Artificial Intelligence Methods to Assist the Diagnosis of Pancreatic Diseases in Radiology

Rebeca Vétil

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2024IPPAT014

Thèse de doctorat

Artificial Intelligence Methods to Assist the Diagnosis of Pancreatic Diseases in Radiology

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom Paris

École doctorale n°626 Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Informatique

Thèse présentée et soutenue à Paris, le 23 avril 2024, par

Rebeca Vétil

Composition du Jury :

Nicholas Ayache Directeur de Recherche, INRIA (EPIONE)	Président/Examinateur
Maria A. Zuluaga Maîtresse de Conférences, EURECOM	Rapportrice
Olivier Bernard Professeur, INSA (CREATIS)	Rapporteur
Loïc Boussel Professeur des Universités - Praticien Hospitalier Université Lyon 1, Hospices Civils de Lyon, INSA (CREATIS)	Examinateur
Isabelle Bloch Professeure, Sorbonne Université CNRS (LIP6) Télécom Paris (LTCI)	Directrice de thèse
Pietro Gori Maître de Conférences, Télécom Paris (LTCI)	Co-directeur de thèse
Marc-Michel Rohé Docteur, Guerbet	Co-encadrant de thèse

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude envers les membres du jury pour avoir accepté d'évaluer mon travail. Je remercie sincèrement Maria A. Zuluaga, Oliver Bernard, Nicholas Ayache et Loïc Boussel pour leur présence. C'était un véritable honneur pour moi de vous exposer mes recherches. Je tiens particulièrement à remercier Maria A. Zuluaga et Oliver Bernard pour avoir pris le temps de lire ce manuscrit et pour avoir partagé leurs précieux commentaires et conseils.

Je voudrais remercier mes encadrants de thèse : Isabelle, Pietro et Marc-Michel. Travailler sous votre direction ces dernières années a été une chance inestimable pour moi. Isabelle et Pietro, je suis extrêmement admirative de votre implication auprès de vos élèves, de vos connaissances scientifiques et surtout de votre bienveillance sans égal. Marc-Michel, un immense merci pour ton regard scientifique toujours très pertinent, pour la grande confiance que tu m'as accordée en me mettant sur ce projet en Juin 2020, et pour la confiance que tu m'accordes encore en me permettant de poursuivre dans l'équipe. Je tiens également à remercier Marie Pierre Vullierme pour m'avoir apporté de précieuses connaissances médicales et pour nos discussions scientifiques passionnantes.

J'aimerais remercier très sincèrement mes deux collègues Clément et Alexandre. Votre aide a été plus que précieuse durant ces dernières années. Alexandre, merci pour ton optimisme sans faille, ton pragmatisme mais aussi et surtout ta grande bienveillance. Tes milles et une idées ont été très précieuses et m'ont remotivée plus d'une fois. Clément, un très grand merci de m'avoir accompagnée au quotidien sur cette thèse grâce à tous tes conseils scientifiques mais aussi ton humour décapant. Merci de toujours avoir essayé de comprendre les résultats de mes 2000 expériences, de t'être plongé avec moi dans les méandres du deep learning plus d'une fois, et également merci pour ton immense aide lors de l'écriture du manuscrit. Encore un sincère merci à vous deux pour avoir rendu cette thèse plus facile.

Je voudrais ensuite remercier mes très chers collègues de Guerbet, et notamment l'équipe IA. Vous avez tous été de précieux piliers lors de ces dernières années. Un merci tout particulier à ma petite sœur de thèse, Emma, et mon grand frère de thèse, Omar, pour leur soutien quotidien. Merci Emma pour m'avoir encouragée et surtout fait rire matin, midi et soir. Merci Omar pour avoir été mon compagnon de galère lors de ces rushs avant les deadlines ISBI et MICCAI. Merci aux deux autres merveilleux membres de la Guerbet Sport Club, Noëlie et Alex, pour toutes ces séances de sport, mais aussi et surtout d'avoir été de très bons amis. Merci à Arnaud pour nos milles et une discussions de bobos parisiens, merci à Sisi pour tes conseils auxquels j'ai fait appel plus d'une fois, et merci à Littisha pour ta gentillesse sans pareil. Plus généralement, merci à tous mes collègues de Guerbet pour tous ces moments du quotidien partagés ensemble depuis bientôt 4 ans, et pour toutes ces discussions sur le très beau canapé doré du 104 boulevard de Sébastopol.

Je voudrais désormais remercier mes collègues et amis de l'équipe IMAGES de Télécom Paris, et tout particulièrement Chloé, Erwan, Antoine, Matthis, Nicolas, Pierrick, Giammarco, Mateus, Raphaël R., Inès, Carlo Alberto, Marie, Zoé, Gwillherm, Camille, Raphaël A., Cristiano, Ali K., et Ali M. Bien que le trajet jusqu'à Palaiseau m'ait démotivée plus d'une fois, je suis très heureuse d'avoir réussi à tisser des liens d'amitié avec vous. Je garderai de très bons souvenirs de ces moments passés ensemble, que ce soit lors de l'Ekiden, au Parc Asterix, à la Butte aux Cailles ou bien lors de ces soirées aux Divans du Monde. Je suis très heureuse d'avoir partagé ces quatre années avec vous, et j'étais très fière de vous accompagner lors de vos soutenances de thèses respectives (et bien sûr, j'ai hâte d'être là pour les soutenances qui restent à venir).

Merci à mes proches pour leur participation à cette thèse, de près ou de loin. En particulier, Anaïs et Olivia, un immense merci pour votre soutien indéfectible ces dernières années. Pedro, muito obrigada pelo carinho e pela luz de sempre. Merci à Chloé pour la force et pour ton amitié plus que précieuse. Enfin, merci à Amine de m'avoir soutenue et accompagnée dans cette aventure riche en enseignements tant sur le plan professionnel que personnel. Je voudrais également remercier ma famille, et tout particulièrement ma maman y mi abuelita Libertad. Merci à mes frères et sœurs, Virginie, Laurent et Eric.

Enfin, je dédie cette thèse à mon merveilleux papa, Michel Vétil, malheureusement parti avant le début de cette thèse, mais dont le goût de l'effort, l'envie d'apprendre et l'infinie générosité me guident chaque jour.

List of Abbreviations

AD	Anomaly Detection
AE	Autoencoder
AI	Artificial Intelligence
ASM	Active Shape Models
AUC	Area Under the Curve
BMI	Body Mass Index
C2F	Coarse-to-Fine
CA	Carbohydrate Antigen
CI	Confidence Interval
CNN	Convolutional Neural Network
CT	Computed Tomography
DL	Deep Learning
DLR	Deep-Learning Radiomics
DNN	Deep Neural Networks
DSC	Dice Similarity Coefficient
ERCP	Endoscopic Retrograde Cholangiopancreatography
EUS	Endoscopic Ultrasound
FCN	Fully Convolutional Network
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
GAN	Generative Adversarial Model
HCR	Hand-Crafted Radiomics
HU	Hounsfield Unit
IPMN	Intraductal Papillary Mucinous Neoplasm
KL	Kullback-Leibler
LDDMM	Large Deformation Diffeomorphic Metric Mapping
MCN	Mucinous Cystic Neoplasm
MI	Mutual Information
ML	Machine Learning
MPD	Main Pancreatic Duct
MRCP	Magnetic Cholangiopancreatography
MRI	Magnetic Resonance Imaging
MSD	Medical Segmentation Decathlon
NN	Neural Network
NPV	Negative Predictive Value
NSD	Normalized Surface Dice
PDAC	Pancreatic Ductal Adenocarcinoma
PNET	Pancreatic Neuroendocrine Tumor

PPV	Positive Predictive Value
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SCN	Serous Cystic Neoplasm
SPN	Solid Pseudo-Papillary Neoplasm
SSM	Statistical Shape Model
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
US	Ultrasound
VAE	Variational Autoencoder

Résumé

Le cancer du pancréas est un enjeu de santé publique mondial croissant. Avec l'augmentation de son incidence et son taux de survie à cinq ans (9%), cette maladie pourrait devenir la troisième cause de décès par cancer d'ici à 2025. Le faible taux de survie s'explique principalement par des diagnostics tardifs, résultant de l'absence de symptômes spécifiques et des difficultés à l'interprétation des images de tomodensitométrie (TDM) en phase portale, qui est la modalité d'imagerie abdominale standard. Sur ces images, les petites lésions sont fréquemment manquées par les radiologues, conduisant à un diagnostic tardif et limitant donc les options thérapeutiques. Les patients diagnostiqués à un stade précoce sont, quant à eux, plus souvent éligibles à une chirurgie du pancréas avec un taux de survie à cinq ans excédant 80%. Ces chiffres soulignent l'importance d'assister les radiologues lors du diagnostic.

Cette thèse, en collaboration avec Télécom Paris et Guerbet, vise à développer des méthodes d'intelligence artificielle (IA) pour assister les radiologues dans le diagnostic du cancer du pancréas. Les objectifs de cette thèse incluent la détection des lésions pancréatiques sur les TDM portales, et la poursuite de la détection précoce en identifiant les signes radiologiques secondaires qui sont visibles avant l'apparition de masses. Ces objectifs visent à faciliter un diagnostic précoce, à augmenter l'éligibilité des patients à la chirurgie et, in fine, à améliorer le taux de survie.

Pour atteindre ces objectifs, un grand nombre d'images de TDM en phase portale, acquises sur des patients atteints de diverses maladies du pancréas, a été collecté. Cela a permis la création de l'une des plus grandes bases de données annotées pour la recherche sur le cancer du pancréas, avec 2 800 examens provenant de plusieurs régions géographiques. Ces données ont ensuite été exploitées pour développer plusieurs méthodes d'IA selon trois pistes de recherche principales :

Premièrement, une méthode de segmentation automatique du pancréas a été développée. La segmentation automatique du pancréas est une étape essentielle pour permettre d'isoler une région d'intérêt dans l'image. Le pancréas présentant une forme allongée et des extrémités difficiles à délimiter, la méthode proposée utilise des informations géométriques pour ajuster localement la sensibilité de la segmentation prédite. Les résultats montrent que cette approche corrige la sous-segmentation systématique des extrémités générée par le nnU-Net, un modèle de segmentation de référence, tout en préservant la qualité globale de la segmentation.

Deuxièmement, une méthode réalise la détection des lésions et de la dilatation du canal pancréatique principal (CPP), deux signes cruciaux du cancer du pancréas. La méthode proposée commence par segmenter le pancréas, les lésions et le CPP à l'aide d'un nnU-Net. Ensuite, des caractéristiques quantitatives sont extraites des segmentations prédites, puis utilisées par deux régressions logistiques pour prédire la présence d'une lésion et la dilatation du CPP. La robustesse de la méthode est démontrée sur une large base externe de 756 patients, et des études d'ablation montrent que l'utilisation des caractéristiques sur le CPP permet d'améliorer la détection de lésions.

Dernièrement, afin de permettre un diagnostic précoce, deux approches sont proposées pour détecter des signes secondaires du cancer du pancréas. La première approche consiste à apprendre un modèle normatif des formes du pancréas. Pour cela, un auto-encodeur variationnel (AEV) est entraîné à partir d'un grand nombre de masques de segmentation de pancréas sains. Ce modèle est ensuite exploité pour détecter des formes anormales, en utilisant des méthodes de détection d'anomalies avec peu ou pas d'exemples d'entraînement. En plus de la détection de formes anormales, le modèle obtenu peut aussi être utilisé pour visualiser la différence de formes entre groupes. La seconde approche s'attaque aux signes secondaires portant sur la forme et la texture. Pour cela, deux types de radiomiques sont utilisés : les radiomiques profonds (RP), extraits automatiquement par des réseaux de neurones profonds, et les radiomiques manuels (RM), calculés analytiquement à partir de formules prédéfinies. La méthode repose sur un AEV qui extrait des RP non redondants par rapport à un ensemble prédéterminé de RM, afin de compléter l'information déjà contenue par ces derniers. Ensuite, les deux types de radiomiques résultants sont concaténés et utilisés par une régression logistique pour prédire la présence de signes secondaires : la forme anormale, l'atrophie, l'infiltration de graisse et la sénilité. Les résultats, obtenus sur une large base de test indépendante, montrent que cette méthode améliore les performances de prédiction par rapport à des approches de combinaison qui ne minimisent pas la redondance. Ces deux approches présentent des perspectives prometteuses pour

détecter les signes secondaires, ce qui pourrait faciliter la détection précoce du cancer du pancréas.

La principale perspective de cette thèse consiste à améliorer la précision du diagnostic en caractérisant les types de lésions et la malignité des signes secondaires, deux tâches difficiles pour les radiologues.

Abstract

Pancreatic cancer is recognized as a growing health concern worldwide. Due to the increase in its incidence and five-year survival rates (9%), this disease could become the third leading cause of cancer-related deaths by 2025. The low five-year survival rate primarily stems from late-stage diagnoses. Such diagnoses result from the lack of specific symptoms, as well as challenges pertaining to the interpretation of portal computed tomography (CT) scans which is the standard abdominal imaging modality for patients with non-specific symptoms. Consequently, radiologists frequently overlook small and subtle lesions, leading to late-stage diagnosis of pancreatic cancer, thereby limiting the treatment options. Yet, patients diagnosed at early stages are more often eligible for pancreatic resection, leading to a five-year survival rate surpassing 80%. This highlights the importance of assisting radiologists at the diagnosis stage.

This thesis, in collaboration with Télécom Paris and Guerbet, aims to develop artificial intelligence (AI) methods to help radiologists in the diagnosis of pancreatic cancer. The objectives include identifying pancreatic lesions on portal CT scans to prevent missed diagnoses, and pursuing early detection by identifying secondary radiological signs, visible before the appearance of distinguishable masses. These objectives seek to facilitate timely diagnosis, increase patient eligibility for resection, and ultimately enhance the survival rate.

To reach these objectives, a large dataset comprising portal CT scans from patients with a variety of pancreatic diseases was created. This effort resulted in one of the largest annotated databases for pancreatic cancer research, currently consisting of over 2,800 examinations from multiple geographical regions. This database was leveraged to accomplish the objectives, focusing on three primary research avenues.

First, a method for the automatic segmentation of the pancreas on portal CT scans was developed. To deal with the specific anatomy of the pancreas, which is characterized by an elongated shape and subtle extremities easily missed, the proposed method relied on local sensitivity adjustments using geometrical priors. Experiments showed that geometrical priors corrected the systematic under-segmentation error pattern generated by the nnU-Net, a state-of-the-art segmentation method. Then, the thesis tackled the detection of pancreatic lesions and main pancreatic duct (MPD) dilatation, both crucial indicators of pancreatic cancer. The proposed method started with the segmentation of the pancreas, the lesion and the MPD. Then, quantitative features were extracted from the predicted segmentations and leveraged to predict the presence of a lesion and the dilatation of the MPD. The method was evaluated on an external test cohort comprising hundreds of patients, demonstrating its robustness.

Continuing towards early diagnosis, two strategies were explored to detect secondary radiological signs of pancreatic cancer. The first approach leveraged large databases of healthy pancreases to learn a normative model of healthy pancreatic shapes, facilitating the identification of anomalies. To this end, volumetric segmentation masks were embedded into a common probabilistic shape space, enabling zero-shot and few-shot abnormal shape detection. The second approach leveraged two types of radiomics : deep learning radiomics (DLR), extracted by deep neural networks, and hand-crafted radiomics (HCR), derived from predefined mathematical formulas. The proposed method sought to extract non-redundant DLR that would complement the information contained in the HCR. Results showed that this method effectively detected four secondary signs of pancreatic cancer : abnormal shape, atrophy, senility, and fat replacement.

Further development could focus on refining diagnostic precision by predicting lesion types and the malignancy of secondary signs, which are challenging tasks for radiologists. From a technical perspective, integrating multi-modal data, such as multi-phase CT scans or electronic health records, can be promising for enhancing the models' efficacy.

Summary

Chapitre 1

Introduction

1.1 Preface

This introductory chapter lays the groundwork for this thesis, which is focused on applying artificial intelligence methods to detect pancreatic cancer. It begins by presenting the origins of this project, highlighting the clinical and technical challenges that motivated its creation. Then, it outlines the main objectives of this work and discusses the technical challenges it addresses. Moreover, the chapter offers an overview of the thesis structure, providing readers with a structured guide to navigate the subsequent chapters. Lastly, the chapter concludes with a comprehensive list of all the scientific works produced during the course of this thesis.

1.2 Genesis

This thesis originated from the initiative of Prof. Vullierme, a radiologist with over 25 years of experience in abdominal imaging. Working at the Beaujon Hospital in Paris, which has a gastroenterology department specializing exclusively in the treatment of pancreatic diseases, she was facing a high frequency of cases with advanced and incurable stages of pancreatic cancer. Detecting these cases earlier would have broadened the range of therapeutic options, thereby potentially increasing the survival rate. However, for many of these cases, opportunities for timely diagnosis were missed : despite previous imaging studies, pancreatic lesions often went unnoticed, either due to their subtle nature or due to the lack of suspicion surrounding asymptomatic early-stage pancreatic cancer. Recognizing the potential of Artificial Intelligence (AI) to systematically assist radiologists in identifying pancreatic cancers during routine clinical examinations, Prof. Vullierme initiated contact with Guerbet during the 2019 edition of the Journées Francophones de Radiologie (JFR). The collaboration aimed at launching a project for the development a computer-aided detection system, empowering radiologists with systematic and accurate pancreas analyses, with a specific emphasis on early pancreatic cancer detection.

This thesis played a pivotal role in the project genesis, contributing significantly to its conceptualization and initiation. Starting from scratch in 2020, the project sought to leverage AI technology to enhance patient outcomes within the complex landscape of pancreatic cancer diagnosis. Throughout this journey, several open questions have emerged : *Which imaging modality should be selected as input to maximize clinical impact ? How can we accurately delineate the pancreas given its variable and elongated shape with indistinct boundaries ? What strategies can be employed to develop a reliable algorithm capable of consistently detecting pancreatic lesions, particularly when they are small and subtle ? Which radiological features should the algorithm analyze ?* We propose to address these crucial questions in this manuscript.

1.3 Thesis objectives

The aim of this CIFRE PhD, in collaboration with the Laboratoire de Traitement et Communication de l'Information (LTCI) of Télécom Paris (Institut Polytechnique de Paris) and Guerbet, is to develop AI-based methods to help radiologists in the diagnosis of pancreatic cancer.

Pancreatic cancer is characterized by a notably low five-year survival rate of 9%, primarily attributable to missed or delayed diagnoses (Rawla et al., 2019). Several reasons explain these diagnostic failures. Given that early stages of pancreatic cancer typically lack symptoms, patients are often diagnosed at late stages, with approximately 90% of cases being diagnosed at stages for which the tumor is non-resectable[1] (Rawla et al., 2019). As surgical intervention stands as the primary therapeutic option, delayed diagnoses significantly correlate with diminished survival rates. Consequently, early detection emerges as the primary avenue for augmenting survival rates. Yet, unlike breast or colon cancers, which benefit from screening routines allowing for early detection, pancreatic cancer lacks recommended screening measures due to the absence of blood tests capable of reliably and efficiently detecting the disease (John Hopkins Medicine, b). Thus, one strategy for identifying pancreatic cancer at early stages involves incidental discovery during routine abdominal imaging (Goodman et al., 2012). However, the identification of cancer during such routine exams is challenging : as the pancreas is not the primary target of investigation, it often lacks a thorough examination. Consequently, small or subtle tumors are prone to being overlooked (Kang et al., 2021). Furthermore, routine abdominal imaging frequently employs the portal Computed Tomography (CT) phase, which is suboptimal for visualizing the pancreas as some lesions exhibit isodense patterns, making them barely visible or even imperceptible on this imaging modality (Shrikhande et al., 2012). To overcome this limitation, radiologists rely on secondary signs, which are radiological signs that may manifest before a tumor becomes clearly discernible (Gonoi et al., 2017). These signs, related to the shape, texture, and appearance of the Main Pancreatic Duct (MPD), play a pivotal role in detecting pancreatic cancer, particularly the isodense lesions. However, without thorough examination, these signs may go unnoticed, contributing to the numerous cases of missed early-stage cancers.

1. A resectable tumor is a tumor that can be removed by surgery.

Therefore, supporting radiologists in the diagnostic process is crucial for identifying potential cases of pancreatic cancer at early stages, ultimately enhancing overall survival rates. In this regard, computer-aided detection systems using AI, which have shown efficacy in diseases such as breast cancer and lung cancer (Fujita, 2020), offer significant promise for pancreatic cancer diagnosis given its challenging detection. Given this context, the general objective of this thesis is to propose AI-based methods designed to assist radiologists in identifying pancreatic cancer on portal CT scans, which serve as the primary imaging modality for abdominal examinations in clinical routines. The primary objectives encompass the systematic identification of pancreatic lesions to prevent missed or delayed diagnoses, and the pursuit of early diagnosis by detecting secondary signs that could be visible before the clear appearance of pancreatic lesions. The aim of these objectives is to facilitate timely diagnosis, increase eligibility for surgical intervention and ultimately enhance patients' survival.

To accomplish these goals, it is imperative to develop image analysis methods that address various methodological questions. These questions include tasks such as segmenting elongated and small objects, particularly under challenging conditions characterized by low contrast and indistinct contours. Additionally, the process involves characterizing both shape and appearance through a combination of features derived from expert prior knowledge and learned features. These questions are instantiated for the pancreas as follows :

- **Pancreas segmentation :** Accurate delineation of the pancreas is essential to establish a reliable Region of Interest (ROI) for subsequent analysis. The pancreas segmentation algorithm should robustly handle the organ's elongated shape and effectively identify its extremities, which is challenging due to the low contrast at its boundaries.
- **Lesion detection :** Developing a lesion detection method is critical for identifying all pancreatic lesions, particularly small and/or isodense ones. Given the difficulty of this task, leveraging radiological features derived from secondary signs should be considered.
- **Secondary signs detection :** Early cancer detection involves identifying secondary radiological signs indicative of pathological abnormality. Identification of these signs would greatly assist radiologists, as suggested by Kim et al. (2017). Various strategies can be explored, depending on the specific secondary signs that need to be detected.

