Automatic detection of pancreatic lesions and main pancreatic duct dilatation on portal venous CT scans using deep learning

Clément Abi Nader, Rebeca Vétil, Laura Kate Wood, Marc-Michel Rohe, Alexandre Bône, Hedvig Karteszi, Marie-Pierre Vullierme

Title

Automatic detection of pancreatic lesions and main pancreatic duct dilatation on portal venous CT scans using deep learning.

Manuscript type: original article.

Authors

- Clément Abi Nader [(1)], PhD
- Rebeca Vetil [(1,2)]
- Laura Kate Wood [(1)], PhD
- Marc-Michel Rohe [(1)], PhD
- Alexandre Bône [(1)], PhD
- Hedvig Karteszi [(3,4)], MD, FRCR
- Marie-Pierre Vullierme [(5)], MD

The corresponding author is Clément Abi Nader.

104 Boulevard de Sébastopol, 75003, Paris, France.

clement.abi-nader@guerbet.com

+33 6 46 40 57 79

Institutions

1. Guerbet Research, France.
2. LTCI, Télécom Paris, Institut Polytechnique de Paris, France.
3. Department of Radiology, University Hospitals Bristol, UK.
4. Weston NHS Foundation Trust, UK.
5. Department of Radiology, Hospital of Annecy-Genevois, Université Paris-Cité, France.

Funding information

Research study funded by Guerbet and BPI France within the scope of the France 2030 project.

Abbreviations

- CT: Computed Tomography.
- DL: Deep Learning.
- MPD: Main Pancreatic Duct.
- PDAC: Pancreatic Ductal Adenocarcinoma.

- NET: Neuroendocrine Tumor.
- IPMN: Intraductal Papillary Mucinous Neoplasm.
- SCA: Serous Cystadenoma.
- NSD: Normalized Surface Dice.
- HU: Hounsfield Unit

Abstract

Objectives

This study proposes and evaluates a deep learning method to detect pancreatic neoplasms and to identify main pancreatic duct (MPD) dilatation on portal venous computed tomography (CT) scans.

Materials and Methods

A total of 2890 portal venous CT scans from 9 institutions were acquired, among which 2185 had a pancreatic neoplasm and 705 were healthy controls. Each scan was reviewed by one in a group of 9 radiologists. Physicians contoured the pancreas, pancreatic lesions if present, and the MPD if visible. They also assessed tumor type and MPD dilatation. Data was split into a training and independent testing set of 2134 and 756 cases, respectively.

A method to detect pancreatic lesions and MPD dilatation was built in three steps. First, a segmentation network was trained in a 5-fold cross-validation manner. Second, outputs of this network were postprocessed to extract imaging features: a normalized lesion risk, the predicted lesion diameter, and the MPD diameter in the head, body, and tail of the pancreas. Third, two logistic regression models were calibrated to predict lesion presence and MPD dilatation, respectively. Performance was assessed on the independent test cohort using receiver operating characteristic analysis. The method was also evaluated on subgroups defined based on lesion types and characteristics.

Results

The area under the curve (AUC) of the model detecting lesion presence in a patient was 0.98 (95% CI: 0.97, 0.99). A sensitivity of 0.94 (469 of 493, 95% CI: 0.92, 0.97) was reported. Similar values were obtained in patients with small (less than 2*cm*) and isodense lesions with a sensitivity of 0.94 (115 of 123, 95% CI: 0.87, 0.98) and 0.95 (53 of 56, 95% CI: 0.87, 1.0), respectively. The model sensitivity was also comparable across lesion types with values of 0.94 (95% CI: 0.91, 0.97), 1.0 (95% CI: 0.98, 1.0), 0.96 (0.97, 1.0) for pancreatic ductal adenocarcinoma, neuroendocrine tumor and intraductal papillary

neoplasm, respectively. Regarding MPD dilatation detection, the model had an AUC of 0.97 (95% CI: 0.96, 0.98).

Conclusions

The proposed approach showed high quantitative performance to identify patients with pancreatic neoplasms and to detect MPD dilatation on an independent test cohort. Performance was robust across subgroups of patients with different lesion characteristics and types. Results confirmed the interest to combine a direct lesion detection approach with secondary features such as the MPD diameter, thus indicating a promising avenue for the detection of pancreatic cancer at early stages.

Key words

Pancreatic neoplasm. Main Pancreatic Duct. Portal venous CT. Lesion detection. Deep Learning.

1. Introduction

Pancreatic cancer is currently the 11th most common cancer and the 7th leading cause of cancer-related deaths worldwide [1] . With a predicted incidence expected to increase up to 78% during the period 2018-2040, it is a major healthcare issue on the rise [1]. Due to the increase of pancreatic cancer incidence, combined with its dramatically low 5-year survival rate of 9%, the disease could become the third leading cause of cancer-related deaths by 2025 [2].

Most patients with early pancreatic cancer present with non-specific symptoms hence, if investigated, often undergo routine computer tomography (CT) examination performed in portal venous phase. Image interpretation might be difficult, as in early stages, pancreatic lesions tend to be small (less than 2*cm*) and isodense, with a reported sensitivity ranging from 58% to 77% [3] [4]. In addition, radiologists' heavy workload as well as their level of expertise and experience might further affect the interpretation of a CT scan. As the disease progresses rapidly [5], pancreas cancer is mostly detected during late stages for which limited therapeutic options are available, thus explaining the observed low 5-year survival rate. To date only 10% [6] of patients undergo pancreatic resection, the sole curative treatment.

However, in the past years the proportion of patients diagnosed with stage 1A pancreatic cancer has increased [7]. At this stage, the patients are more frequently eligible for pancreatic resection and adjuvant chemotherapy and as such, the 5-year survival rate of these patients now exceeds 80% [7]. This highlights however, the importance for the earliest possible detection of pancreas cancer. To identify findings that should alert radiologists about the potential presence of pancreatic cancer, studies retrospectively analyzed CT scans of patients with pancreatic cancer before the histopathological

diagnosis [8 9 10 11]. A consensus emerged on the fact that subtle secondary signs, such as main pancreatic duct (MPD) dilatation, were often visible up to 1 year before the cancer diagnosis. This is due to the fact that pancreas cancer is mainly a ductal adenocarcinoma and the malignant tumor is responsible for a stricture of the MPD, thus causing upstream dilatation. Dilatation is usually defined as a duct larger than 3*mm* in the head and 2*mm* in the body and tail of the pancreas. In addition, enlargement upstream to a stenosis, reported to be a focal disappearing of MPD lumen, is also a pathologic pattern.

