using non-contrast CT and 15 pancreas specialists using contrast-enhanced CT for lesion detection. **d**,**e**, Balanced accuracy improvement in radiologists with different levels of expertise for lesion detection (**d**) and PDAC identification (**e**). **f**, Examples of early-stage PDACs and a case of autoimmune pancreatitis (AIP) missed by readers on non-contrast CT and on contrast CT but detected by PANDA.

https://doi.org/10.1038/s41591-023-02640-w

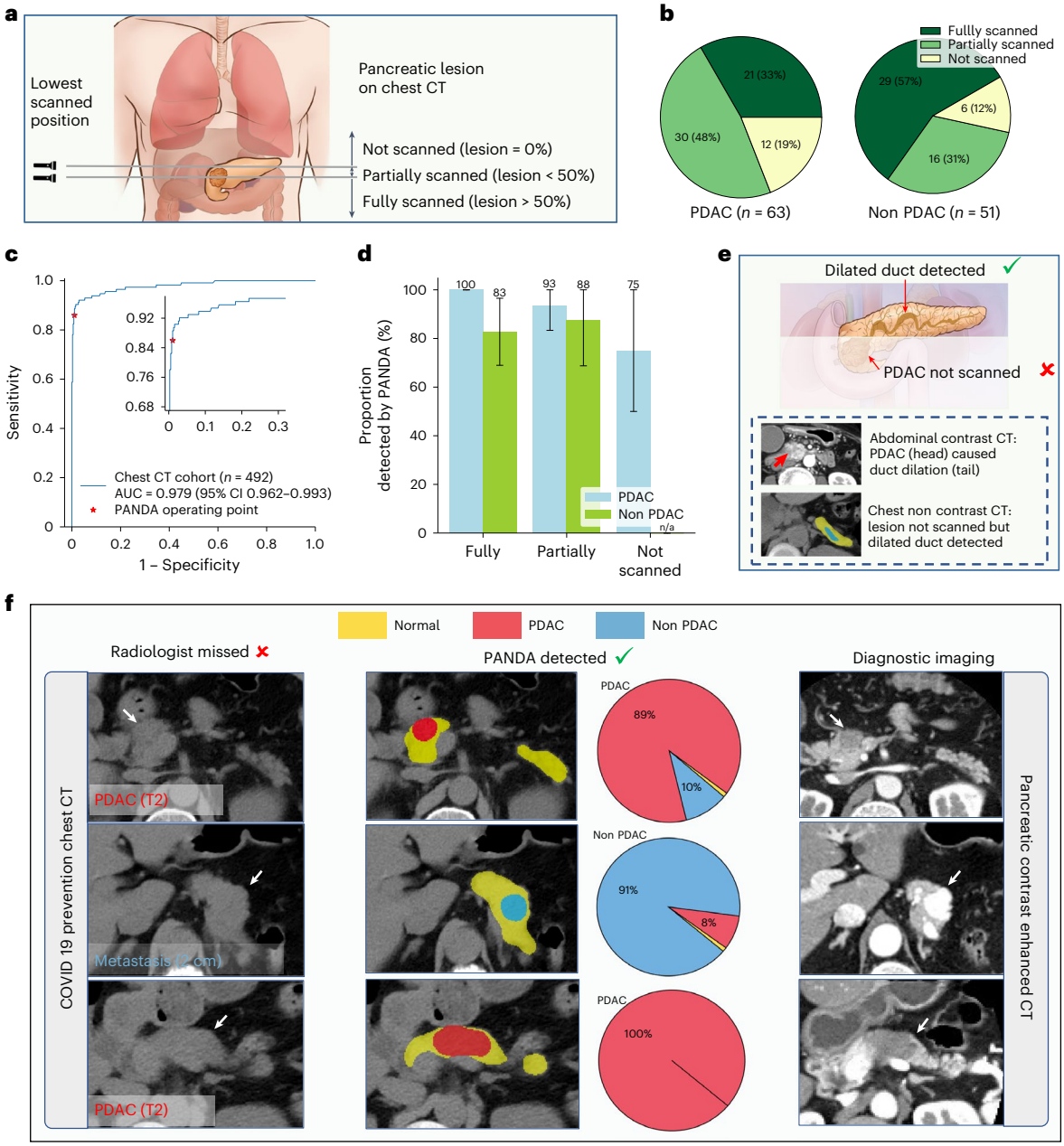


Fig. 4 | Validation on chest non-contrast CT. a, Schematic diagram of the proportion of the pancreatic lesion scanned in chest non-contrast CT. We categorize all cases into three categories, that is, lesion not scanned, lesion partially scanned, and lesion fully scanned, based on the relative position of the lowest scanned slice and the lesion. **b**, The proportion of the three categories in PDAC and non-PDAC cases. **c**, ROC curve for lesion detection on non-contrast chest CT. **d**, Proportion of lesions detected by PANDA in the PDAC ($n = 63$) and non-PDAC cases ($n = 51$). Error bars indicate 95% CI. The center shows the computed mean of the metric specified by the respective axis labels. The results of subgroups with too few samples to be studied reliably (≤10) are omitted and marked as 'n/a'. **e**, Illustration of how PANDA can detect lesions that are not scanned in chest CT. Two scans of the same patient showing that PANDA can detect dilated pancreatic duct (usually caused by PDAC) even when the PDAC is not scanned. **f**, PANDA can detect early-stage PDACs and metastatic cancer that was initially misdetected by the radiologists on chest non-contrast CT (COVID-19 prevention CT).

via secondary signs of the disease such as dilation of the pancreatic duct (Fig. 4e).

Real-world clinical evaluation

The above experiments validate the clinical utility of PANDA, but they are limited to pathology-confirmed pancreatic lesions (thus with higher risk) and a moderate number of normal cases. It is unclear whether PANDA could be generalized well to the real-world population, including patients with lesions of lower risk (for example, chronic pancreatitis and branch-duct IPMN) and the large, diverse set of subjects with normal pancreas. To close the clinical translation gap, evaluation in real-world application settings is required to answer the following critical questions: first, what is the true performance of PANDA when used for consecutive real-world patient populations, possibly containing unseen lesion subtypes, collected from varying CT imaging protocols and clinical scenarios (that is, physical examination, emergency, outpatient, and inpatient); second, can the tool detect malignancies that were not previously detected by the standard of care clinical diagnosis; third, can patients benefit from such detection (for example, if the malignancy was detected at its surgically resectable stage[26]); fourth,

can the tool be clinically safe and efficient, without a large number of false-positive findings that require unnecessary follow-up tests and extra time for being ruled out; and last, can the benchtop-derived AI be further improved according to bedside clinical requirements[27].

We deployed PANDA at the SIPD by seamlessly integrating it into the existing clinical infrastructure and workflow (Supplementary Fig. 9 and Methods 'Real-world deployment'), and performed two rounds of large-scale, real-world, retrospective studies enrolling consecutive patients (clinical trial ChiCTR2200064645, chictr.org.cn; includes both studies) (Fig. 5a, Extended Data Fig. 5 and Extended Data Fig. 6; Methods). Due to the retrospective nature of the study we used two timeframes for the standard of truth for each patient (Fig. 5a): the initial standard of care, that is, the clinical diagnosis at the initial visit when the non-contrast CT was acquired; and the follow-up standard of care, that is, the clinical diagnosis obtained at follow-up (after the initial visit and before the PANDA evaluation study).

First real-world evaluation cohort. For the first real-world evaluation cohort (RW1, $n = 16{,}420$), the first four questions were assessed: performance; change in standard of care diagnosis; patient benefit; and safety and efficiency.

Performance. RW1 included 44 PDACs and 135 non-PDACs. For lesion detection, PANDA achieved an overall sensitivity of 84.6% (95% CI 79.4%–89.9%) and specificity of 99.0% (95% CI 98.9%–99.2%) (Fig. 5b), and for PDAC identification PANDA achieved an overall sensitivity of 95.5% (95% CI 89.3%–100%), specificity of 99.8% (95% CI 99.7%–99.9%) and positive predictive value (PPV) of 56.0% (95% CI 44.8%–67.2%) (Fig. 5c). Of the four scenarios (that is, physical examination, emergency, outpatient, and inpatient), inpatient had the highest sensitivity of 88.6% (95% CI 78.0%–99.1%) and physical examination had the highest specificity of 99.8% (95% CI 99.7%–99.9%), for lesion detection. The multi-disciplinary team found that 51% (80 of 156) of the false positives by AI were actually (peri-)pancreatic diseases (Fig. 5h and Supplementary Fig. 8) requiring attention from radiologists[28]. Considering that these findings might be signs of pathology, they were excluded from the results, which were adjusted as below: for lesion detection the overall adjusted specificity increased to 99.5% (Fig. 5b), and for PDAC identification the adjusted specificity increased to 99.9% and the adjusted PPV, to 68.9% (Fig. 5c). More detailed results are shown in Supplementary Figs. 10–14.

Change in standard of care diagnosis. PANDA detected 26 pancreatic lesions that were not detected by the initial standard of care (Fig. 5i and Extended Data Table 3), consisting of 1 PDAC, 1 PNET, 3 IPMNs, 1 metastatic cancer, 6 cases of pancreatitis, 1 peri-pancreatic tumor and 13 SCN/cysts (10–33 mm). The opportunistic screening with PANDA could advance the early detection of (peri-)pancreatic malignancies and high-risk lesions.

Patient benefit. Of the aforementioned 26 lesions first detected by PANDA, eight were detected by follow-up standard of care before this retrospective study, including one T2 stage PDAC and one aneurysm (Fig. 5i and Extended Data Table 3). Nevertheless, earlier detection of some of these lesions by PANDA might benefit the patients' management and treatment. The remaining patients were invited to undergo magnetic resonance imaging (MRI) but only one (a 57-year-old) complied (due to the COVID-19 pandemic), undergoing contrast-enhanced MRI followed by minimally invasive surgery with curative intent (Extended Data Fig. 7). The surgical pathology confirmed the lesion as a G1 PNET with a size of 1.5 cm.

Safety and efficiency. Only 0.5% of patients ($n = 76$) had false-positive AI findings (Supplementary Table 14), of which 92% (70 of 76) were easy to rule out by the radiologists. Of the false negatives ($n = 28$), 89% were benign cysts ($n = 25$), most of which ($n = 19$) were < 10 mm in diameter; the remaining three consisted of a PNET, a case of chronic pancreatitis, and a lesion of undetermined type.

Second real-world evaluation cohort. To further optimize PANDA for real-world usage, that is, reduce false positives and enable the detection of previously unseen disease types (for example, acute pancreatitis in the emergency scenario), we utilized hard example mining and incremental learning to upgrade PANDA (the resulting model is named PANDA Plus; see Methods). PANDA Plus was evaluated on the second real-world evaluation cohort (RW2, $n = 4{,}110$) to assess the fifth question regarding improvement of the model.

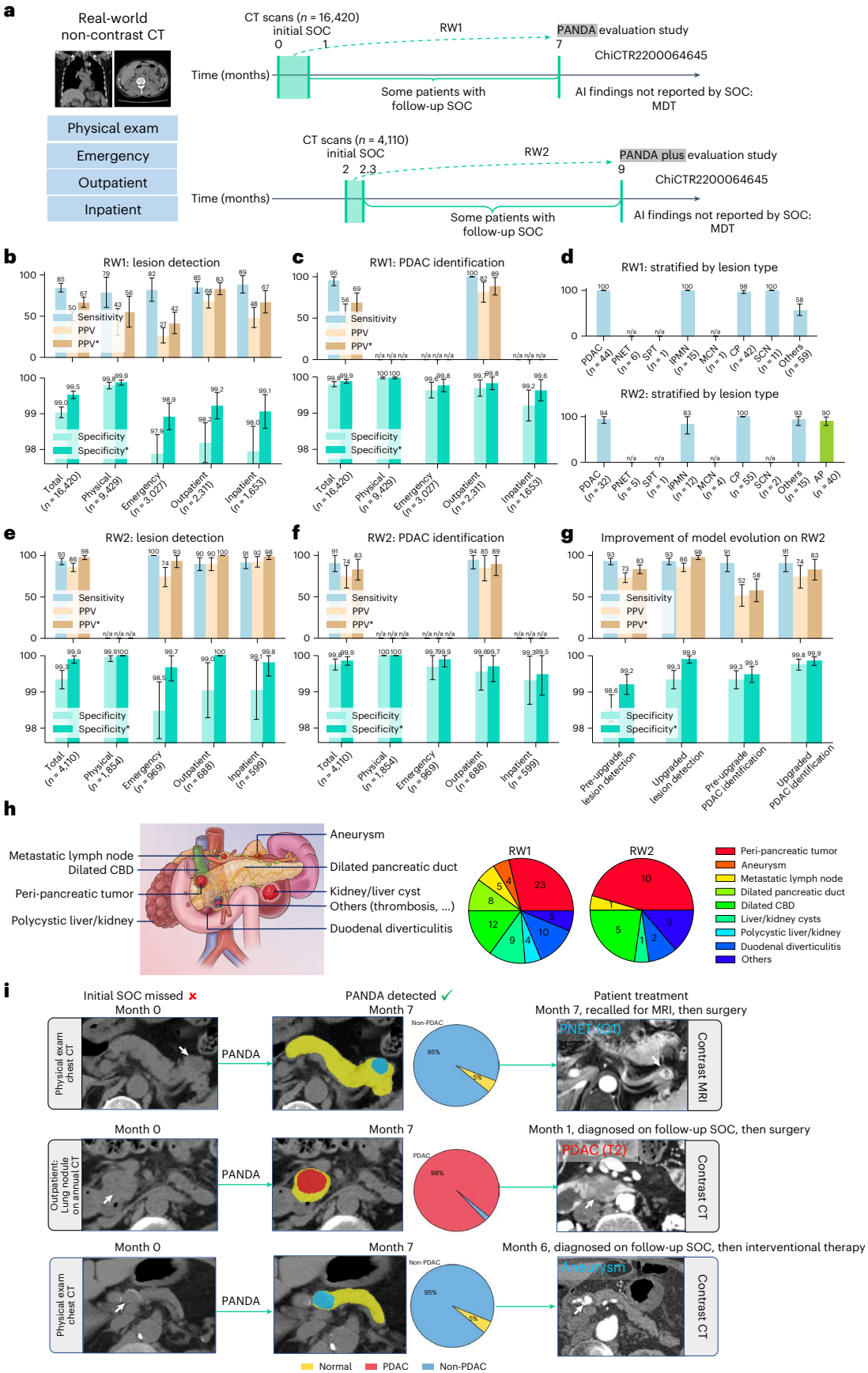
Model evolution. RW2 included 32 PDACs and 134 non-PDACs. PANDA Plus retained the same sensitivity as PANDA but significantly reduced the false positives by more than 80%, reaching an adjusted specificity of 99.9% for both lesion detection (95% CI 99.8%–100%) and PDAC identification (95% CI 99.7%–100%) (Fig. 5e–g and Supplementary Figs. 15–19). In addition, for the newly learned disease type (that is, acute pancreatitis), the sensitivity for detection was 90.0% (95% CI 80.7%–99.3%) for the 40 patients with acute pancreatitis (Fig. 5d). PANDA Plus detected five pancreatic lesions that were missed by the initial standard of care, consisting of 1 PDAC (T2 stage), 1 PNET and 3 cysts (10–32 mm) (Extended Data Table 3). In addition, the real-world evaluation showed that PANDA maintained robust performance in low-risk lesions despite being originally trained on surgical pathology-confirmed lesions. Specifically, our model had a sensitivity of 92.6% for detecting IPMN and 99.0% for chronic pancreatitis in RW1 and RW2 combined (Fig. 5d), although 22 (81%) of the 27 IPMNs and 94 (97%) of the 97 cases of chronic pancreatitis were not biopsied or resected.