These methods aim to empower both expert and non-expert radiologists, ensuring the timely identification of pancreatic cancer lesions and early detection of cases. They would not only mitigate diagnostic errors but also alleviate the workload of radiologists while increasing their confidence diagnostic through a "second opinion". Most crucially, such methods could substantially increase the number of cases diagnosed at early stages, thereby contributing to an overall rise in survival rates (Barat et al., 2021). In order to develop these methods and guarantee their robustness and reliability, the creation of a comprehensive database emerged as a central component of this thesis. It involved outlining precise criteria for the dataset (e.g., data type, clinical and geographical scope), alongside specifying the requisite annotations (e.g., segmentations, image-level labeling).

1.4 Thesis outline and main contributions

This manuscript is organized as follows :

In Chapter 2, a comprehensive exploration of the clinical landscape sets the stage for an in-depth understanding of the challenges at hand. Epidemiological figures underscore the aggressive nature of pancreatic cancer, and an exploration of its imaging characteristics elucidates the reasons behind delayed diagnoses. The chapter concludes by outlining the current landscape of pancreatic cancer, emphasizing the potential of AI for automated pancreas analysis.

Chapter 3 presents the current state of Deep Learning (DL) applications in pancreatic analysis. After a brief introduction to fundamental concepts of DL, a thorough exploration of the state-of-the-art in DL for pancreatic cancer is provided. This encompasses pancreas segmentation, lesion segmentation and detection, and feature extraction. An exhaustive review of existing methods is conducted, concluding with identified limitations and the proposed solutions outlined in this thesis.

Chapter 4 provides an in-depth overview of the datasets used throughout this study. It specifically details $\mathcal{D}_{CAPA}$ dataset, Guerbet's proprietary dataset for pancreatic research. The process of data collection and the definition of annotation protocols for $\mathcal{D}_{CAPA}$ is part of the scientific work of this thesis. **It required collaboration and expertise in AI and radiology in order to get the necessary data for training algorithms that could address significant clinical challenges. This resulted in the creation of one of the largest annotated databases for pancreatic cancer research**, currently comprising over $2,800$ portal CT scans from several geographical regions.

Following the examination of the clinical, technical, and material aspects within this thesis, Chapters 5, 6, and 7 outline the main scientific contributions of this work :

Chapter 5 focuses on the first methodological step of this work, which concerns pancreas segmentation on portal CT scans. **A novel method is proposed to overcome the difficulties encountered when segmenting elongated organs with ambiguous contours. This method, based on geometrical priors that locally adjust the sensitivity of the segmentation, is illustrated in the case of the pancreas.** This work was selected for an oral presentation at ISBI 2022 [I1], and for a poster presentation at GRETSI 2022 [N1]. The proposed segmentation model serves as a fundamental methodological tool enabling the exploration of more intricate challenges, including the identification of pancreatic abnormalities. Consequently, upon establishing this segmentation method, diverse research avenues were identified to detect different key findings that contribute to predicting the development of pancreatic cancer.

Chapter 6 introduces a **robust method for detecting two primary signs of pancreatic cancer : the presence of a lesion and the dilatation of the main pancreatic duct (MPD). The method relies on features extracted from the automatic segmentation of the pancreas, the lesions and the MPD. It is trained on multi-centric cohorts, and extensively validated on a large external dataset.** To the best of our knowledge, this method represents the first AI-driven approach for detecting MPD dilatation. Moreover, it demonstrates that enhancing lesion detection is achievable through the incorporation of explicit features derived from the MPD segmentation. This work has been presented at the JFR 2022 [N2] and published in Investigative Radiology [J1].

The next chapter work tackles early diagnosis through the detection of secondary cancer signs for which only image-level labels were available. Therefore, different strategies are explored to leverage these labels. Chapter 7 starts with the detection of abnormal pancreatic shapes. **To this end, a method to detect abnormal organ shapes through a normative model is proposed. This model is learned on large databases of healthy shapes in an unsupervised setting. Once the model is trained, the resulting distribution is leveraged to enable the detection of abnormal shapes in both zero-shot and few-shot settings. The resulting model can also be used to visualize the morphological differences between the organ shapes of different clinical groups.** The method is illustrated for the detection of abnormal pancreatic shapes. This led to a patent application [P1] and a poster presentation at MICCAI's main conference in 2022 [I2].

In the second part of the chapter, the focus extends to the use of radiomics for detecting four secondary signs of pancreatic cancer : general shape, atrophy, fat replacement, and senility. **To maximize the use**

of both existing types of radiomics, namely Deep-Learning Radiomics (DLR) and Hand-Crafted Radiomics (HCR), a novel method is introduced to extract DLR features that are non-redundant with a predefined set of HCR. The two types of radiomic features obtained can then be combined and leveraged by a classifier to predict secondary signs of cancer from an image. This work was selected for an oral presentation at the CaPTion workshop of MICCAI 2023 [I3], winning the best oral presentation award [A2], and a poster presentation at IABM 2024 [N3].

Chapter 8 concludes this thesis. It provides a summary and a discussion of key findings, and outlines future perspectives.

1.5 Publications

1.5.1 International conference proceedings

I1 R. Vétil, A. Bône, M-M. Rohé, M-P. Vullierme, P. Gori, and I. Bloch, "Improving the Automatic Segmentation of Elongated Organs using Geometrical Priors". Accepted for an oral presentation at *IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, Kolkata, India.*

I2 R. Vétil, C. Abi-Nader, A. Bône, M-M. Rohé, M-P. Vullierme, P. Gori, and I. Bloch, "Learning shape distributions from large databases of healthy organs : applications to zero-shot and few-shot abnormal pancreas detection". Accepted for a poster presentation at *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2022, Singapore.*

I3 R. Vétil, C. Abi-Nader, A. Bône, M-M. Rohé, M-P. Vullierme, P. Gori, and I. Bloch, "Non-Redundant Combination of Hand-Crafted and Deep Learning Radiomics : Application to the Early Detection of Pancreatic Cancer". Accepted for an oral presentation at *Cancer Prevention Through Early Detection Workshop (CaPTion), held in conjunction with MICCAI 2023, Vancouver, Canada.*

1.5.2 National conference proceedings

N1 R. Vétil, A. Bône, M-M. Rohé, M-P. Vullierme, P. Gori, and I. Bloch, "Utilisation d'informations géométriques pour l'amélioration de la segmentation automatique d'organes allongés". Accepted for a poster presentation at *28° Colloque sur le traitement du signal et des images, Groupe de Recherche en Traitement du Signal et des Images (GRETSI), 2022, Nancy, France.*

N2 R. Vétil, C. Abi-Nader, L-K. Wood, M-M. Rohé, F. Nicolas, P. Gori, I. Bloch, and M-P. Vullierme, "Détection automatique de lésions pancréatiques par apprentissage profond". Accepted for an oral presentation at *70èmes Journées Francophones de Radiologie (JFR), 2022, Paris, France.*

N3 R. Vétil, C. Abi-Nader, A. Bône, M-M. Rohé, M-P. Vullierme, P. Gori, and I. Bloch, "Combinaison non-redondante de radiomiques manuels et profonds : application à la détection précoce du cancer du pancréas". Accepted for a poster presentation at *Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale (IABM), 2024, Grenoble, France.*

1.5.3 International journals

J1 C. Abi-Nader, R. Vétil, L-K. Wood, M-M. Rohé, A. Bône, H. Karteszi, and M-P. Vullierme, "Automatic Detection of Pancreatic Lesions and Main Pancreatic Duct Dilatation on Portal Venous CT Scans Using Deep Learning" published in *Investigative Radiology*, 2023, (hal-04122865).

1.5.4 Patents

P1 R. Vétil, C. Abi-Nader, A. Bône, M-M. Rohé, P. Gori, and I. Bloch, "Method for characterizing an organ of a patient in a medical image", filled with the *European Patent Office, no. : EP22194886.2.*

1.5.5 Awards

A1 A *Student Travel Award* was awarded at the conference MICCAI 2022 for the paper : "Learning shape distributions from large databases of healthy organs : applications to zero-shot and few-shot abnormal pancreas detection" [I2].

A2 A *Best Oral Presentation Award* was awarded at the workshop CaPTion of the conference MICCAI 2023 for the paper : "Non-Redundant Combination of Hand-Crafted and Deep Learning Radiomics : Application to the Early Detection of Pancreatic Cancer" [I3].

Chapitre 2

Medical Context

2.1 Preface

We begin this thesis by providing the clinical context for pancreatic cancer. Our introduction begins with an exploration of the pancreas, emphasizing its anatomical characteristics. Then, we focus on pancreatic cancer, using epidemiological figures to understand the magnitude of the issue and emphasize the importance of early detection. We elaborate on diagnostic and therapeutic approaches, highlighting both the importance and the challenges of early detection. Next, we look at the radiological signs that help with early detection through imaging. Finally, we outline the current and future challenges of pancreatic cancer, paving the way for the integration of artificial intelligence tools.

2.2 Pancreas

2.2.1 Anatomy

2.2.1.1 General anatomy

The pancreas is a vital human organ, along with the brain, heart, lungs, liver, kidneys and skin. It is a flat, oblong organ, about 18 cm long, 3 cm wide and 4 cm high. It lies at the heart of the abdomen, between the stomach, the lumbar vertebrae, the duodenum[1] and the spleen. As shown in Figure 2.1, the pancreas is divided into five distinct parts : the uncinate process, the head, the neck, the body and the tail. The pancreas is traversed by two ducts : the main pancreatic duct, also referred to as the Wirsung duct, and the accessory pancreatic duct, alternatively known as the Santorini duct (Bockman, 1993).

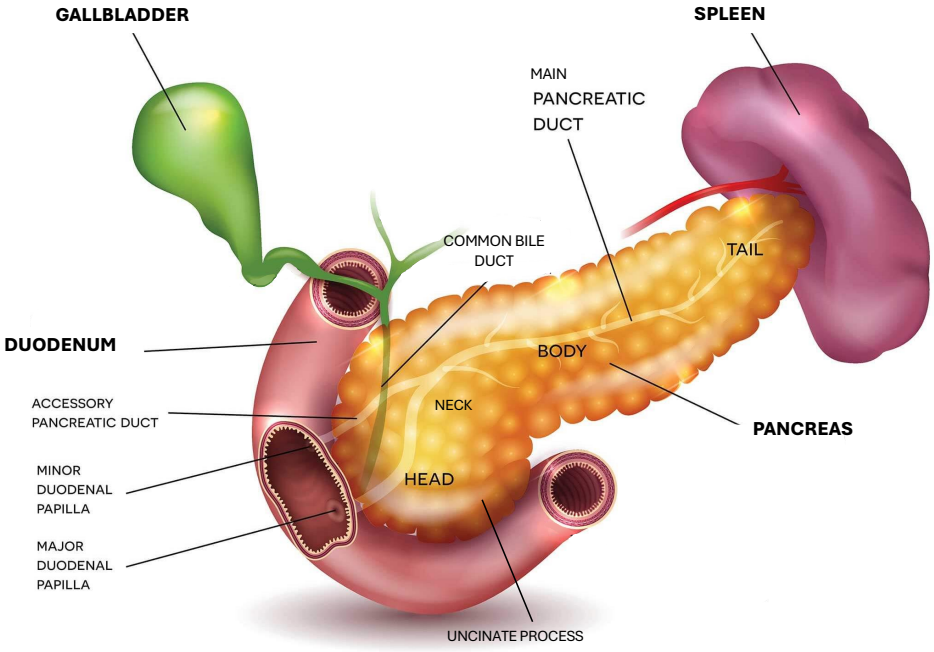


Figure 2.1 – Anatomy of the pancreas and adjacent organs. The image provides a comprehensive depiction of the pancreas, featuring its five distinct regions : uncinate process, head, neck, body, and tail. Adjacent structures such as the gallbladder, duodenum, spleen, and the pancreatic ducts are also highlighted. The main pancreatic duct initiates at the tail, traverses through the various regions, and culminates at the head. Notably, the main duct exhibits a gradual increase in diameter along its course and merges with the common bile duct (also known as choledochus) before opening into the duodenum through the major duodenal papilla. Additionally, the accessory pancreatic duct, which is thinner, runs across the head of the pancreas and terminates in the duodenum through the minor duodenal papilla, positioned approximately 2 cm above the main duct. Illustration adapted from John Hopkins Medicine (a).

2.2.1.2 Variations

While Figure 2.1 illustrates a conventional depiction of pancreatic anatomy, it is important to acknowledge that the pancreas exhibits considerable inter- and intra-individual variability (Borghei et al., 2013). Radiologists need to be aware of these variations to effectively distinguish between normal and pathological conditions. Considering these anatomical differences is particularly crucial in surgical interventions and diagnostic procedures within the abdominal region (Kozu et al., 1995). Such variations may arise from congenital factors originating from embryological development, or result from physiological changes.

1. Upper part of the small intestine.

Congenital variations The embryological development of the pancreas is complex and leads to numerous congenital anatomical variations, mainly affecting the pancreatic duct system (Kozu et al., 1995). Various classification systems exist to characterize the configuration of pancreatic ducts and the bile-pancreatic junction (Dimitriou et al., 2018). Figure 2.2 illustrates a five-group classification, notably featuring pancreas divisum. This anomaly, estimated to have a prevalence between 3-14%, arises from an abnormal fusion of the two pancreatic ducts (Dimitriou et al., 2018) and is visible on imaging (see Figure 2.3.B). Anomalies in the embryological development can also affect the entire gland or the vascular system. For instance, ectopic pancreas is characterized by the presence of pancreatic tissue in an abnormal location, unrelated anatomically to the main gland (see Figure 2.3.C). Its prevalence is estimated between 1-10% (Yu et al., 2006). Additionally, pancreatic variations can be vascular, related to the position of the splenic artery and hepatic portal vein concerning the pancreas (Orellana-Donoso et al., 2023).

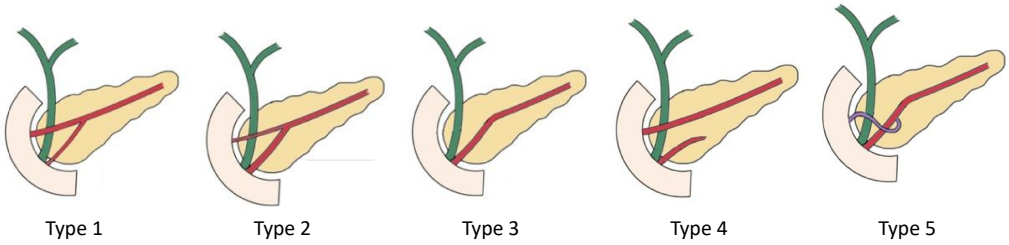


Figure 2.2 – Different pancreatic ducts configurations. Types 1 to 3 represent so-called "normal" configurations : type 1 exhibits a configuration with Wirsung as the dominant duct, while type 2 features a configuration with Santorini as the dominant duct. Type 3 is marked by a rudimentary, non-draining, or absent Santorini duct. Type 4 corresponds to pancreas divisum. Type 5, known as ansa pancreatica, is a rare variation where the accessory pancreatic duct takes a sinuous curve before merging with the main pancreatic duct. Adapted from Dimitriou et al. (2018).

Figure 2.3 – Congenital variations of the pancreas seen on portal CT scans. A : Typical anatomy of the pancreas. The pancreas is delineated in red, while the pancreatic ducts remain imperceptible. Image from the $\mathcal{D}_{NIH}$ public dataset (Roth et al., 2016a), presented in Chapter 4. **B : Pancreas divisum.** 46-year-old male with pancreas divisum. The image illustrates the two pancreatic ducts (indicated by long white and black arrows) and the common bile duct (short black arrow). Adapted from Soto et al. (2005). **C : Ectopic pancreas.** Soft-tissue attenuation focus (white arrows) originating from the pancreas, exhibiting similar attenuation to the tissue in the pancreatic head (P). Adapted from Yu et al. (2006). *CT : Computed tomography, $\mathcal{D}_{NIH}$: National Institute of Health dataset.*

Other variations Beyond congenital variations, the pancreas displays significant inter- and intra-individual variability in terms of size, shape, and contours. While the head and body have fixed positions, the left pancreas is mobile, and the pancreatic tail can exhibit varying degrees of mobility within the abdomen (Parent, 2012). This can result in diverse shapes, as illustrated in Figure 2.4. Additionally, although the normal pancreatic head typically features a smooth and continuous contour, some healthy pancreases may exhibit heterogeneous contours in the head (Yu et al., 2006). Moreover, the pancreas can undergo physiological changes, such as the replacement of pancreatic tissue with adipose cells. Elderly individuals (over 60 years)

may experience fatty involution of the gland with progressive atrophy and harmonious dilatation of the main duct : these pancreases are known as senile or ageing pancreases (Parent, 2012; Löhr et al., 2018).

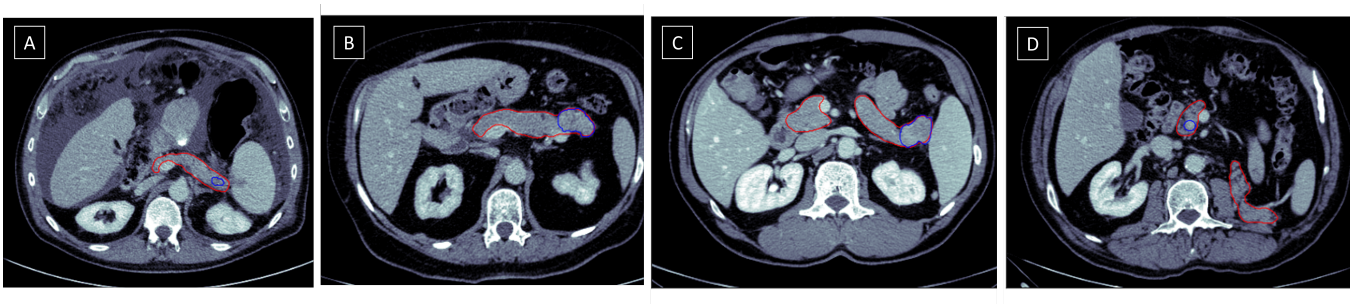


Figure 2.4 – Variability of the pancreas in terms of size, shape and positions. Portal CT images showing the pancreas and lesion delinated in red and blue, respectively. **A :** 67-year-old man with a hypo-enhancing mass. **B :** 61-year-old woman with a heterogeneous pseudocyst. **C :** Patient with a hyper-enhancing PNET tumor. **D :** 73-year-old man with a hyper-enhancing PNET tumor. The pancreas extends up to the position of the absent left kidney. Images A, B, D are sourced from the $\mathcal{D}_{CAPA}$ dataset, while image C is from the $\mathcal{D}_{MSD}$ dataset (datasets detailed in Chapter 4). *$\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, CT : Computed tomography, PNET : Pancreatic neuroendocrine tumor, $\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset.*

2.2.2 Function

The pancreas functions as a gland, producing and secreting substances such as hormones and digestive enzymes. As a mixed gland, the pancreas has both exocrine and endocrine functions, meaning that it releases substances into ducts and directly into the bloodstream.

Its primary exocrine role involves producing digestive enzymes, which travel through the main pancreatic duct into the duodenum. To safeguard against potential self-digestion of pancreatic tissue, these enzymes are released in an inactive state within the gland, with activation occurring exclusively upon reaching the digestive tract.

The pancreas also performs a minor endocrine function, producing hormones like glucagon and insulin that regulate blood glucose levels.

The exocrine function, which is managed by serous acini[2], is crucial for digestion, while the endocrine function is handled by islet cells and is vital for hormonal regulation. Despite the endocrine cells constituting only 1% of the pancreas, both of these functions (endocrine and exocrine) are intricately intertwined. Figure 2.5 illustrates these dual roles of the pancreas.

2.3 Pancreatic cancer

The term “pancreatic cancer” is commonly used as a synonym for Pancreatic Ductal Adenocarcinoma (PDAC) due to its high prevalence, accounting for 80% of all pancreatic cancers (Krejs, 2010). However, it is important to note that pancreatic cancer includes a broader spectrum of malignancies originating from either the exocrine or endocrine tissues of the pancreas. In addition to PDACs, approximately 10% of tumors originate from the exocrine pancreas, encompassing various histological types such as cystadenocarcinoma, giant-cell carcinoma, and intraductal papillary mucinous neoplasms (a type of cystic lesion discussed in Section 2.4.1). Beyond these exocrine pancreatic cancers, there exists a distinct subset constituting 10% of all pancreatic cancers, known as Pancreatic Neuroendocrine Tumor (PNET). Emerging from the endocrine cells of the pancreas, PNET represent a unique category with distinctive characteristics and clinical behaviors (Krejs, 2010).