Given this context, Deep Learning (DL) methods could play an important role to assist with diagnosis, by issuing alerts for patients at risk of pancreatic cancer. Promising results have already been obtained for several diseases such as bone tumors [12], prostate cancer [13], kidney cancer [14]. In addition, DL models could help radiologists in their daily practice. Recently, an algorithm for breast cancer detection significantly decreased the false positive and false negative rates on two large datasets, while substantially reducing the radiologists' workload [15]. Other studies applied DL on Positron Emission Tomography (PET)/CT scans to detect lesions in the whole body, thus allowing to compute a global tumor burden potentially predictive of a patient response to treatment [16 17 18 19]. Efforts have also been directed towards pancreas cancer [20] with attempts to use DL for lesion detection [21 22]. These works proposed DL models to detect pancreatic neoplasms and validated them on independent databases of patients. While they showed promising results, these approaches focused on lesion detection and did not propose to detect key secondary findings such as the MPD dilatation.

In this work, a DL pipeline allowing to predict patients at risk of pancreas cancer is proposed and evaluated. The contribution is twofold: (i) a lesion segmentation algorithm is combined with information related to the lesion and MPD size, thus improving pancreatic neoplasm detection performance compared to current state-of-the-art DL models; (ii) the proposed method can also be used to predict MPD dilatation. To demonstrate the reliability of the approach the model is validated on an external test cohort.

2. Materials and Methods

2.1. Dataset

Data was collected from 9 medical centers in Europe, USA, and Brazil. Inclusion criteria were as follows: (i) presence of a portal venous CT scan; (ii) maximum slice thickness of 3*mm*; (iii) for patient with confirmed pancreatic neoplasm, studies were acquired prior to any treatment or surgery. This led to a total of 2890 cases that were further split in a training and independent testing set of 2134 and 756 subjects, respectively (see Figure 1).

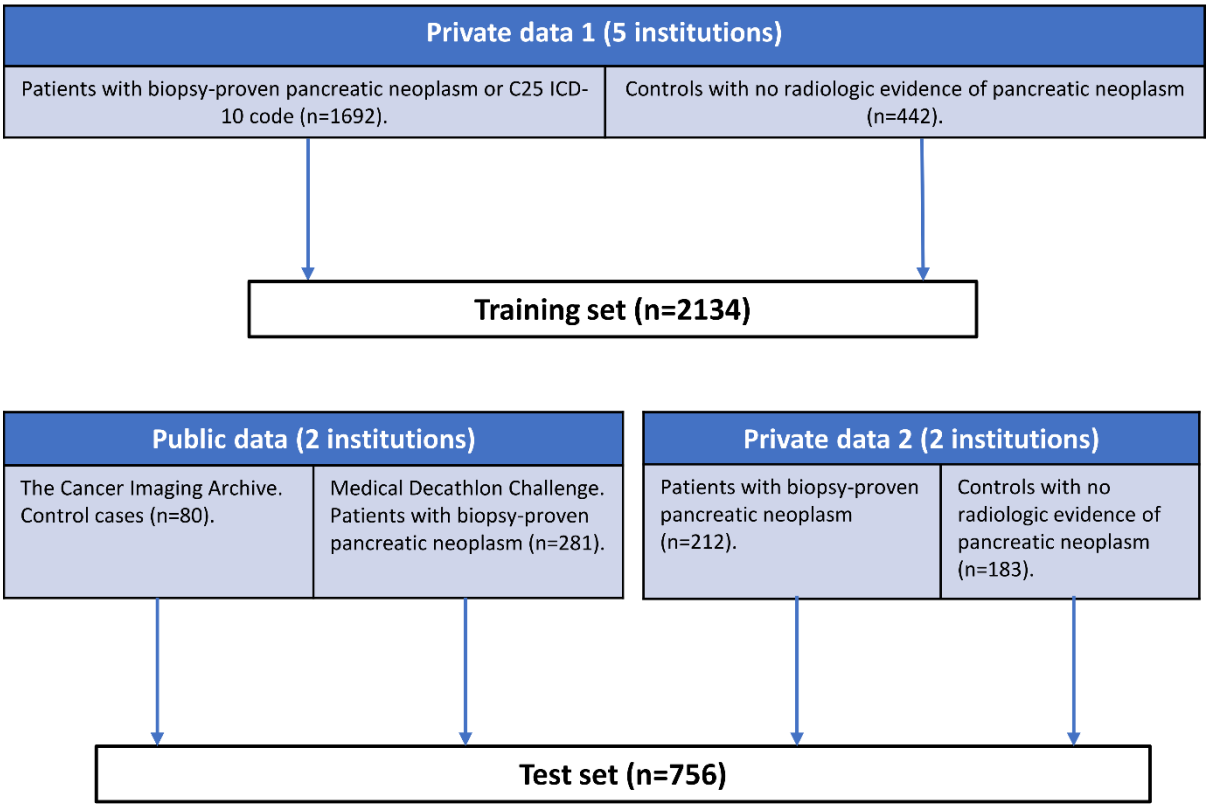

Figure 1: *Composition of the training and testing sets. Data in the training and testing sets come from different institutions. ICD: International Classification of Diseases.*

The training set was composed of portal venous CT scans of 2134 patients from 5 institutions, among which 1692 had a pancreatic neoplasm. Diagnosis was either obtained via the biopsy report for 78% of the subjects, or through the C25 code related to pancreatic neoplasms in the International Classification of Diseases (ICD-10) for the rest of the subjects. In addition, control portal CT scans were extracted from a private cohort of patients with bone lesions. The radiological reports of these patients were inspected to only keep subjects without abnormalities in the pancreas. Then, their portal CT scans were further reviewed to ensure that no pancreatic lesion was visible, thus leading to a total of 442 controls. Table 1 describes the patients characteristics. The 2134 subjects were composed of 1174 (55%) women and 960 men (45%) and showed a median age of 64 years old (range [56, 73] years). 1184 patients had Pancreatic Ductal Adenocarcinoma (PDAC), 134 had Neuroendocrine Tumor (NET) and 158 had unclassified solid lesion. There were also 81 subjects with Intraductal Papillary Mucinous Neoplasm (IPMN), 34 with Mucinous Cystic Neoplasm (MCN) and 42 with Serous Cystadenoma (SCA), as well as 59 subjects with unclassified cystic lesion. Finally, 43% (907) of the subjects had dilated MPD.

The external test set included 756 subjects collected from both public and private data in 4 institutions (cf. Figure 1). These institutions were different from the ones used in the training set. Public data contained 361 portal venous CT scans among which 281 had a pancreatic lesion and 80 were healthy

	Training set (n=2134)	Test set (n=756)
Age* (year)	64 [56, 73]	NA
Sex		NA
M	960 (45)	
F	1174 (55)	
CT manufacturer		NA
Philips	1113 (52)	
Siemens	168 (8)	
GE	309 (14)	
Toshiba	114 (5)	
Unknown	430 (21)	
Pancreatic lesions		
Yes	1692 (80)	493 (65)
No	442 (20)	263 (35)
Solid lesions	1476 (87)	421 (85)
PDAC	1184 (80)	360 (85)
NET	134 (9)	49 (12)
Unclassified	158 (11)	12 (3)
Cystic lesions	216 (13)	72 (15)
IPMN	81 (38)	18 (25)
MCN	34 (16)	0 (0)
SCN	42 (20)	1 (1)
Unclassified	59 (26)	53 (74)
Lesions size (cm)*	3.4 [2.3, 4.9]	2.4 [2.0, 3.2]
Main Pancreatic Duct		
Visible	1256 (59)	408 (54)
Dilated	907 (43)	276 (36)
Non dilated	1227 (57)	572 (64)
Diameter* (mm)	4.5 [3.5, 6.5]	6.0 [4.1, 8.8]