Discussion

We present PANDA, an AI model that detects the seven most common pancreatic lesions and 'other', and diagnoses the lesion subtypes in routine non-contrast CT scans. This task has long been considered impossible for radiologists and, as such, contrast-enhanced CT and/or MRI and endoscopic ultrasound (EUS) have been used as the recognized and recommended diagnostic imaging modalities. We show that by curating a large dataset covering common pancreatic lesion types confirmed by pathology, transferring lesion annotations from contrast-enhanced to non-contrast CT, designing a deep learning approach that incorporates a cascade network architecture for lesion detection and a memory transformer for pancreas lesion diagnostic information modeling, and learning from the real-world feedback, PANDA, which uses only non-contrast CT as input, achieves high sensitivity and exceptionally high specificity in the detection of pancreatic lesions, with a significantly higher accuracy than radiologists in the primary diagnosis between PDAC and non-PDAC, and non-inferior

Fig. 5 | Real-world clinical evaluation. a, The data collection process of two real-world datasets, that is, RW1 and RW2, for the original PANDA model and the upgraded PANDA Plus model, respectively. SOC, standard of care. **b**,**c**,**e**,**f**, The sensitivity, specificity and PPV on RW1 ($n = 16{,}420$) and RW2 ($n = 4{,}110$). The superscript * represents adjusted results if we exclude cases of (peri-)pancreatic findings. **d**, Proportion of different lesion types detected in RW1 ($n = 179$) and RW2 ($n = 166$). **g**, The comparison between PANDA and PANDA Plus on RW2 ($n = 4{,}110$). Error bars indicate 95% CI. The center shows the computed mean of the metric specified by the respective axis labels. The results of subgroups with too few samples to be studied reliably (≤10) are omitted and marked as 'n/a'. **h**, Examples of (peri-)pancreatic findings (left) and the number detected by PANDA (right). CBD, common bile duct. **i**, Examples of cases in which the lesion was missed by the initial SOC but was detected by PANDA.

https://doi.org/10.1038/s41591-023-02640-w



accuracy to radiology reports in the differential diagnosis of the eight aforementioned pancreatic lesion subtypes.

PANDA exhibits effective generalizability to external centers, varying imaging protocols (Extended Data Table 1) and real-world populations. The favorable generalizability of PANDA can be attributed to the following factors. First, the training data are from a high-volume tertiary hospital, encompassing a diverse representation of the Chinese population. Second, non-contrast CT is likely to be a more generalizable modality for AI models than contrast-enhanced CT. Third, our approach combines segmentation (capturing the local pathological basis) and classification, reducing the overfitting risk of pure classification-based AI models. Fourth, the model has been tuned to yield a 99% specificity during cross-validation on the large training set (n = 3,208), to achieve reliable control of false positives. Fifth, the AI model's continual learning[27] enhances specificity to 99.9% by fine-tuning with false positives from external centers and the real world. And last, regarding training data, the cases and controls have similar CT imaging protocols (for example, slice thickness, CT dose index, oral water), thereby forcing the model to focus on the primary learning objectives rather than fitting to shortcuts or confounders.

PANDA exceeds the performance upper bound of human expert radiologists when reading only in non-contrast CT. This can be attributed to two main reasons. First, during its learning, PANDA is equipped with two informative supervisions that do not exist in non-contrast CT, however, radiologists have not been systematically trained for lesion detection and diagnosis in non-contrast CT. Specifically, one supervision consists of our curated expert lesion annotations transferred from contrast-enhanced CT; the other is the pathology-confirmed lesion types. Second, deep learning algorithms are more sensitive to subtle imaging grayscale intensity changes than human eyes, which are better at using color rather than intensity changes to interpret images[29]. Unlike generative deep learning methods to synthesize contrast or color[18–21], we train supervised learning models, which effectively capture subtle image details and directly learn downstream lesion detection and diagnosis tasks based on these detailed characteristics. Therefore, PANDA outperforms or matches radiologists on contrast-enhanced CT, the performance of which is in concordance with recent studies[30–32].

PANDA is an interpretable deep model that outputs the lesion boundaries and lesion subtype probabilities. Although radiologists usually do not diagnose pancreatic lesions from non-contrast CT alone, when assisted by PANDA their performance could be drastically increased regardless of experience, especially for the task of PDAC identification. Radiology residents with less experience benefit the most from PANDA's assistance, and can reach a level comparable with pancreas specialists. Although general radiologists might still doubt the AI results, their performance could be improved to a level close to that of pancreas specialists. Note that non-contrast CT is widely performed in non-tertiary hospitals and physical examination centers, where radiologists are usually less experienced or not specialized in pancreas imaging diagnosis. In tertiary hospitals, non-contrast CT is commonly performed as well, such as chest CT for lung nodule detection and abdominal CT in the emergency room. Taken together, PANDA could be widely used to increase the level of pancreas cancer diagnosis expertise in medical centers, especially by detecting more pancreatic malignancies at an earlier stage.

To assess the added value of PANDA for real-world clinical misdetection, we used the stricter standard of care clinical diagnosis as the standard of truth, which accounted for the entire patient management scenario, beyond the radiology report alone. Even so, of the 20,530 consecutive patients evaluated retrospectively, PANDA detected five cancers and 26 other pancreatic lesions that were missed by the initial standard of care, and enabled curative treatment of one patient with PNET.

Despite its high mortality rate, PDAC is relatively uncommon. Screening for PDAC in the asymptomatic population was not recommended because existing diagnostic methods would lead to a large number of false positives, resulting in considerable ramifications and costs. Although AI advancement in the areas of pancreatic lesion detection and diagnosis has occurred with the use of contrast-enhanced CT and EUS[30,33,34], the level of specificity is insufficient, and applying these imaging techniques to the general population is impractical due to their invasiveness, cost, and the need for iodine contrast. Liquid biopsy for cancer detection[26,35,36] has shown specificities of more than 99% but the sensitivity for early-stage pancreatic cancer detection is only satisfactory (approx. 50–60%, refs. 35,36). PANDA Plus (hereinafter referred to as PANDA) was highly sensitive (>96%) for early-stage PDAC and yielded an exceptional specificity of 99.9% in the large-scale real-world evaluation, which equates to approximately one false positive out of 1,000 tests. On the one hand, such a performance enables opportunistic screening in asymptomatic populations. Considering the prevalence of PDAC (13 cases per 100,000 adults), the PPV for PDAC identification will be approximately 10% (11 true positives and 100 false positives in 100,000 tests). This is even higher than the PPVs of some other cancer screening tests currently recommended by the Preventive Services Task Force (USPSTF), for example, mammography for breast cancer, with a PPV of 4.4% (ref. 37), stool DNA for colorectal cancer, with a PPV of 3.7% (ref. 38), and low-dose CT for lung cancer, with a PPV of 3.8% (ref. 39). Our experiments also show that when PANDA was applied in routine multi-scenario CT examinations, PDAC detection in asymptomatic adults could potentially be considered at no additional cost, with no extra examination or radiation exposure. Ideally, even if 10 AI-identified patients with PDAC underwent follow-up exams to confirm one PDAC at a 10% PPV, the overall cost per PDAC found remains manageable. For example, the price ranges from US$1,264 to US$1,685 in Shanghai, China, for 10 exams, depending on the specific type of follow-up exam, that is, contrast-enhanced CT, MRI or EUS, although the price could be higher in Western countries. Nevertheless, further prospective studies are needed to assess the risk–benefit ratio and cost-effectiveness in the future. On the other hand, PANDA could also be used in designed screening in high-risk populations[40] (Supplementary Methods 1.4). In such a scenario, the sensitivity of (particularly early-stage) PDAC identification can be further improved by adjusting the model threshold at the cost of a slight decrease in specificity. In both opportunistic and designed screening scenarios, PANDA is meant to be used in screening, a pre-step before diagnosis, and not to replace existing diagnostic imaging modalities. Nevertheless, PANDA's reliable initial diagnosis can better assist physicians in triaging and managing patients with pancreatic lesions, a frequent dilemma in clinical practice[41].

PANDA is trained on a continual learning approach using multicenter data, but includes only limited data outside the East Asian population and hospitals. The model should be further validated in external real-world centers, more international cohorts, and prospective studies. PANDA exhibited relatively low accuracy for PNET. PNET tumors are rare and highly diverse in appearance, and the model may primarily miss some cases with very low image contrast in non-contrast CT.

PANDA has already demonstrated its potential for accurate detection of other cancers, especially cancer types (esophagus[42], liver[43], stomach[44]) for which no guideline-recommended screening tests are available for average-risk individuals. This opens up an exciting possibility of universal cancer detection at both high sensitivity and high specificity levels, while requiring only a non-invasive, low-cost, widely adopted non-contrast CT scanning procedure. We hope that PANDA and its variations will help transform the current cancer-detection paradigm from late-stage diagnosis, when symptoms first present, to early-stage screening in which cancers can be detected before symptoms appear.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-023-02640-w.

References

1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Vasen, H. et al. Benefit of surveillance for pancreatic cancer in high-risk individuals: outcome of long-term prospective follow-up studies from three European expert centers. *J. Clin. Oncol.* **34**, 2010–2019 (2016).
3. Singhi, A. D., Koay, E. J., Chari, S. T. & Maitra, A. Early detection of pancreatic cancer: opportunities and challenges. *Gastroenterology* **156**, 2024–2040 (2019).
4. Pereira, S. P. et al. Early detection of pancreatic cancer. *Lancet Gastroenterol. Hepatol.* **5**, 698–710 (2020).
5. Klatte, D. C. F. et al. Pancreatic cancer surveillance in carriers of a germline *CDKN2A* pathogenic variant: yield and outcomes of a 20-year prospective follow-up. *J. Clin. Oncol.* **40**, 3267–3277 (2022).
6. Dbouk, M. et al. The multicenter Cancer of Pancreas Screening study: impact on stage and survival. *J. Clin. Oncol.* **40**, 3257–3266 (2022).
7. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33 (2021).
8. Gonda, T. A. et al. Recommendations for a more organized and effective approach to the early detection of pancreatic cancer from the PRECEDE (Pancreatic Cancer Early Detection) Consortium. *Gastroenterology* **161**, 1751–1757 (2021).
9. Grossberg, A. J. et al. Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma. *CA Cancer J. Clin.* **70**, 375–403 (2020).
10. Lucas, A. L. & Kastrinos, F. Screening for pancreatic cancer. *JAMA* **322**, 407–408 (2019).
11. Sodickson, A. et al. Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. *Radiology* **251**, 175–184 (2009).
12. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
13. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).
14. Ardila, D. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **25**, 954–961 (2019).
15. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
16. McKinney, S. M. et al. International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94 (2020).
17. Lotter, W. et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat. Med.* **27**, 244–249 (2021).
18. Preetha, C. J. et al. Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. *Lancet Digit. Health* **3**, 784–794 (2021).
19. Zhang, Q. et al. Toward replacing late gadolinium enhancement with artificial intelligence virtual native enhancement for gadolinium-free cardiovascular magnetic resonance tissue characterization in hypertrophic cardiomyopathy. *Circulation* **144**, 589–599 (2021).
20. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917–920 (2018).
21. Rivenson, Y. et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* **3**, 466–477 (2019).
22. Pickhardt, P. J. Value-added opportunistic CT screening: state of the art. *Radiology* **303**, 241–254 (2022).
23. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
24. Dosovitskiy, A. et al. An image is worth 16x16 words: transformers for image recognition at scale. In *9th International Conference on Learning Representations* (2021).
25. Wang, H., Zhu, Y., Adam, H., Yuille, A. & Chen, L.-C. Max-deeplab: end-to-end panoptic segmentation with mask transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5463–5474 (2021).
26. Cohen, J. D. et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
27. Rajpurkar, P. & Lungren, M. P. The current and future state of AI interpretation of medical images. *N. Engl. J. Med.* **388**, 1981–1990 (2023).
28. To'o, K. J. et al. Pancreatic and peripancreatic diseases mimicking primary pancreatic neoplasia. *Radiographics* **25**, 949–965 (2005).
29. Balaur, E. et al. Colorimetric histology using plasmonically active microscope slides. *Nature* **598**, 65–71 (2021).
30. Park, H. J. et al. Deep learning-based detection of solid and cystic pancreatic neoplasms at contrast-enhanced CT. *Radiology* **306**, 140–149 (2023).
31. Liu, K.-L. 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* **2**, 303–313 (2020).
32. LeBlanc, M., Kang, J. & Costa, A. F. Can we rely on contrast-enhanced CT to identify pancreatic ductal adenocarcinoma? A population-based study in sensitivity and factors associated with false negatives. *Eur. Radiol.* https://doi.org/10.1007/s00330-023-09758-y (2023).
33. Marya, N. B. et al. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. *Gut* **70**, 1335–1344 (2021).
34. Xia, Y. et al. The FELIX project: deep networks to detect pancreatic neoplasms. Preprint at *bioRxiv* https://doi.org/10.1101/2022.09.24.22280071 (2022).
35. Klein, E. et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann. Oncol.* **32**, 1167–1177 (2021).
36. Fahrmann, J. F. et al. Lead-time trajectory of CA19-9 as an anchor marker for pancreatic cancer early detection. *Gastroenterology* **160**, 1373–1383 (2021).
37. Lehman, C. D. et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. *Radiology* **283**, 49–58 (2017).
38. Imperiale, T. F. et al. Multitarget stool DNA testing for colorectal-cancer screening. *N. Engl. J. Med.* **370**, 1287–1297 (2014).
39. Pinsky, P. F. et al. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. *Ann. Intern. Med.* **162**, 485–491 (2015).
40. Goggins, M. et al. Management of patients with increased risk for familial pancreatic cancer: updated recommendations from the International Cancer of the Pancreas Screening (CAPS) Consortium. *Gut* **69**, 7–17 (2020).

41. Springer, S. et al. A multimodality test to guide the management of patients with a pancreatic cyst. *Sci. Transl. Med.* **11**, eaav4772 (2019).
42. Yao, J. et al. Effective opportunistic esophageal cancer screening using noncontrast CT imaging. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2022* (eds Wang, L. et al.), Lecture Notes in Computer Science, Vol. 13433, pp. 344–354 (Springer, 2022).
43. Yan, K. et al. Liver tumor screening and diagnosis in CT with pixel-lesion-patient network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* in press (2023).
44. Yuan, M. et al. Cluster-induced mask transformers for effective opportunistic gastric cancer screening on non-contrast CT scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* in press (2023).

[1]Department of Radiology, Shanghai Institution of Pancreatic Disease, Shanghai, China. [2]DAMO Academy, Alibaba Group, New York, NY, USA. [3]Hupan Laboratory, Hangzhou, China. [4]Damo Academy, Alibaba Group, Hangzhou, China. [5]Department of Hepatobiliary and Pancreatic Surgery, First Affiliated Hospital of Zhejiang University, Hangzhou, China. [6]Department of Radiology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic. [7]Department of Radiology, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [8]Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China. [9]Department of Surgery, Shanghai Institution of Pancreatic Disease, Shanghai, China. [10]Department of Pathology, Shanghai Institution of Pancreatic Disease, Shanghai, China. [11]Department of Biostatistics, Harvard University T.H. Chan School of Public Health, Cambridge, MA, USA. [12]Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, China. [13]Department of Invasive Cardiology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic. [14]Department of Oncology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic. [15]Department of Radiology, Guangdong Provincial People's Hospital, Guangzhou, China. [16]Department of Radiology, Sun Yat-Sen University Cancer Center, Guangzhou, China. [17]Department of Radiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China. [18]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [19]These authors contributed equally: Kai Cao, Yingda Xia, Jiawen Yao, Xu Han, Lukas Lambert, Tingting Zhang, Wei Tang. ✉e-mail: cwshao@sina.com; 18940259980@163.com; qi.zhang@zju.edu.cn; liangtingbo@zju.edu.cn; ling.z@alibaba-inc.com; cjr.lujianping@vip.163.com

Methods

Ethics approval

The retrospective collection of the patient datasets in each cohort was approved by the institutional review board (IRB) at each institution with a waiver for informed consent: the Shanghai Institution of Pancreatic Diseases (SIPD) IRB, Shengjing Hospital of China Medical University (SHCMU) IRB, First Affiliated Hospital of Zhejiang University (FAHZU) IRB, Xinhua Hospital (XH) of Shanghai Jiao Tong University School of Medicine IRB, Fudan University Shanghai Cancer Center (FUSCC) IRB, Tianjin Medical University Cancer Institute and Hospital (TMUCIH) IRB, Sun Yat-Sen University Cancer Center (SYUCC) IRB, Guangdong Provincial People's Hospital (GPPH) IRB, Linkou Chang Gung Memorial Hospital (CGMH) IRB, and General University Hospital in Prague (GUHP) IRB. All data in this study were de-identified prior to model training, testing and reader studies.