2. A grouping of serous cells that secrete serous fluid.

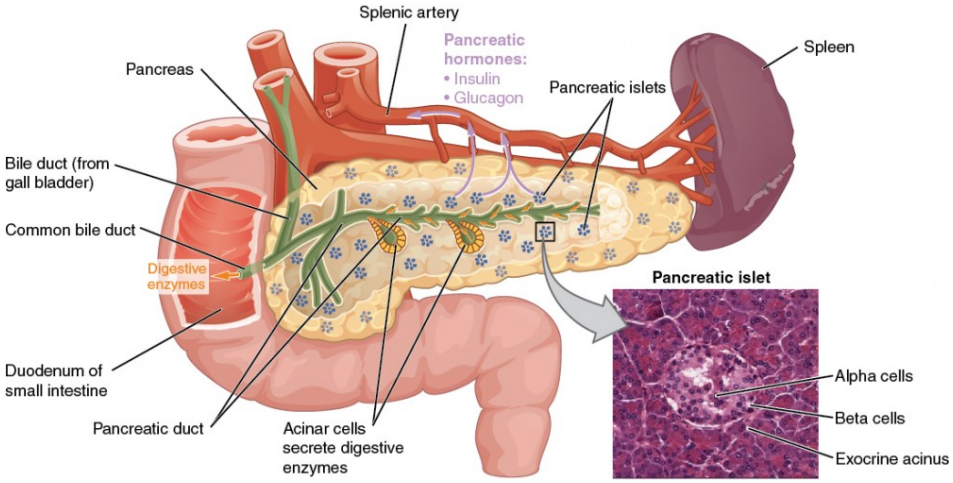


Figure 2.5 – Dual role of pancreatic function. Exocrine function is characterized by acinar cells releasing digestive enzymes transported through the pancreatic duct into the small intestine. Endocrine function involves pancreatic islets secreting insulin and glucagon, regulating glucose metabolism in the body. Micrograph provided by the Regents of the University of Michigan Medical School, 2012. Adapted from Human Anatomy and Physiology (Betts et al., 2013).

2.3.1 Epidemiology

Incidence, Prevalence, Mortality In 2017, there were 447,665 new cases of pancreatic cancer worldwide. The prevalence was estimated at 49.8 cases per million inhabitants, and the mortality at 55.7 cases per million inhabitants (Lippi and Mattiuzzi, 2020). Among all cancers, pancreatic cancer ranks 14th in incidence, 27th in prevalence, but 9th in Disability-Adjusted Life Years[3] and 6th in cancer-related deaths (Lippi and Mattiuzzi, 2020). These figures categorize pancreatic cancer as an uncommon yet highly aggressive disease. Figure 2.6.A illustrates the close parallel between pancreatic cancer incidence and mortality rates, underscoring its highly fatal nature with a mortality/incidence ratio of 94% (Rawla et al., 2019). Incidence, prevalence, and mortality increase with age, peaking after 70 years (Bosetti et al., 2012; Malvezzi et al., 2016), rendering it a disease primarily affecting elderly populations.

Survival The five-year survival rate for all types of pancreatic cancers stands at 9% (Rawla et al., 2019). Nevertheless, diverse factors, including age, gender, lifestyle, treatment approaches, and healthcare systems, have notable influences on survival rates. Table 2.1 illustrates the pivotal role of cancer type and staging at the time of diagnosis in determining survival rates. PDAC manifests a considerably grim prognosis compared to PNET, with five-year survival rates ranging from 13% to 1% for the former and from 61% to 16% for the latter (Rawla et al., 2019).

Clinical Stage	**Five-year survival (%)**	
	PDAC	*PNET treated with surgery*
I	13	61
II	7	52
III	3	41
IV	1	16

Table 2.1 – Five-year survival rates according to the clinical stage of pancreatic cancer. Adapted from Rawla et al. (2019). *PDAC : Pancreatic ductal adenocarcinoma, PNET : Pancreatic neuroendocrine tumor.*

Past and future trends Over the past 25 years, data revealed a substantial rise in both the incidence (+55%) and mortality (+53%) rates of pancreatic cancer (Lippi and Mattiuzzi, 2020). By 2020, pancreatic cancer accounted for 0.79% of all deaths, compared to 0.44% in 1992. Projections extrapolated from the

3. Disability-adjusted life-years (DALYs) are calculated as the sum of years of life lost and years lost due to disability.

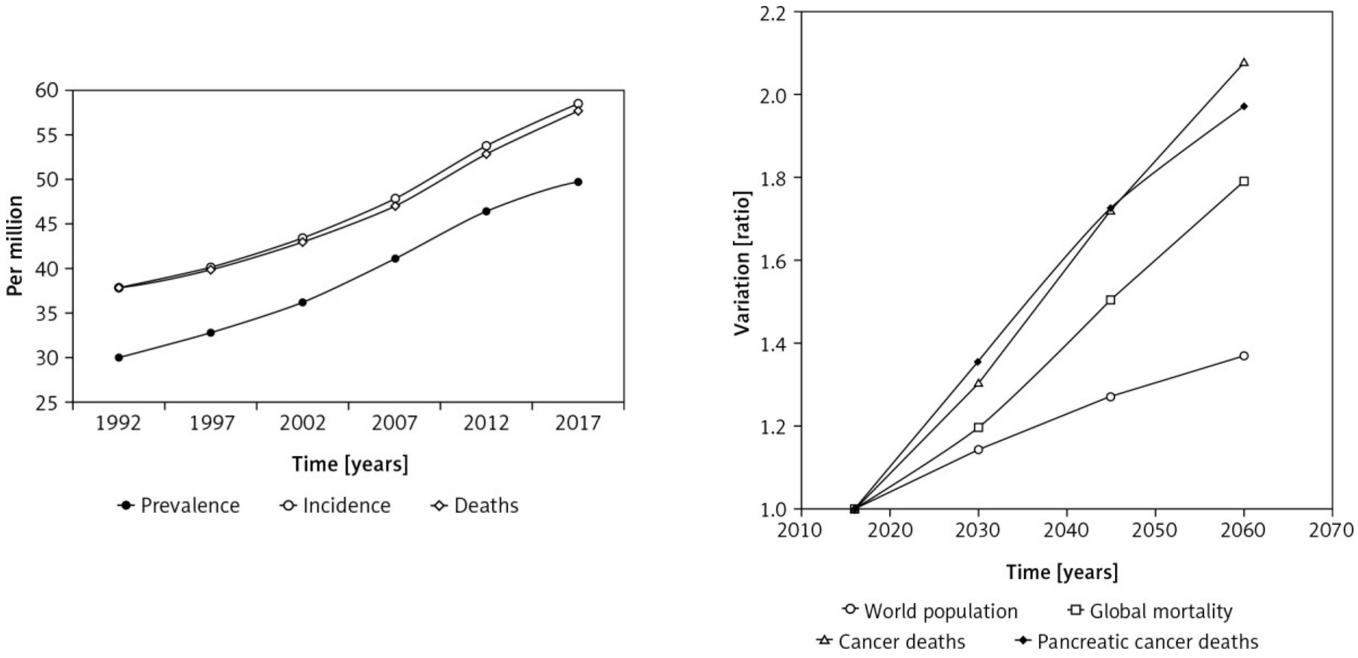


Figure 2.6 – **Left :** Worldwide prevalence, incidence, and mortality of pancreatic cancer during the last 25 years. **Right** : Future trends of mortality for all causes, all types of cancers and pancreatic cancers. Charts from Lippi and Mattiuzzi (2020).

World Health Organization (WHO) databases reveal a continuing upward trajectory in both mortality and incidence rates. Figure 2.6.B illustrates these projections, emphasizing an expected 1.97-fold rise in pancreatic cancer deaths over the next four decades. Similarly, according to the GLOBOCAN 2018[4], the incidence is projected to surge by +77.7% from 2018 to 2040 (Rawla et al., 2019) and similar patterns emerge for mortality, with a global expected increase of +79.9%. The exact causes of this significant increase are not thoroughly understood ; however, the surge in obesity rates and sedentary lifestyles are considered primary factors (Mizrahi et al., 2020; Rawla et al., 2019). Risk factors will be further explored in the following section.

2.3.2 Risk factors

To date, the causes of pancreatic cancer are still insufficiently known (Ilic and Ilic, 2016; Rawla et al., 2019). Nevertheless, risk factors have been identified and can be classified into modifiable and non-modifiable factors (Midha et al., 2016).

2.3.2.1 Modifiable risk factors

Modifiable risk factors associated with development of pancreatic cancer include obesity, diabetes, tobacco and alcohol use.

Obesity There is a well-established association between obesity and an elevated risk of pancreatic cancer (Mizrahi et al., 2020; Rawla et al., 2019). A comprehensive study conducted by the National Institute of Health, involving a substantial cohort exceeding 500,000 individuals, provided strong evidence that individuals classified as overweight or obese (Body Mass Index (BMI) ≥ 30 kg/m^2) face an increased likelihood of developing pancreatic cancer compared to those within the normal BMI range with hazard ratios ranging from 1.15 to 1.53 (Stolzenberg-Solomon et al., 2013).

4. GLOBOCAN 2018 is an online database providing estimates of incidence and mortality in 185 countries for 36 types of cancer, and for all cancer sites combined (Globocan, 2018).

Diabetes Both type I and type II diabetes have consistently been linked to an augmented risk of pancreatic cancer (Mizrahi et al., 2020; Pereira et al., 2020; Rawla et al., 2019). Notably, new-onset diabetes is recognized as a potential early indicator and a significant risk factor for the development of pancreatic cancer (Mizrahi et al., 2020; Pereira et al., 2020; Rawla et al., 2019). Conversely, pancreatic cancer is also considered a risk factor for the subsequent development of diabetes (Mizrahi et al., 2020; Pereira et al., 2020).

Alcohol and tobacco High alcohol intake (more than three drinks per day) and smoking are associated with an elevated risk of pancreatic cancer (Rawla et al., 2019). Smoking poses a two-fold higher risk for pancreatic cancer in smokers compared to non-smokers, with increased risks observed in former and passive smokers (Ezzati et al., 2005; Rawla et al., 2019). The complex relationship between alcohol and smoking makes establishing alcohol as an independent risk factor challenging (Rawla et al., 2019). In particular, Rahman et al. (2015) noted that smoking may modulate the association between alcohol and pancreatic cancer, with alcohol significantly increasing the risk among current smokers but not among non-smokers.

Other modifiable factors Additional modifiable risk factors include occupational exposures to substances such as nickel, cadmium, and arsenic, contributing to 12% of pancreatic cancer risk (Rawla et al., 2019). Dietary factors also play a role, with meta-analyses linking red meat consumption to an elevated risk of pancreatic cancer (Larsson and Wolk, 2012; Paluszkiewicz et al., 2012). More generally, adopting healthy lifestyles has been consistently associated with a reduced probability of developing pancreatic cancer, independent of tobacco consumption (Naudin et al., 2020). This is particularly significant given the increasing prevalence of metabolic syndromes[5] associated with obesity and sedentary lifestyles, contributing to the global rise in pancreatic cancer incidence despite declining smoking rates (Mizrahi et al., 2020).

2.3.2.2 Non modifiable risk factors

Non-modifiable risk factors for pancreatic cancer include gender, age, ethnicity, family history of pancreatic cancer, genetic factors, and chronic pancreatitis.

Gender, age, ethnicity Epidemiological figures highlight that men face a higher risk than women (incidence of 5.5 vs. 4.0 per 100,000 (Bray et al., 2018)). This gender discrepancy may be attributed to environmental risk factors or habits such as alcohol or tobacco consumption. Yet, undiscovered genetic factors may contribute to the variance in incidence and mortality between males and females. Age is an incontestable risk factor, with the likelihood of developing pancreatic cancer increasing with age, typically diagnosed at a median age of 70 (Bray et al., 2018). Ethnicity is another non-modifiable risk factor, with the highest incidence rates observed in African-Americans and the lowest in Asian-Americans and Pacific Islanders (American Cancer Society, 2014). While modifiable risk factors may explain these disparities, population-based studies suggest additional contributions from genetic factors (Arnold et al., 2009; Pernick et al., 2003).

Family history Approximately 5-10% of individuals with pancreatic cancer report familial pancreatic cancer history, indicating a first-degree relative (parent, sibling, or child) diagnosed with pancreatic cancer (Rawla et al., 2019). A prospective analysis reveals that first-degree relatives of individuals with familial pancreatic cancer face a nine-fold increased risk, doubling with two affected relatives and rising to 32-fold with three or more. The highest risk is observed in families with a case of young-onset pancreatic cancer (age < 50 years) (Brune et al., 2010).

5. Metabolic syndrome (also called insulin resistance syndrome) is a group of conditions that together raise the risk of coronary heart disease, diabetes, stroke, and other health problems.

Genetic factors Approximately 10% of all pancreatic cancers exhibit genetic variation or mutation (Ghiorzo, 2014). Several genetic mutations and familial cancer syndromes have been identified, including hereditary breast and ovarian cancer syndrome, familial pancreatitis and Peutz-Jeghers syndrome[6] (Mizrahi et al., 2020).

Pancreatitis Numerous studies establish longstanding chronic pancreatitis (inflammation of the pancreas, further discussed in Section 2.4.2) as a risk factor for pancreatic cancer (Rawla et al., 2019). However, a significant proportion of pancreatitis cases are considered to be the result of a ductal obstruction caused by a tumor, highlighting its dual role as both a risk factor and an early disease indicator (Raimondi et al., 2010). Furthermore, patients with hereditary pancreatitis syndromes exhibit a higher risk of developing pancreatic cancer compared to those with other forms of pancreatitis or individuals without pancreatitis (Mizrahi et al., 2020).

2.3.3 Diagnosis

2.3.3.1 Clinical signs

Pancreatic cancer often manifests insidiously, frequently exhibiting minimal or no symptoms until it progresses to an advanced stage (Mizrahi et al., 2020). When symptomatic, patients typically present with nonspecific manifestations, such as epigastric[7] or back pain, nausea, bloating, abdominal fullness, or alterations in stool consistency. Yet, these symptoms are commonly attributed to benign causes, leading to a potential delay in diagnosis (Macdonald et al., 2006; Walter et al., 2016).

Upon diagnosis, prevalent symptoms include abdominal pain (40-60%), abnormal liver function tests (50%), jaundice (30%), new-onset diabetes (13-20%), dyspepsia[8] (20%), nausea (16%), back pain (12%), and weight loss (10%) (Schmidt-Hansen et al., 2016). The manifestation of these signs and symptoms depends on the tumor's location within the pancreas : tumors situated in the body or tail of the pancreas typically induce epigastric or back pain, while those located in the head and neck obstruct the main pancreatic duct, resulting in jaundice. Additional symptoms may include gastric outlet or bowel obstruction, weight loss, anorexia, depression, venous thrombosis, pancreatitis, or indications of pancreatic enzyme insufficiency.

A comprehensive study, involving more than 3,400,000 patients in primary care, identified jaundice as the only high-risk feature for pancreatic cancer, with a Positive Predictive Value (PPV) ranging from 4 to 13% (Schmidt-Hansen et al., 2016). Conversely, the PPVs for other individual symptoms were notably low, with the highest PPV being 1% for repeated attendance with abdominal pain in patients aged 60 years or older (Schmidt-Hansen et al., 2016). This underscores the challenges associated with identifying pancreatic cancer based solely on isolated symptoms, emphasizing the need for a comprehensive diagnostic approach.

2.3.3.2 Imaging

Both diagnostic and interventional imaging play crucial roles in diagnosing pancreatic cancer. Presently, pancreatic imaging modalities primarily include Ultrasound (US), Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Endoscopic Ultrasound (EUS), and Endoscopic Retrograde Cholangiopancreatography (ERCP)[9] (see examples in Figure 2.7).

Diagnostic imaging

US stands as the most widely used imaging modality due to its cost-effectiveness and non-ionizing radiation properties. However, suboptimal visualization of the pancreas may occur due to intestinal gases,

6. Peutz-Jeghers syndrome is a rare disorder in which growths called polyps form in the intestines.
7. Epigastric pain refers to pain below the ribs in the upper abdomen area.
8. Recurring symptoms of an upset stomach that have no obvious cause. Symptoms resemble those of an ulcer and include pain in the upper abdomen, bloating, belching and nausea.
9. Invasive procedure that combines upper gastrointestinal endoscopy and x-rays to treat problems of the bile and pancreatic ducts.

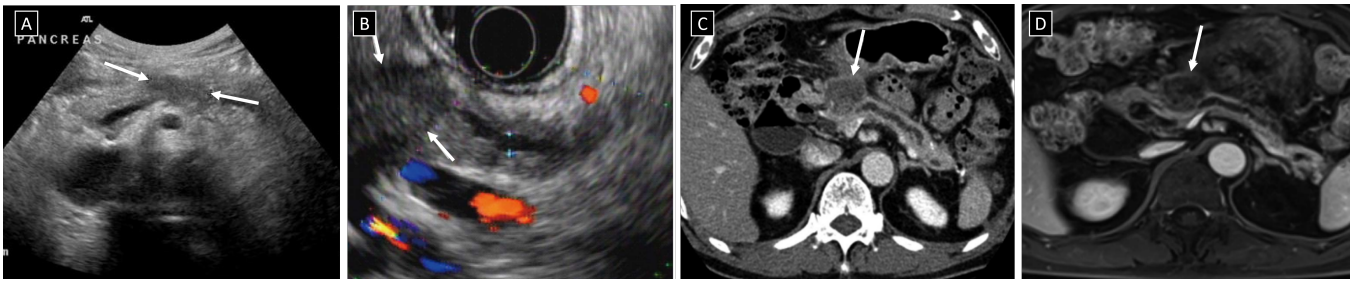


Figure 2.7 – A : US in 70-year-old male with pathologically-proven pancreatic cancer. Hypoechoic lesion (arrows) with irregular borders in the body of the pancreas. **B : EUS in 58-year-ole male patient with pancreatic cancer.** The EUS shows an approximately 3 cm, hypoechoic mass (arrows) in the pancreatic body. **C, D : CT and MRI in a 73-year-old male with pathologically-proven pancreatic head cancer. C :** CT scan showing a low attenuating mass (arrow) at the pancreatic head. **D :** After contrast media administration in a T1-weighted gradient echo sequence of MRI, the pancreatic head cancer (arrow) has poor enhancement. Image A adapted from Radiopaedia (a), images B, C, and D adapted from Lee and Lee (2014). *CT : Computed tomography ; EUS : Endoscopic ultrasounds, MRI : Magnetic resonance imaging ; US : Ultrasounds.*

and the examination remains operator-dependent. Various techniques such as respiratory maneuvers, different windows, patient positions, and oral contrast agents may enhance visualization, yet the procedure remains inherently operator-dependent (Hanbidge et al., 2002). The sensitivity of US for pancreatic cancer detection ranges from 75% to 89% (Elbanna et al., 2020).

CT has emerged as the preferred method for pancreatic examination, particularly with the advent of multidetector scanners. This technology offers improved spatial resolution and allows detailed analysis of the parenchyma[10], biliopancreatic ducts, and vessels. CT imaging of the pancreas can be performed with and without contrast media. When contrast media is injected, several acquisitions are made over time to capture variations in contrast enhancement as the contrast media circulates through the body. Therefore, CT imaging typically involves an unenhanced acquisition, an arterial phase acquisition (20-25 seconds after contrast injection), and a portal venous phase acquisition (70-80 seconds after contrast injection). Sometimes, a pancreatic phase acquisition is taken approximately 40-45 seconds after contrast injection (or 30-35 seconds for PNET) (Parent, 2012). This pancreatic phase is designed for the exploration of the pancreas as it maximizes the pancreatic glandular enhancement. Arterial phase and particularly pancreatic phase provide superior differentiation between normal pancreatic tissue and potential tumors, making it more suitable for detecting pancreatic cancer. Nevertheless, the portal phase remains the standard modality for routine abdominal examinations, making it a primary diagnostic tool for pancreatic cancer (Al-Hawary et al., 2014). PDAC typically appear hypodense, particularly during the arterial phase, although a proportion (11%) may be isodense (Shrikhande et al., 2012). The different enhancement patterns of PDAC on portal CT scans can be seen in Figure 2.8. As isodense lesions are barely or not at all visible, their detection is based on secondary signs. These signs, detailed in Section 2.3.5, are visible even before the clear appearance of the tumor and are of primary importance for the detection of isodense cancers (Gonoi et al., 2017; Kim et al., 2010). Sensitivity and specificity estimates for multi-phase CT vary, with reported ranges of 75-100% and 70-100%, respectively, as indicated by several studies (Costache et al., 2017; Elbanna et al., 2020; Valls et al., 2002). However, sensitivity decreases notably for small tumors ($\leq$ 2 cm), with estimates ranging from 58 to 77% across studies (Costache et al., 2017; Elbanna et al., 2020).

MRI, specifically Magnetic Cholangiopancreatography (MRCP)[11], is well-suited for detecting small non-contour-deforming tumors and assessing the surrounding structures, including liver lesions and metastases (Vachiranubhap et al., 2009). Sensitivity and specificity for MRI are reported at 81-99% and 70-93%, respectively (Costache et al., 2017). MRI and MRCP present notable advantages, including the absence of radiation exposure, enhanced contrast resolution, and the non-invasive exploration of the pancreatobiliary system facilitated by MRCP. However, their cost limits widespread use (Shrikhande et al., 2012). Two meta-analyses found that CT and MRI/MRCP exhibit comparable sensitivity and specificity rates for pancreatic cancer diagnosis and staging (Shrikhande et al., 2012; Treadwell et al., 2016).

10. The parenchyma is the functional parts of an organ.
11. A special type of MRI that uses computer software to image the pancreatic and bile ducts, areas where tumors often form.

Interventional imaging

EUS is recommended for assessing lesions that are not clearly detected but are suspected based on CT/MRI findings. It is also employed in tumors considered to be at the "borderline resectability" threshold to evaluate vascular involvement and identify lymph nodes. Pancreatic cancer manifests as a hypoechoic (darker) lesion compared to normal pancreatic tissue. EUS, coupled with fine-needle aspiration[12], can offer tissue confirmation for patients with potentially resectable masses. EUS provides the highest rates of precision and sensitivity for lesions under 2 cm (91% and 94% for EUS vs. 69% and 73% for CT, 82% and 84% for MRI) (Costache et al., 2017). However, its added value depends on the operator's skill (Lévy et al., 2017). Additionally, this imaging modality cannot be used in isolation as it does not allow for the detection of extra-abdominal metastases.