Table 1: *Demographic and clinical information for the different datasets. If not specified data are number of patients with percentages in parentheses. * Data are medians with interquartile range in square brackets. CT: Computed Tomography. GE: General Electric. PDAC: Pancreatic Ductal Adenocarcinoma. NET: Neuroendocrine Tumor. IPMN: Intraductal Papillary Mucinous Neoplasm. MCN: Mucinous Cystic Neoplasm. SCN: Serous Cystadenoma. NA: Not Available.*

cases. Cancer cases were retrieved from the Medical Decathlon Challenge [23]. Control subjects were accessed through the Cancer Imaging Archive pancreas CT dataset [24]. Only general information about this dataset was available: "Seventeen of the subjects are healthy kidney donors scanned prior to nephrectomy. The remaining 65 patients were selected by a radiologist from patients who neither had major abdominal pathologies nor pancreatic cancer lesions. Subjects' ages range from 18 to 76 years with a mean age of 46.8 ± 16.7" [24] [25]. The private dataset was composed of 212 portal venous CT scans of patients with histopathological confirmation of pancreatic neoplasm. In addition, routine portal CT scans from the gastroenterology department of a private institution were acquired. Scans were further reviewed by the radiologists pool to confirm that they did not have a pancreatic lesion, leading to a total of 183 control cases. Demographic information was not available for the subjects from the test set. This database was composed of patients with PDAC (n=360), NET (n=48) and unclassified solid lesions (n=12), as well as IPMN (n=18) and unclassified cystic lesions (n=53) (cf. Table 1). Finally, 276 subjects had dilated MPD.

2.2. Annotation Protocol

Each portal venous CT scan was reviewed and annotated by one in a group of 9 radiologists. The annotators segmented the pancreas on each image. When a lesion was identified by the radiologist it was systematically segmented. The radiologists characterized the tumor type based on the CT scan. Possible types were: PDAC, NET, IPMN, MCN and SCA. Cases for which the lesion type could not be determined were labeled as unclassified. In addition, lesion density was characterized with four possible answers: hyper-enhancing, hypo-enhancing, isodense, and unclassified. A lesion was considered isodense when its density was similar to the one of the adjacent pancreatic parenchyma and there was a secondary sign, such as abrupt cutoff of the MPD or the common bile duct, allowing to confirm lesion presence. In total, 2526 lesions were segmented. Lesion segmentation variability was investigated in Appendix 1. This analysis estimated that the average inter-annotator variability ranged from 0 to 0.12 Dice score points, which is a reasonable variability threshold for pancreatic lesions whose contours can be hard to define. Finally, the radiologists segmented the MPD when it was visible and visually assessed MPD dilatation. In practice, CT scans were annotated using a custom module of the 3D Slicer [26] [27] software that was developed internally. For each case, the ground truth 3D segmentation was generated by segmenting the pancreas, lesions, and MPD slice by slice. The customized version of 3D Slicer also allowed the radiologists to answer questions regarding lesions' characteristics and MPD dilatation.

2.3. Detection Pipeline

2.3.1 Segmentation

A 3D nnUNet [28] was trained to segment the pancreas, lesions and the MPD in a 5-fold cross-validation setting. The nnUNet is a self-configuring segmentation framework which automatically determines for a given dataset a suited image preprocessing pipeline, as well as an adapted UNet architecture. In this case, images were resampled to a target spacing of (0.79, 0.79, 2) *mm/voxel* using 3rd order bspline interpolation. Based on the training cases, a mean and standard deviation intensity of 68.7 Hounsfield Unit (HU) and 51.0 HU were computed and used to perform a z-normalization on the images. A batch size of 2 and a patch size of (192, 128, 56) along the x, y, and z axes, were selected. The loss function consisted in the sum of the Cross-entropy and Soft Dice loss functions. The model was optimized using Stochastic Gradient Descent with a learning rate of 0.01 and polynomial decay. Details on the heuristics allowing to infer the parameters values are provided in Appendix 2, as well as a figure of the resulting UNet architecture.

When applied on a new image, the trained nnUNet generated a segmentation of the pancreas, pancreatic lesion and MPD, as well as a lesion probability map assigning to each voxel a probability to be a lesion (cf. Figure 2).

2.3.2 Feature extraction

Using the output of the nnUNet, features were extracted as follows: (i) a lesion risk between 0 and 1, computed based on the 3D probability map given by the nnUNet. To do so, all the candidate lesions outside the pancreas were eliminated using the segmentation map. Then for each connected component a lesion risk between 0 and 1 was computed by averaging the probabilities of all its voxels. Connected components with a lesion risk lower than 0.05 were automatically removed; (ii) the maximum diameter of the lesion segmented by the nnUNet, computed as the maximum 2D Feret diameter [29] of the lesion segmented by the nnUNet. If no lesion was segmented it was set to 0; (iii) the maximum MPD diameter in the head, body, and tail of the pancreas. Computation was carried out using the IMEA [30] library. Details on how the head, body, and tail of the pancreas were identified are provided in Appendix 3. The maximum value of the MPD diameter in the pancreas was also computed by taking the maximum diameter between head, body, and tail. For each region the MPD diameter was set to 0 if there was no segmentation.

2.3.3 Logistic Regression

We proposed to predict lesion presence in a patient as well as MPD dilatation using two logistic regression models relying on the features previously defined. In practice, the nnUNet was applied on