Dataset description

This multicenter retrospective study involved five patient cohorts: an internal training cohort, on which the AI models were built; an internal test cohort, on which the model performance and reader study were assessed (together with an additional internal differential diagnosis cohort to increase statistical power for the evaluation of the model's performance on differential diagnosis); an external multicenter ($n = 9$) test cohort, on which the generalization across multiple centers was assessed; a chest non-contrast CT test cohort, on which the generalization to chest CT scans was assessed; and a real-world clinical evaluation cohort, on which critical questions about the clinical translation were assessed.

PDAC and seven non-PDAC lesion subtypes (PNET, SPT, IPMN, MCN, SCN, chronic pancreatitis and 'other')[33,41,45] were targeted in this study. In the first four cohorts, PDAC and non-PDAC lesions were confirmed by surgical or biopsy histopathology. The patient-level label of the surgical pathology was determined based on the 2019 *World Health Organization Classification of Tumors - 5th edition, Digestive System Tumors*. For biopsy pathology, definitive evidence was required for diagnosis. Patients with mixed neoplasms were not included. The normal controls were confirmed as being free of pancreatic or peri-pancreatic disease at 2 year follow-up (details of the collection process are given in Supplementary Methods 1.1.1). Patients with acute pancreatitis and a history of abdominal treatment were excluded. In the real-world cohort, pathology or the standard of care clinical diagnosis was used as the ground truth. All of the patients in the five cohorts were staged according to the eighth edition of the AJCC (American Joint Committee on Cancer) cancer staging system. The characteristics of the study participants are listed in Extended Data Table 1 (patient and CT characteristics), Supplementary Table 2 (reference standard of lesion types) and Supplementary Table 3 (lesion size stratified by lesion type). More details of the datasets included in this study are given below and in Supplementary Methods 1.1.2–1.1.7.

Internal training cohort. The internal training cohort consisted of 3,208 patients (1,431 with PDAC, 140 with PNET, 98 with SPT, 254 with IPMN (163 with main/mixed-duct IPMN and 91 with branch-duct IPMN), 37 with MCN, 110 with chronic pancreatitis, 134 with SCN, 66 with 'other' (Supplementary Table 1) and 938 normal controls) who had been treated between January 2015 and October 2020 at the SIPD, China. Consecutive patients (except for those who had chest CT before surgery, refer to the 'Chest computed tomography test cohort' section) with pancreatic lesions confirmed on surgical pathology were included.

Lesion and pancreas annotation. Besides the patient-level label, we also annotated pixel-level segmentation masks of the lesion and pancreas. We required only manual annotation of the lesion masks. Due to the difficulty of, and issues with reliability regarding, direct lesion annotation by radiologists using only non-contrast images, we additionally collected paired contrast-enhanced CT scans for annotation purposes. Pancreatic lesion annotations on non-contrast CT images were obtained by image registration from an experienced radiologist's manual annotations on the contrast-enhanced CT phase images, where tumors were more visible. The pancreas annotations were obtained via an improved version of our annotation-efficient semi-supervised learning approach[46], which uses only publicly available pancreas annotations (Supplementary Methods 1.1.3).

Internal test and differential diagnosis cohorts. We used the testing set of our prior work[47] as the source of the internal test cohort of the current study, given that interpretations on this set by 11 readers had been collected. Furthermore, we excluded ampullary and common bile duct cancer cases because they were usually not categorized as pancreatic lesions in the literature[41,45]. In addition, one normal participant was re-categorized as having chronic pancreatitis (actually autoimmune pancreatitis, but treated as chronic pancreatitis in our study) after carefully checking the patient records; and one normal participant was excluded due to a severe pancreatic duct dilation. As a result, the internal test cohort contained CT scans of 291 patients randomly collected between December 2015 and June 2018 at the SIPD, China, consisting of 108 with PDAC, 9 with SPT, 5 with PNET, 22 with IPMN (11 with main or mixed-duct IPMN and 11 with branch-duct IPMN), 2 with MCN, 10 with SCN, 13 with chronic pancreatitis, 6 with 'other', and 116 normal controls.

To enhance the statistical power of the differential diagnosis evaluation, we also collected an internal addition cohort consisting of 611 consecutive patients who underwent surgery between November 2020 and October 2021 at SIPD (367 with PDAC, 53 with PNET, 30 with SPT, 65 with IPMN (40 with main or mixed-duct IPMN and 25 with branch-duct IPMN), 21 with MCN, 32 with chronic pancreatitis, 19 with SCN, and 24 with 'other'). These 611 patients, and the 175 patients with pancreatic lesions in the internal test cohort, constitute the internal differential diagnosis cohort ($n = 786$). All patients underwent multi-phase CT, including non-contrast, arterial, venous, and delay. We used only the non-contrast phase for PANDA testing and the first reader study. The multi-phase CT scans of the internal test cohort were used for the second reader study.

External multicenter test cohorts. The external test cohorts were collected from nine centers, of which seven were located in China, one in Taiwan ROC (CGMH, Site H), and one in the Czech Republic (GUHP, Site I). The seven centers from China are distributed widely in geographical area: one in the northeast (SHCMU, Site A), four in the east (FAHZU, Site B; XH, Site C; FUSCC, Site D; TMUCIH, Site E), and two in the south (SYUCC, Site F; GPPH, Site G). Inclusion criteria were as follows: non-contrast abdominal CT fully covering the pancreas region before treatment; ground truth lesion type confirmed on either surgical or biopsy pathology; and normal control confirmed on at least 2 years of follow-up. Normal controls in most centers were randomly selected from the same time period as that of lesion collection. Patients with low image quality due to artifacts caused by metal in stents or drastic motion during imaging were excluded. The multicenter test cohort, consisting of non-contrast CT scans of 5,337 patients (2,737 with PDAC, 932 with non-PDAC, and 1,668 normal), was used for independent validation when no model parameters were tuned or adjusted.

Chest computed tomography test cohort. To evaluate the model's generalizability to chest CT, we collected a non-contrast chest CT test cohort with pathology-confirmed PDAC and non-PDAC and normal controls confirmed on 2 year follow-up, from SIPD, which is affiliated with a major tertiary hospital. Specifically, for patients with PDAC or non-PDAC confirmed by surgical pathology, we searched for their nearest chest CT images for up to 1 year before surgery. For patients with chest CT reports of normal pancreas, we searched for their follow-up

records of normal pancreas for at least 2 years. By doing so, we collected a cohort of 63 patients with PDAC, 51 with non-PDAC, and 378 normal controls spanning from November 2015 to May 2022 at SIPD. These non-contrast chest CT scans of PDAC and non-PDAC were acquired 4 days (range, −20 to 191 days) before the contrast-enhanced abdominal CT diagnosis, and most of them were acquired during the COVID-19 pandemic for prevention purposes in this tertiary hospital. We ensured that all patients were independent of the patients in the training cohort.

Real-world evaluation cohorts. The real-world, retrospective studies consisted of two rounds (RW1 and RW2) of evaluations between July 2022 and October 2022 at the SIPD. The clinical trial was complete and registered with http://www.chictr.org.cn, ChiCTR2200064645, and included both RW1 and RW2. PANDA was evaluated on RW1, and PANDA Plus (that is, the upgrade of PANDA by learning from the internal, external and RW1 feedback) was evaluated on RW2. The inclusion criterion was the availability of a non-contrast CT scan covering the pancreas region, for example, lung, esophagus, liver or kidney CT. Patients with acute pancreatitis (in RW1), abdominal cancer treatment, severe ascites, abdominal trauma, and low imaging quality were excluded. The process of the standard of truth determination is described in Extended Data Figs. 5 and 6.

Our real-world data were collected from four scenarios, that is, physical examination, emergency, inpatient, and outpatient department (Supplementary Methods 1.1.7). Because the patient indications, the CT image background complexity, the pancreatic lesion prevalence, and the experience of the (first-line) radiologists varied widely between these four scenarios, we conducted separate evaluations to determine the feasibility of opportunistic screening using PANDA. These results can serve as a valuable reference when applied to different countries or institutions based on the sources of patients.

The original RW1 consisted of 18,654 consecutive individuals whose non-contrast CT scans were examined between 1 and 31 December 2021, from four different clinical scenarios at the SIPD. After exclusion (n = 2,234, 12%), 16,420 individuals remained (that is, 9,429, 3,027, 2,311 and 1,653 from the physical examination, emergency, outpatient and inpatient scenarios, respectively). RW1 included 44 PDACs, 6 PNETs, 1 SPT, 15 IPMNs, 1 MCN, 42 cases of chronic pancreatitis, 11 SCNs and 59 cases of 'other' (mostly benign cysts).

The original RW2 consisted of 4,815 consecutive individuals between 1 and 10 February 2022, from the four clinical scenarios at the SIPD. The exclusion criteria were the same as for RW1, except that we included acute pancreatitis for RW2. After exclusion (n = 705, 15%), 4,110 individuals remained (1,854, 969, 688 and 599 from the physical examination, emergency, outpatient, and inpatient scenarios, respectively). RW2 included 32 PDACs, 5 PNETs, 1 SPT, 12 IPMNs, 4 MCNs, 55 cases of chronic pancreatitis, 2 SCNs, 15 cases of 'other', and 40 cases of acute pancreatitis.

AI model: PANDA

PANDA consists of three stages (Extended Data Fig. 1) and was trained by supervised machine learning. Given the input of a non-contrast CT scan, we first localize the pancreas, then detect possible lesions (PDAC or non-PDAC), and finally classify the subtype of the detected lesion if any. The output of PANDA consists of two components, that is, the segmentation mask of the pancreas and the potential lesion, and the classification of the potential lesion associated with probabilities of each class.

Pancreas localization. The aim of the first stage (Stage 1) is to localize the pancreas. Because the pancreatic lesion is usually a small region in the CT scan, the localization of the pancreas can accelerate the lesion finding process and prune out unrelated information for the specialized training of the pancreatic region. In this stage we train an nnU-Net[23] to segment the whole pancreas (the union mask of healthy pancreas tissue and the potential lesions) from the input non-contrast CT scan. Specifically, the three-dimensional (3D) low-resolution nnU-Net, which trains UNet on downsampled images, is used as the architecture because of its efficiency in inference. The model training is supervised by the voxel-wise annotated masks of the pancreas and lesion. More details on the training and inference for PANDA Stage 1 are given in Supplementary Methods 1.2.1.

Lesion detection. The aim of the second stage (Stage 2) is to detect the lesion (PDAC or non-PDAC). We trained a joint segmentation and classification network to simultaneously segment the pancreas and potential lesion, as well as classify the patient-level abnormality label, that is, abnormal or normal. The benefit of the classification branch is to enforce global-level supervision and produce a patient-level probability score, which is absent in semantic segmentation models. Similar designs had been used in previous studies, such as for cancer detection[47,48] and outcome prediction[49]. The network architecture is shown in Extended Data Fig. 1b. This is a joint segmentation and classification network with a full-resolution nnU-Net[23] backbone (left part in Extended Data Fig. 1b). We extract five levels of deep network features, apply global max-pooling, and concatenate the features before carrying out the final classification. We output both the segmentation mask of the potential lesion and pancreas, and the probabilities of abnormal or normal for enhanced interpretability. This network was supervised by a combination of segmentation loss and classification loss:

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \alpha\mathcal{L}_{\text{cls}} \quad (1)$$

where the segmentation loss $\mathcal{L}_{\text{seg}}$ was an even mixture of Dice loss and voxel-wise cross-entropy loss, and the classification loss was the cross-entropy loss. α was set to 0.3 to balance the contribution of the two loss functions. More details on the training and inference of PANDA Stage 2 are given in Supplementary Methods 1.2.2.

Differential diagnosis. The aim of the third stage network (Stage 3) is the differential diagnosis of pancreatic lesion type, which is formulated as the classification of eight sub-classes, that is, PDAC, PNET, SPT, IPMN, MCN, chronic pancreatitis, SCN and 'other'. Due to the subtle texture change in pancreatic diseases, especially on non-contrast CT scans, we incorporate a separate memory path network that interacts with the UNet path to enhance the ability to model global contextual information, which is usually associated with the diagnosis of pancreatic lesions by radiologists. As shown in Extended Data Fig. 1c, we use a dual-path memory transformer network. This design is inspired by Max-Deeplab[25]. The architecture of the UNet branch is the same as that of Stage 2, implemented as a full-resolution nnU-Net. The UNet branch takes the input of the cropped 3D pancreas bounding box, which is cropped with a fixed input size of (160, 256, 40). The memory branch starts with learnable memories designed to store both positional and texture-related prototypes of the eight types of pancreatic lesion, and is initialized as 200 tokens with 320 channels. The memory path iteratively interacts with multi-level UNet features (plus a shared learnable positional embedding across layers) via cross-attention and self-attention layers. Through this process the memory vectors were automatically updated to encode both the texture-related information from the UNet features and the positional information from the learnable positional embedding, for example, relative positions of the pancreatic lesion inside the pancreas, resulting in distinguishable descriptors for each type of pancreatic lesion.

The mechanism of the cross-attention and self-attention used in the model is formally described in Supplementary Methods 1.2.3, together with more details on model instantiation, training and inference of PANDA Stage 3.

Additionally, we trained an IPMN subtype classifier in a cascaded fashion following PANDA Stage 3, with the aim of binary classification

between main or mixed-duct IPMN and branch-duct IPMN (Supplementary Methods 1.2.3).

Generalization of PANDA to chest computed tomography. One major difference between chest CT and abdominal CT is that the pancreatic and lesion regions are sometimes partially scanned in chest CT, depending on the different scanning ranges of the protocol and the anatomy of the patient. This difference could induce domain shift issues for machine learning models if our AI model was trained only on abdominal CT scans. To address this issue we propose a data augmentation method that randomly (with a probability) cuts off the pancreas region in the axial plane to simulate the imaging scenario in which the pancreas is not fully scanned in the chest CT. This data augmentation is applied to the training process of Stages 2 and 3. This simple simulation of the chest CT effectively helps our model generalize to chest non-contrast CT without the addition of any chest CT data to the training set, while maintaining high performance on abdominal non-contrast CT.

Real-world deployment and model evolution. In the real-world clinical evaluation, PANDA was deployed at SIPD by integrating it into the clinical infrastructure and workflow (Supplementary Fig. 9). The deployment facilitates large-scale retrospective real-world studies in the hospital environment by securing data privacy, efficiently utilizing computational resources, and accelerating the process of large data inference and clinical evaluation. Specifically, we deploy PANDA in a local server located in the hospital (Supplementary Methods 1.2.4), which enables radiologists to visualize each case using our user-friendly DAMO Intelligent Medical Imaging user interface (IMI UI; Supplementary Fig. 9), easily review all results and access necessary information from their daily work environment. After RW1 we again collected non-contrast CT data of false positives and negatives and cases of acute pancreatitis from the internal, external and RW1 cohorts. In the field of machine learning this is known as hard example mining and incremental learning. The evolved model was named PANDA Plus and tested on RW2. The collection and annotation of these new training data and the fine-tuning schedule are described in Supplementary Methods 1.2.5.

Evaluation metrics

Lesion detection metrics. Lesion detection is a binary classification task to distinguish whether the patient has a pancreatic lesion or not. Having a lesion is defined as the 'positive' class for calculation of the AUC, sensitivity, specificity, accuracy and balanced accuracy. In addition, we evaluate the lesion detection rates stratified by lesion type. Particularly for the PDAC cases, we assess the sensitivity for detection stratified by cancer stage (stages I–IV) and tumor stage (T1–4).

Primary diagnosis metrics. Primary diagnosis is a three-class classification task to distinguish PDAC versus non-PDAC versus normal. We use the top-1 accuracy and three-class balanced accuracy to present the detailed results of the three-class classification. In addition, we define a PDAC identification task because PDAC is a unique lesion type with the most dismal prognosis. Distinguishing it from other types, that is, PDAC versus non-PDAC + normal, is always the primary question to answer for doctors and is the key task for cancer screening. Having a PDAC is defined as the 'positive' class for calculation of the AUC, sensitivity, specificity, PPV, accuracy and balanced accuracy.

Differential diagnosis metrics. Differential diagnosis is an eight-class classification task for the seven most common pancreatic lesion types and 'other', following the pancreatic tumor–cyst classification task[41,45], without normal patients included and with each patient having a lesion type assigned. The confusion matrices are used to present the detailed classification results. We report the overall top-1 accuracy and multi-class balanced accuracy for the classification of all of the lesion types, to facilitate the comparison of the AI model's performance with second-reader radiology reports and across external multiple centers. The second-reader radiology report is a secondary analysis of a primary standard of care clinical radiology report, in which radiologists have complete access to the patient's clinical history (for example, contrast-enhanced CT examination indicated for chronic pancreatitis follow-up), and the results of other clinical examinations (for example, tumor biomarkers). In addition, we also report the performance of the full pipeline (lesion detection + differential diagnosis), that is, nine-class classification consisting of normal and eight lesion types.