In cases of biliary or pancreatic duct obstruction, ERCP can be performed to place a biliary stent, which is a medical device inserted into the bile duct to allow the flow of bile. Despite its superior sensitivity and specificity exceeding 90% for pancreatic cancer diagnosis (Niederau and Grendell, 1992), this invasive imaging technique is reserved for situations presenting ductal obstruction due to its high risk of complications.

In the majority of cases, accurate interpretation of cross-sectional imaging eliminates the need for invasive techniques like EUS. A well-performed CT scan with slices less than 2 mm, unenhanced and contrast-enhanced phases, and digestive contrast with water can resolve a significant number of diagnostic challenges (Lévy et al., 2017). Diagnostic tools and protocols vary among countries, contributing to differences in incidence and mortality rates, especially between high and low human development index regions (Rawla et al., 2019). Consequently, high incidence rates may result from accurate diagnostic protocols rather than factors related to the causes or origins of pancreatic cancer.

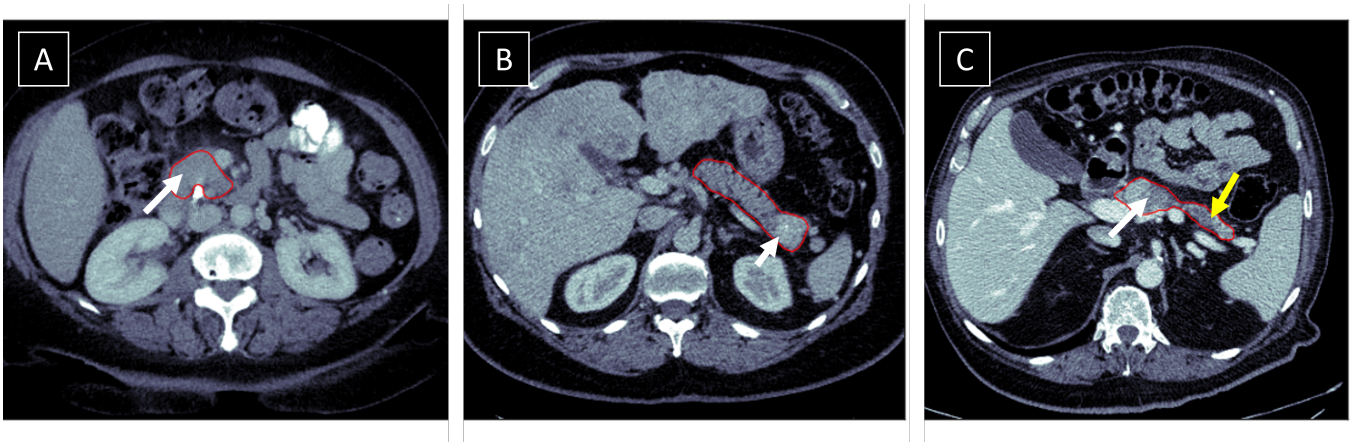


Figure 2.8 – Enhancement patterns of pancreatic cancer on portal CT scans. A : Iso-enhancing PDAC. 70-year-old woman with iso-enhancing PDAC (white arrow) in the head of the pancreas. **B : Hyper-enhancing PDAC**. Patient with hyper-enhancing PDAC (white arrow) in the tail of the pancreas. **C : Hypo-enhancing PDAC**. 60-year-old woman with iso-enhancing PDAC (white arrow) in the body of the pancreas. Upstream dilatation of the main pancreatic duct is indicated by a yellow arrow. On each image, the pancreas is delineated in red. Images A and B are sourced from the $\mathcal{D}_{CAPA}$ dataset, while image C is from the $\mathcal{D}_{MSD}$ dataset (datasets detailed in Chapter 4). *$\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, CT : Computed tomography, $\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset, PDAC : Pancreatic ductal adenocarcinoma.*

2.3.3.3 Screening

The low incidence of pancreatic cancer in the general population excludes the recommendation of screening for asymptomatic individuals (Owens et al., 2019). However, research has demonstrated the efficacy of screening for high-risk individuals with an inherited predisposition to pancreatic cancer (Lu et al., 2015). A 2012 American study revealed that screening 225 asymptomatic high-risk individuals led to the detection of neoplasms[13] in 85 patients, with results suggesting that MRI and EUS were more

12. Minimally invasive way to obtain a cell sample to confirm a diagnosis or guide treatment.
13. Any abnormal tissue growth that results in a benign or malignant tumor.

sensitive than CT for detecting pancreatic abnormalities (Canto et al., 2012). In 2014, the International Cancer of the Pancreas Screening Consortium recommended screening for high-risk individuals, concurring that initial screening should involve MRI or EUS or a combination of both (Canto et al., 2013). However, there was discord regarding optimal screening modalities, intervals for follow-up imaging, and criteria for recommending surgery based on screening abnormalities. The role of screening is also under investigation for the surveillance of patients with chronic pancreatitis or pancreatic cystic neoplasms (Del Chiaro et al., 2014). Presently, several screening programs for high-risk individuals are in operation, even though the precise screening protocols vary. They typically involve initial cross-sectional imaging and blood tests, followed by annual non-radiating imaging, such as MRI or EUS (Pereira et al., 2020).

2.3.3.4 Incidental findings

In contrast to cases identified through systematic screening, where efforts are expressly directed at detecting cancers, a significant number of cases are accidentally discovered in asymptomatic individuals or those whose symptoms do not specifically align with pancreatic cancer manifestations (Goodman et al., 2012). These incidental findings, termed pancreatic incidentalomas, have witnessed a surge in prevalence, likely attributable to the increased frequency and enhanced quality of abdominal imaging examinations (Fitzgerald et al., 2003; Santo and Bar-Yishay, 2017). The prevalence of these incidental pancreatic lesions varies depending on the lesion type and the imaging modality employed (Santo and Bar-Yishay, 2017). For solid incidentalomas, diverse studies report prevalences of 0.6% among healthy kidney donors or 0.49% in patients undergoing FDG-PET[14] for various non-pancreatic-related reasons (Santo and Bar-Yishay, 2017). These figures surpass official estimates as the GLOBOCAN 2018 database projected a maximum incidence of 7.7 per 100,000 in Europe (Rawla et al., 2019). The primary reasons for the initial scans leading to the discovery of incidentalomas predominantly involve the monitoring of extra-pancreatic malignancies, evaluations for hematuria[15], and examinations related to abdominal or pelvic pain (Fitzgerald et al., 2003; Goodman et al., 2012; Sachs et al., 2009). Although the portal phase is suboptimal for detecting pancreatic cancer, most incidental findings are detected using this modality, given its role as the standard procedure for routine abdominal examinations (Al-Hawary et al., 2014).

2.3.3.5 Serum biomarkers

Concerning biomarkers, Carbohydrate Antigen (CA)19-9 exhibits a sensitivity of 79-81% and specificity of 82-90% for diagnosing pancreatic cancer in symptomatic patients (Ballehaninna and Chamberlain, 2012). The monitoring of CA19-9 levels plays a pivotal role in evaluating the response to systemic treatment[16] and often serves as an early indicator of treatment response observed on imaging (Mizrahi et al., 2020). While CA19-9 levels assist in identifying patients at a higher risk of surgery, they may also offer prognostic value in inoperable pancreatic cancer patients. However, the diagnosis of pancreatic cancer cannot rely solely on a single CA19-9 measurement as other benign conditions, such as biliary obstructions, can lead to elevated CA19-9 levels (Mann et al., 2000).

2.3.4 Staging and therapeutic options

Staging is crucial to assess the extent of cancer, assessing factors like tumor size, invasion into adjacent tissues, and the presence of distant metastasis. In pancreatic cancer, staging revolves around the concept of tumor resectability, using a four-stage classification system (Mizrahi et al., 2020). Therapeutic options are stage-dependent, with surgery being the only curative treatment (Millikan et al., 1999; Okasha et al., 2017).

14. A fluorodeoxyglucose (FDG) Positron Emission Tomography (PET) scan is an imaging test that is commonly used to look for cancer in the body.
15. Presence of blood in the urine.
16. Systemic treatment is any medication that travels through the body in the bloodstream to find, damage or destroy cancer cells. It includes chemotherapy, immunotherapy, hormone therapy or targeted therapy.

Stage I and II Stage I involves resectable cancer confined to the pancreas, with a size less than 2cm (stage IA) or than 4 cm (stage IB), while stage II includes borderline resectable cancer exceeding 4 cm but limited to the pancreas (stage IIA) or nearby lymph nodes (stage IIB). Surgical resection, the primary treatment, yields five-year survival rates ranging between 10 and 25% (Mizrahi et al., 2020). Tumors in the head undergo resection with a pancreaticoduodenectomy, involving removal of the pancreatic head, duodenum, proximal jejunum, common bile duct, gall bladder, and a segment of the stomach (Mizrahi et al., 2020). Tumors in the body or the tail can be addressed with a distal pancreatectomy, often combined with a partial or complete removal of the spleen when necessary. To ensure negative surgical margins[17], venous and arterial resections are frequently performed, but the latter are associated with increased post-operative morbidity (Mizrahi et al., 2020). Post-operative systemic therapy has shown improved overall survival in phase 3[18] studies (Conroy et al., 2018; Neoptolemos et al., 2001), and neoadjuvant therapy[19] is increasingly used in high-volume centers to enhance resection outcomes and identify patients unfit for surgery (Mizrahi et al., 2020).

Stage III Stage III involves locally advanced cancer extending to adjacent blood vessels or nerves without distant metastasis, affecting over 30% of patients. Primary treatment includes systemic chemotherapy, supported by multiple studies (Versteijne et al., 2020). Surgery might be considered for a small subset of responders, but for most, the cancer remains incurable (Philip et al., 2020). The role of chemoradiation in these cases is debated due to conflicting study outcomes (Chauffert et al., 2008; Loehrer Sr et al., 2011).

Stage IV Finally, stage IV denotes cancer that has metastasized to other organs (Rawla et al., 2019) and concerns approximately 50% of patients with diagnosed pancreatic cancer. For these patients, systemic chemotherapy is the primary treatment, aiming to alleviate cancer-related symptoms and prolong life. Supportive care, including interventions like metal stent placement for biliary obstruction, is a crucial additional component of treatment. Various symptoms such as pancreatic insufficiency, depression, anxiety, anorexia, and weight loss also need to be addressed with pharmacologic treatment, and palliative care involvement can be considered (Mizrahi et al., 2020).

2.3.5 Secondary signs

Numerous studies have highlighted the presence of visible radiological indicators well before the clear onset of tumors. Detecting these secondary signs is crucial as it can lead to early diagnosis of pancreatic cancer. These indicators primarily concern the main pancreatic duct but also the appearance of the pancreatic parenchyma.

2.3.5.1 Secondary signs affecting the main pancreatic duct

Dilatation of the main pancreatic duct The most extensively studied early marker is the dilatation of the pancreatic duct, caused by small, invisible tumors blocking and dilating the pancreatic duct (Vasen et al., 2019). Several retrospective studies of longitudinally monitored patients have observed the dilatation of the pancreatic duct (Tanaka et al., 2010; Vasen et al., 2019). Notably, a case-control study (Tanaka et al., 2002), retrospectively analyzing 39 pre-diagnostic US scans of cancer patients and 10244 control cases, revealed that the proportion of cases with slight dilatation ($\geq$ 2 mm in diameter) of the main pancreatic duct was 65% in the precancerous group, compared to 5.35% in age-matched control subjects. This dilatation was visible more than 4 years before pancreatic cancer resection, with this proportion increasing over time among pathological cases. Another case-control study (Gonoi et al., 2017) showed similar results, finding that the incidence of ductal dilatation was significantly higher in the cancer group than in the healthy group and gradually increased over the 15 months preceding the diagnosis. These findings collectively underscore

17. Negative surgical margin means that there are no cancer cells found at the edge of the tissue that was removed.
18. A Phase 3 clinical study is a large-scale, randomized, and controlled investigation conducted on a broad patient population to assess the efficacy, safety, and overall performance of a new medical intervention before regulatory approval.
19. Neoadjuvant therapy refers to any treatment that is given for cancer before the main treatment, with the goal of making the main treatment more likely to be successful.

the potential of pancreatic duct dilatation as a valuable marker for early detection and risk assessment of pancreatic cancer.

Other abnormalities of the main pancreatic duct In addition to duct dilatation, other anomalies of the pancreatic duct have been identified. Case-control studies report that the dilatation of the pancreatic duct is often accompanied by the abrupt interruption of the pancreatic duct (Gangi et al., 2004). Ahn et al. (2009) estimated that the sensitivity and specificity of abrupt interruption are 45% and 82%, respectively. Stenosis has also been reported as a pre-diagnostic sign with a predictive value estimated at 47% (Kanno et al., 2019; Yamao et al., 2020). However, the radiological characteristics of the stenosis are insufficient to distinguish between benign and malignant stenosis, and histological confirmation through pancreatic juice cytology is the best way to characterize ductal stenosis (Kanno et al., 2019).

2.3.5.2 Secondary signs affecting the pancreatic parenchyma

Furthermore, other secondary signs related to the pancreatic parenchyma in general have been identified. These signs include parenchymal atrophy, fatty infiltration, and parenchymal heterogeneity.

Parenchymal atrophy Numerous studies report that pancreatic atrophy is an indicator of pancreatic cancer (Chen et al., 2020b; Miura et al., 2020; Yamao et al., 2020), with its sensitivity and specificity being estimated at 45% and 96%, respectively (Ahn et al., 2009). The atrophy originates by the thinning of the cell layer due to the disruption of flow in the pancreatic duct caused by cancer. This thinning is followed by apoptosis[20] and/or cell death through necrosis. The loss of cells is replaced by fibrotic changes and the infiltration of immune cells, visualized as atrophy. This atrophy can also result from chronic inflammation (Yamao et al., 2020).

Fat infiltration Fatty infiltration, which can be evaluated through histological examination and medical imaging (CT, MRI, EUS), is suspected to alter the cellular microenvironment of the pancreas and promote tumor development (Khoury et al., 2017). Several studies demonstrated that the degree of pancreatic fat infiltration is positively associated with cancer, even after adjusting for BMI, diabetes prevalence, and other confounding factors (Khoury et al., 2017; Miura et al., 2020; Takahashi et al., 2018).

Parenchymal inhomogeneity The heterogeneous appearance of the parenchyma arises from inflammation associated with the lesion (often small and non-visible) and desmoplasia[21]. This alters the microenvironment by decreasing the quantity of blood vessels and the oxygen supply, causing heterogeneity throughout the parenchyma (Gonoi et al., 2017). Gonoi et al. (2017) reported that pancreatic parenchymal inhomogeneity seen in any CT phase is a pre-diagnostic factor identifiable up to 34 months before diagnosis. Similarly, Kim et al. (2017) demonstrated that malignant ductal dilatations are characterized by a difference in parenchymal density at the site of dilatation.

We provide, in Figure 2.9, examples of the different secondary signs as seen on portal CT.

2.4 Other pancreatic diseases

2.4.1 Pancreatic cystic neoplasms

Pancreatic cystic neoplasms are prevalent conditions : depending on the studies, their prevalence in the general population undergoing MRI has been estimated at 13.5-20%, and consistent findings indicate that this prevalence increases with age (Lee et al., 2010; Zhang et al., 2002). The category of cystic lesions in the pancreas includes more than 20 entities, with four major types, namely Intraductal Papillary Mucinous

20. The programmed death of some of an organism's cells as part of its natural growth and development.
21. Formation of excessive, dense, and often disorganized tissue around cancer cells.

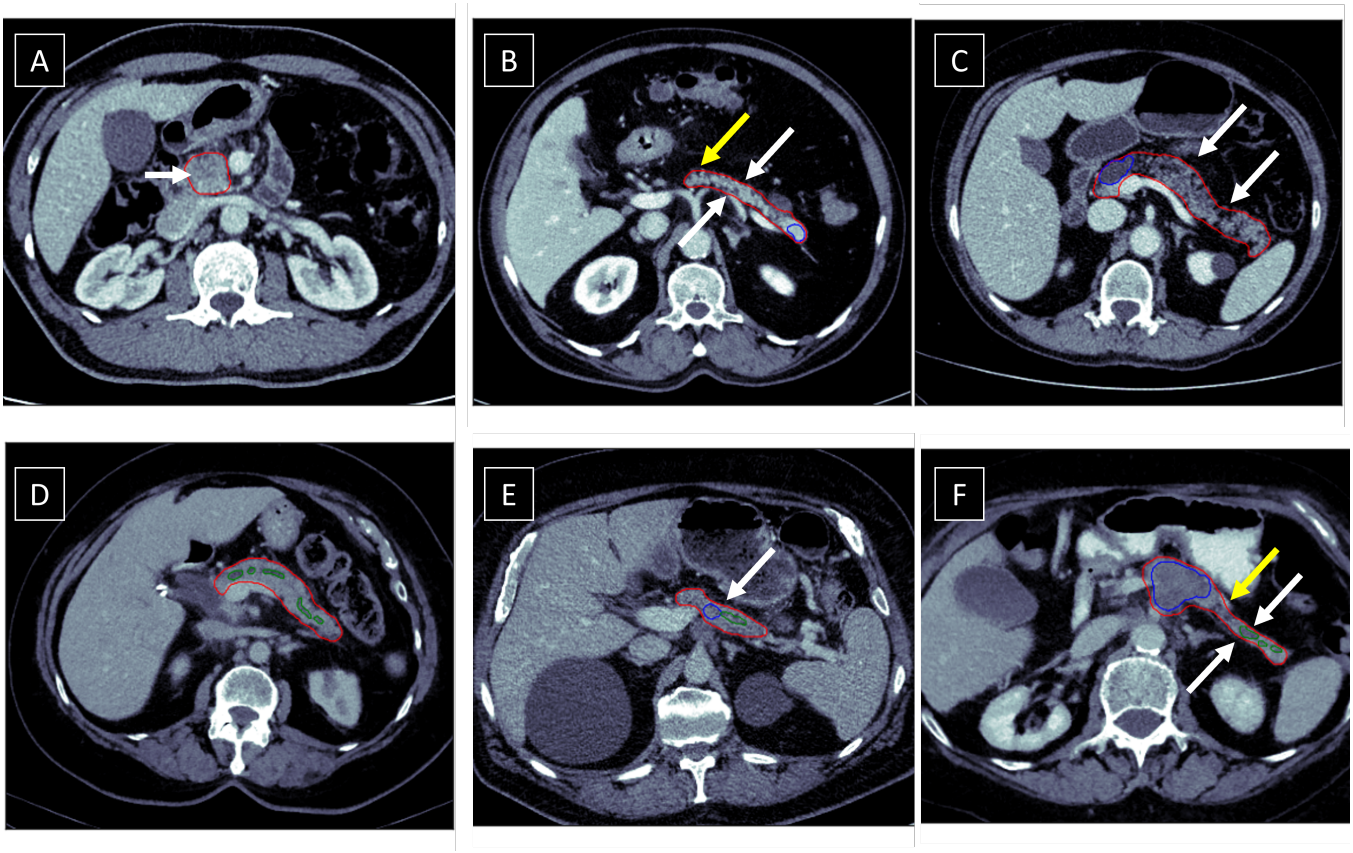


Figure 2.9 – Visualization of secondary signs of pancreatic cancer on portal CT. When visible, the pancreas, the lesion, and the MPD are delinated in red, blue, and green, respectively. **A : Parenchymal inhomogeneity.** Patient with hypo-enhancing PDAC in the head of the pancreas, exhibiting inhomogeneous texture (white arrow). **B : Parenchymal atrophy.** 77-year-old man with iso-enhancing PNET. The pancreas exhibits atrophy (white arrows) and fat infiltration (yellow arrow). **C : Fat infiltration.** 60-year-old woman with hypo-enhancing IPMN. The pancreas exhibits fat replacement (white arrows). **D : MPD dilatation.** 76-year-old woman showing a dilatation of the MPD. **E : MPD dilatation and stenosis.** Patient showing an iso-enhancing mass blocking the MPD (white arrow). **F : MPD dilatation, stenosis and atrophy.** Patient with a hypo-enhancing mass following the stenosis (yellow arrow) of the MPD. The MPD is dilated and the pancreas is atrophic (white arrows). All images are sourced from the $\mathcal{D}_{CAPA}$ dataset detailed in Chapter 4. *$\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, CT : Computed tomography, IPMN : Intraductal papillary mucinous neoplasms, MPD : Main pancreatic duct, PDAC : Pancreatic ductal adenocarcinoma, PNET : Pancreatic neuroendocrine tumor.*

Neoplasm (IPMN), Mucinous Cystic Neoplasm (MCN), Serous Cystic Neoplasm (SCN), and Solid Pseudo-Papillary Neoplasm (SPN), constituting about 90% of all cystic tumors (Del Chiaro et al., 2013). While some are benign, others exhibit a potentially malignant behavior or even serve as precursors to cancer, such as IPMN or MCN. Specifically, the risk of cancer in IPMN varies from 42% to 60%, depending on the type of involvement in the pancreatic ductal system (Del Chiaro et al., 2014). Given their substantial prevalence (4.2% for those under 60 and 11.4% for those over 60), IPMN poses a significant public health concern (Lévy et al., 2017). CT and/or MRI are recommended for all patients with cystic lesions of the pancreas to facilitate differential diagnosis and identify signs suggestive of malignancy (Del Chiaro et al., 2013).