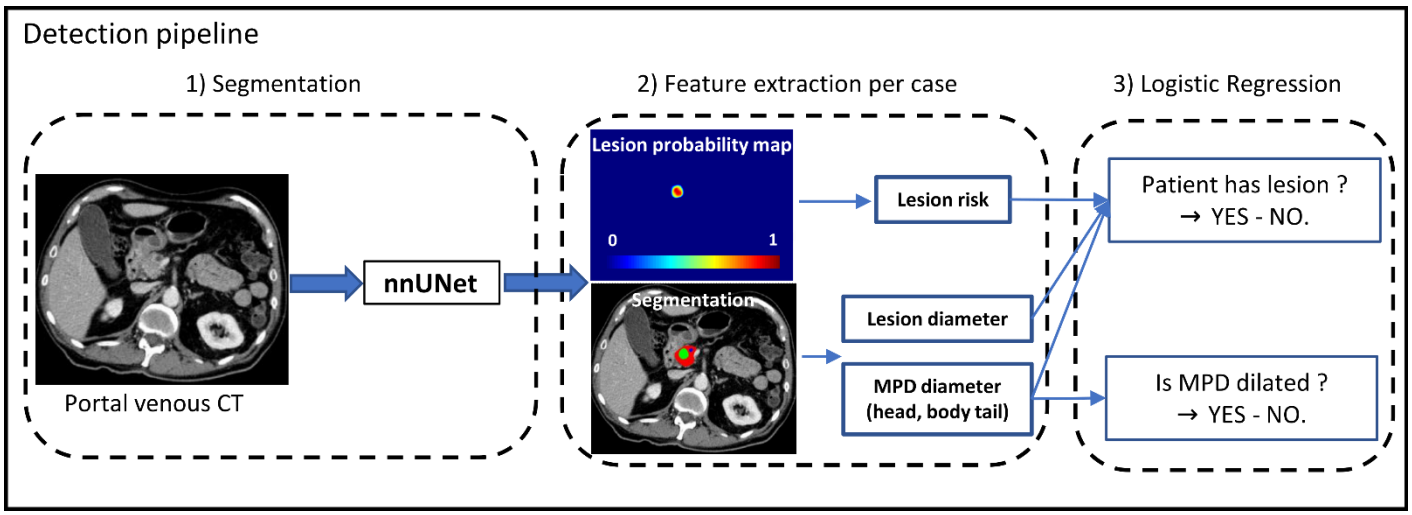

Figure 2: *The algorithm is composed of three steps: 1) Based on a portal venous CT scan, the nnUNet generates a lesion probability map as well as a segmentation of the pancreas (red), lesions (green) and MPD (blue). 2) Lesion and MPD features are extracted. 3) Lesion presence and MPD dilatation are predicted using two logistic regression models. MPD: Main Pancreatic Duct. CT: Computed Tomography.*

the validation set of each fold, leading to a total of 2134 3D probability maps and segmentations, from which features were extracted. Two logistic regression models were trained using the Scikit-Learn library [31]. To reduce the potential impact of class imbalance, the two logistic regression models were regularized by balancing the cost function with the inverse class frequencies. The first logistic regression predicted lesion presence based on three features: 1) lesion risk; 2) lesion diameter; 3) maximum MPD diameter in the pancreas. These features were extracted for all candidate lesions of each case. Candidate lesions matching a ground truth lesion (Dice score greater than 0.1) were considered as true positives, while candidate lesions that did not match a ground truth lesion were considered as false positives. Based on these target classes the logistic regression could be calibrated on the training set.

The second logistic regression model predicted MPD dilatation using the MPD diameter in the head, body, and tail of the pancreas.

Once the two logistic regressions were trained, evaluation of a test subject comprised three steps: 1) computing the lesion probability map as well as the segmentation of the pancreas, lesion and MPD using the nnUNet; 2) features extraction; 3) applying the two logistic regression models previously trained to predict lesion presence as well as MPD dilatation (see Figure 2).

2.4 Statistical Analysis

To evaluate performance, the receiver operating characteristic (ROC) curve was constructed by plotting sensitivity against false positive rate at different thresholds. The area under the curve (AUC) was measured, as well as sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) at the operating point maximizing the balanced accuracy. The evaluation was performed at the case-level for both lesion presence and MPD dilatation. For instance, in the case of lesion detection, sensitivity was defined as the ratio of the number of patients correctly detected with a lesion by the model to the total number of patients with a pancreatic lesion. Computation of the other metrics was defined accordingly. Details on how each metric was computed are provided in Appendix 4 and 5.

Bootstrap sampling was used to provide median values and 95% confidence intervals (CI) for AUC, sensitivity, specificity, PPV and NPV.

Finally, segmentation performance was evaluated by calculating the Dice and the Normalized Surface Dice (NSD) scores, which measure the overlap between the reference and predicted segmentation for each test patient. The NSD allows a tolerance error between the reference and predicted segmentation [32]. This metric is well-suited to evaluate structures such as pancreatic lesions and the MPD which can be hard to accurately delineate. In this study, the tolerance error was set to 2*mm* along each spatial dimension. To distinguish between segmentation error types, voxel-level False Positive Rate (FPR) and False Negative Rate (FNR) were computed. The FPR was defined for each class as the number of wrongly segmented voxels to the total number of segmented voxels. The FNR was computed accordingly.

3. Results

3.1. Detecting Patients with Pancreatic Neoplasms

The model performance on the test set is reported in Figure 3 and Table 2. The model reached an AUC of 0.98 (95% CI: 0.97, 0.99) as well as a sensitivity of 0.94 (469 of 493, 95% CI: 0.92, 0.97) and a specificity of 0.95 (246 of 262, 95% CI: 0.92, 0.98). A boxplot showing the evaluation metrics distributions is provided in Appendix 6.

In order to compare with other methods, the model performance on public data (defined in Figure 1) is also reported in Table 2. Evaluation metrics remained similar to the ones obtained on the whole test set.

The model was also evaluated on subjects with specific lesions characteristics and types. Five subgroups of the cohort were further investigated: (i) patients with lesions less than 2*cm* in diameter; (ii) patients with isodense lesions; (iii) patients with PDAC; (iv) patients with NET; (v) patients with IPMN. Performances obtained on these subsets are reported in Table 2. The AUC, sensitivity, and specificity remained consistent across the subgroups and equivalent to the ones obtained on the whole test set. The model performed best on the NET subgroup with a sensitivity of 1.0 (95% CI: 0.98, 1.0). Specificity was slightly lower on small lesions compared to the other subsets.

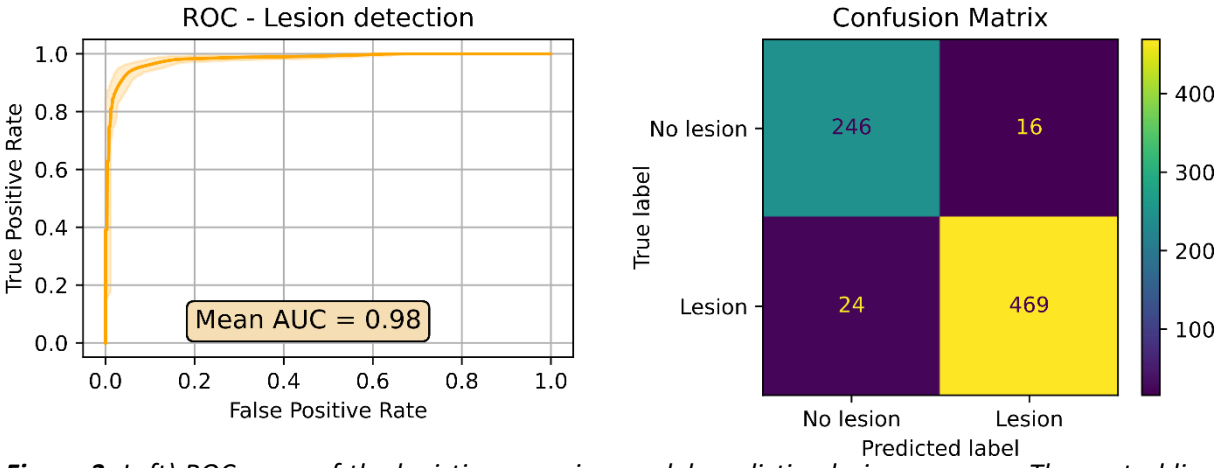

Figure 3: *Left) ROC curve of the logistic regression model predicting lesion presence. The central line represents the average curve and shadowed areas the 95% confidence interval. Right) Confusion matrix of the model obtained at the operating point maximizing the balanced accuracy. ROC: Receiver Operating Characteristic. AUC: Area Under the Curve.*