Ablation studies

We perform three ablation studies. For PANDA Stage 2 we compare our multi-task CNN model with a volume-based classifier on the nnU-Net segmentation model (Extended Data Fig. 2a). This baseline model uses the volume of the segmented lesion by an nnU-Net as an indicator for the existence of the lesion. For PANDA Stage 3 we compare our dual-path transformer model with the Stage 2 multi-task CNN model. In addition, we demonstrate the importance of the quantity of training data on different tasks of our problem (Extended Data Fig. 3). We first retrain the PANDA model under four settings, using 10%, 25%, 50% and 75% of the training dataset, respectively, and then test the model in each setting on the internal and external test cohorts.

Reader studies

Two groups of readers participated in two independent reader studies.

Reader study on non-contrast computed tomography. The aim of the first reader study was to assess the readers' performance in detecting pancreatic lesions and diagnosing whether the lesion was a PDAC on non-contrast CT. The study was conducted in two sessions. The first session compared PANDA's performance with that of radiologists with varying levels of expertise in pancreatic imaging. The second session investigated whether PANDA would be capable of assisting radiologists. There was a washout period of at least 1 month between the two rounds for each reader.

A total of 33 readers from 12 institutions were recruited in this study, consisting of 11 pancreatic imaging specialists, 11 general radiologists who are not specialized in pancreatic imaging, and 11 radiology residents. These readers had practiced for an average of 8.3 years (range, 2–31 years) in various radiology departments, and had read an average of 510 pancreatic CT scans (range, 100–2,600) in the year before the reader study (Extended Data Table 2).

In the first session each reader was trained to use the ITK-SNAP software[50] for the visualization of the CT images. Basic functions of this software include but are not limited to HU (Hounsfield unit) windowing, zooming in and out, and axial, sagittal and coronal view simultaneous display. In interpreting the 291 randomly ordered cases from the internal test cohort, non-contrast CT images and information on age and sex were provided. The readers were informed that the study dataset was enriched with more positive patients than the standard prevalence of pancreatic lesions in daily practice. However, they were not informed about the proportions of each class. Each reader interpreted the image without time constraints and classified each case as PDAC, non-PDAC or normal. In addition to the patient-level label, each reader also recorded the location of the detected tumor in the format of pancreatic head/uncinate, neck, and body/tail. The performance of each reader is listed in Supplementary Tables 6 and 8.

In the second session the same group of readers interpreted the 291 cases again using ITK-SNAP. In addition to the non-contrast CT images and the information on age and sex, the readers were provided with PANDA's case-level prediction probability of PDAC, non-PDAC or normal, as well as the corresponding lesion segmentation masks. Some examples of the provided PANDA predictions (in interactive video format) are shown in Supplementary Fig. 3. The improvement of each reader between the two sessions is measured.

Reader study on contrast-enhanced computed tomography. The second reader study compared PANDA's (non-contrast CT) performance with that of pancreatic imaging specialists' readings on contrast-enhanced CT. A total of 15 additional pancreatic imaging specialists from a high-volume pancreatic cancer institution (SIPD) were recruited in this study. These readers had practiced for an average of 9.5 years (range, 6–19 years) in the radiology department at SIPD, and had read an average of 907 pancreatic CT scans (range, 400–3,000) in the year prior to the reader study (Extended Data Table 2).

Each reader was first trained to use the same software for visualizing multi-phase CT images. Next, they were provided with the non-contrast, arterial and venous phase CT images of the same 291 patients from the internal test cohort, as well as information on age and sex. The interpretation rules were the same as those of the first reader study. We also measured individual differences between non-contrast CT and contrast-enhanced CT (Supplementary Methods 1.3.1).

Interpretability of the AI model

Our AI model jointly outputs the probability of the abnormality, the prediction of the subtype classification (if any abnormality is detected), and the segmentation mask of the detected abnormality lesion. Unlike other AI-based classification models[51,52] that require the visualization of the network feature map to acquire the abnormality's positional cues, our model directly outputs the segmentation mask of the detected mass together with the patient-level probability, which provides straightforward and advanced interpretability. The correspondence between the segmented lesion and the ground truth lesion was evaluated using the Dice coefficient (DSC) and the 95th percentile of Hausdorf distance (HD95). The segmentation performance of the pancreas and each type of pancreatic lesion was evaluated.

In addition, we visualized the heatmap of the convolutional feature map of PANDA Stage 2 classification using Grad-CAM[53] (Extended Data Fig. 4a), to understand which part of the feature map contributed most to lesion detection. For PANDA Stage 3 lesion differential diagnosis, we plotted the attention map of the memory tokens, which showed the activation of the top activated tokens (Extended Data Fig. 4b) to interpret the model's attention.

Statistical analysis

The performance of the binary classification task was evaluated using the AUC, sensitivity, specificity, PPV, accuracy and balanced accuracy metrics. The performance of the multi-class classification task was evaluated using accuracy and balanced accuracy. Cohen's kappa coefficient κ was also computed between the AI prediction and the standard of truth for differential diagnosis. The confidence intervals were calculated based on 1,000 bootstrap replications of the data. The significance comparisons of sensitivity, specificity, accuracy and balanced accuracy were conducted using permutation tests to calculate two-sided P values with 10,000 permutations. For non-inferiority comparisons, a 5% absolute margin was pre-specified before the test set was inspected. The significance of the difference between the AUCs of the AI model and nnU-Net was assessed using the Delong test. The threshold to determine statistical significance is $P < 0.05$. Data analysis was conducted in Python using the numpy (v1.20.3), scipy (v1.8.1) and scikit-learn (v0.24.2) packages.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sample data and an interactive demonstration are given at http://panda.medofmind.com/. The remaining datasets used in this study are currently not permitted for public release by the respective institutional review boards. Requests for access to aggregate data and supporting clinical documents will be reviewed and approved by an independent review panel on the basis of scientific merit. All data provided are anonymized to protect the privacy of the patients who participated in the studies, in line with applicable laws and regulations. Data requests pertaining to the study may be made to the first author (Kai Cao; mdkaicao163@163.com). Requests will be processed within 6 weeks.

Code availability

The code used for the implementation of PANDA has dependencies on internal tooling and infrastructure, is under patent protection (application numbers: CN 202210575258.9, US 18046405), and thus is not able to be publicly released. All experiments and implementation details are described in sufficient detail in the Methods and Supplementary Information (Methods) sections to support replication with non-proprietary libraries. Several major components of our work are available in open-source repositories: PyTorch (https://pytorch.org/) and nnU-Net (https://github.com/MIC-DKFZ/nnUNet).

References

45. Chu, L. C. et al. Classification of pancreatic cystic neoplasms using radiomic feature analysis is equivalent to an experienced academic radiologist: a step toward computer-augmented diagnostics for radiologists. *Abdom. Radiol.* **47**, 4139–4150 (2022).
46. Yao, J. et al. Deep learning for fully automated prediction of overall survival in patients undergoing resection for pancreatic cancer: a retrospective multicenter study. *Ann. Surg.* **278**, 68–79 (2023).
47. Xia, Y. et al. Effective pancreatic cancer screening on non-contrast CT scans via anatomy-aware transformers. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021* (eds de Bruijne, M et al.), Lecture Notes in Computer Science, Vol. 12905, pp. 259–269 (Springer 2021).
48. Luo, H. et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol.* **20**, 1645–1654 (2019).
49. Jin, C. et al. Predicting treatment response from longitudinal images using multi-task deep learning. *Nat. Commun.* **12**, 1851 (2021).
50. Yushkevich, P. A. et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116–1128 (2006).
51. Qian, X. et al. Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning. *Nat. Biomed. Eng.* **5**, 522–532 (2021).
52. Lu, M. Y. et al. AI-based pathology predicts origins for cancers of unknown primary. *Nature* **594**, 106–110 (2021).
53. Selvaraju, R. R. et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision* (*ICCV*) 618–626 (IEEE, 2017).

Acknowledgements

The authors acknowledge external clinical validation contributions provided by Chang Gung Memorial Hospital (CGMH), and T.-C. Yen at CGMH for guiding the manuscript's revision. The authors thank R. M. Summers at the National Institutes of Health Clinical Center for many valuable discussions on cancer screening via imaging tools. The authors also acknowledge inspiration from E. Fishman at Johns Hopkins Hospital (JHH) on the pancreatic cancer screening initiative, and B. Vogelstein's group at JHH for their pioneering work on 'CancerSeek'. K.C. is supported by the National Natural Science Foundation of China (grant 82372045).

Author contributions

K.C., L.Z. and Y.X. conceived the study. L.Z., K.C., Y.X., L. Lu and Z.L. designed the study. K.C., X.H., L. La., T.Z., W.T., G.J., H.J., X.F., X.L., Y.W., M.Q., Y.H., T.K., M.V., Y.L., Y.C., X.C., Z.L., J.Z., C.X., R.Z., H.L., C.S., Y.S., Q.Z., T.L. and J.L. carried out data acquisition. Y.X., J.Y., L.Z., K.C., Y.W., M.Q. and W.F. carried out the data preprocessing. Y.X. developed the AI model. K.C., Y.X., L.Z., J.Y., X.H., L. La., Z.L., L. Lu, A.L.Y., G.D.H., Q.Z., T.L. and J.L. analyzed and interpreted the data. W.G., K.C., J.Y. and Y.X. carried out the clinical deployment. Y.X., J.Y., L.Z. and I.N. carried out the statistical analysis. Y.X., L.Z., K.C., J.Y., L. Lu and I.N. wrote and revised the paper.

Competing interests

Alibaba group has filed for patent protection (application numbers: CN 202210575258.9, US 18046405) on behalf of Y.X., L.Z., J.Y., L. Lu and X. Hua for the work related to the methods of detection of pancreatic cancer in non-contrast CT. Y.X., J.Y., W.G., Y.W., W.F., M.Q., L. Lu and L.Z. are employees of Alibaba Group and own Alibaba stock as part of the standard compensation package. All other authors have no competing interests.

Additional information

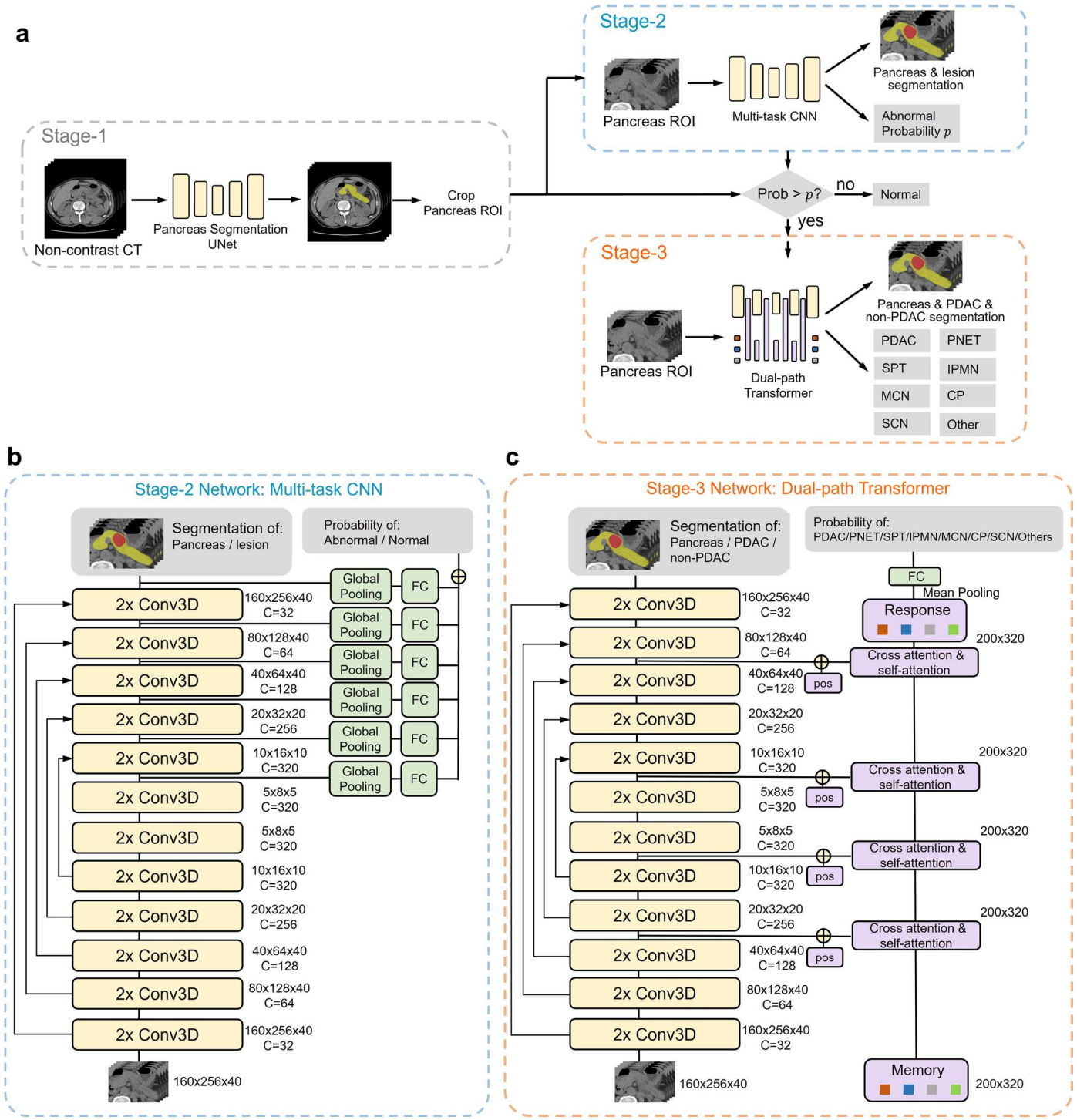
Extended data are available for this paper at https://doi.org/10.1038/s41591-023-02640-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-023-02640-w.

Correspondence and requests for materials should be addressed to Chengwei Shao, Yu Shi, Qi Zhang, Tingbo Liang, Ling Zhang or Jianping Lu.

Peer review information *Nature Medicine* thanks Jörg Kleeff, Ruijiang Li, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Lorenzo Righetto, in collaboration with the *Nature Medicine* team.