2.4.2 Pancreatitis

Acute and chronic pancreatitis are currently recognized as distinct manifestations of pancreatic inflammation (Whitcomb, 2004). The former is characterized by sudden inflammation of the pancreas, often presenting with severe abdominal pain, while the latter involves recurrent inflammation leading to long-term structural changes, including fibrosis and pancreatic dysfunction (Mitchell et al., 2003). During a pancreatitis episode, digestive enzymes become active within the pancreas before being released into the duodenum. As a result, these enzymes attack the pancreas, initiating an inflammatory process that can

ultimately lead to complete pancreatic deterioration (Rawla et al., 2019). Pancreatitis can be visualized with medical imaging techniques, and its various characteristics are assessed to diagnose and understand the condition. As shown in Figure 2.10, specific features observed can vary between acute and chronic pancreatitis. The causes of pancreatitis are multifaceted, with tobacco playing a significant role and alcoholism acting as an exacerbating factor that accelerates disease progression. Pancreatitis can stem from various factors, including hereditary, idiopathic, autoimmune, post-endoscopic, and even tumor-related, concealing a malignant cyst or cancer (1% of acute pancreatitis cases) (Lee and Papachristou, 2019; Lévy et al., 2017). Consequently, any pancreatitis evaluation should include comprehensive imaging (CT or MRI), primarily focusing on the pancreatic duct. This approach could help in identifying cases of undiagnosed cancers that are operable at the initial stage of the first acute pancreatitis (Lévy et al., 2017).

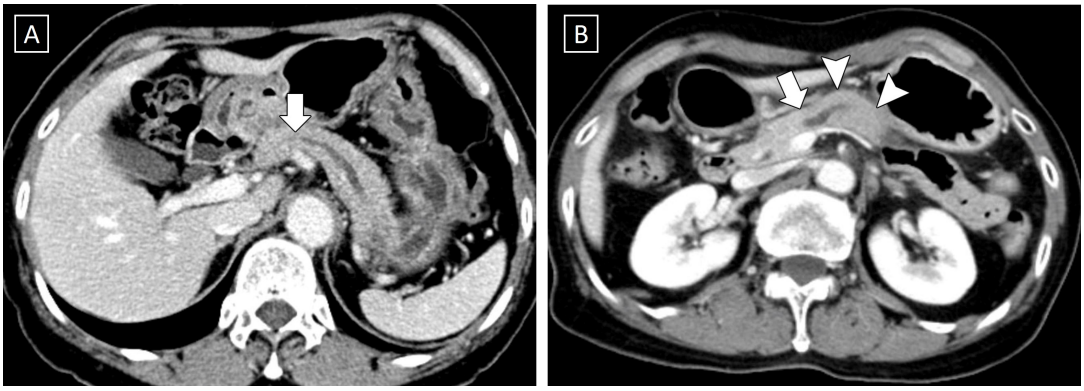


Figure 2.10 – Pancreatitis seen on axial contrast enhanced portal venous CT scan. A : Chronic pancreatitis. Coarse pancreatic calcifications are seen accompanied by duct dilatation, inferring chronic pancreatitis. **B : Acute pancreatitis.** Pancreatic swelling and peripancreatic stranding. The pancreas is enhancing homogeneously, with no signs of necrosis on CT and no pancreatic duct dilatation. Images from Radiopaedia (b). *CT : Computed tomography.*

2.5 Current and emerging challenges in the landscape of pancreatic cancer

2.5.1 Current landscape of pancreatic cancer

Clinical challenges Pancreatic cancer poses significant challenges in the field of oncology due to its aggressive nature, high mortality rates, and low five-year survival rates. The primary reason for these statistics is the frequent occurrence of late-stage diagnoses, when the cancer has already spread beyond resectable limits. Therefore, 80-90% of patients have unresectable tumors at the time of diagnosis (Rawla et al., 2019). Late diagnosis of PDAC is a consequence of lack of specific symptoms coupled with hard identification of pancreatic lesions, which is caused by several factors :

- First, the location of the pancreas, combined with its small size and variable shape, makes its examination challenging. Parts of the pancreas can be easily overlooked, especially during image readings primarily focused on exploring other organs.
- Secondly, portal-phase CT, which is the first-line imaging modality for clinical routine of the abdomen, does not maximize the density difference between the parenchyma and pancreatic lesions (Brennan et al., 2007). Thus, for early and/or small lesions, this modality is less sensitive (58–77% for lesions under 2 cm) (Shrikhande et al., 2012). Pancreatic-phase CT scans or EUS would be more appropriate, yet their application is limited to suspected pancreatic cancer cases.
- The third difficulty related to pancreas imaging arises from isodense lesions for which the prevalence is estimated at 27-33% (Elbanna et al., 2020; Goodman et al., 2012). Detection of these lesions can only occur through secondary signs, but these can be missed if not actively investigated. Moreover, these secondary signs are not always present, with studies reporting the absence of secondary signs in 10-14% of cases (Tamada et al., 2016; Yoon et al., 2011).

In addition to these difficulties which are specific to pancreatic imaging, radiologists also face more general challenges such as heavy workload and varying levels of expertise which might further affect the interpretation of CT scans (Mazur et al., 2014).

Improvements over the past year Despite these challenges, there have been notable improvements in pancreatic cancer management. Technological enhancements in scanners, including improved resolution and multi-phase imaging, have substantially improved diagnostic capabilities (Costache et al., 2017). Surgical advancements, including refined techniques and preoperative chemotherapy, have contributed to improved survival rates, especially in less-advanced cases (Lévy et al., 2017; Sauvanet et al., 2014). As depicted in Figure 2.11, the overall survival following a diagnosis of PDAC exhibited improvement from 2004 to 2016 across all tumor stages. This enhanced survival outcome was facilitated by surgical progress, including refined techniques and preoperative chemotherapy, particularly benefiting less-advanced cases (Lévy et al., 2017; Sauvanet et al., 2014). Specifically, five-year overall survival for stage IA cases improved from 44.7% (95% Confidence Interval (CI) = 34.1% to 63.7%) in 2004 to 83.7% (95% CI = 78.6% to 89.2%) in 2012 (see Figure 2.11 from Blackford et al. (2020)).

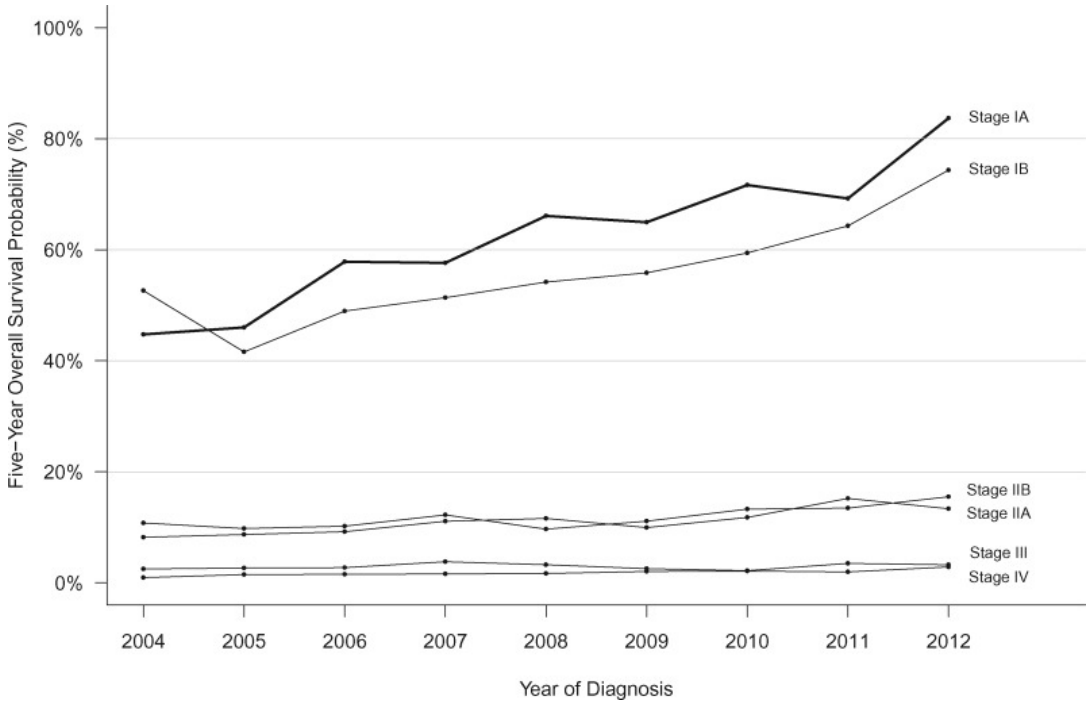


Figure 2.11 – Trends in five-year overall survival probabilities according to pancreatic cancer stage (2004–2012). Adapted from Blackford et al. (2020).

2.5.2 Future directions

In spite of these improvements, there is a need for continued efforts. The anticipated surge in pancreatic cancer incidence in the coming years necessitates a dual focus on prevention and early detection.

Prevention Emphasizing the role of modifiable factors, such as tobacco use, obesity, alcohol consumption, and sedentary lifestyle, holds promise in reducing pancreatic cancer risk. Prevention strategies targeting tobacco use alone could potentially avert 30% of pancreatic cancers (Rawla et al., 2019).

Early detection Research efforts should prioritize early detection, as most patients diagnosed with advanced or metastatic tumors have a significantly worse prognosis than those diagnosed with less advanced

tumors (Gonoi et al., 2017; Mizrahi et al., 2020) : as depicted in Figure 2.11, the five-year survival rate surpasses 70% for stage I tumors, whereas later stages do not exceed 20%. Therefore, facilitating early diagnosis, which in turn allows for surgical intervention, represents the most effective approach to enhance the five-year survival rate (Gonoi et al., 2017).

To this end, some studies explore the use of less invasive liquid biopsies, but these methods face challenges in terms of low specificity and sensitivity since tumor DNA is detectable in only 50% of less advanced tumors (Bettegowda et al., 2014). Thus, imaging emerges as a critical area to further explore, with significant potential interest in Artificial Intelligence (AI)-based methods for this purpose.

2.5.3 Role of artificial intelligence

Image interpretation could be facilitated through AI methods that systematically review images in a clinical-routine scenario (Pereira et al., 2020). Such AI-based computer-aided detection systems have been successfully applied to several diseases such as breast cancer or lung cancer (Fujita, 2020), and their application to pancreatic cancer would be of the greatest importance due to the difficulty of its detection (Barat et al., 2021). However, applying AI in the context of pancreatic cancer presents several key challenges that need to be addressed :

- **Data collection and annotation :** Collecting and annotating the necessary data are paramount for building AI models. This involves acquiring appropriate datasets that cover a diverse range of cases, and establishing comprehensive yet effective annotation process to facilitate model training and evaluation. Notably, decisions regarding the type of annotation, such as determining which questions warrant manual segmentation and which suffice with case-level labels, are crucial. This will be addressed in Chapter 4.
- **Accurate pancreas segmentation :** A precise delineation of the pancreas is essential to establish a reliable Region of Interest (ROI) for subsequent analysis. This necessitates the development of a pancreas segmentation method capable of robustly handling the elongated shape of the organ. Specifically, the method must effectively identify the pancreas extremities, which often pose challenges due to their low contrast with surrounding tissues. This will be addressed in Chapter 5.
- **Lesion detection :** Building a robust lesion detection algorithm is crucial, especially for identifying small and/or isodense lesions which are characteristic of early-stage pancreatic cancer. State-of-the-art segmentation models combined with a large database including diverse pancreatic diseases, various geographical origins, and multiple scanner manufacturers could help building such a robust algorithm. In addition, the use of radiological secondary signs like Main Pancreatic Duct (MPD) dilatation can enhance detection but it also introduces the challenge of efficiently extracting and integrating this information. This will be addressed in Chapter 6.
- **Secondary signs detection :** In order to address early cancer detection, identifying secondary signs suggestive of malignancy is crucial (Kim et al., 2017). However, while the dilatation of the MPD is a clearly identifiable sign as it relies on the diameter of the MPD, the other secondary signs lack of quantitative measure. Atrophy, abnormal shape, senility and fat replacement are crucial secondary signs which are visually assessed, leading to variability between radiologists. Thus, predicting these signs requires distinct methods compared to detecting MPD dilatation, as there is not explicit anatomical information. Instead, we need to rely on strategies that extract discriminating features from the image. This will be addressed in Chapter 7.

All the methods developed in this thesis are designed to use portal phase CT scans. As discussed in Section 2.3.3.2, while arterial phase or pancreatic phase scans offer superior differentiation between normal parenchyma and potential tumors, they are typically prescribed only when there is suspicion of pancreatic disease, usually in symptomatic patients. Consequently, providing assistance to radiologists at this stage may offer limited utility as thorough pancreas analysis is standard practice in such cases. Conversely, focusing on portal phase scans, which are routinely conducted during abdominal examinations, could provide substantial value to radiologists. This approach is underscored by findings by Kang et al. (2021), which indicate that a majority of missed pancreatic cancer diagnoses occur in single-phase examinations during

the portal venous phase. Enhancing radiologists' capabilities during this critical phase, characterized by reduced contrast, could help with difficult diagnoses and increase the likelihood of incidental findings—i.e., discoveries in patients not suspected of having pancreatic cancer, as detailed in Section 2.3.3.4. This, in turn, could significantly elevate the chances of detecting early-stage cancers and improving patient survival rates (Barat et al., 2021). For these reasons, this thesis centers on portal CT scans. Nevertheless, the potential use of arterial scans during the training phase is explored as a prospective avenue in Section 8.

Chapitre 3

Deep Learning Methods for Pancreatic Cancer

3.1 Preface

The field of medical image analysis has witnessed remarkable progress with the advent of deep learning techniques. In particular, the application of deep learning in pancreatic analysis has emerged as a critical area of research, offering unprecedented opportunities for improved diagnostic. This chapter aims to provide a comprehensive overview of the current landscape of deep learning applications in pancreas analysis.

We start by introducing fundamental concepts in deep learning, which are essential for a comprehensive understanding of our literature review. This introduction unfolds through two primary tasks, namely classification and segmentation. For each of these tasks, we will elucidate the principles, the main methods and their application in medical image analysis.

Having laid down these aspects, we will explore the state-of-the-art of deep learning methods applied to pancreas imaging. This section will be divided into three parts : pancreas segmentation, lesion segmentation and detection, and feature extraction for pancreas analysis.

3.2 Deep learning fundamentals

3.2.1 Introduction

Deep Learning (DL) is a subfield of Machine Learning (ML), itself a subset of Artificial Intelligence (AI). DL encompasses methods using multi-layered Neural Network (NN)s, also known as Deep Neural Networks (DNN)s. NNs consist of interconnected nodes (neurons) organized into layers to process input data, such as images (LeCun et al., 2015). Each connection between neurons is associated with a weight, adjusted during model training to achieve desired predictions. DL employs DNNs with multiple layers to learn hierarchical representations, essential for addressing complex problems like image understanding (Cai et al., 2020; LeCun et al., 2015). NN and DNN are depicted in Figure 3.1.

In the context of pancreatic analysis, DL has been employed for various purposes. We categorize these applications into two tasks : classification and segmentation. We provide a concise overview of the fundamental principles and methods for both of these tasks.

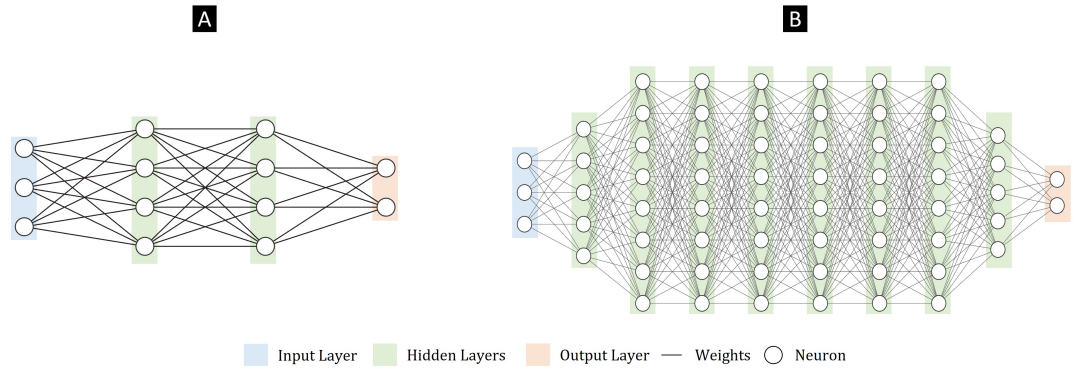


Figure 3.1 – Deep learning architectures. A : Illustration of a Neural Network (NN). B : Illustration of a Deep Neural Network (DNN). The NN, on the left, consists of a two hidden layers, while the DNN, on the right, comprises multiple hidden layers. The additional depth allows the DNN to capture more complex hierarchical features.

Convolutional Neural Network (CNN)s are a type of DNNs. They are particularly powerful for feature extraction from images as they automatically learn hierarchical representations of visual data, capturing complex patterns and features at different scales (Albawi et al., 2017; LeCun et al., 1995). The learned features from early to deep layers capture both low-level details and complex, abstract representations, providing a rich and discriminative set of features for downstream tasks. As depicted in Figure 3.2, CNNs use convolutional layers that apply filters to small input regions and preserve local spatial relationships in images. Alongside convolutions, CNNs employ operations like pooling to reduce spatial dimensions and extract relevant features. The convolutional and pooling layers are often followed by activation functions that add non-linearity to the neural networks, enabling them to learn complex patterns in the data. The succession of these operations serve to encode the information, and as such, CNNs are commonly referred to as *encoders*. The resulting vector from this encoding is often referred to as the *latent space* : a compressed, abstract, and often continuous representation of the input data.

3.2.2 Classification

3.2.2.1 Principle

Image classification involves assigning a class or label to an image. Typically, this process is carried out using features that help summarize the information contained in the image. These features are extracted from the images and then used to train a classification model. Through supervised training[1], the model learns the relationships between these features and the different classes, enabling it to accurately classify new images based on their distinctive features.

1. Supervised machine learning relies on labeled input and output training data. It is opposed to unsupervised learning that processes unlabeled or raw data.

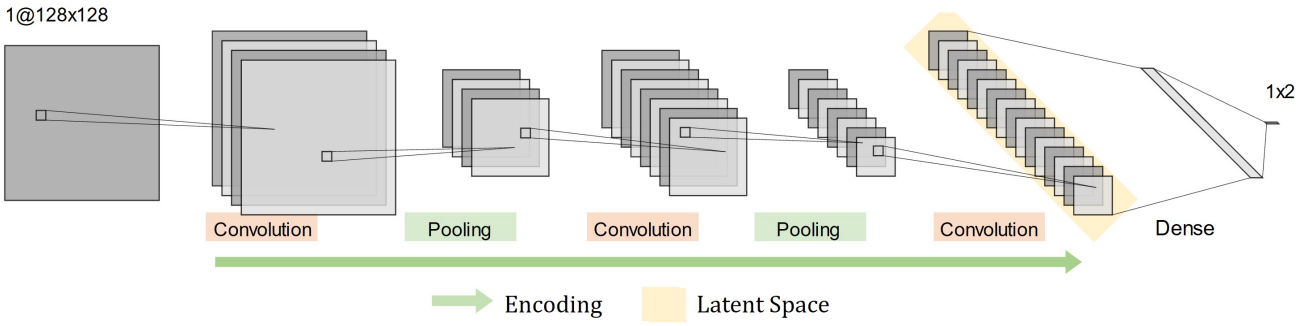


Figure 3.2 – Illustration of a Convolutional Neural Network (CNN) for end-to-end image classification. Layers in a CNN include convolutional layers, pooling layers, and dense layers. Through the application of filters to small input regions, convolutional layers preserve local spatial relationships in images and enable the network to learn hierarchical features such as edges, textures, and complex patterns. Pooling layers reduce the size of the feature maps, making the model more computationally efficient. Dense layers are typically found after convolutional and pooling layers of a CNN, they connect all the neurons in a layer to all the neurons in the next layer, allowing the model to learn possible non-linear combinations of the features learned. The successive layers contribute to the formation of a latent space representation (in yellow), enhancing the ability of the model to process and classify images.

3.2.2.2 Methods

We distinguish two approaches to perform classification :

(i) Feature extraction followed by a down-stream supervised classification model. Typical ML algorithms like support vector machines, decision trees, or random forests are commonly employed for this classification task due to their interpretability, efficiency, and effectiveness (Pedregosa et al., 2011).

(ii) Feature extraction and classification performed jointly by a single end-to-end DL model.

During the feature extraction phase, the goal is to distill a subset of features that effectively captures the distinctive characteristics of the image, thereby simplifying the subsequent classification task. This process is essential for converting raw image data into a more concise and informative representation, which is crucial for accurate and efficient classification. In this manuscript we will consider two categories of features :

- **Hand-Crafted Radiomics (HCR)**, which are are pre-designed features, based on intensity, shape, and texture (Kumar et al., 2012; Lambin et al., 2012). They have demonstrated diverse applications in cancer-related tasks such as diagnosis, malignancy prediction, or survival prediction (Marti-Bonmati et al., 2022). Numerous Python libraries or specialized software enable the computation of these HCR ; however, their calculation is usually not entirely automatic, involving several steps like segmentation, preprocessing, and feature selection (Van Griethuysen et al., 2017).
- **Deep-Learning Radiomics (DLR)**, which can be obtained using DL models trained to extract highly informative features in a fully automatic manner (Kumar et al., 2015; Shafiee et al., 2017). CNNs are particularly well-suited for this task as they excel at processing structured grid data like images, hence offering the potential to unveil complex patterns (Scapicchio et al., 2021). Therefore, latent spaces of CNNs are typically used as DLR feature vector (Afshar et al., 2019).

3.2.2.3 Optimization

Classification models are typically trained by optimizing a suitable loss function. In the realm of possibilities, two commonly employed loss functions for classification tasks are binary cross-entropy and categorical cross-entropy.

Binary cross-entropy In binary classification, where the model assigns samples to one of two classes, binary cross-entropy measures the dissimilarity between the true binary labels and the predicted probabi-

lities :

$$\text{Binary Cross-Entropy Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \cdot \log(p_i) + (1-y_i)\cdot\log(1-p_i)]$$

where N represents the number of samples in the dataset, y_i is the true label (either 0 or 1) for the i-th sample, and $p_i \in]0,1[$ is the predicted probability that the i-th sample belongs to class 1.