	AUC	Sensitivity	Specificity	PPV	NPV
Test set (493/756)	0.98 [0.97, 0.99]	0.94 [0.92, 0.97]	0.95 [0.92, 0.98]	0.97 [0.96, 0.99]	0.90 [0.85, 0.94]
Public data (281/361)	0.99 [0.98, 0.99]	0.94 [0.88, 0.98]	0.95 [0.90, 1.0]	0.99 [0.97, 1.0]	0.82 [0.67, 0.94]
Lesions characteristics					
D ≤2*cm* (123/386)	0.97 [0.95, 0.98]	0.94 [0.87, 0.98]	0.91 [0.85, 0.96]	0.83 [0.74, 0.92]	0.97 [0.94, 0.99]
Isodense (56/319)	0.97 [0.95, 0.99]	0.95 [0.87, 1.0]	0.94 [0.85, 0.98]	0.78 [0.57, 0.90]	0.99 [0.97, 1.0]
Lesions type					
PDAC (418/681)	0.98 [0.98, 0.99]	0.94 [0.91, 0.97]	0.96 [0.92, 0.98]	0.97 [0.95, 0.99]	0.92 [0.88, 0.96]
NET (49/311)	0.99 [0.98, 1.0]	1.0 [0.98, 1.0]	0.96 [0.92, 0.98]	0.81 [0.69, 0.92]	1.0 [1.0, 1.0]
IPMN (18/290)	0.98 [0.95, 0.99]	0.96 [0.87, 1.0]	0.94 [0.91, 0.98]	0.56 [0.29, 0.8]	1.0 [0.99, 1.0]

Table 2: *Evaluation metrics obtained for lesion detection on the test set as well as on specific subsets of patients. In parenthesis for each group, number of patients with pancreatic neoplasm to the total number of patients. Data are median values with 95% confidence interval in square brackets. AUC: Area*

Under the Curve. PPV: Positive Predictive Value. NPV: Negative Predictive Value. D: Diameter. PDAC: Pancreatic Ductal Adenocarcinoma. NET: Neuroendocrine Tumor. IPMN: Intraductal Papillary Mucinous Neoplasm.

3.2. Feature Importance on Lesion Detection Sensitivity

The logistic regression model used to predict lesion presence in a patient was based on three features: lesion risk, lesion diameter and MPD diameter (cf. Section 2.3.2). To evaluate the effect of combining these features on performance, an ablation study was carried out. Two additional logistic regression models were trained: a first one with two features, lesion risk and lesion diameter; and a second one using only lesion risk. Table 3 reports the sensitivity of these three models on the test set, as well as on the subsets previously defined in Section 3.1.

Using the MPD diameter and the lesion diameter systematically improved lesion detection sensitivity across all groups. On the whole test set, adding the MPD diameter and lesion diameter led to a sensitivity improvement of 4% compared to the baseline model using the lesion risk only. The effect of these two features was particularly strong on isodense lesions with a sensitivity gain of 10% with respect to the model with one feature only.

	Lesion risk, lesion diameter, MPD diameter	Lesion risk, lesion diameter	Lesion risk
Test set (493/756)	0.94 [0.92, 0.97]	0.91 [0.86, 0.95]	0.90 [0.84, 0.94]
Public data (281/361)	0.94 [0.88, 0.98]	0.92 [0.86, 0.95]	0.91 [0.86, 0.95]
Lesions characteristics			
D ≤2*cm* (123/386)	0.94 [0.87, 0.98]	0.91 [0.80, 0.97]	0.87 [0.77, 0.95]
Isodense (56/319)	0.95 [0.87, 1.0]	0.90 [0.91, 0.98]	0.85 [0.75, 0.96]
Lesions type			
PDAC (418/681)	0.94 [0.91, 0.97]	0.91 [0.88, 0.95]	0.89 [0.85, 0.94]
NET (49/311)	1.0 [0.98, 1.0]	0.98 [0.93, 1.0]	0.98 [0.93, 1.0]
IPMN (18/290)	0.96 [0.87, 1.0]	0.95 [0.83, 1.0]	0.90 [0.75, 1.0]

Table 3: *Sensitivity of the logistic regression predicting lesion presence depending on the features used to train it. In parenthesis for each group, number of patients with pancreatic neoplasm to the total number of patients Median values and 95% confidence intervals in square brackets. D: Diameter. PDAC: Pancreatic Ductal Adenocarcinoma. NET: Neuroendocrine Tumor. IPMN: Intraductal Papillary Mucinous Neoplasm.*

3.3. MPD Dilatation Detection Performance

Regarding MPD dilatation performance, the model of Section 2.3.3 was also evaluated on the test set. Results are reported in Table 4. An AUC of 0.97 (95% CI: 0.96, 0.98) was reached as well as a sensitivity of 0.94 (259 of 276, 95%: CI: 0.89, 0.97) and a specificity of 0.90 (432 of 480, 95%CI: 0.86, 0.94). A boxplot showing the evaluation metrics distributions is provided in Appendix 6.

	AUC	Sensitivity	Specificity	PPV	NPV
Test set (n=756)	0.97 [0.96, 0.98]	0.94 [0.89, 0.97]	0.90 [0.86, 0.94]	0.85 [0.79, 0.90]	0.96 [0.93, 0.98]

Table 4: *Evaluation metrics obtained by the logistic regression predicting MPD dilatation. Data are median values with 95% confidence interval in square brackets. MPD: Main Pancreatic Duct. AUC: Area Under the Curve. PPV: Positive Predictive Value. NPV: Negative Predictive Value.*

3.4. Segmentation Performance

Segmentations predicted by the nnUNet of Section 2.3.1 were evaluated both quantitatively and qualitatively. The Dice score between the ground truth and the nnUNet segmentation maps is reported in Table 5 for the pancreas, lesions, and the MPD. The FPR and FNR were also computed for the three structures. The NSD score was computed for lesions and the MPD. These metrics were also measured on public data only in order to compare our segmentation network to other works. Qualitative examples of the model segmentations are provided in Figure 4. Additional cases are presented in Appendix 7.

On the whole test set, the mean Dice score was 0.91 (± 0.06), 0.69 (± 0.34) and 0.58 (± 0.37) for the pancreas, lesions, and MPD, respectively. The mean NSD score was 0.71 (± 0.39) and 0.77 (± 0.33) for the MPD and lesions, respectively. While the FPR and FNR were similar for the pancreas on both the whole test set and public data, the FNR was systematically higher than the FPR for lesions and the MPD. A boxplot illustrating segmentation performance over the complete cohort is provided in Appendix 8.