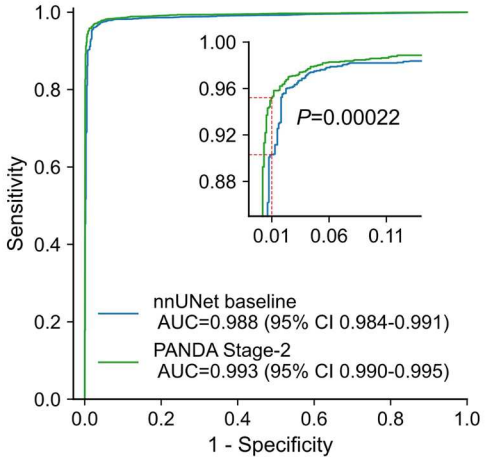
https://doi.org/10.1038/s41591-023-02640-w



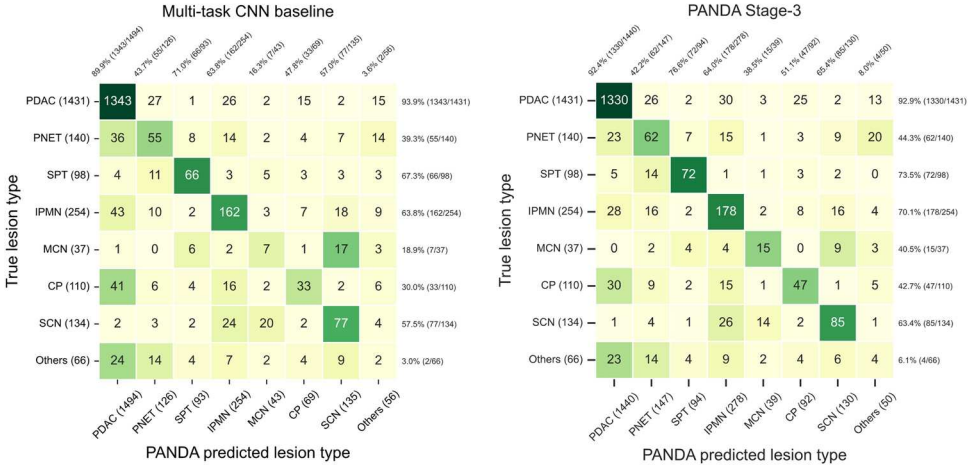
Extended Data Fig. 1 | Network architecture. a, Overview. Our deep learning framework consists of three stages: pancreas localization using a segmentation UNet, abnormality detection using a multi-task CNN, and lesion subtype classification using a dual-path transformer. b, Architecture of the multi-task CNN for Stage-2. We extract multi-level features from the segmentation UNet, and concatenate the features after global pooling for abnormal and normal classification. c, Architecture of the dual-path transformer for Stage-3. Lesion-related features are encoded into the learnable memory vectors from the UNet features and the learnable positional embeddings using cross-attention and self-attention. The response vectors of this procedure are then used for the classification of PDAC and seven non-PDAC subtypes.

https://doi.org/10.1038/s41591-023-02640-w

a



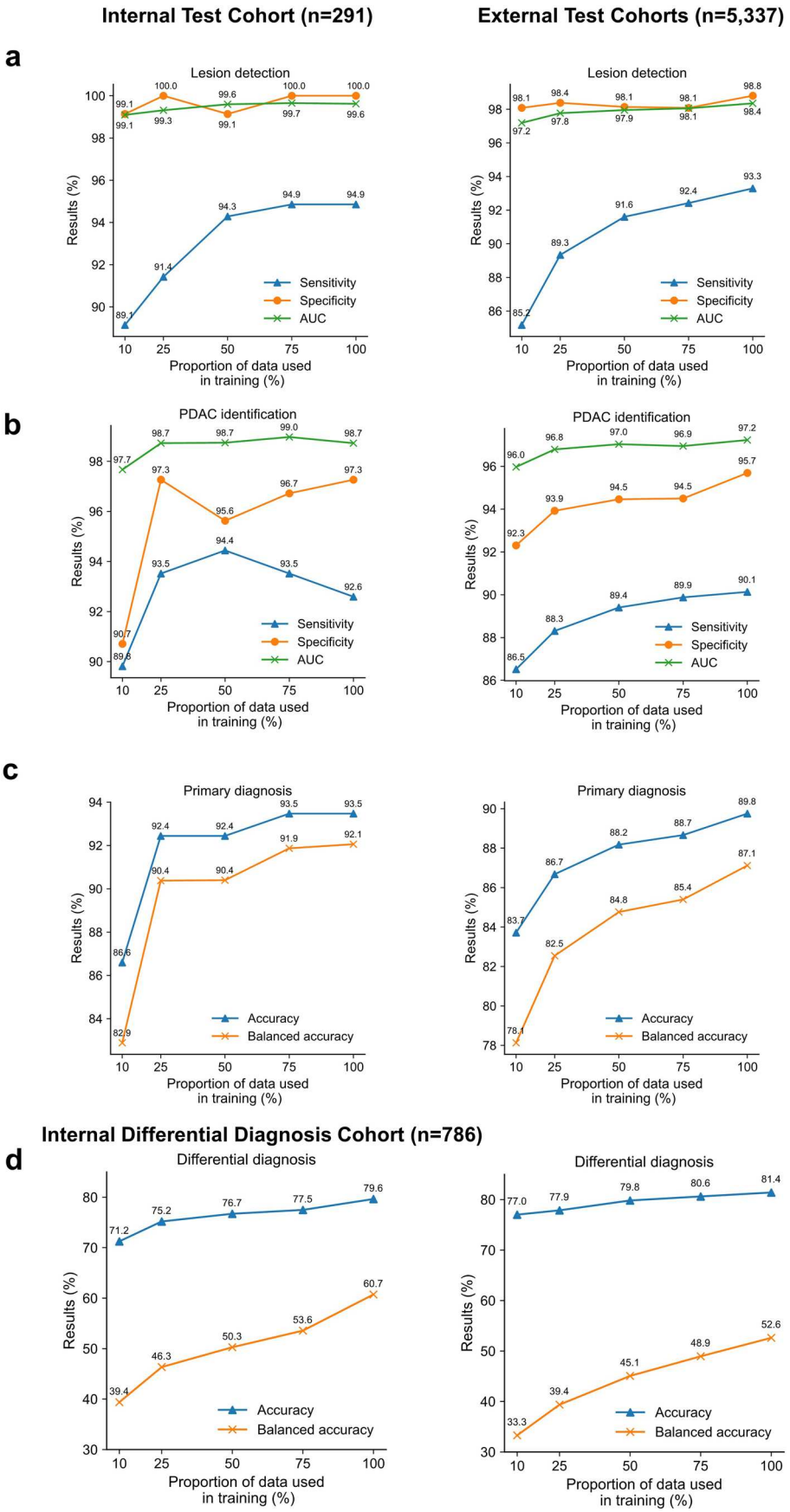
b



	Acc. (%)	Bal. acc. (%)
Multi-task CNN	76.9	46.7
PANDA	79.0	54.2
p-value	0.0034	0.0002

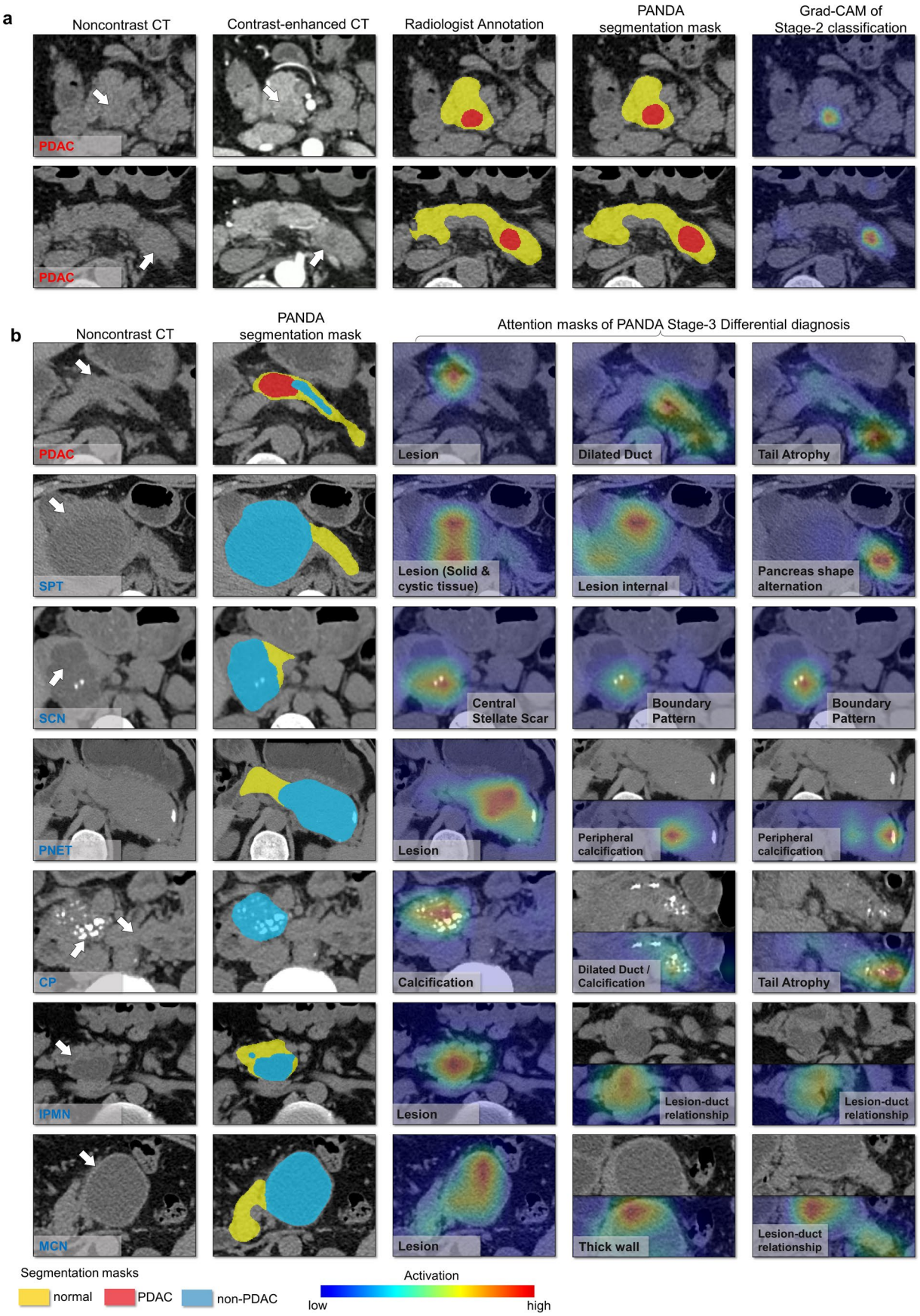
Extended Data Fig. 2 | Ablation studies of the 5-fold cross-validation on the training set (n = 3,208). a, nnUNet vs. PANDA Stage-2 network (multi-task CNN) for lesion detection, where PANDA achieved significant improvement in AUC score (P = 0.00022). At the same (desired) specificity level of 99.0%, PANDA Stage-2 outperformed nnUNet in sensitivity by 4.9% (95.2% vs. 90.3%) (marked in red dotted line). b, Multi-task CNN baseline (same as PANDA Stage-2 network with nnUNet backbone and classification head) vs. PANDA Stage-3 (dual-path transformer) for differential diagnosis, where PANDA achieved significant improvement in both accuracy (Acc.) and balanced accuracy (Bal. acc.). The significance test comparing the AUCs of the AI model and nnUNet is conducted using the Delong test. Two-sided permutation tests were used to compute the statistical differences of accuracy and balanced accuracy.

https://doi.org/10.1038/s41591-023-02640-w



Extended Data Fig. 3 | Influence of the proportion of training data. Influence of the proportion of training data tested on the internal test cohort (left) and the external test cohorts (right) on the task of a, lesion detection b, PDAC identification c, primary diagnosis d, differential diagnosis.