Categorical cross-entropy For multi-class classification, categorical cross-entropy extends the concept of binary cross-entropy to handle multiple classes. It quantifies the difference between the true class distribution and the predicted probabilities across all classes :

$$\text{Categorical Cross-Entropy Loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{C} y_i^k \cdot \log(p_i^k)$$

where N is the number of samples in the dataset, C is the number of classes, y_i^k is an indicator variable that equals 1 if the i-th sample belongs to class k and 0 otherwise, $p_i^k \in]0,1[$ is the predicted probability that the i-th sample belongs to class k.

Minimization of the loss function Once an appropriate choice of the loss function is made, the next step in training a classification model involves minimizing this function to achieve a well-performing model. Some classical classification models such as logistic regression or linear discriminant analysis can be optimized analytically. However, as the complexity of the models increases, numerical optimization methods become necessary. For instance, in the case of DL classification networks, the widely used gradient-based optimization methods such as stochastic gradient descent (Bottou, 2010) and Adam (Kingma and Ba, 2014) are employed. This optimization process also entails selecting hyperparameters, such as learning rate or batch size, that significantly influence the model performance.

3.2.2.4 Evaluation measures

In the evaluation of classification models, several key evaluation measures are employed to quantify their performance.

Accuracy Accuracy is a fundamental measure representing the ratio of correctly classified instances (true positives and true negatives) to the total number of instances :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Sensitivity Sensitivity, also known as recall, measures the model ability in identifying positive instances :

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity Specificity measures the model ability in identifying negative instances :

$$Specificity = \frac{TN}{TN + FP}$$

Area under the receiver operating characteristic curve The Area Under the Curve (AUC)-Receiver Operating Characteristic (ROC) serves as a comprehensive measure for assessing a model discriminative power across varying classification thresholds. This measure represents the area under the ROC curve, a plot of the true positive rate (other name for sensitivity) against the false positive rate ($1-$specificity), with values closer to 1 indicating superior performance. The AUC-ROC metric captures the trade-off between sensitivity and specificity, providing insights into the robustness and reliability of a model.

3.2.2.5 Standard DL networks for classification

We introduce several conventional DL networks designed for classification tasks. We start with the autoencoder and its variational variant, known as the variational autoencoder. These architectures enable the extraction of features that can be further used by a downstream supervised classification model. We then introduce the ResNet, a well-established CNN for end-to-end classification

Autoencoder Autoencoder (AE), depicted in Figure 3.3, are a type of NN designed for unsupervised learning that extract meaningful features from input data (Bank et al., 2023). The key principle of an autoencoder involves training the network to encode the input data into a latent space representation and then reconstruct the input as accurately as possible. AEs consist of an encoder and a decoder : the former compresses the input data into a lower-dimensional representation, while the latter attempts to reconstruct the original input from this compressed representation. As the AEs learn to capture essential features during the training process, they are commonly used for feature extraction purposes.

The most common loss function for training an AE is the Mean Squared Error (MSE) which measures the mean squared difference between the input data and the reconstructed output :

$$\text{MSE loss function} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2$$

where N represents the number of input samples, x_i is the original input, and $\hat{x}_i$ is the reconstructed output.

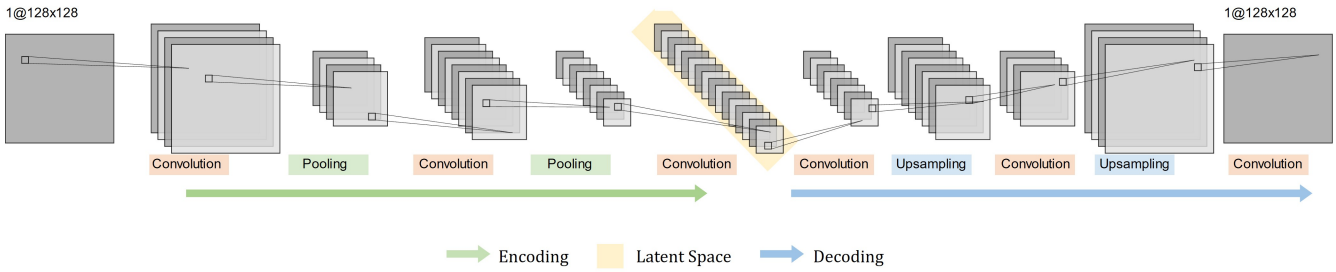


Figure 3.3 – Illustration of an Autoencoder (AE). The AE is trained to perform two key processes : an encoding process compresses input data into a lower-dimensional latent space (depicted in yellow), and a decoding process reconstructs the input from this latent space. The resulting latent space vector serves as a compressed representation of the input data, offering a useful resource for classification purposes.

After training, mapping an input image into the latent space of the AE allows for the extraction of a feature vector that encapsulates the unique characteristics of the image. This resultant feature vector can then be employed to address a classification problem.

Variational autoencoder Variational Autoencoder (VAE)s represent a powerful variant of AEs that introduces probabilistic principles into the encoding-decoding framework (Kingma and Welling, 2014). The fundamental principle of VAEs involves learning a probabilistic mapping from input data to latent space, emphasizing the statistical distribution of latent representations. Unlike traditional AEs, VAEs generate a

distribution of possible latent space representations for each input, allowing for more robust and expressive feature extraction. The encoder in a VAE maps input data to the parameters of a probability distribution, and the decoder samples from this distribution to reconstruct the input.

This probabilistic nature enables VAEs to capture the uncertainty inherent in real-world data. VAEs are particularly beneficial for feature extraction as they not only provide a compact and meaningful representation but also offer a continuous and smooth latent space, enabling interpolation between different data points. This capability is crucial in applications where understanding the variability and uncertainty in data representations is essential.

The loss function for training a VAE consists of two main components : the reconstruction loss (RL) function and the Kullback-Leibler (KL) divergence term. The former measures how well the model can reconstruct the input data, it is often formulated as the negative log-likelihood of the data given the parameters predicted by the model. The latter regularizes the latent space by encouraging it to follow a specific distribution, typically a multivariate normal distribution. This term helps in making the latent space more continuous and encourages the model to generate diverse and meaningful samples. Mathematically, the VAE loss function can be expressed as :

$$\text{VAE loss function} = -\frac{1}{N}\sum_{i=1}^{N}\left[\text{RL}(x_i, \hat{x}_i) + \text{KL}(q(z_i|x_i)\|p(z_i))\right]$$

where N is the number of input samples, x_i is the original input, $\hat{x}_i$ is the reconstructed output, $q(z_i|x_i)$ is the distribution of the latent variable z_i given the input data x_i, $p(z_i)$ is the prior distribution of the latent variable z_i. The specific form of the RL function and the KL divergence term depends on the nature of the data (continuous or binary) and the chosen distributions for the latent space.

Similar to the AE, projecting an input image into the latent space of the VAE allows for the extraction of a feature vector which can subsequently be used to tackle a classification problem.

ResNet Within the CNNs performing end-to-end classification, ResNet (short for Residual Networks) is a notable architecture. Proposed by He et al. (2016), ResNet introduces the concept of residual learning, which involves constructing blocks of layers that learn residual functions to be applied to the input. The key innovation lies in the use of skip connections, or shortcuts, which allow the network to skip one or more layers during training. This mitigates the vanishing gradient problem[2], making it easier to train extremely deep networks.

For classification tasks, ResNet has demonstrated exceptional performance. Its ability to train very deep networks helps in learning hierarchical features and representations, making it particularly effective for image classification (Jiang et al., 2019; Sarwinda et al., 2021). The skip connections help with the smooth flow of gradients during backpropagation, enabling efficient training of deep neural networks. As a result, ResNet has become a popular choice for various computer vision applications (Baltruschat et al., 2019).

3.2.3 Segmentation

3.2.3.1 Principle

The goal of segmentation is to correctly identify and delineate the boundaries of different objects present in an image. In pratice, it involves the assignment of a predefined class to each pixel (or voxel in 3D) in an image, as illustrated in Figure 3.4. These labeled pixels delineate either foreground objects or background, collectively constituting the image (Sharma and Aggarwal, 2010). In classification, a class is linked to the entire image or an object within it, whereas in segmentation, a class is assigned to each individual pixel. This distinction underscores the finer level of detail achieved by segmentation, as it precisely identifies the

2. Challenge encountered during the training of very deep networks. It occurs when the gradients of the loss function with respect to the parameters of the network become extremely small as they are backpropagated through the layers during the training process. It prevents the effective update of weights in the early layers.

classification for every pixel in the image. The resulting segmentation map provides detailed information on the spatial distribution of objects in images, which is crucial for a precise semantic understanding of the images.

Segmentation is a task of major interest in medical imaging (Sharma and Aggarwal, 2010). It enables the precise identification and localization of complex anatomical structures such as organs or tumors, thereby facilitating the detection of certain pathologies. The detailed anatomical understanding it provides aligns with the development of personalized medicine and paves the way for more targeted approaches to surgical interventions or radiotherapies (Wang et al., 2022). Furthermore, segmentation facilitates the quantitative analysis of images as it allows the calculation of certain measures such as the size or volume of specific structures. These quantitative measures are essential for monitoring disease progression or evaluating the effectiveness of certain treatments (Djuric-Stefanovic et al., 2012; Yoon et al., 2020).

The advent of computer-aided diagnostic tools has automated the segmentation of medical images, significantly reducing the workload of radiologists. Whether directly used by physicians or serving as intermediaries for more complex methods, automatic segmentations are indispensable tools for the analysis of medical images. Consequently, this is a highly active research area, and numerous methods and applications have been proposed.

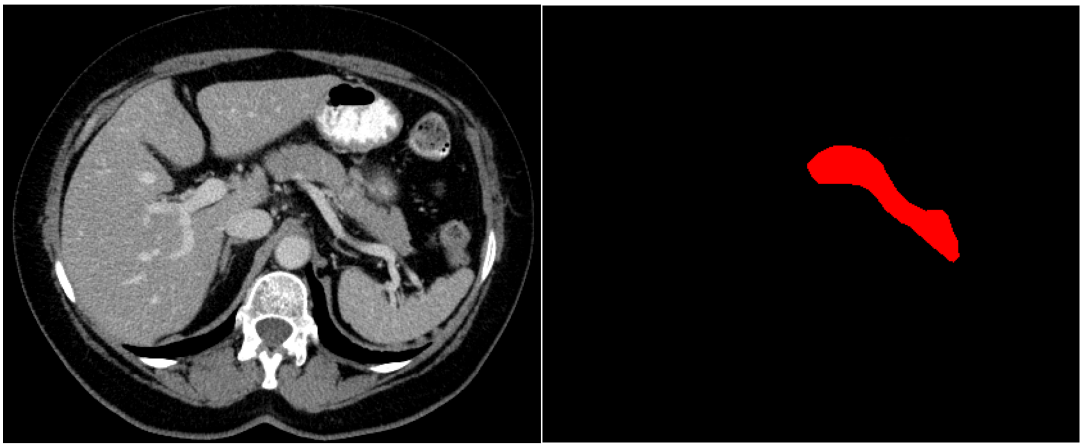


Figure 3.4 – Pancreas Segmentation on Portal CT Scan. Left : Original image. **Right** : Segmentation outcome. The pancreas voxels are highlighted in red (label = 1), while the background is represented in black (label = 0). *CT : Computed tomography.*

3.2.3.2 Methods

Early segmentation methods include distinct categories, such as grayscale-based techniques, texture features-based methods or atlas-based methods (Sharma and Aggarwal, 2010; Yao et al., 2020a). With the explosion of computing capabilities and data storage, ML has gradually emerged as a tool that can be used to replace or complement early segmentation methods to enhance their performance. While more traditional machine learning methods (such as support vector machine or random forests) have been employed, DL methods are now the most prevalent (Asgari Taghanaki et al., 2021). These methods use NN with multiple layers, enabling the recognition of complex patterns.

Many DL approaches have been presented in the literature, with many relying on Fully Convolutional Network (FCN)s. FCN, a type of CNN, is specifically designed to process input data of arbitrary sizes and generate outputs of corresponding dimensions (Long et al., 2015). FCNs comprise an encoder complemented by a decoder with upsampling layers to restore the original resolution. The encoder captures image context, increasing semantic understanding, while the decoder precisely locates contextual information obtained from the encoder. FCNs are widely adopted for image segmentation tasks, and numerous reviews detail variations in model types, cost functions, architectures, supervision modes, or training methodologies (Asgari Taghanaki et al., 2021; Sharma and Aggarwal, 2010; Wang et al., 2022).

3.2.3.3 Optimization

Two loss functions widely employed for segmentation tasks are the cross entropy loss function and the soft Dice loss function (Ma et al., 2021).

Cross entropy loss function Segmenting an image can be viewed as a pixel-level classification problem. Thus, the cross-entropy loss function, presented in Equation 3.2.2.3 for a multi-class classification problem, can be adapted for a segmentation task :

$$\text{Categorical Cross-Entropy loss function} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{C} y_i^k \cdot \log(p_i^k)$$

where N is the number of voxels in the image, C is the number of classes, y_i^k is an indicator variable that equals 1 if the i-th voxel belongs to class k and 0 otherwise, p_i^k is the predicted segmentation probability for the i-th voxel belonging to class k.

Soft Dice loss function The soft Dice loss function is another commonly used loss function for segmentation, aiming to maximize the Dice Similarity Coefficient (DSC) (Dice, 1945). The DSC measures the similarity between a reference segmentation Y and a predicted segmentation P :

$$DSC = 2 \times \frac{|Y \cap P|}{|Y| + |P|} \tag{3.1}$$

The DSC ranges from 0 to 1, with a DSC of 1 indicating a perfect overlap between the two segmentations. The soft Dice loss function seeks to maximize a differentiable reformulation of the DSC and is defined as :

$$\text{Soft Dice loss function} = 1 - \frac{2 \times \sum_{i=1}^{N}\sum_{k=1}^{C} y_i^k p_i^k}{\sum_{i=1}^{N}\sum_{k=1}^{C} y_i^k + \sum_{i=1}^{N}\sum_{k=1}^{C} p_i^k}$$

Some studies combine the soft Dice loss function with the cross-entropy loss function, as this combination helps achieve a balanced trade-off between accurate pixel-wise predictions and spatially coherent segmentation (Isensee et al., 2021a; Ma et al., 2021).

Minimization of the loss function Similar to classification models, when an appropriate loss function is selected, the subsequent step in training a segmentation model is to minimize this function. DL-based segmentation models commonly use gradient-based optimization methods, including stochastic gradient descent (Bottou, 2010) and Adam (Kingma and Ba, 2014). As in classification tasks, this optimization process involves tuning hyperparameters such as learning rate or batch size.

3.2.3.4 Evaluation measures

Although there are numerous evaluation measures available to quantify the performance of a segmentation model, providing a detailed overview of all these measures is beyond the scope of this section. Therefore, we will focus on the aforementioned DSC and the normalized surface Dice.

Dice similarity coefficient The DSC, defined in Equation 3.1), can also be defined using the definition of true positive (TP), false positive (FP), and false negative (FN) as :

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.2}$$

Figure 3.5 provides an illustration of the DSC calculation.

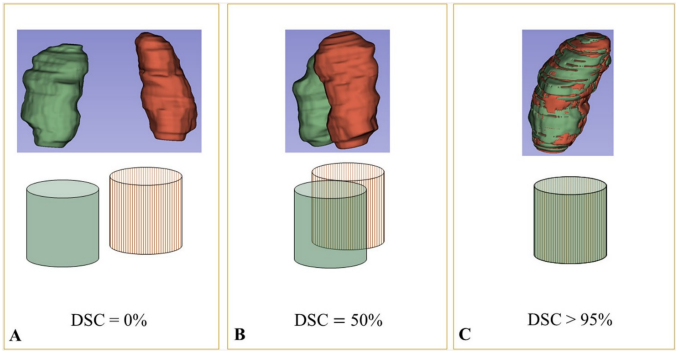


Figure 3.5 – Visualization of the Dice similarity coefficient (DSC) calculation. A higher DSC indicates a greater overlap between the two segmentations masks (depicted in red and green). Panel A depicts no overlap, resulting in a 0% DSC. Panel B illustrates partial overlap, leading to a 50% DSC. In Panel C, complete overlap is observed, resulting in a 100% DSC. Illustration from Barat et al. (2021).

Normalized surface Dice Unlike the DSC which assesses the overlap of two volumes, the Normalized Surface Dice (NSD) (Nikolov et al., 2018) assesses the overlap of two surfaces at a specified tolerance τ. More specifically, the NSD measures the fraction of a segmentation boundary which is correctly predicted. A boundary element is considered correctly predicted if the closest distance to the reference boundary is smaller than or equal to the specified threshold τ, related to the acceptable amount of deviation in voxels. It is calculated as follows.

For all the voxels belonging to the boundary of the predicted segmentation P, the nearest-neighbor distances to the boundary of the reference segmentation Y are computed, resulting in a multiset of distances $\mathcal{D}_P$. Then, $\mathcal{D}'_P$ is determined as the multiset of distances that are smaller or equal to the acceptable distance τ :

$$\mathcal{D}'_P = \{\!\{d \in \mathcal{D}_P |\ d \leq \tau\}\!\}$$

Then, the procedure is symmetrically repeated for Y, yielding $\mathcal{D}_Y$ and $\mathcal{D}'_Y$. The NSD is then computed as :

$$NSD = \frac{|\mathcal{D}'_Y| + |\mathcal{D}'_P|}{|\mathcal{D}_Y| + |\mathcal{D}_P|} \tag{3.3}$$

The NSD is bounded between 0 and 1, with 0 indicating that either the boundary is completely off, or that the class is present in the image but has not been predicted. A NSD of 1 indicates that all the distances are smaller than the acceptable deviation τ (Seidlitz et al., 2022). As illustrated in Figure 3.6, the NSD can be thought as the fraction of a segmentation boundary that would have to be redrawn to correct for segmentation errors. In this respect, it differs from the DSC which weighs all regions of misplaced delineation equally and independently of their distance from the surface.

3.2.3.5 Standard DL networks for segmentation

U-Net The U-Net, introduced by Ronneberger et al. (2015), is a specific type of FCN featuring an encoder with consecutive convolutional blocks and downsampling steps, and a symmetric decoder with upsampling layers. The architecture, depicted in Figure 3.7, resembles a U-shape, and skip connections connect corresponding layers of the encoder and decoder, facilitating information flow between layers with varying semantic richness and spatial resolution.

Since its introduction in 2015, the U-Net architecture has emerged as one of the most widely adopted models for medical image segmentation (Antonelli et al., 2022; Kelly et al., 2022; Ma et al., 2021). Several variants have been proposed, demonstrating not only the adaptability and versatility of the original design but also its inherent robustness. Some notable variants include Res-UNet (Xiao et al., 2018), Attention U-Net (Oktay et al., 2022), UNETR (Hatamizadeh et al., 2022), U-Net 3+ (Huang et al., 2020), and Swin-Unet Cao et al. (2022).

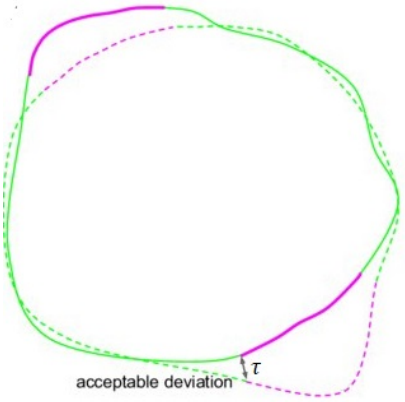


Figure 3.6 – Visualization of the normalized surface Dice (NSD) calculation. *Continuous line* : predicted surface. *Dashed line* : reference surface. *Black arrow* : the maximum margin of deviation τ which may be tolerated without penalty. *Green* : acceptable surface parts (distance between surfaces $\leq \tau$). *Pink* : unacceptable regions of the surfaces (distance between surfaces $> \tau$). The NSD reports the correct surface parts compared to the total surface. Illustration from Nikolov et al. (2018).

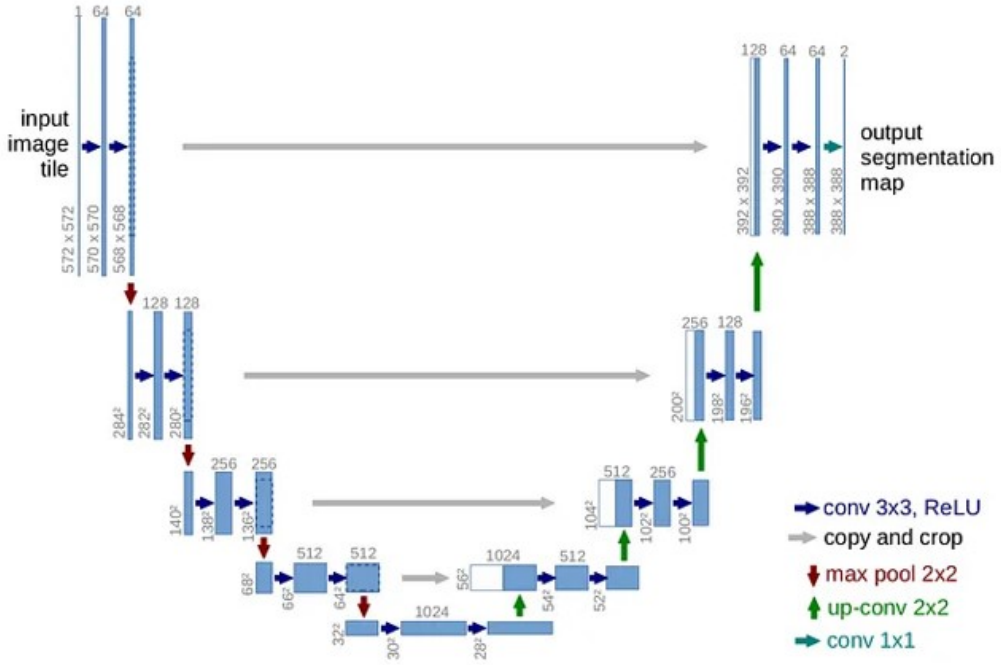


Figure 3.7 – U-Net architecture as introduced by Ronneberger et al. (2015)

nnU-Net The nnU-Net, short for "no-new-Net," is a self-adapting deep learning framework for medical image segmentation proposed by Isensee et al. (2021a). Built upon the traditional U-Net architecture, the nnU-Net stands out by its automatic adaptation and configuration based on the unique properties of a given dataset. When faced with a new dataset, the nnU-Net systematically analyzes the training cases and generates a unique "dataset fingerprint" captures parameters related to the image size, the spacing between the different voxels, the type of imaging modality, the total number of classes, and the total number of training cases. The segmentation pipeline is then defined based on three fundamental principles :

— Fixed parameters, such as data augmentation or learning rate, remain unaltered as the authors have identified a robust configuration that proves effective across diverse scenarios.