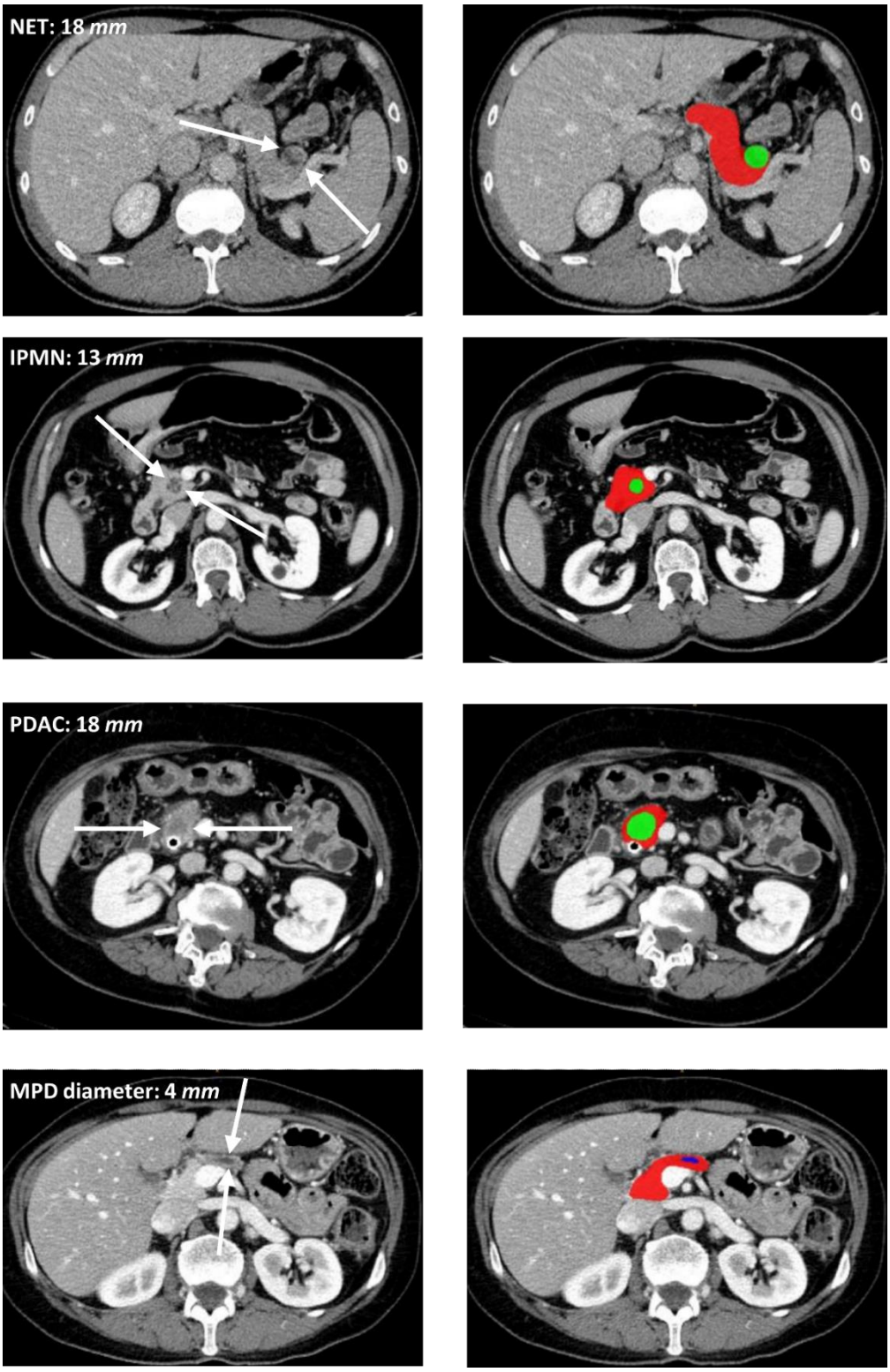

Figure 4: *Left column: Axial portal venous CT slices of four patients. White arrows indicate lesions location for the first three rows and the MPD for the last row. Right column: Segmentation of pancreas (red), lesions (green) and MPD (blue) output by our model. PDAC: Pancreatic Ductal Adenocarcinoma.*

IPMN: Intraductal Papillary Mucinous Neoplasm. NET: Neuroendocrine Tumor. MPD: Main Pancreatic Duct.

	Test set (n=756)			Public data (n=361)		
	Pancreas	Lesions	Main Pancreatic Duct	Pancreas	Lesions	Main Pancreatic Duct
Dice	0.91 ± 0.06	0.69 ± 0.34	0.58 ± 0.37	0.87 ± 0.06	0.63 ± 0.33	0.53 ± 0.37
NSD	NA	0.77 ± 0.33	0.71 ± 0.39	NA	0.73 ± 0.32	0.65 ± 0.40
FPR	0.12 ± 0.09	0.20 ± 0.31	0.32 ± 0.38	0.15 ± 0.10	0.23 ± 0.31	0.37 ± 0.40
FNR	0.09 ±0.09	0.39 ± 0.33	0.49 ± 0.20	0.12 ± 0.10	0.45 ± 0.31	0.51 ± 0.20

Table 5: *Dice score, NSD score, FPR, and FNR measured on the whole test set and public data. The Dice score, the FPR, and the FNR were computed for the pancreas, lesions, and the main pancreatic duct. The NSD score was measured for lesion and the main pancreatic duct only. Data are mean values plus/minus standard deviation. NSD: Normalized Surface Dice. FPR: False Positive Rate. FNR: False Negative Rate.*

4. Discussion

In this study, we presented a method to automatically detect patients with pancreatic neoplasms and to identify cases with MPD dilatation. The proposed approach was validated on an independent cohort of 756 subjects. We showed how using the MPD dilatation information could improve sensitivity for lesion detection compared to a baseline approach solely relying on a segmentation network output. Finally, we evaluated the ability of our model to correctly localize the pancreas, lesions, and the MPD, by assessing its segmentation performance.

The model was assessed on subgroups of patients based on lesion characteristics and types. Similar AUC, sensitivity and specificity were observed across the different subgroups, thus highlighting the robustness of our approach. A greater variability was observed in the case of PPV and NPV, mostly due to the imbalance between healthy and diseased subjects in some subgroups such as NET (49 of 312) and IPMN (18 of 290). More generally, the test set was mostly composed of subjects with pancreatic lesions (65% of the cases). Given this distribution, a model that tends to over-detect lesions could be favored compared to an unbiased model. However, it can be noticed in Table 2 that class imbalances change depending on the sub-analysis. In the isodense subgroup only 18% of cases had a pancreatic lesion while the rest were controls. Conversely, the public data subgroup was mostly composed of pathological cases (78%). Yet, a median AUC of 0.97 and 0.99, a median sensitivity of 0.95 and 0.94, and a median specificity of 0.94 and 0.95 were reported for the isodense and public data subgroups,

respectively. This result shows that on two subsets with radically different class distribution, the lesion detection model led to comparable results, highlighting its robustness to class imbalance.