https://doi.org/10.1038/s41591-023-02640-w

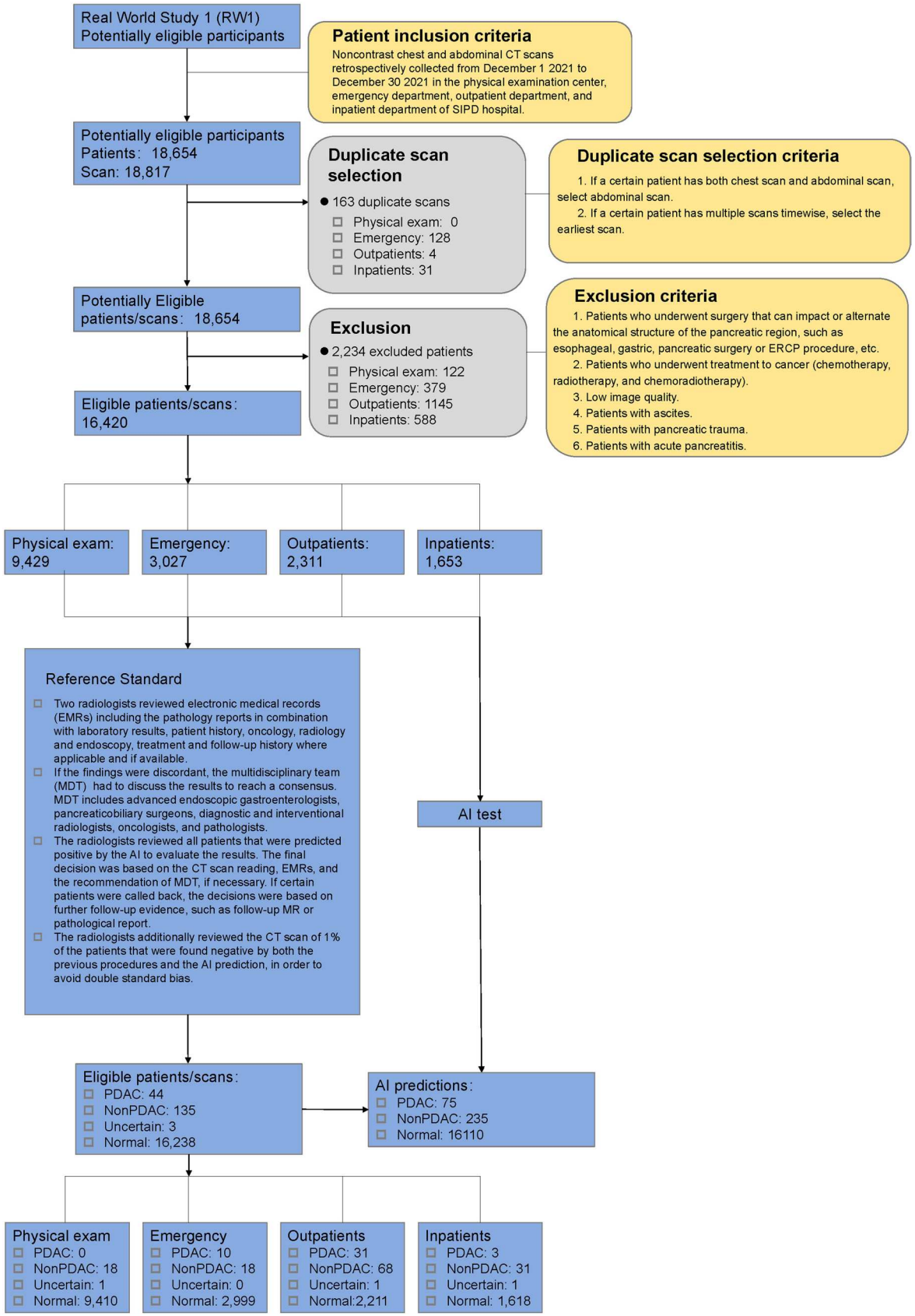


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41591-023-02640-w

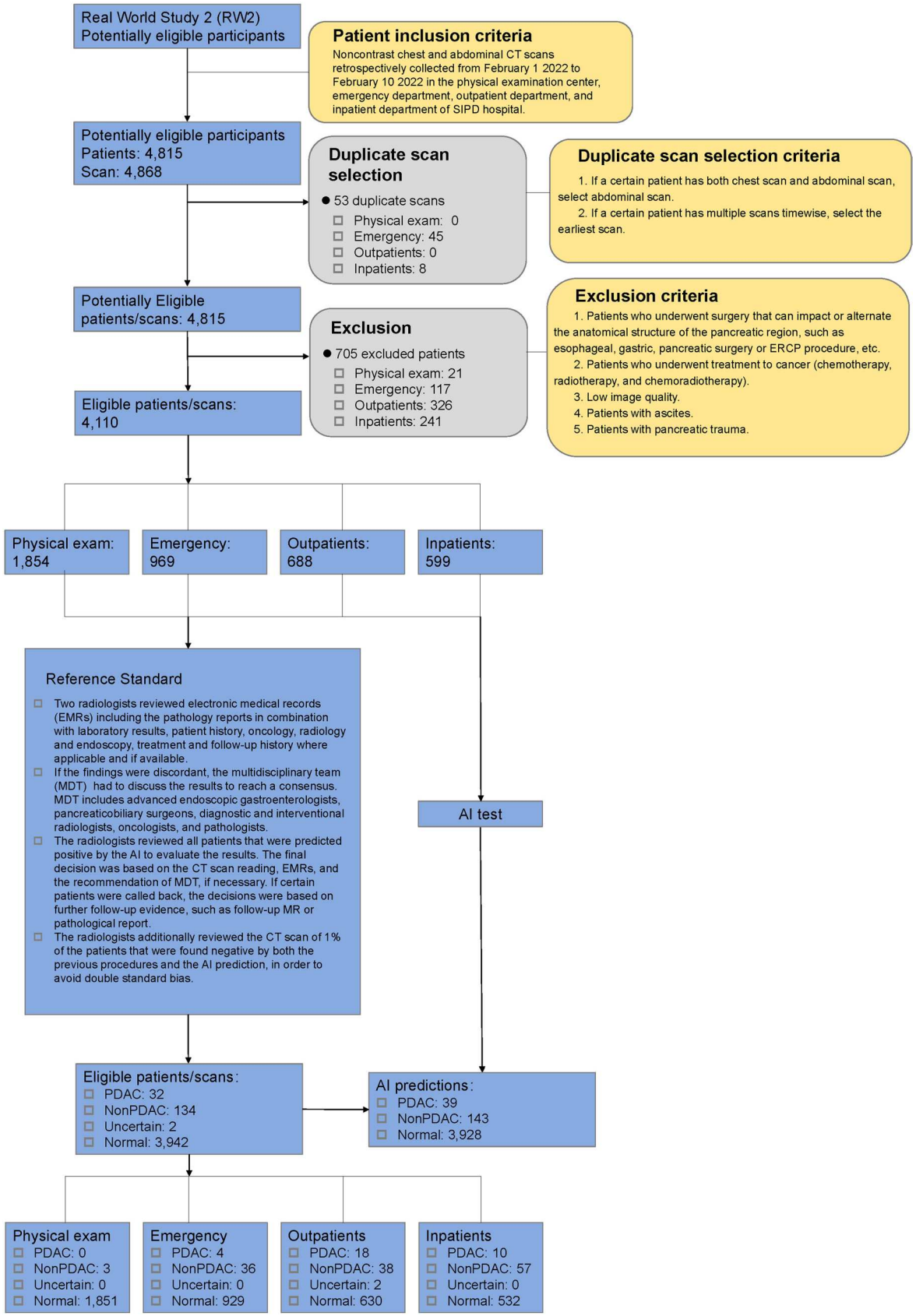
Extended Data Fig. 4 | Analysis of interpretability. a, we visualize the noncontrast CT, contrast-enhanced CT, and the radiologist's annotated mask and compare them with the PANDA segmentation map and the Grad-CAM heatmap of PANDA Stage-2 classification for lesion detection. PANDA correctly predicted the position of the PDAC (PANDA segmentation map) and made positive classification based on the local features of the PDAC (Grad-CAM heatmap). b, we visualize the top activated attention maps of the Transformer branch of PANDA Stage-3 to interpret how PANDA classified the lesions. The memory tokens of the Transformer not only attended to the lesion locations but also considered the secondary signs for lesion diagnosis as utilized by the radiologists. E.g. A PDAC caused pancreatic duct dilation and pancreatic atrophy; A SPT was circumscribed with the heterogeneity of both solid and cystic regions; A SCN had a pattern of central stellate scar and so-called honeycomb pattern; A PNET had isoattenuating mass and peripheral calcification; A CP was associated with calcification, dilated duct, and pancreatic atrophy; An IPMN lesion was connected to the pancreatic duct; A MCN had the thick cystic wall and no visual connection with the pancreatic duct. The heatmaps of multiple slices were displayed for the CP, IPMN, and MCN.

https://doi.org/10.1038/s41591-023-02640-w



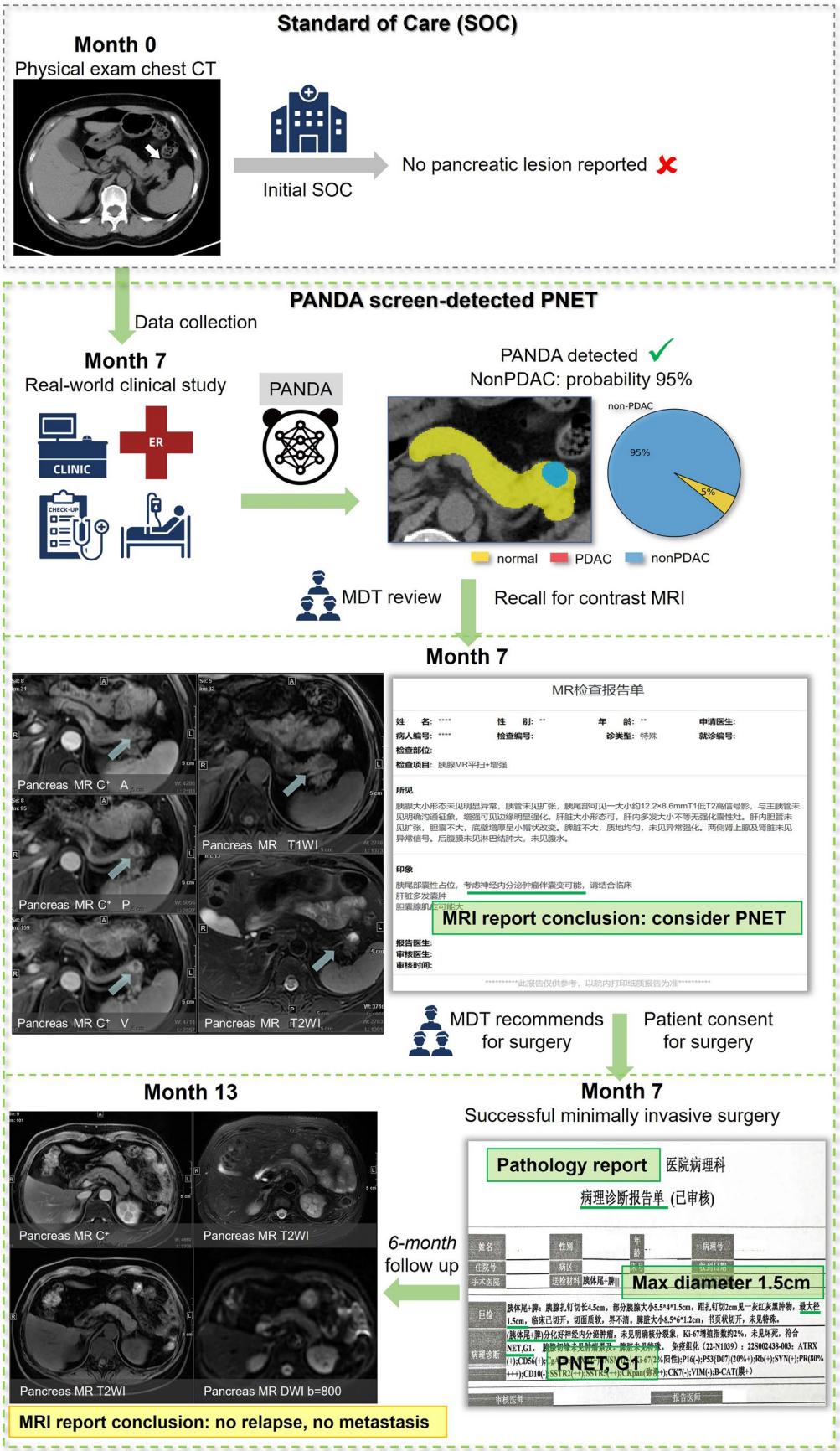
Extended Data Fig. 5 | Overview of the workflow of the first real-world study (RW1).

https://doi.org/10.1038/s41591-023-02640-w



Extended Data Fig. 6 | Overview of the workflow of the second real-world study (RW2).

https://doi.org/10.1038/s41591-023-02640-w



Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41591-023-02640-w

Extended Data Fig. 7 | Flowchart describing the successful discovery and intervention of a patient with pancreatic neuroendocrine tumors (PNET) in the real-world clinical evaluation. Noncontrast chest CT was performed on this patient in the physical examination center in Month 0, where the standard of care did not report any pancreatic findings. This patient was included in the real-world study in Month 7 and was reported as non-PDAC (95% probability) by PANDA. After the case was reviewed by MDT, the patient was recalled for contrast-enhanced MRI and was considered as PNET in the radiology report. The patient consented to surgery, which was later successfully performed in Month 7. The post-surgical pathology report confirmed an early-stage PNET (G1, 1.5cm). The 6 month follow-up (Month 13) showed no relapse or metastasis. The English translation of the MRI and pathology reports' key results are provided in green boxes.

https://doi.org/10.1038/s41591-023-02640-w

Extended Data Table 1 | Dataset characteristics

	Internal Train (n=3,208)	Internal Test (n=291)	Internal Addition (n=611)	Site A (SHCMU) (n=1,769)	Site B (FAHZU) (n=2,019)	Site C (XH) (n=370)	Site D (FUSCC) (n=292)	Site E (TMUCIH) (n=60)	Site F (SYUCC) (n=173)	Site G (GPPH) (n=92)	Site H (CGMH) (n=382)	Site I (GUHP) (n=180)	Chest (n=492)	RW1 (n=16,420)	RW2 (n=4,110)
Patient Characteristics															
Lesion types															
Normal, no.(%)	938(29)	116(40)	0(0)	495(28)	513(25)	194(52)	38(13)	0(0)	0(0)	49(53)	292(76)	87(48)	378(77)	16241(99)	3944(96)
PDAC, no.(%)	1431(45)	108(37)	367(60)	1023(58)	983(49)	115(31)	157(54)	60(100)	173(100)	43(47)	90(24)	93(52)	63(13)	44(0)	32(1)
PNET, no.(%)	140(4)	5(2)	53(9)	25(1)	86(4)	11(3)	38(13)	0(0)	0(0)	0(0)	0(0)	0(0)	11(2)	6(0)	5(0)
SPT, no.(%)	98(3)	9(3)	30(5)	20(1)	61(3)	4(1)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	6(1)	1(0)	1(0)
IPMN, no.(%)	254(8)	22(8)	65(11)	30(2)	118(6)	6(2)	18(6)	0(0)	0(0)	0(0)	0(0)	0(0)	16(3)	15(0)	12(0)
MCN, no.(%)	37(1)	2(1)	21(3)	24(1)	49(2)	5(1)	12(4)	0(0)	0(0)	0(0)	0(0)	0(0)	2(0)	1(0)	4(0)
CP, no.(%)	110(3)	13(4)	32(5)	27(2)	87(4)	15(4)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	7(1)	42(0)	55(1)
SCN, no.(%)	134(4)	10(3)	19(3)	92(5)	78(4)	11(3)	29(10)	0(0)	0(0)	0(0)	0(0)	0(0)	7(1)	11(0)	2(0)
other, no.(%)	66(2)	6(2)	24(4)	33(2)	44(2)	9(2)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	2(0)	59(0)	15(0)
AP, no.(%)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	40(1)
Reference standard on lesions															
Surgical pathology, no.(%)	2270(100)	175(100)	611(100)	1274(100)	918(61)	151(86)	254(100)	60(100)	173(100)	43(100)	90(100)	4(4)	114(100)	23(13)	18(11)
Biopsy pathology, no.(%)	0(0)	0(0)	0(0)	0(0)	588(39)	25(14)	0(0)	0(0)	0(0)	0(0)	0(0)	89(96)	0(0)	18(10)	20(12)
Clinical diagnosis, no.(%)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	138(77)	128(77)
Normal															
Female, no.(%)	483(51)	65(56)	0(-)	236(48)	255(50)	96(49)	38(100)	0(-)	0(-)	15(31)	102(35)	33(38)	162(43)	5915(36)	1391(35)
Age(IQR)	49 (39-60)	50 (38-60)	- (-)	52 (38-60)	57 (46-66)	46 (34-60)	49 (39-55)	- (-)	- (-)	60 (54-66)	53 (42-63)	73 (66-78)	37 (32-45)	38 (33-47)	58 (50-68)
Non-PDAC															
Female, no.(%)	414(49)	34(51)	121(50)	149(59)	280(54)	33(54)	61(63)	0(-)	0(-)	0(-)	0(-)	0(-)	26(51)	56(41)	38(28)
Age(IQR)	55 (44-64)	51 (43-64)	57 (45-67)	56 (46-64)	56 (45-66)	60 (50-67)	54 (42-62)	- (-)	- (-)	- (-)	- (-)	- (-)	59 (51-65)	61 (51-72)	48 (37-64)
PDAC															
Female, no.(%)	498(35)	48(44)	160(44)	415(41)	422(43)	46(40)	76(48)	29(48)	68(39)	20(47)	41(46)	38(41)	25(40)	19(43)	8(25)
Age(IQR)	63 (55-69)	64 (57-70)	64 (56-71)	61 (55-68)	65 (58-71)	65 (60-70)	62 (56-68)	58 (51-62)	59 (51-66)	61 (54-69)	57 (51-65)	64 (59-70)	65 (56-70)	66 (51-72)	60 (57-69)
T stage															
T1, no.(%)	146(10)	14(13)	49(13)	118(12)	80(8)	14(12)	16(10)	11(18)	24(14)	3(7)	17(19)	0(0)	6(10)	10(23)	4(13)
T2, no.(%)	659(46)	63(58)	217(59)	443(43)	381(39)	61(53)	59(38)	31(52)	97(56)	16(37)	60(67)	17(18)	40(63)	15(34)	12(38)
T3, no.(%)	433(30)	24(22)	98(27)	147(14)	153(16)	15(13)	35(22)	5(8)	44(25)	6(14)	9(10)	30(32)	16(25)	6(14)	3(9)
T4, no.(%)	15(1)	4(4)	3(1)	240(23)	320(33)	5(4)	46(29)	13(22)	3(2)	18(42)	3(3)	44(47)	1(2)	13(30)	12(38)
Missing data, no.(%)	178(12)	3(3)	0(0)	75(7)	49(5)	20(17)	1(1)	0(0)	5(3)	0(0)	1(1)	2(2)	0(0)	0(0)	1(3)
TNM stage															
I, no.(%)	316(22)	35(32)	94(26)	364(36)	156(16)	44(38)	40(25)	21(35)	50(29)	4(9)	24(27)	4(4)	20(32)	11(25)	5(17)
II, no.(%)	601(42)	52(48)	181(49)	276(27)	241(25)	30(26)	66(42)	19(32)	86(50)	7(16)	46(51)	8(9)	29(46)	13(30)	4(13)
III, no.(%)	196(14)	13(12)	84(23)	267(26)	254(26)	7(6)	41(26)	19(32)	27(16)	25(58)	19(21)	13(14)	11(17)	10(23)	7(22)
IV, no.(%)	140(10)	5(5)	8(2)	41(4)	283(29)	14(12)	9(6)	1(2)	5(3)	7(16)	0(0)	66(71)	3(5)	10(23)	15(47)
Missing data, no.(%)	178(12)	3(3)	0(0)	75(7)	49(5)	20(17)	1(1)	0(0)	5(3)	0(0)	1(1)	2(2)	0(0)	0(0)	1(3)
Location															
Head/uncinate, no.(%)	776(54)	54(50)	226(62)	700(68)	410(42)	38(33)	80(51)	57(95)	114(66)	20(47)	0(0)	31(33)	41(65)	25(57)	18(56)
Neck, no.(%)	17(1)	1(1)	4(1)	8(1)	97(10)	0(0)	8(5)	1(2)	13(8)	0(0)	0(0)	0(0)	0(0)	1(2)	2(6)
Body/tail, no.(%)	497(35)	40(37)	135(37)	245(24)	370(38)	25(22)	68(43)	2(3)	39(23)	23(53)	0(0)	59(63)	22(35)	18(41)	12(38)
Whole gland, no.(%)	0(0)	0(0)	0(0)	0(0)	38(4)	1(1)	1(1)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
Missing data, no.(%)	141(10)	13(12)	2(1)	70(7)	68(7)	51(44)	0(0)	0(0)	5(3)	0(0)	90(100)	2(2)	0(0)	0(0)	0()
IPMN types															
Main/mixed-duct, no.(%)	163(64)	11(50)	40(62)	14(47)	58(49)	4(67)	6(33)	-(-)	-(-)	-(-)	-(-)	-(-)	10(62)	3(20)	2(17)
Branch-duct, no.(%)	91(36)	11(50)	25(38)	16(53)	60(51)	2(33)	12(67)	-(-)	-(-)	-(-)	-(-)	-(-)	6(38)	12(80)	10(83)
Missing data, no.(%)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	-(-)	-(-)	-(-)	-(-)	-(-)	0(0)	0(0)	0(0)
CT characteristics															
PixelSpacing (mm)															
Lesion (IQR)	0.680 (0.647-0.717)	0.688 (0.648-0.727)	0.671 (0.644-0.720)	0.724 (0.684-0.772)	0.723 (0.684-0.764)	0.684 (0.648-0.733)	0.730 (0.662-0.780)	0.781 (0.781-0.781)	0.686 (0.661-0.742)	0.723 (0.684-0.772)	0.717 (0.682-0.771)	0.734 (0.680-0.800)	0.754 (0.695-0.811)	0.698 (0.678-0.745)	0.692 (0.654-0.741)
Normal (IQR)	0.684 (0.666-0.721)	0.680 (0.668-0.708)	- (-)	0.755 (0.703-0.805)	0.746 (0.703-0.800)	0.695 (0.682-0.740)	0.727 (0.634-0.770)	- (-)	- (-)	0.755 (0.703-0.805)	0.707 (0.684-0.759)	0.764 (0.733-0.842)	0.818 (0.771-0.859)	0.773 (0.703-0.834)	0.754 (0.689-0.814)
Slice thickness (mm)															
Lesion (IQR)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (2.0-5.0)	3.0 (3.0-3.0)	5.0 (5.0-6.0)	3.0 (3.0-3.0)	1.2 (1.2-2.0)	5.0 (5.0-5.0)	1.25 (1.0-1.25)	5.0 (5.0-5.0)	5.0 (5.0-8.0)	5.0 (5.0-5.0)	3.0 (3.0-5.0)	3.0 (3.0-5.0)
Normal (IQR)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	- (-)	3.0 (3.0-5.0)	5.0 (5.0-5.0)	6.0 (5.0-6.0)	5.0 (5.0-5.0)	- (-)	- (-)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	3.0 (3.0-5.0)
Z-axis extent (mm)															
Lesion (IQR)	198 (183-213)	195 (178-209)	210 (195-228)	215 (168-445)	450 (426-474)	285 (275-326)	261 (240-282)	233 (232-261)	250 (215-540)	414 (211-454)	261 (217-301)	216 (192-278)	315 (295-340)	225 (195-320)	248 (192-320)
Normal (IQR)	186 (168-201)	180 (168-195)	- (-)	250 (175-461)	465 (445-490)	310 (275-404)	268 (245-294)	- (-)	- (-)	415 (370-471)	210 (195-235)	450 (430-483)	320 (300-340)	330 (300-360)	320 (285-360)
$CTDI_{vol}$ (mGy)															
Lesion (IQR)	16.5 (14.2-16.5)	16.5 (16.5-16.5)	16.5 (14.6-16.5)	15.6 (11.6-21.1)	15.3 (12.9-18.2)	9.9 (6.7-12.2)	10.1 (8.6-15.0)	9.0 (8.2-11.5)	9.0 (4.8-12.0)	12.0 (10.6-13.8)	10.4 (7.9-14.2)	8.0 (6.0-11.6)	14.1 (10.8-15.3)	15.3 (13.2-16.5)	16.5 (15.2-17.6)
Normal (IQR)	15.4 (13.4-16.5)	16.5 (13.8-16.5)	- (-)	15.7 (11.8-20.9)	16.0 (12.7-18.3)	8.4 (5.2-12.6)	10.0 (8.6-12.3)	- (-)	- (-)	13.6 (11.3-17.9)	10.4 (8.0-14.4)	4.1 (3.3-5.5)	12.4 (11.2-13.8)	14.1 (13.2-16.8)	14.1 (12.7-16.5)
Kvp															
Lesion (IQR)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-120)	100 (90-120)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-120)
Normal (IQR)	120 (120-120)	120 (120-120)	- (-)	120 (120-120)	120 (120-120)	120 (120-120)	120 (120-123)	- (-)	- (-)	120 (120-120)	120 (120-120)	130 (130-130)	120 (120-120)	120 (120-120)	120 (120-120)
Tube Current (mA)															
Lesion (IQR)	300 (300-353)	300 (300-353)	300 (300-339)	284 (213-371)	281 (247-362)	270 (164-368)	211 (151-365)	211 (162-250)	267 (200-336)	318 (266-459)	240 (168-336)	247 (160-266)	300 (250-360)	332 (300-385)	332 (300-422)
Normal (IQR)	338 (329-430)	300 (300-432)	- (-)	311 (243-387)	271 (230-352)	191 (128-270)	193 (183-233)	- (-)	- (-)	378 (300-496)	240 (189-346)	49 (39-66)	264 (196-264)	264 (250-313)	300 (250-343)
Exposure (mAs)															
Lesion (IQR)	150 (150-190)	150 (150-190)	150 (150-181)	184 (147-234)	206 (149-255)	200 (98-305)	157 (113-200)	132 (120-169)	150 (112-177)	180 (156-215)	136 (65-160)	123 (81-134)	186 (150-200)	187 (150-221)	203 (174-250)
Normal (IQR)	181 (179-214)	150 (150-180)	- (-)	192 (150-244)	217 (118-258)	140 (71-200)	140 (114-160)	- (-)	- (-)	165 (26-222)	140 (73-160)	36 (29-49)	149 (129-148)	150 (149-200)	156 (149-200)
Enteric contrast															
Lesion	water	water	water	water	water	water	water	water	no	water	no	no	no	water/no	water/no
Normal	water	water	-(-)	water/no	water/no	water/no	water/no	-	-	water/no	no	no	no	water/no	water/no