— Rule-based parameters leverage the dataset fingerprint to adapt specific properties of the segmentation pipeline through predefined heuristic rules. For instance, network topology, patch size, and batch size

are jointly adjusted considering GPU memory constraints.

— Empirical parameters involve a trial-and-error approach, including tasks like selecting the optimal U-Net configuration (2D, 3D full resolution, 3D low resolution, 3D cascade) and refining the postprocessing strategy.

The nnU-Net also distinguishes itself through its image preprocessing pipeline, automatically configured based on the rule-based parameters derived from the data fingerprint. Two automatic preprocessing steps are defined :

1. **Intensity normalization** : The preprocessing pipeline includes intensity normalization, with two supported schemes. For all modalities except Computed Tomography (CT) images, each image undergoes independent normalization by subtracting its mean and dividing by its standard deviation (z-score normalization). For CT images, where intensity values are in Hounsfield Unit (HU), reflecting physical tissue properties, a global normalization is applied. To retain quantitative information, nnU-Net uses the 0.5 and 99.5 percentiles of foreground voxels for clipping, along with the global foreground mean and standard deviation for normalization across all images.
2. **Resampling to a target spacing** : Due to heterogeneous voxel spacing across diverse datasets, nnU-Net resamples all images to a common target spacing, ensuring the CNN model operates on voxels with the same physical spacing. For the 3D full-resolution U-Net, the default target spacing is the median value of spacings in the training cases, computed independently for each axis. In anisotropic datasets (maximum axis spacing / minimum axis spacing > 3), the default rule may lead to interpolation artifacts or significant information loss. Therefore, the target spacing for the lowest resolution axis is set to the 10th percentile of spacings. Once the target spacing is set, images and segmentation maps are resampled to this spacing, and default resampling uses third-order spline interpolation. For anisotropic images, out-of-plane axes are treated differently to reduce resampling artifacts : in-plane resampling employs third-order spline, while out-of-plane interpolation uses nearest neighbor. Resampling of the segmentation maps starts by converting them into one-hot encodings. Then, each channel is interpolated using linear interpolation, and the segmentation map is obtained through an argmax operation. Again, anisotropic cases are interpolated using "nearest neighbor" on the low-resolution axis.

Thanks to this holistic approach, the nnU-Net has gained attention for its ability to achieve state-of-the-art segmentation results in various medical imaging tasks, demonstrating its potential as a robust tool for accurate and efficient medical image segmentation. Today, the nnU-Net is considered as a baseline and method development framework : 9 out of 10 challenge winners at MICCAI 2020 (Antonelli et al., 2022) and 5 out of 7 in MICCAI 2021 built their methods on top of nnU-Net (Isensee et al., 2021b).

3.2.3.6 Applications in medical imaging

Due to its important role in the development of diagnostic assistance tools, segmentation has been applied across various domains and types of medical images. The most popular applications include liver and hepatic tumor segmentation, brain and brain tumor segmentation, optic disc segmentation, cell segmentation, lung and pulmonary nodule segmentation, and heart segmentation (Wang et al., 2022). The most commonly used modalities are X-rays, CT images, Magnetic Resonance Imaging (MRI) images, and ultrasound images (Wang et al., 2022).

Certain applications have gained popularity through challenges (such as BRATS[3], KiTS[4], or LiTS[5]) that provide public datasets and allow for quantitative comparisons of different proposed methods. Unlike most of these challenges, which are specific to a segmentation task (an organ on a modality), the Medical Segmentation Decathlon is a challenge including multiple tasks and modalities (Antonelli et al., 2022). Launched in November 2018, it aims to identify a versatile algorithm for medical image segmentation, capable of generalization and high performance across a multitude of organs. The challenge results confirm

3. Brain Tumor Segmentation challenge (BRATS)
4. Kidney Tumor Segmentation challenge (KiTS)
5. The Liver Tumor Segmentation benchmark (Bilic et al., 2023)

the hypothesis that a method capable of achieving good results across multiple tasks generalizes well to a new task, and may even outperform a method designed specifically for that application. Tables 3.1a and 3.1b show the DSC obtained by the top five methods for organ segmentation and lesion segmentation tasks, respectively. Task-specific performance reflect the intrinsic difficulty of different segmentation tasks. For organ segmentations (see Table 3.1a), performance for the liver and spleen are high (DSC around 95% and 97%, respectively) as these are large organs, easy to detect and delineate. By contrast, the methods for segmenting the pancreas and the prostate achieve lower performance (DSC around 82% and 76%, respectively), reflecting the difficulty of these tasks. For lesion segmentations (see Table 3.1b), lesions such as non-enhancing brain lesions, pancreatic cancers, and colon cancers have been the most challenging to segment (DSC around 49%, 59%, and 60%, respectively). These regions, characterized by small, heterogeneous, and subtle masses, are challenging for both algorithms and radiologists, showing high inter-annotator variability on these tasks (Re et al., 2011; Sirinukunwattana et al., 2017).

General rank	Team	Created	Mean Position	**Heart** *MRI*	**Liver** *ce-CT*	**Hippocampus** *MRI*		**Prostate** *multiparametric MRI*		**Pancreas** *ce-CT*	**Spleen*** *ce-CT*
						Anterior	Posterior	Peripheral zone	Transition zone		
1st	universal_model	Feb. 2023	5.6	0.93 (1)	0.95 (10)	0.90 (2)	0.89 (2)	0.76 (3)	0.76 (3)	0.83 (1)	0.97 (8)
2nd	Swin_UNETR	Nov. 2021	10.1	0.93 (29)	0.95 (11)	0.90 (17)	0.88 (18)	0.76 (7)	0.76 (7)	0.82 (6)	0.97 (14)
3rd	ahatamiz2	Nov. 2021	10.1	0.93 (29)	0.95 (11)	0.90 (17)	0.88 (18)	0.76 (7)	0.76 (7)	0.82 (6)	0.97 (14)
4th	nnU-Net	Dec. 2019	12.3	0.93 (4)	0.96 (1)	0.90 (5)	0.89 (7)	0.77 (1)	0.77 (1)	0.82 (10)	0.97 (1)
5th	AndyL	Nov. 2022	12.5	0.93 (27)	0.95 (15)	0.90 (16)	0.88 (16)	0.76 (5)	0.76 (5)	0.82 (2)	0.97 (20)

(a) Organ segmentation tasks

General rank	Team	Created	Mean Position	**Brain** *multiparametric MRI*			**Liver** *ce-CT*	**Lung** *CT*	**Pancreas** *ce-CT*	**Liver-2*** *ce-CT*	**Colon*** *ce-CT*
				Edema	Non-enhancing	Enhancing					
1st	universal_model	Feb. 2023	5.6	0.70 (5)	0.46 (30)	0.68 (16)	0.79 (1)	0.80 (1)	0.62 (1)	0.76 (1)	0.63 (1)
2nd	Swin_UNETR	Nov. 2021	10.1	0.70 (3)	0.53 (1)	0.71 (1)	0.76 (12)	0.77 (7)	0.58 (4)	0.72 (4)	0.59 (5)
3rd	ahatamiz2	Nov. 2021	10.1	0.70 (4)	0.53 (2)	0.71 (2)	0.76 (12)	0.77 (7)	0.58 (4)	0.72 (4)	0.59 (5)
4th	nnU-Net	Dec. 2019	12.3	0.68 (20)	0.47 (25)	0.68 (7)	0.76 (11)	0.74 (21)	0.53 (23)	0.72 (9)	0.58 (13)
5th	AndyL	Nov. 2022	12.5	0.67 (39)	0.47 (15)	0.68 (25)	0.76 (16)	0.78 (2)	0.62 (2)	0.75 (2)	0.61 (2)

(b) Lesion segmentation tasks

Table 3.1 – Dice Similarity Coefficient values of the five best methods (out of 55) of the Medical Segmentation Decathlon for the different organ and lesion segmentation tasks. * indicates a task and dataset that was not available during training. Rank for each task is indicated in brackets. From post-challenge leaderboard, accessed on December 2023. *ce : Contrast-enhanced, CT : Computed tomography, MRI : Magnetic resonance imaging.*

3.3 Application to pancreas

This section aims to give an exhaustive overview of existing DL methods for pancreas analysis, focusing on three key tasks : organ segmentation, lesion segmentation and detection, and feature extraction for pancreas analysis. Through these three axes, we explore the main methodological approaches that have contributed to the evolving landscape of DL-based solutions for pancreatic image analysis. Finally, we will conclude this bibliographic review with a concise review of more clinically oriented research.

3.3.1 Pancreas segmentation

As illustrated by the results of the Medical Segmentation Decathlon (see Table 3.1a), pancreas segmentation is a complex task, relatively more challenging than the segmentation of other abdominal organs such as the liver or spleen (Yao et al., 2020a). This complexity is attributed to the anatomical characteristics of the pancreas. First, the pancreas is a small organ nested in the abdomen and surrounded by numerous organs and vessels. Its anatomical position makes localization particularly challenging. Second, its delineation is complicated by its lobulated and indistinct contours. These indistinct contours result from limited contrast at the borders, explained by the close imaging-related physical properties of the pancreas and its neighboring organs. For instance, the CT attenuation coefficients of the head and tail of the pancreas are similar to those of the duodenum and spleen, respectively, making their differentiation difficult. Finally, the pancreas exhibits significant inter- and intra-individual variability due to numerous

anatomical and physiological variations, as discussed in Section 2.2.1.2. These characteristics, illustrated on Figure 3.8, render automatic pancreas segmentation challenging, leading to the proposal of various methods to account for the complex geometry and anatomy of this small organ (Kumar et al., 2019a; Yao et al., 2020a). Numerous methods have tried to adress these fundamental challenges, and DNNs are widely employed today, as these methods achieve the best performance in terms of DSC (Kumar et al., 2019a). A meta-analysis estimates that the average DSC for pancreas segmentation is 82.3 (95% confidence interval (CI) 80.7-84.0%) for methods based on DNNs compared to an average DSC of 74.4 (95% CI 70.9-77.8%) for all methods combined (Kumar et al., 2019a).

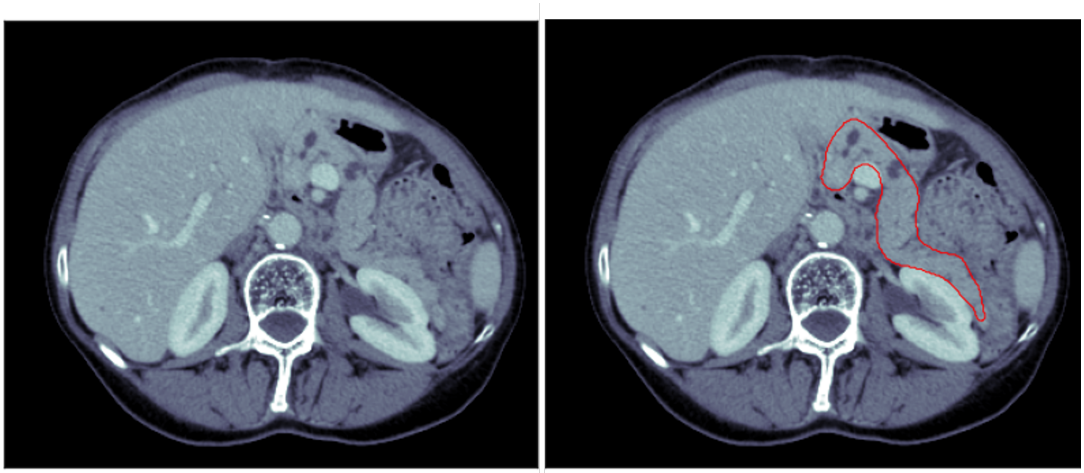


Figure 3.8 – Challenging case in pancreas segmentation. Left : Original portal CT scan. The pancreas is positioned among adjacent organs with comparable CT density, resulting in ambiguous contours. **Right** : The outlined pancreas in red. *CT : Computed tomography.*

We delve into the state-of-the-art approaches employed for pancreas segmentation using DL, as other methods are beyond the scope of this work. We examine the trends, highlights and key methodological innovations developed to achieve precise and reliable pancreas segmentation from medical imaging data. This presentation will be conducted in two parts : first, the methods using CNNs, followed by those employing transformers.

3.3.1.1 Methods using CNNs

CNNs have established themselves as a prominent DL method for automatic pancreas segmentation. Most approaches use the U-Net (Ronneberger et al., 2015) as a starting point and subsequently apply numerous technical or methodological variations, each addressing specific challenges. Thus, we categorize these approaches based on the primary focus of the methodology : we sequentially explore methods centered on dimension, those leveraging coarse-to-fine strategies, approaches optimizing coarse-to-fine strategies collectively, methods incorporating attention mechanisms, and techniques integrating multi-scale information.

Dimensionality Initially, much research focused on the dimensionality to adopt. Since medical images are originally 3D, slicing them into 2D reduces the algorithmic complexity and the memory demand. However, this transformation also alters the data use dynamics : in the 3D context, a data sample corresponds to a collection of slices, whereas in the 2D context, one data sample corresponds to one individual slice. This simplification leads to a significant loss of spatial context which is crucial for the pancreas due to its elongated and variable nature. Consequently, 2D methods often require post-processing steps. Yet, transitioning to 3D does not fully capture spatial information because computational limitations enforce the use of 3D patches. To address this, some methods propose a compromise between 2D and 3D. Zhou et al. (2017b) introduce a 2.5D multi-view model : three CNNs (axial, coronal, sagittal) are trained on

inputs containing three consecutive slices and the results of the three models are merged. Li et al. (2020) opt for seven consecutive slices and add an inter-slice DSC regularization term to the cost function. Fang et al. (2019) propose a network that learns local 3D features and global 2D features : the former are extracted through the encoding of 2.5D patches, while the latter are extracted through the encoding of a down-sampled 2D slice. To enforce spatial smoothness consistency among successive slices, Cai et al. (2017) train a Long Short-Term Memory [6] (Sherstinsky, 2020) network on segmentations generated by a primary 2D CNN. Similarly, Li et al. (2021b) combine inter and intra-slice information to leverage 3D spatial information while preserving low computational costs. Yan and Zhang (2021) propose a model that combines 3D and 2D convolutions from a 2.5D input. Finally, Zhu et al. (2019a) employ a neural architecture to automatically search for the best operation (2D or 3D convolution) at each layer. Presently, the optimal dimension for pancreas segmentation remains an open question, even though current advancements in computational and memory capacities allow for training fully 3D networks. Each method, characterized by its distinct advantages and disadvantages, requires thorough examination in the context of computational capabilities and available data.

Coarse-to-fine strategy for improved localization One of the challenges in automatic pancreas segmentation arises from the small volume fraction occupied by this organ in the abdomen : pancreatic voxels represent less than 1.5% in a 2D image and less than 0.5% in a 3D image (Zhang et al., 2021b). In such configurations, CNNs can be distracted by the extensive background containing abundant information, leading to less accurate segmentation. To address this difficulty, numerous studies have tackled the pancreas segmentation problem in two steps : first, a localization step from the entire image identifies a Region of Interest (ROI) containing the pancreas. Second, a segmentation step from this ROI segments the pancreas. This method, often referred to as Coarse-to-Fine (C2F) or cascaded networks, has been implemented in various works (Liu et al., 2019d; Roth et al., 2016b; Zhao et al., 2019; Zhou et al., 2017b; Zhu et al., 2017) and is depicted in Figure 3.9. Zhao et al. (2019) leverage image resolution for localization : the initial coarse segmentation step works on a down-sampled 3D volume, and the subsequent fine segmentation step uses the ROI at its original resolution. Roth et al. (2015) also perform segmentation in two steps but in a "bottom-up" fashion. Starting from superpixels, an initial CNN generates a probability map obtained from highly localized patches. Then, a second CNN segments larger areas using information provided by the first CNN. Instead of using a CNN for the coarse step, Man et al. (2019) learn pancreas localization by using a deep reinforcement learning strategy while Dogan et al. (2021) use a Mask R-CNN [7] (He et al., 2017). Some methods introduce one or more additional steps to the localization and segmentation steps. For instance, Zhang et al. (2021b) propose a three-step approach where the last step refines the segmentation by smoothing it with a 3D level-set post-processing. Fang et al. (2022) introduce a final step composed of an adversarial classifier which is trained to recognize the pancreas in both masked and unmasked images. When the classifier successfully identifies the pancreas in masked images, the fine segmentation network is penalized, indicating insufficient precision. This adversarial process, aimed at enhancing segmentation accuracy, is also adopted by Li et al. (2021c) who introduce a discriminator that seeks to distinguish a predicted segmentation from a manual segmentation.

While this two-step segmentation strategy allows obtaining a smaller input focused on the pancreas (the ROI), it suffers from several drawbacks. First, these methods heavily rely on the coarse segmentation step : if it fails to detect a suitable ROI, the final segmentation is entirely compromised. Second, these methods exhibit an inconsistency between training and inference : the optimization of the two steps occurs individually, whereas in inference, both models are used jointly. A third drawback of two-stage models stems from their lack of efficiency : the number of parameters is doubled, and some low-level features are extracted redundantly.

Joint optimization of coarse-to-fine networks As described above, C2F methods consist of two models but lack a global cost function that would enable their joint optimization. To address this drawback, Yu

6. A type of recurrent neural network (RNN) aimed to deal with the vanishing gradient problem present in traditional RNNs.
7. Mask R-CNN is a deep learning model that combines object detection and instance segmentation

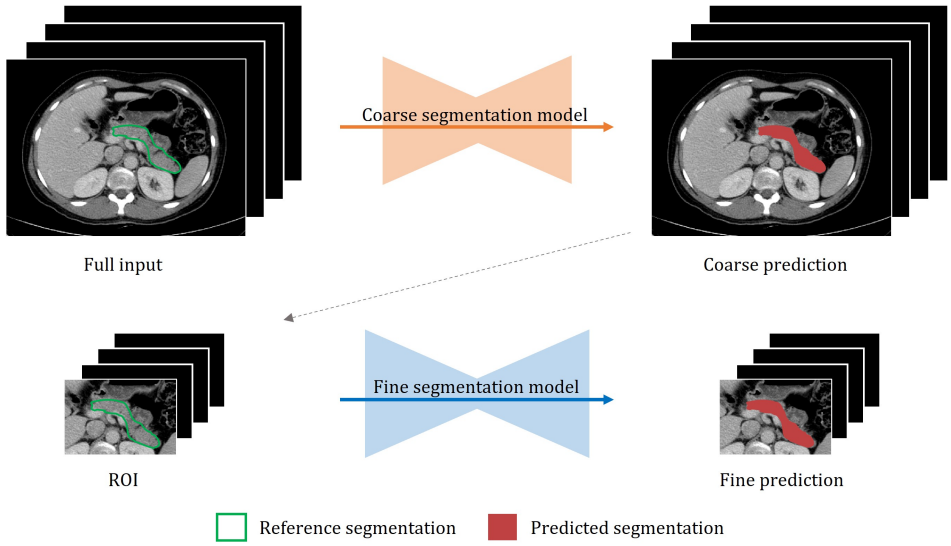


Figure 3.9 – Illustration of the coarse-to-fine segmentation process. Initially, a first segmentation model processes the entire image input, producing a preliminary coarse segmentation. This initial output allows for the isolation of a Region of Interest (ROI). Then, a second segmentation model refines this initial segmentation, producing a fine-grained segmentation.

et al. (2018) propose connecting these two stages using a recurrent neural network : for iteration t, the network takes the input image multiplied by a function of the output from iteration $(t-1)$ as input. This function, among other things, crops the input image. Optimization occurs jointly across all iterations. In practice, recurrence is constrained by GPU memory : the number of iterations cannot exceed five. Similarly, a model where intermediate features are optimized iteratively is suggested by Mo et al. (2020) ; however, the number of iterations is also limited by GPU memory.

Localizing the pancreas using attention To emphasize the learning of relevant characteristics, some approaches propose the use of attention modules. These attention mechanisms have proven effective in concentrating the learning of discriminative features, mimicking the human perception process. Attention modules are trained to modulate the signal and retain only that from regions relevant to pancreas localization and segmentation. Lu et al. (2019), Oktay et al. (2022), Wang et al. (2021) and Yan and Zhang (2021) eliminate the localization model and incorporate attention modules into a U-Net. These attention modules can take various forms : Oktay et al. (2022) filter features from skip connections by coarser features in order to remove irrelevant and noisy signal contained in skip connections. Lu et al. (2019) use an attention module on channels to consolidate the problem of learning *what*, while a spatial attention module consolidates the problem of learning *where*.