The results on public data can be compared to competing DL models that used this database to test their approach. In [33] the authors reported an AUC of 0.91 and in [34] they obtained an AUC of 0.92 (95% CI: 0.89, 0.95). Our algorithm showed notably higher performance with an AUC of 0.99 (95% CI: 0.98, 0.99). This might be explained by the size of our training database (more than 2000 cases), which is larger than what competing DL approaches reported so far [21] [22], thus allowing better generalization on independent cohorts. Moreover, generalization was also helped by the diversity of the training data which was composed of various types of tumors such as PDAC, NET, or IPMN.

However, improved performance can also be attributed to the combination of specific features to predict lesion presence. While recent approaches tend to either rely on a pure DL segmentation model or combined segmentation-classification networks, we used the lesion risk, the MPD diameter, and lesion diameter, to predict lesion presence via a separate logistic regression. The advantage of this approach compared to end-to-end DL methods is twofold: (i) it does not require to modify the UNet architecture by incorporating a classification module which may overfit; (ii) it relies on both a small number of interpretable features and an explainable classification method, thus better allowing to understand the model predictions. To evaluate the effect of combining features, the sensitivity of the logistic regression was measured according to the features used to train it. The logistic regression model using only the lesion risk, the one most closely resembling state-of-the-art approaches, had a lower sensitivity on all subsets compared to the models using two or three features. In particular, using the MPD diameter systematically improved sensitivity, especially for isodense lesions (plus 5%). This is probably due to the fact that MPD dilatation is often observed on CT scans during early stages of pancreatic cancer, thus helping to identify iso-attenuating tumors [35]. Improvement of sensitivity for PDAC detection (plus 5%) when using the lesion and MPD diameter was also observed. Regarding pancreas NET, a milder effect was observed (plus 2% sensitivity). However, as NETs are not linked to MPD dilatation, effect on sensitivity in patient with NET was not anticipated. Finally, even though adding MPD diameter and lesion diameter strongly improved sensitivity in the case of IPMN (plus 6%), the low number of cases (18 of 290) have prevented from drawing any conclusion.

The segmentation network was also leveraged to design a logistic regression model predicting MPD dilatation based on its diameter in the head, body, and tail of the pancreas (see Section 2.3.3). An AUC of 0.97 (95% CI: 0.96, 0.98) was reported. We highlight that even if other DL methods allow to segment the MPD [36] [37], none of them took advantage of the MPD segmentation to provide an alert on its potential dilatation, which is a key finding for radiologists when assessing the pancreas.

Segmentation performance of the algorithm on the pancreas, lesions, and MPD was also assessed. Regarding the pancreas, there were no competing DL methods which evaluated the Dice score on the same datasets as in this study. However, three DL models which reported a mean Dice score of 0.87 on their test set were found in [38] [39] [40]. The segmentation network presented in this work obtained a similar result on public data and a higher mean Dice score on the whole test set. As for lesion segmentation, a mean Dice score of 0.63 was obtained on public data. This is a 9% improvement over the nnUNet trained in [28] which reported a Dice score of 0.54 on this dataset. Finally, we could not find Dice scores obtained by other DL approaches for the MPD. However, given the small size of the MPD and compared to the Dice score obtained on lesions, the model seemed to show satisfactory performance corroborated by an NSD score of 0.71 on the test set. While higher than in published work, mean Dice score values remained below 0.7 for lesions and the MPD. The reason for this is twofold: (i) for false negative and false positive cases, the associated Dice score will be 0, thus strongly deteriorating the mean value. We show in Appendix 8 that using the median instead of the mean allows to be more robust to these extreme cases. A median Dice score of 0.83 and 0.69 were reported for lesions and the MPD, respectively; (ii) Lesions and the MPD are difficult to segment, with potentially few voxels, and contours hard to define. The Dice score strongly penalizes segmentation errors in this scenario. Using the NSD better highlights the predicted segmentation quality for these structures. Compared to the Dice score, a gain of 0.08 and 0.13 points were observed for the NSD in the case of lesions and the MPD, respectively.

There were several limitations to this study. First, tumor type was visually assessed by the radiologists, which, due to some cases being indeterminate, led to a significant number of unclassified lesion types: 218 in the training set and 65 in the test set. These cases were not included for the subgroup analysis in Tables 2 and 3. Second, the number of subjects with specific lesions characteristics or tumor types was rather limited for some subgroups, especially in the case of isodense lesions (56 of 319), patients with NETs (49 of 312), and patients with IPMN (18 out of 290). Future work should focus on creating more diverse test cohorts to evaluate our algorithm. Third, MPD dilatation was assessed visually and not based on a quantitative measure of the MPD diameter.

In conclusion, the presented DL approach showed promising results to detect subjects with pancreatic neoplasm and to identify cases with MPD dilatation. In particular, the combination of the output of a segmentation network with secondary features increased lesion detection performance. Future efforts should focus on acquiring additional data with greater diversity of lesions characteristics to increase the model robustness. The model should also be further validated on a test cohort more representative of typical missed cases (small and isodense lesions). A different study design could also be proposed, by applying the model on a longitudinal cohort of patients which were eventually diagnosed with

pancreatic neoplasm. This type of study would be of great interest to assess the ability of the model to detect early-stage pancreas cancer. Such an evaluation could be conducted via a retrospective study in collaboration with clinical institutions having sufficiently large volume of patients.

Currently, the model can be used as a pipeline taking a portal CT scan as input and generating a segmentation of the pancreas, lesion if detected, and the MPD if visible. It also generates an alert stating if a lesion is present and if the MPD is dilated. However, this pipeline has not been industrialized yet and can only be used in a research environment. Therefore, a software development effort needs to be undertaken to reach clinical deployment. Ultimately, we aim to provide a diagnostic tool assisting radiologists in routine clinical practice by issuing an alert when a pancreatic lesion is suspected by the model, with an emphasis on detection of early-stage tumors to improve long-term outcome for patients with pancreatic cancer.

References

1. Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: Global trends, etiology and risk factors. *World J. Oncol.* February 2019;10:10–27.

2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. February 2021;71:209–249.

3. Elbanna KY, Jang HJ, Kim TK. Imaging diagnosis and staging of pancreatic ductal adenocarcinoma: a comprehensive review. *Insights into Imaging*. April 2020;11:58.

4. COSTACHE MI, COSTACHE CORNELIAALEXANDRA. Which is the Best Imaging Method in Pancreatic Adenocarcinoma Diagnosis and Staging. *Current Health Sciences Journal*. July 2017:132–136.

5. Raman SP, Reddy S, Weiss MJ, et al. Impact of the time interval between MDCT imaging and surgery on the accuracy of identifying metastatic disease in patients with pancreatic cancer. *AJR Am. J. Roentgenol.* January 2015;204:W37–42.