T stage (AJCC eighth edition), TNM stage (AJCC eighth edition), and location are displayed for the PDAC patients. no., number; IQR, interquartile range; Internal Addition, additional cohort to enhance the evaluation of differential diagnosis, together with the 175 patients with pancreatic lesions in the internal test cohort, constitute the internal differential diagnosis cohort (n=786); SHCMU, Shengjing Hospital of China Medical University; FAHZU, First Affiliated Hospital of Zhejiang University; XH, Xinhua Hospital; FUSCC, Fudan University Shanghai Cancer Center; TMUCIH, Tianjin Medical University Cancer Institute and Hospital; SYUCC, Sun Yat-Sen University Cancer Center; GPPH, Guangdong Provincial People's Hospital; CGMH, Chang Gung Memorial Hospital; GUHP, General University Hospital in Prague; RW, real-world cohort.

https://doi.org/10.1038/s41591-023-02640-w

Extended Data Table 2 | Reader experience

Reader ID	Experience (yr)	CT read per year	Pancreatic CT read per year	Traning/Expertise
Specialist 1 (S1)	17	7,500	950	Pancreatic radiology
Specialist 2 (S2)	14	3,000	550	Pancreatic radiology
Specialist 3 (S3)	14	15,000	1,500	Pancreatic radiology
Specialist 4 (S4)	7	20,000	2,000	Pancreatic radiology
Specialist 5 (S5)	7	12,000	460	Pancreatic radiology
Specialist 6 (S6)	7	12,000	1000	Pancreatic radiology
Specialist 7 (S7)	9	7500	340	Pancreatic radiology
Specialist 8 (S8)	12	11,000	450	Pancreatic radiology
Specialist 9 (S9)	13	16,565	2600	Pancreatic radiology
Specialist 10 (S10)	8	15,000	560	Pancreatic radiology
Specialist 11 (S11)	8	8000	1000	Pancreatic radiology
General 1 (G1)	13	3,000	150	General radiology
General 2 (G2)	31	5,000	300	General radiology
General 3 (G3)	9	13,000	200	General radiology
General 4 (G4)	9	3800	170	General radiology
General 5 (G5)	8	1,800	100	General radiology
General 6 (G6)	8	20,000	500	General radiology
General 7 (G7)	8	1500	100	General radiology
General 8 (G8)	10	15,000	300	General radiology
General 9 (G9)	9	3200	150	General radiology
General 10 (G10)	10	18,000	200	General radiology
General 11 (G11)	9	3000	150	General radiology
Resident 1 (R1)	2	4,500	300	General radiology
Resident 2 (R2)	3	5,000	350	General radiology
Resident 3 (R3)	2	1,000	200	General radiology
Resident 4 (R4)	2	12,000	1,000	General radiology
Resident 5 (R5)	2	500	100	General radiology
Resident 6 (R6)	4	6500	200	General radiology
Resident 7 (R7)	2	300	100	General radiology
Resident 8 (R8)	8	12,000	350	General radiology
Resident 9 (R9)	4	6000	200	General radiology
Resident 10 (R10)	2	1200	100	General radiology
Resident 11 (R11)	4	6000	200	General radiology
Specialist 12 (S12)	6	16,000	400	Pancreatic radiology
Specialist 13 (S13)	7	17,000	400	Pancreatic radiology
Specialist 14 (S14)	7	15,000	500	Pancreatic radiology
Specialist 15 (S15)	12	17,000	2,000	Pancreatic radiology
Specialist 16 (S16)	8	25,000	500	Pancreatic radiology
Specialist 17 (S17)	10	17,000	1,000	Pancreatic radiology
Specialist 18 (S18)	6	23,000	500	Pancreatic radiology
Specialist 19 (S19)	12	20,000	2,000	Pancreatic radiology
Specialist 20 (S20)	12	30,000	3,000	Pancreatic radiology
Specialist 21 (S21)	6	17,000	400	Pancreatic radiology
Specialist 22 (S22)	7	15,000	1,000	Pancreatic radiology
Specialist 23 (S23)	19	20,000	450	Pancreatic radiology
Specialist 24 (S24)	10	20,000	450	Pancreatic radiology
Specialist 25 (S25)	10	20,000	500	Pancreatic radiology
Specialist 26 (S26)	10	21,000	500	Pancreatic radiology

Specialists were radiologists who had ≥5 years of experience in pancreatic imaging. Specialists 4 and 20 were highly regarded for their excellence within a high-volume pancreatic cancer institution. General 1 -General 11 were general radiologists who were practicing at community hospitals or other level hospitals and undergoing a refresher program in pancreatic radiology at the SIPD center at the time of the reader study. Resident 2 was a radiology resident whose research interest was pancreatic imaging.

https://doi.org/10.1038/s41591-023-02640-w

Extended Data Table 3 | Cases that were misdetected by the initial standard of care (SOC) but were successfully detected by PANDA in the real-world clinical evaluations

Patient	MDT diagnosis	Age at initial SOC	CT method at initial SOC	Clinical scenario	Diagnosed by follow-up SOC?	Maximum diameter at initial SOC / follow-up SOC (mm)	Method at follow-up SOC	Initial and follow-up SOC interval (months)	PANDA prediction/ probability	Contact/ Recall MRI	Surgery type	Pathology/ TNM stage AJCC 8th edition	Outcome
Real-world Study 1													
1	PDAC (head)	64	Chest CT w/o contrast	Outpatient	Y	25.0/26.3	Pancreas CT w/ contrast	1	PDAC/98%	N/-	PD	PDAC/ T2N2M0	Alive
2	SCN (head)	67	Rib three-dimensional CT	Outpatient	N	20.3/-	-	-	nonPDAC/98%	Y/N	-	-	Alive
3	Pancreatic cyst (body)	80	Renal CT w/o contrast	Outpatient	Y	8.0/10.0	Liver CT w/ contrast	1	nonPDAC/76%	N/-	-	-	Deceased†
4	Pancreatic cyst (body)	52	Middle abdominal CT w/o contrast	Outpatient	N	17.6/-	-	-	nonPDAC/76%	Y/N	-	-	Alive
5	Pancreatic cyst (body)	68	Chest CT w/o contrast	Outpatient	Y	13.1/14.2	Whole-body CT angiography	3	nonPDAC/78%	N/-	-	-	Alive
6	Pancreatic metastasis (body)	65	Chest CT w/ contrast	Inpatient	Y	20.3/33.0	Chest CT w/ contrast	2	PDAC/54%	N/-	-	-	Deceased*
7	BD-IPMN (head)	62	Renal CT angiography	Inpatient	N	21.5/-	-	-	nonPDAC/91%	Y/N	-	-	Alive
8	CP	92	Upper abdominal CT w/o contrast	Inpatient	N	-/-	-	-	nonPDAC/74%	Y/N	-	-	Alive
9	CP	63	Chest CT w/o contrast	Inpatient	N	-/-	-	-	nonPDAC/79%	Y/N	-	-	Alive
10	CP: Calcification (head)	52	Chest CT w/o contrast	Inpatient	N	-/-	-	-	nonPDAC/58%	Y/N	-	-	Alive
11	CP: Calcification (tail)	64	Chest CT w/o contrast	Inpatient	N	-/-	-	-	nonPDAC/84%	Y/N	-	-	Alive
12	Pancreatic cyst (neck)	75	Abdominal CT angiography	Inpatient	N	11.2/-	-	-	nonPDAC/49%	Y/N	-	-	Alive
13	Pancreatic cyst (head)	85	Renal CT angiography	Inpatient	N	12.2/-	-	-	nonPDAC/56%	Y/N	-	-	Alive
14	Pancreatic cyst (body)	63	Chest CT w/o contrast	Inpatient	Y	19.3/19.0	Pancreas MR w/ contrast	7	nonPDAC/100%	N/-	-	-	Alive
15	Pancreatic cyst (head)	77	Whole-body CT angiography	Inpatient	N	33.0/-	-	-	nonPDAC/59%	Y/N	-	-	Alive
16	Low malignant pancreatic mass (tail)	56	Chest CT w/o contrast	Physical exam	N	12.2/-	-	-	nonPDAC/95%	Y/Y	DP	PNET G1/-	Alive
17	Hepatic arterial aneurysm	93	Chest CT w/o contrast	Physical exam	Y	19.8/23.0	Spleen CT angiography	6	nonPDAC/95%	N/-	Inter-ventional therapy	Hepatic arterial aneurysm /-	Alive
18	BD-IPMN (head)	96	Chest CT w/o contrast	Physical exam	N	16.5/-	-	-	nonPDAC/57%	Y/N	-	-	Alive
19	BD-IPMN (body)	90	Chest CT w/o contrast	Physical exam	Y	8.0/12.0	Pancreas MR w/ contrast	2	nonPDAC/50%	N/-	-	-	Alive
20	CP: AIP	63	Chest CT w/o contrast	Physical exam	Y	-/-	Pancreas CT w/ contrast	2	PDAC/54%	N/-	-	-	Alive
21	CP	53	Chest CT w/o contrast	Physical exam	N	-/-	-	-	nonPDAC/57%	Y/N	-	-	Alive
22	Pancreatic cyst (tail)	92	Chest CT w/o contrast	Physical exam	Y	17.5/18.0	MRCP	3	nonPDAC/64%	N/-	-	-	Alive
23	Pancreatic cyst (head)	48	Chest CT w/o contrast	Physical exam	N	17.6/-	-	-	nonPDAC/74%	Y/N	-	-	Alive
24	Pancreatic cyst (head)	78	Chest CT w/o contrast	Physical exam	N	12.9/-	-	-	nonPDAC/81%	Y/N	-	-	Alive
25	Pancreatic cyst (head)	54	Chest CT w/o contrast	Physical exam	N	10.1/-	-	-	nonPDAC/96%	Y/N	-	-	Alive
26	Pancreatic cyst (body)	75	Chest CT w/o contrast	Physical exam	N	14.6/-	-	-	nonPDAC/58%	Y/N	-	-	Alive
Real-world Study 2													
1	PNET	73	Thymus CT w/ contrast	Outpatient	N	24.2/-	-	-	nonPDAC/61%	Y/N	-	-	Alive
2	PDAC (body)	67	Chest CT w/o contrast	Inpatient	N	21.0/-	-	3	PDAC/99%	Y/N	DP	PDAC/ T2N0Mx	Alive‡
3	Pancreatic cyst (head)	63	Chest CT w/o contrast	Inpatient	Y	14.1/15.3	Chest CT w/ contrast	1	nonPDAC/98%	N/-	-	-	Deceased±
4	Pancreatic cyst (body)	30	Chest CT w/o contrast	Physical exam	N	9.7/-	-	-	nonPDAC/95%	Y/N	-	-	Alive
5	Pancreatic cyst (head)	77	Chest CT w/o contrast	Emergency	N	32.0/-	-	-	nonPDAC/87%	Y/N	-	-	Alive

The last outcome follow-up is in February 2023. *Died due to metastases originating from lung cancer. †Died due to lung cancer. ‡Underwent surgery in another hospital. ± Died due to lung cancer. MDT, multi-disciplinary team; AJCC, American Joint Committee on Cancer; PDAC, pancreatic ductal adenocarcinoma; BD-IPMN, branch duct intraductal papillary mucinous neoplasm; AIP, autoimmune pancreatitis; CP: Chronic pancreatitis; MCN, mucinous cystic neoplasm; PNET, pancreatic neuroendocrine tumor; MRCP, magnetic resonance cholangiopancreatography; DP, distal pancreatectomy; PD, pancreatoduodenectomy; Y, yes; N, no.

nature portfolio

Corresponding author(s): Chengwei Shao, Yu Shi, Qi Zhang, Tingbo Liang, Ling Zhang, and Jianping Lu