6. Lucas AL, Kastrinos F. Screening for Pancreatic Cancer. *JAMA*. August 2019;322:407-408.

7. Blackford AL, Canto MI, Klein AP, Hruban RH, Goggins M. Recent trends in the incidence and survival of stage 1A pancreatic cancer: A surveillance, Epidemiology, and End Results analysis. *J. Natl. Cancer Inst.* November 2020;112:1162–1169.

8. Gangi S, Fletcher JG, Nathan MA, et al. Time interval between abnormalities seen on CT and the clinical diagnosis of pancreatic cancer: retrospective review of CT scans obtained before diagnosis. *AJR Am. J. Roentgenol.* April 2004;182:897–903.

9. Gonoi W, Hayashi TY, Okuma H, et al. Development of pancreatic cancer is predictable well in advance using contrast-enhanced CT: a case-cohort study. *Eur. Radiol.* December 2017;27:4941–4950.

10. Toshima F, Watanabe R, Inoue D, et al. CT abnormalities of the pancreas associated with the subsequent diagnosis of clinical stage I pancreatic ductal adenocarcinoma more than 1 year later: A case-control study. *AJR Am. J. Roentgenol.* December 2021;217:1353–1364.

11. Hoogenboom SA, Engels MML, Chuprin AV, et al. Prevalence, features, and explanations of missed and misinterpreted pancreatic cancer on imaging: a matched case-control study. *Abdom. Radiol. (NY)*. December 2022;47:4160–4172.

12. Fritz B, Yi PH, Kijowski R, Fritz J. Radiomics and Deep Learning for Disease Detection in Musculoskeletal Radiology. *Investigative Radiology*. September 2022;58:3–13.

13. Netzer N, Weißer C, Schelb P, et al. Fully Automatic Deep Learning in Bi-institutional Prostate Magnetic Resonance Imaging. *Investigative Radiology*. June 2021;56:799–808.

14. Toda N, Hashimoto M, Arita Y, et al. Deep Learning Algorithm for Fully Automated Detection of Small (≤4 cm) Renal Cell Carcinoma in Contrast-Enhanced Computed Tomography Using a Multicenter Database. *Investigative Radiology*. December 2021;57:327–333.

15. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. January 2020;577:89-94.

16. Bi L, Fulham M, Li N, et al. Recurrent feature fusion learning for multi-modality pet-ct tumor segmentation. *Computer Methods and Programs in Biomedicine*. 2021;203:106043.

17. Capobianco N, Meignan M, Cottereau AS, et al. Deep-learning 18F-FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-cell lymphoma. *J. Nucl. Med.* January 2021;62:30–36.

18. Gatidis S, Hepp T, Früh M, et al. A whole-body FDG-PET/CT Dataset with manually annotated Tumor Lesions. *Sci. Data*. October 2022;9:601.

19. Jemaa S, Fredrickson J, Carano RAD, Nielsen T, de Crespigny A, Bengtsson T. Tumor segmentation and feature extraction from whole-body FDG-PET/CT using cascaded 2D and 3D convolutional neural networks. *J. Digit. Imaging*. August 2020;33:888–894.

20. Kenner B, Chari ST, Kelsen D, et al. Artificial Intelligence and Early Detection of Pancreatic Cancer: 2020 Summative Review. *Pancreas*. March 2021;50:251–279.

21. Chen PT, Wu T, Wang P, et al. Pancreatic Cancer Detection on CT Scans with Deep Learning: A Nationwide Population-based Study. *Radiology*. January 2023;306:172–182.

22. Park HJ, Shin K, You MW, et al. Deep Learning–based Detection of Solid and Cystic Pancreatic Neoplasms at Contrast-enhanced CT. *Radiology*. January 2023;306:140–149.

23. Simpson AL, Antonelli M, Bakas S, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *CoRR*. 2019;abs/1902.09063.

24. Roth HR, Lu L, Farag A, et al. DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation. Paper presented at: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5-9, 2015, Proceedings, Part I, 2015.

25. Roth H, Farag A, Turkbey EB, Lu L, Liu J, Summers RM. Data From Pancreas-CT

26. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging*. November 2012;30:1323–1341.

27. Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In: Jolesz FA, ed. *Intraoperative Imaging and Image-Guided Therapy*. New: Springer New York; 2014.

28. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*. February 2021;18:203-211.

29. Van der Walt S, Schönberger JL, Nunez-Iglesias J, et al. scikit-image: image processing in Python. *PeerJ*. 2014;2:e453.

30. Kroell N. imea: A Python package for extracting 2D and 3D shape measurements from images. *Journal of Open Source Software*. 2021;6:3091.

31. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. 2012.

32. Ostmeier S, Axelrod B, Bertels J, et al. Evaluation of Medical Image Segmentation Models for Uncertain, Small or Empty Reference Annotations. *ArXiv*. 2022;abs/2209.13008.

33. Alves N, Schuurmans M, Litjens G, Bosma JS, Hermans J, Huisman H. Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography. *Cancers*. 2022;14.

34. Liu KL, Wu T, Chen PT, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. *Lancet Digit. Health*. June 2020;2:e303–e313.

35. Yoon SH, Lee JM, Cho JY, et al. Small (< 20 mm) Pancreatic Adenocarcinomas: Analysis of Enhancement Patterns and Secondary Signs with Multiphasic Multidetector CT. *Radiology*. 2011;259:442-452.

36. Xia Y, Yu Q, Chu L, et al. The FELIX Project: Deep Networks To Detect Pancreatic Neoplasms. Paper presented at: medRxiv, 2022.

37. Viviers CGA, Ramaekers M, de With PHN, et al. Improved Pancreatic Tumor Detection by Utilizing Clinically-Relevant Secondary Features. Paper presented at: CaPTion@MICCAI, 2022.

38. Zhou Y, Li Y, Zhang Z, et al. Hyper-Pairing Network for Multi-Phase Pancreatic Ductal Adenocarcinoma Segmentation, 2019; Berlin.

39. Zhu Z, Xia Y, Xie L, Fishman EK, Yuille AL. Multi-scale Coarse-to-Fine Segmentation for Screening Pancreatic Ductal Adenocarcinoma. Paper presented at: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 2019; Cham.

40. Zhu Z, Lu Y, Shen W, Fishman EK, Yuille AL. Segmentation for Classification of Screening Pancreatic Neuroendocrine Tumors. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 2020:3395-3401.

41. Yu J, Blackford AL, Dal Molin M, Wolfgang CL, Goggins M. Time to progression of pancreatic ductal adenocarcinoma from low-to-high tumour stages. *Gut*. November 2015;64:1783–1789.

42. van Erning FN, Mackay TM, van der Geest LGM, et al. Association of the location of pancreatic ductal adenocarcinoma (head, body, tail) with tumor stage, treatment, and survival: a population-based analysis. *Acta Oncol.* December 2018;57:1655–1662.

43. Kang JD, Clarke SE, Costa AF. Factors associated with missed and misinterpreted cases of pancreatic ductal adenocarcinoma. *Eur. Radiol.* April 2021;31:2422–2432.

44. Ahualli J. The Double Duct Sign. *Radiology*. 2007;244:314-315.