Last updated by author(s): Oct 9, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Dicom files were handled with the open source libraries Pydicom (https://pydicom.github.io/, version 2.2.2), SimpleITK (https://simpleitk.org/, version 2.0.2), and NiBabel (https://nipy.org/nibabel/, version 3.2.1). Custom Python (version 3.9.7) script was developed for data de-identification.
Data analysis	The code used for the implementation of PANDA has dependencies on internal tooling and infrastructure, is under patent protection (application numbers: CN 202210575258.9, US 18046405), and thus is not feasible to be publicly released. All experiments and implementation details are described in sufficient detail in the Methods and Supplementary Methods sections to support replication with non-proprietary libraries. Several major components of our work are available in open-source repositories: PyTorch (https://pytorch.org/); nnUNet (https://github.com/MIC-DKFZ/nnUNet). Data analysis was conducted in Python using the numpy (version 1.20.3), scipy (version 1.8.1), and scikit-learn (version 0.24.2) packages. The calculation of people needed to screen in the high-risk population was based on Test for One-Sample Sensitivity and Specificity via PASS software (version 15).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sample data and interactive demo are displayed in the webpage (http://panda.medofmind.com/). The remaining datasets used in this study are currently not permitted for public release by the respective Institutional Review Boards. Requests for access to aggregate data and supporting clinical documents will be reviewed and approved by an independent review panel on the basis of scientific merit. All data provided are anonymized to respect the privacy of patients who have participated in the studies, in line with applicable laws and regulations. Data requests pertaining to the manuscript may be made to the first author (Kai Cao; mdkaicao163@163.com). Requests will be processed within 6 weeks.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We are using only retrospective data collected through clinical practice. Sex was assigned based on the government-issued ID. The datasets used in the internal training and test cohorts, and the external multi-center test cohorts have sex distributions reported in the paper. Sex-based analysis was not reported because sex was unrelated to model implementation or deployment. Self-identification gender was not collected from the patients.
Reporting on race, ethnicity, or other socially relevant groupings	We are using only retrospective data collected through clinical practice. Race, ethnicity, and other socially relevant groupings were not collected from the patients and were unrelated to model implementation or deployment.
Population characteristics	This retrospective study included five patient cohorts: an internal training cohort , an internal test cohort (together with an additional internal differential diagnosis cohort), an external international multicenter test cohort, a chest noncontrast CT test cohort, and a real-world clinical test cohort. In the first four cohorts, the pancreatic ductal adenocarcinoma (PDAC) and nonPDAC lesions were confirmed by surgical or biopsy histopathology, which we used as the ground truth label for each patient. The normal controls were confirmed without pancreatic or peri-pancreatic disease by a two-year follow-up. Patients with acute pancreatitis and a history of abdominal treatment were excluded. The internal training test cohort (normal controls: median age 49 years [IQR 39-60], nonPDAC: median age 55 years [IQR 44-64], PDAC: median age 63 [IQR 55-69]), internal test cohort (normal controls: median age 50 [IQR 38-60], nonPDAC: median age 51 [IQR 43-64], PDAC: median age 64 [IQR 57-70]), internal additional test cohort (nonPDAC: median age 57 [IQR 45-67], PDAC: median age 64 [IQR 56-71]), chest noncontrast CT test cohort (normal controls: median age 37 [IQR 32-45], nonPDAC: median age 59 [IQR 51-65], PDAC: median age 65 [IQR 56-70]), and the real-world clinical test cohort (RW1-normal controls: median age 38 [IQR 33-47], nonPDAC: median age 61 [IQR 51-72], PDAC: median age 66 [IQR 51-72]; RW2-normal controls: median age 58 [IQR 50-68], nonPDAC: median age 48 [IQR 37-64], PDAC: median age 60 [IQR 57-69]) were collected from the internal center in Shanghai, China (Shanghai Institution of Pancreatic Diseases [SIPD]). The external test cohorts were collected from nine centers, of which seven were located in China, one in Taiwan ROC (Linkou Chang Gung Memorial Hospital [CGMH], normal controls: median age 53 [IQR 42-63], PDAC: median age 41 [IQR 51-65] -- Site H), and one in the Czech Republic (General University Hospital in Prague [GUHP], normal controls: median age 73 [IQR 66-78], PDAC: median age 64 [IQR 59-70] -- Site I). The seven centers from China are distributed widely in geographical areas: one in the Northeast (Shengjing Hospital of China Medical University [SHCMU], normal controls: median age 52 [IQR 38-60], nonPDAC: median age 56 [IQR 46-64], PDAC: median age 61 [IQR 55-68] -- Site A), four in the East (First Affiliated Hospital of Zhejiang University [FAHZU], normal controls: median age 57 [IQR 46-66], nonPDAC: median age 56 [IQR 45-66], PDAC: median age 65 [IQR 58-71] -- Site B; Xinhua Hospital [XH], normal controls: median age 46 [IQR 34-60], nonPDAC: median age 60 [IQR 50-67], PDAC: median age 65 [IQR 60-70] --Site C; Fudan University Shanghai Cancer Center [FUSCC], normal controls: median age 49 [IQR 39-55], nonPDAC: median age 54 [IQR 42-62], PDAC: median age 62 [IQR 56-68]-- Site D; and Tianjin Medical University Cancer Institute and Hospital[TMUCIH], PDAC: median age 58 [IQR 51-62]--Site E), and two in the South (Sun Vat-sen University Cancer Center [SYUCC], PDAC: median age 59 [IQR 51-66] -- Site F; and Guangdong Provincial People's Hospital [GPPH], normal controls: median age 60 [IQR 54-66], PDAC: median age 61 [IQR 54-69] -- Site G). For all patients included in the multicenter test cohort, additional metadata for data characteristic was available, including patient age and sex. For the patients with PDAC, the T stage and TNM stage (AJCC eighth edition) and the location of the lesion are available. For example, 707 PDAC patients (25.8%) and 779 PDAC patients (28.5%) in the external test cohorts were TNM stage I cancer and stage II cancer, respectively. Further details are provided in the extended data.
Recruitment	The internal training cohort included 3,208 patients (1,431 PDAC, 140 pancreatic neuroendocrine tumor [PNET], 98 solid pseudopapillary tumor [SPT], 254 intraductal papillary mucinous neoplasm [IPMN], 37 mucinous cystic neoplasm [MCN], 110 chronic pancreatitis [CP], 134 serous cystic neoplasm [SCN], 66 'other', and 938 normal controls) who had been treated between January 2015 to October 2020 at the Shanghai Institution of Pancreatic Diseases (SIPD), China. Consecutive patients (except for who had chest CT before surgery) with pancreatic lesions confirmed by surgical pathology were included. Normal controls confirmed by at least 2 years of follow-up were randomly selected from the same time period. All cases had preoperative multi-phase contrast-enhanced CT images acquired by Philips, Siemens, Toshiba, or Vital scanners. The internal test cohort contained CT scans of 291 patients randomly collected between December 2015 and June 2018 at

the SIPD, China, including 108 PDAC, 9 SPT, 5 PNET, 22 IPMN, 2 MCN, 10 SCN, 13 CP, 6 'other', and 116 normal controls. We additionally collected an internal addition cohort consisting of 611 consecutive patients who underwent surgery between November 2020 and October 2021 at SIPD, including 367 PDAC, 53 PNET, 30 SPT, 65 IPMN, 21 MCN, 32 CP, 19 SCN, and 24 'other'. These 611 patients, together with the 175 patients with pancreatic lesions in the internal test cohort, constitute the internal differential diagnosis cohort (n=786). All patients took multi-phase CT including noncontrast, arterial, venous, and delay.

In the multicenter test cohorts , the noncontrast CT scans of 5,337 patients, including 2,737 PDAC, 932 nonPDAC, and 1,668 normal, were collected from these centers. Site A, SHCMU, is a tertiary hospital in China. We consecutively collected 1,023 patients with PDAC and 251 patients with nonPDAC, and randomly selected 495 normal controls from January 2010 to May 2020. Site B, FAHZU, is a tertiary hospital in China. We consecutively collected 983 patients with PDAC and 523 patients with nonPDAC from May 2020 July 2022, and randomly collected 513 normal controls from Dec 1 2021 to Dec 31 2021. Site C, XH, is a tertiary hospital in China. We consecutively collected 115 patients with PDAC and 61 patients with nonPDAC, and randomly selected 194 normal controls from January 2019 to December 2020. Site D, FUSCC, is a tertiary hospital in China. We collected 157 PDAC, 97 nonPDAC, and 38 normal controls from November 2016 to November 2020. Site E, TMUCIH, is a tertiary hospital in China. We collected 60 patients with PDAC from January 2010 and November 2019. Site F, SYUCC, is a tertiary hospital in China. We consecutively collected 173 patients with PDAC from March 2010 to April 2020. Site G, GPPH, is a tertiary hospital in China. We collected 43 patients with PDAC and randomly selected 49 normal controls from January 2011 and August 2015. Site H, CGMH, is a hospital in Taiwan, ROC. Doctors from CGMH consecutively collected 90 patients with PDAC and randomly selected 292 normal controls from March 2009 to November 2015. Site I, GUHP, is a hospital in the Czech Republic. We consecutively collected 93 patients with PDAC and randomly selected 87 normal controls from August 2005 to March 2022.

We collected noncontrast chest CT test cohort with pathology-confirmed PDAC and nonPDAC and two-year follow-up confirmed normal controls. Specifically, for patients with PDAC or nonPDAC confirmed by surgical pathology, we searched for their nearest chest CT images for up to one year before surgery. For patients with chest CT reports of normal pancreas, we searched for their follow-up records of normal pancreas for at least two years. By doing so, we collected 63 PDAC, 51 nonPDAC, and 378 normal controls spanning from November 2015 and May 2022 at SIPD. These noncontrast CTs of PDAC and nonPDAC were acquired before a mean of 7 days (range: -20--191 days) from the contrast-enhanced abdominal CT diagnosis. We ensured that all patients were independent of the patients in the training cohort.

Potential Bias: The above experiments validated the clinical utility of our novel tool PANDA, but are limited to pathology-confirmed pancreatic lesions (thus with higher risk) and a moderate number of normal cases. It is unclear by now whether PANDA could generalize well on the real-world population, including patients with lesions of lower risk and the large, diverse set of subjects with normal pancreas.

The real-world, retrospective clinical trial was complete and was registered with http://www.chictr.org.cn, ChiCTR2200064645. We collected two sub-cohorts (real-world-1 [RW1] and real-world-2 [RW2]) at the SIPD. Inclusion criteria was the availability of a noncontrast CT scan covering the pancreas region, e.g., lung, esophagus, liver, and kidney CT Patients with acute pancreatitis (AP) (in RW1), abdominal cancer treatment, severe ascites, abdominal trauma, and low imaging quality were excluded. The original RW1 consisted of 18,654 consecutive individuals whose noncontrast CT scans were examined from December 1, 2021, to December 31, 2021, from four different clinical scenarios at the SIPD. After exclusion (n=2,234, 12%), 16,420 individuals remained, including 9,429, 3,027, 2,311, and 1,653 from the physical examination, emergency, outpatient, and inpatient department, respectively. RW1 included 44 PDAC, 6 PNET, 1 SPT, 15 IPMN, 1 MCN, 42 CP, 11 SCN, and 59 other (mostly benign cysts).The original RW2 consisted of 4,815 consecutive individuals from February 1, 2022, to February 10, 2022, from the four clinical scenarios at the SIPD. The exclusion criteria was same as RW1 except that we included AP for RW2. After exclusion (n=705, 15%), 4,110
individuals remained, including 1,854, 969, 688, and 599 from the physical examination, emergency, outpatient, and inpatient department, respectively. RW2 included 32 PDAC, 5 PNET, 1 SPT, 12 IPMN, 4 MCN, 55 CP, 2 SCN, 15 other, and 40 AP.

Ethics oversight

The retrospective collection of the patient datasets in each cohort was approved by the Institutional Review Board (IRB) at each institution with a waiver for informed consent. The following review boards were used for each dataset: Site SIPD: Shanghai Institution of Pancreatic Diseases IRB, Site A: Shengjing Hospital of China Medical University IRB, Site B: First Affiliated Hospital of Zhejiang University IRB, Site C: Xinhua Hospital of Shanghai Jiao Tong University School of Medicine IRB, Site D: Fudan University Shanghai Cancer Center IRB, Site E: Tianjin Medical University Cancer Institute and Hospital IRB, Site F: Sun Yat-sen University Cancer Center IRB, Site G: Guangdong Provincial People's Hospital IRB, Site H: Linkou Chang Gung Memorial Hospital IRB, Site I: Charles University and General University Hospital IRB. All data in this study were de-identified prior to model training, testing, and reader studies. The investigators followed the requirements of the Declaration of Helsinki throughout the study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The internal training cohort includes 3,208 patients. We apply 5-fold cross validation where each fold randomly selected 80% for training and 20% for the validation purpose. This scheme follows machine learning convention for model tuning and hyperparameter selection. The internal test cohort includes 291 patients, which is a random selection of patients and independent from the training cohort. The size was selected due to time and budgetary constraints for the reader study on the same data, while maintaining sufficient positive and negative patients to power statistical comparisons on the metric of sensitivity, specificity, and accuracy. This test cohort selection is also based on prior work, as 11 readers' initial interpretations on this set had already been collected. The internal addition cohort include 611 patients. These patients, along with the original internal test cohort patients, constitute a new cohort named the internal differential diagnosis cohort. The minimal number of lesions (i.e., MCN) is 23, which is sufficient for the evaluation of differential diagnosis. The external test cohorts include 5,337 patients is a larger independent test set and include a more representative population.
Data exclusions	Patients who comply one or more of the following criteria were excluded from the studies: (1) patients who underwent surgery that can impact or alternate the anatomical structure of the pancreatic region, such as esophageal, gastric, pancreatic surgery or endoscopic retrograde cholangiopancreatography procedure, etc; (2) patients who underwent treatment to cancer (chemotherapy, radiotherapy, and chemoradiotherapy); (3) patients with low image quality due to artifacts caused by metal in stents or drastic motion during imaging; (4) patients with ascites; (5) patients with pancreatic trauma; (6) patients with acute pancreatitis (except for those in the second real-world clinical evaluation).
Replication	All attempts at replication were successful. The performance of PANDA was consistent across the internal center and 9 external centers across population (Asian and European), equipment manufacture (GE, Philips, Siemens, and Toshiba CT scanners), scanning protocols (abdominal noncontrast CT and chest noncontrast CT), and application scenarios (physical examination centers, emergency department, inpatient department, and outpatient department). In both of the reader studies, comparison between PANDA and human performance revealed consistent trend.
Randomization	For the dataset in the internal training cohort and the internal test cohort, patients were randomly assigned into training and test splits. In the internal training cohort, patients were randomly assigned to training and validation in the process of the cross-validation.
Blinding	The internal test cohort , the external international multicenter test cohort, the chest noncontrast CT test cohort, and the real-world clinical test cohort were not used for the development of PANDA. The second subset of real-world clinical test cohort (RW2) were not used for the development of PANDA Plus. In the reader studies, readers were blinded to pathology results and other clinical information, except for patient age and sex. Readers were also blinded to the data collection, exact ratio of the positive patients, and blinded to other readers. Readers were blinded to the ground-truth labels and their performance after the study. In the real-world study, the two radiologists who were responsible for the patients' record review were blinded to the results of AI.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	ChiCTR(chictr.org.cn): ChiCTR2200064645
Study protocol	ChiCTR clinical trial protocols: https://www.chictr.org.cn/showprojen.aspx?proj=169295
Data collection	We retrospectively collected two sub-cohorts (real-world-1 [RW1] and real-world-2 [RW2]) at the Shanghai Institution of Pancreatic

Data collection	Diseases (SIPD). Inclusion criteria was the availability of a noncontrast CT scan covering the pancreas region, e.g., lung, esophagus, liver, and kidney CT. Patients with acute pancreatitis (AP) (in RW1), abdominal cancer treatment, severe ascites, abdominal trauma, and low imaging quality were excluded. The original RW1 consisted consecutive individuals whose noncontrast CT scans were examined from December 1, 2021, to December 31, 2021, from the physical exam center, emergency department, inpatient department, and outpatient department at the SIPD. The original RW2 consisted of 4,815 consecutive individuals from February 1, 2022, to February 10, 2022, from the same four clinical scenarios at the SIPD.
Outcomes	The primary outcomes were the AUCs, sensitivity, and specificity of the AI models. The secondary outcomes included the analysis of number of false positives (safety), and detection of misdetection of standard-of-care (patient benefit) of the AI model under four real-world clinical scenarios.