Integrating multi-scale information Although the localization step allows obtaining a smaller input more adapted to the volume of the pancreas, it also deprives the fine segmentation model of a significant portion of spatial context. Consequently, methods seek to modify the architectures of their CNNs to learn, within a single model, features of variable sizes and shapes. A strategy involves increasing the receptive fields of networks to incorporate maximum contextual information. Thus, Huang et al. (2021) propose using a deformable U-Net where the convolutional receptive fields are learned to be adjusted to the size and shape of the pancreas. As huge receptive fields come with high computational demand, Heinrich and Oktay (2017) employ sparse binary convolutions to avoid excessive model complexity. Another strategy consists in using feature maps from various spatial resolution (Li et al., 2021c; Hu et al., 2020b; Proietto Salanitri et al., 2021). This allows analyzing the original image with filters of complementary fields of view, capturing the pancreas and the relevant context at various scales. In the same vein, Fu et al. (2018) replace up-sampling layers with multi-level up-sampling layers : instead of a single up-sampling directly to the desired resolution, multiple up-sampling operations are performed to better integrate information from different

scales. However, all these strategies come at a cost as they introduce a large number of parameters.

3.3.1.2 Methods using transformers

To overcome the limitations of convolutional operations, recent studies have focused on leveraging self-attention models based on transformers. Unlike CNNs, which are constrained by their receptive fields, transformers can model long-distance spatial relationships, a crucial aspect in medical image analysis. Initially proposed for natural language processing tasks, transformers are now widely used for computer vision tasks through Vision Transformers (ViT), which use 2D patches and positional embeddings (Dosovitskiy et al., 2020). Notably, the Swin-Transformer (Liu et al., 2021b) has emerged as a foundational architecture, implementing a hierarchical transformer operating on shifted windows (Swin actually coming from "shifted window"). By calculating self-attention only on local windows, this shifted window scheme achieves greater efficiency while enabling connections between windows. Consequently, ViTs have demonstrated excellent performance on both natural and medical images, although their applications for pancreas segmentation remain limited. Cao et al. (2022) propose the Swin-Unet, where the encoder is replaced by a series of Swin-Transformer blocks, and the decoder is symmetrically constructed. While this architecture effectively captures global context, it struggles to model local information, crucial for accurate pancreas segmentation due to its complex geometric shape. Therefore, various methods attempt to combine the advantages of a CNN and a transformer. Qiu et al. (2023) introduce a two-step method. Initially, a residual transformer block is used to identify the pancreas position, leveraging its ability to extract multi-scale features from a global perspective. Then, a double convolution downsampling block is employed to obtain precise pancreas shape and size features. However, this method only uses transformers as an external module instead of seeking to integrate attention into CNN learning. Thus, Chen and Wan (2023) deploy a transformer on skip connections to coordinate the learning of global features, while the rest of the U-Net architecture is retained. To learn sequential representations of the input image, Hatamizadeh et al. (2022) opts to substitute only the encoder with a Swin-Transformer. Fang et al. (2023) integrate a residual transformer structure at the bottleneck of an autoencoder. In summary, transformers enable better modeling of interactions between distant pixels and capturing global context. However, these methods struggle to learn local features, leading numerous studies to combine transformers with CNNs. Moreover, these methods face significant computational complexity (quadratic complexity based on input length). To address this, some methods propose reformulating the self-attention mechanism using a linear approximation (Zhang and Bagci, 2022).

3.3.1.3 Summary

In Table 3.2, we provide a summary of the methods presented, along with the evaluation measures reported by the respective authors. While acknowledging that these performance measures may not be entirely comparable, it is important to note that the majority of these methods underwent training and evaluation on a consistent public dataset comprising 80 portal phase CT scans ($\mathcal{D}_{NIH}$ dataset (Roth et al., 2016a), presented in Section 4). The majority of methods employing this dataset were evaluated using a four-fold cross-validation procedure.

3.3.2 Lesion detection

The ability to identify lesions is essential for early diagnosis and effective treatment planning. The detection of pancreatic lesions introduces additional challenges compared to pancreatic segmentation. In this section, we explore the latest advancements in DL techniques applied to pancreatic lesion detection. We begin by examining methodologies that focus on lesion segmentation, as it represents the prevalent initial step for automated detection tools. Then, we elucidate how this segmentation task is translated into detection.

Ref	Summary	Dataset	Split	DSC (%)
Roth et al. (2015)	C2F strategy using 2.5D CNN and a bottom-up approach from superpixels.	$\mathcal{D}_{NIH}$	4 CV	**71.8 ± 10.7**
Zhou et al. (2017b)	C2F strategy using multi-view 2.5D CNNs.	$\mathcal{D}_{NIH}$	4 CV	**82.37 ± 5.68**
Cai et al. (2017)	Enforce spatial consistency on successive slices using a LSTM.	$\mathcal{D}_{NIH}$	NA	82.4 ± 6.7
Zhu et al. (2017)	C2F strategy using 3D networks on 3D patches.	$\mathcal{D}_{NIH}$	4 CV	**84.59 ± 4.86**
Heinrich and Oktay (2017)	2D U-Net using binary sparse convolutions to increase the receptive field.	SMO_{30}	6 CV	64.5
Roth et al. (2016b)	C2F strategy using multi-view 2.5D HNNs.	$\mathcal{D}_{NIH}$	4 CV	**81.27 ± 6.27**
Oktay et al. (2022)	3D U-Net with attention gates.	$\mathcal{D}_{NIH}$	61/21	82.1 ± 5.7
Yu et al. (2018)	Multi-view 2.5D recurrent CNN.	$\mathcal{D}_{NIH}$	4 CV	**84.50 ± 4.97**
Fu et al. (2018)	Multi-layer upsampling CNN.	X_{59}	5 CV	76.4 ± 14.3
Li et al. (2021b)	Bi-directional recurrent 2.5D U-Net to optimize local 3D information.	$\mathcal{D}_{NIH}$	4 CV	**84.19 ± 5.73**
Zhao et al. (2019)	C2F strategy using 3D networks on downsampled volumes.	$\mathcal{D}_{NIH}$	4 CV	**85.99 ± 4.5**
Man et al. (2019)	Localization policy to produce a bbox of the pancreas, then segment using a deformable multi-view 2D U-Net.	$\mathcal{D}_{NIH}$	4 CV	**86.93 ± 4.92**
Li et al. (2020)	2.5D U-Net using 7 consecutive slices + inter-slice regularization.	$\mathcal{D}_{NIH}$	4 CV	**85.7 ± 4.1**
Lu et al. (2019)	Use of channel and spatial attention to enhance the information exchange between the encoder and decoder of a 2D U-Net.	$\mathcal{D}_{NIH}$	10 CV	88.32 ± 2.84
Zhu et al. (2019a)	Neural architecture search to automatically find the best network architecture.	$\mathcal{D}_{NIH}$	4 CV	85.15 ± 4.55
Liu et al. (2019d)	C2F strategy, with fine segmentation done by ensembling five 2.5 CNNs trained with different loss function.	$\mathcal{D}_{NIH}$	4 CV	**84.1 ± 4.9**
Fang et al. (2019)	Two encoders extract 3D local and 2D global features, respectively.	$\mathcal{D}_{NIH}$	53/29	85.46 ± 4.80
Zhang et al. (2021b)	Three-stages framework : coarse (multi-atlas registration), fine (3D + multi-view 2.5D CNNs), refine (level-set).	$\mathcal{D}_{NIH}$	4 CV	**84.61 ± 5.21**
Nishio et al. (2020)	2D Deep U-Net using intense data-augmentation (Random Image Cropping and Patching and Mixup).	$\mathcal{D}_{NIH}$	4 CV	**78.9 ± 8.3**
Hu et al. (2020b)	Coarse segmentation using Dense Atrous Spatial Pyramid Pooling, and saliency-aware module to fully use the coarse segmentation in the fine segmentation step.	$\mathcal{D}_{NIH}$	NA	85.49 ± 4.77
Mo et al. (2020)	A 3D U-Net with iterative refinement of features at multiple levels.	$\mathcal{D}_{NIH}$	4 CV	**82.47 ± 5.50**
Zhang and Bagci (2022)	Reduce the complexity of the transformer by using linear approximation.	$\mathcal{D}_{NIH}$	72/10	85.5 ± 3.7
Huang et al. (2021)	Deformable 2D U-Net (learnable receptive field).	$\mathcal{D}_{NIH}$	4 CV	**87.25 ± 3.27**
Yan and Zhang (2021)	A attention U-Net mixing 3D and 2D convolutions.	$\mathcal{D}_{NIH}$	4 CV	**86.6 ± 3.5**
Li et al. (2021c)	A 2D U-Net with multi-level pyramidal pooling and aversarial network.	$\mathcal{D}_{NIH}$	4 CV	**81.36**
Proietto Salanitri et al. (2021)	A 3D FCN which extracts features at different scales and decodes them hierarchically.	$\mathcal{D}_{NIH}$	4 CV	**88.01 ± 4.74**
Dogan et al. (2021)	C2F strategy using a Mask R-CNN for localization.	$\mathcal{D}_{NIH}$	4 CV	**86.15 ± 4.45**
Wang et al. (2021)	A dual-input FNC, which receives CT and images processed by contrast-specific visual saliency, and combines attention.	$\mathcal{D}_{NIH}$	4 CV	**87.4 ± 6.8**
Fang et al. (2022)	Enrich the C2F strategy by introducing an adversarial network.	$\mathcal{D}_{NIH}$	NA	87.50 ± 4.89
Chen and Wan (2023)	Deploy transformer module on the skip connections of a U-Net.	$\mathcal{D}_{NIH}$	4 CV	**86.8 ± 4.1**
Hatamizadeh et al. (2022)	Replace the encoder of a 3D U-Net by a transformer sequence.	SMO_{30}	5 CV	76.7
Qiu et al. (2023)	C2F strategy using a transformer for the location of the pancreas and a CNN for the segmentation.	$\mathcal{D}_{NIH}$	NA	86.25
Fang et al. (2023)	Integrate the transformer structure into a 3D U-Net.	$\mathcal{D}_{MSD}$	225/56	77.36 ± 0.11
Cao et al. (2022)	Combine Swin transformer with a 2D U-Net.	SMO_{30}	18/12	56.58

Table 3.2 – Summary of deep learning-based methods for pancreas segmentation. Color corresponds to the primary focus of the methodology : dimensionality , Coarse-to-Fine (C2F) network , joint optimization of C2F network , attention , multi-scale strategy , transformers . For each article, we report a summary of the approach, the dataset, the data split, and the mean ± standard deviation of the DSC on the test data as reported by the authors. f CV means that performance was evaluated with a f fold cross-validation procedure. t/v means that dataset was split into t and v for train and test, respectively. NA means the information was not reported. $\mathcal{D}_{NIH}$ and $\mathcal{D}_{MSD}$ refer to public datasets containing 80 and 281 abdominal CT scans, respectively. These datasets are detailed in Section 4. SMO_{30} refers to the public Synapse Multi Organ dataset (2013) containing 30 abdominal CT scans with 13 organs annotated, including pancreas. X_{59} refers to a private dataset containing 59 abdominal CT scans. Performance in **bold** design performance reported using the common test four fold cross-validation procedure on $\mathcal{D}_{NIH}$. *C2F : Coarse-to-fine, CNN : Convolutional neural network, CT : Computed tomography, CV : Cross validation, DSC : Dice similarity coefficient, FCN : Fully convolutional neural network, HNN : Holistically nested network, LSTM : Long short-term memory,* $\mathcal{D}_{MSD}$ *: Medical Segmentation Decathlon dataset,* $\mathcal{D}_{NIH}$ *: National Institute of Health dataset.*

3.3.2.1 Lesion segmentation

Like pancreas segmentation, lesion segmentation methods can be categorized based on their primary methodological approach. We initiate by introducing methods similar to those developed for pancreas segmentation, specifically those exploring dimensionality and employing C2F strategies. We then present strategies devised to address data scarcity, strategies combining various methodologies, and finally, strategies integrating additional information via anatomical clues or multiple phases combination.

Dimensionality As for pancreas segmentation, the search for the optimal architecture and dimensionality for lesion segmentation has also been explored. Yu et al. (2020) and Zhu et al. (2019a) employ neural architecture search to discover the most effective topology and combination of convolutional layers (2D or 3D). Isensee et al. (2021a) dynamically adjust its U-Net architecture based on input image geometry using heuristics. Chen et al. (2021) project 3D images onto a 2D plane using a spiral transformation. This transformation enables the use of a 2D model while preserving spatial correlation among voxels.

Coarse-to-fine C2F methods have also been examined for the segmentation of cancers (Alves et al., 2022; Turečková et al., 2020; Xia et al., 2022) or cysts (Xia et al., 2022; Zhu et al., 2019a) : initially, a first pancreas segmentation model establishes a ROI, which is then input into a second algorithm for lesion segmentation. In a similar vein, to address variations in sizes and textures exhibited by these lesions, multi-scale strategies have been proposed (Xia et al., 2022; Zhang et al., 2020b; Zhu et al., 2019b). Specifically, Zhu et al. (2019b) develop multiple models to segment the tumor from inputs of varying sizes, and average the results : this allows for the combination of outcomes from models exhibiting high sensitivity (small input) with those demonstrating high specificity (large input).

Data scarcity Numerous methods aim to address the data scarcity. While conventional data augmentation techniques (rotations, flip, crop, scaling, mirroring, etc.) are commonly used (Isensee et al., 2021a), other studies explore more innovative operations. For instance, Chen et al. (2021) employ a spiral transformation to generate diverse 2D examples for a given 3D image. Xia et al. (2022) also demonstrate that augmentations on the physical characteristics of scanners, such as slice thickness, enhance generalization power. In addition to data augmentation, a strategy involves leveraging different types of data : Li et al. (2022) employ MRI data to augment the training dataset for a CT lesion segmentation model. For this, a three-step framework is proposed : first, a model generates intermediate images to continuously model the differences between CT and MRI images. Then, a segmentation model is trained on all images (MRI, CT, and intermediates) to enable the model to learn common pancreatic cancer characteristics across these images. Finally, a last step focuses on learning distinctive lesion features in CT images. Moreover, Zhang et al. (2020a) train a lesion segmentation model using multiple annotated and non-annotated databases. Several teaching models are trained based on the available annotations, and these models are then used to generate pseudo-annotations for non-annotated images. Ultimately, a student model is trained using the entire dataset, including annotated and pseudo-annotated data. The authors demonstrate that this student model achieves a 6.3% improvement in DSC compared to a model trained solely with annotated data (Zhang et al., 2020a).

Combining other approaches Other methods integrate various modules and/or approaches. For instance, Guo et al. (2018) show the value of combining graph-based segmentation methods with deep learning : initially, a U-Net generates tumor segmentation, which is then used to initialize a segmentation graph. To jointly segment the pancreas and the lesion, Li et al. (2023) employ a temperature module to prevent the model from overly focusing on learning the pancreas. Lastly, like done for pancreas segmentation, Qu et al. (2023b) propose to combine a transformer and a CNN through a dual-encoder segmentation model : the first being a CNN extracting local features, while the second is a transformer extracting global features. Both types of features are progressively merged in a decoder.

Adding information via anatomical cues Some approaches aim to explicitly incorporate additional anatomical information, often conveyed through the segmentation of adjacent structures. Inspired by radiologists who use ductal structures for pancreatic cancer diagnosis, Viviers et al. (2022) integrate visual information in the form of segmentations into a lesion segmentation model. Therefore, their 3D U-Net takes as input the image, pancreas segmentation, main pancreatic duct segmentation, and common bile duct segmentation. Adding these segmentation masks enhances the model sensitivity, increasing from 0.98 ± 0.03 (with the image alone) to 1.00 ± 0.00 (when the masks are provided as input) on their test split containing 59 patients. Liu et al. (2019c) propose reinforcing the use of shape and texture by incorporating both the image and the pancreas segmentation mask generated by a preliminary segmentation model. Alves et al.

Ref	Summary	Input	Train	Test	DSC (in %)
Zhou et al. (2017a)	C2F strategy : first segment the pancreas, then the cysts.	*portal CT*	131 cysts	4f CV	64.44
Guo et al. (2018)	Build a graph-based 3D segmentation model initiated by a mask generated by a 2.5D U-Net.	*arterial CT*	30 tumor ROIs	Loc : 21 tumor ROIs	83.2 ± 7.8
Zhou et al. (2019a)	Segment pancreas, PDAC and main pancreatic duct using a model with 2 branches (one per phase) optimized jointly.	*arterial + portal CT, aligned*	239 PDACs	3f CV	63.98 ± 22.74
Tureckova et al. (2020)	Ensemble two 3D CNNs using attention and trained on images with different resolutions.	*portal CT*	$\mathcal{D}_{MSD}$	5f CV	52.99 ± 2.05
Zhang et al. (2020a)	Increase the training dataset of a PDAC segmentation model using a mix of annotated and unannotated venous or multi-phase CT images.	*multiphase CT*	1071 ($\mathcal{D}_{MSD}$+ 700 PDACs)	5f CV	70.9 ± 15.9
Chen et al. (2021)	Transform 3D volumes into 2D images using a spiral transformation and segment PDAC tumors using a 2D U-Net.	*portal CT*	$\mathcal{D}_{MSD}$	5f CV	66.62 ± 16.37
Yu et al. (2020)	C2F strategy combined with neural architecture search to perform PDAC segmentation.	*portal CT*	$\mathcal{D}_{MSD}$	5f CV	54.41
Isensee et al. (2021a)	Automatic U-Net configuration based on the properties of the training dataset, utilizing fixed, rule-based, and empirical parameters.	*portal CT*	$\mathcal{D}_{MSD}$	$\mathcal{D}^{test}_{MSD}$	52.27
Li et al. (2022)	Make use of MRI data to improve a pancreatic cancer segmentation model operating on CT.	*portal CT*	$\mathcal{D}_{MSD}$	4f CV	57.62
Li et al. (2023)	Leverage temperature guided modules to improve the joint segmentation of pancreas and lesions.	*portal CT*	$\mathcal{D}_{MSD}$	5f CV	59.16 ± 28.12

Table 3.3 – Summary of deep learning-based methods for the segmentation of pancreatic lesions. For each article, we report a summary of the approach, the modality used as input, the train dataset, the test sets and the evaluation measures on the test sets as reported by the authors. f CV indicates that the results were obtained doing a f fold cross-validation procedure. A local (loc) set consists of data originating from the same source as the training dataset, usually a validation or test hold-out set. $\mathcal{D}_{MSD}$ and $\mathcal{D}^{test}_{MSD}$ refer to the training and testing splits of the public MSD dataset containing 281 and 139 abdominal CT scans, respectively. The $\mathcal{D}_{MSD}$ dataset is detailed in Section 4. When reported by the authors, standard deviations are indicated by ±. *C2F : Coarse-to-fine, CNN : Convolutional neural network, CT : Computed tomography, MRI : Magnetic resonance imaging,* $\mathcal{D}_{MSD}$ *: Medical Segmentation Decathlon dataset, PDAC : Pancreatic ductal adenocarcinoma, ROI : Region of interest.*

(2022) investigate the benefits of segmenting not only the lesion but also the pancreas and other nearby anatomical structures, such as veins, arteries, the main pancreatic duct, and the bile duct. They notably demonstrate that, at the patient level, there is a clear advantage in segmenting the pancreas and the lesion together. As for the segmentation of other nearby structures, it primarily helps to accurately localize lesions, particularly through the segmentation of the main pancreatic duct.

Adding information via multi-phase CT Another strategy to add information involves using multiple phases, as each of these phases carries complementary information. Thus, for lesion segmentation from CT images, numerous studies explore the joint use of portal and arterial phases (Xia et al., 2020, 2022; Zhou et al., 2019a; Zhu et al., 2021). The most straightforward way to use these two phases is to align the images and employ a two-channel input. Xia et al. (2020) introduce two additional combination methods : late and slow fusion. In the case of late fusion, each phase passes through an encoder, and the resulting two latent spaces are merged and fed into a single decoder. Slow fusion also employs two encoders, but the feature maps are gradually aligned at each downsampling step of the encoder. These slow or late fusions enable coupling the two phases while overcoming the often-imperfect registration. Another strategy involves training two segmentation models : one operating from the portal phase and the other from the arterial phase. Zhu et al. (2021) train these two models independently, whereas Zhou et al. (2019a) connect the two models by hyper-connections and train them jointly.

Table 3.3 compiles the methodology and outcomes of studies focused on the segmentation of pancreatic lesions.

3.3.2.2 From lesion segmentation to cancer detection

Typically, lesion segmentation models are used to categorize patients into healthy or pathological, with certain studies opting for a more nuanced approach, involving the identification of the lesion type. Thus, lesion segmentations are converted into cancer detection at patient-level. For this purpose, most methods apply a simple heuristic to go from segmentation to detection. These heuristics may consider the volume or the number of main connected components of the predicted segmentation. For example, Zhu et al. (2021)

Ref	Summary	Input	Train	Test	Results
Zhu et al. (2019b)	Ensemble three C2F segmentation models with varying input sizes and apply a simple rule to convert the segmentation into detection.	*portal CT*	239 (136 PDACS)	4f CV on PDACs + 200 controls	- **DSC** : 57.3 ± 28.1 - **Sen** : 94.1 - **Spe** : 98.5
Xia et al. (2020)	Ensemble three models with different strategies to combine portal and arterial data.	*arterial + portal CT, aligned*	439 (136 PDACS)	3f CV	- **DSC** : 83.3 ± 8.2 - **Sen** : 97.1 - **Spe** : 96
Abel et al. (2021)	C2F strategy to segment cysts, then a heuristic transforms the segmentation into a detection.	*portal CT*	221 cysts	5f CV	**Sen** : 78.8 ± 0.1
Zhu et al. (2021)	Train two segmentation models (one per phase), apply a heuristic and fuse the results to differentiate PNETs from healthy patients.	*arterial + portal CT, aligned*	376 (228 PNETs)	4f CV	- **DSC** : 43.11 - **Sen** : 89.47 - **Spe** : 81.08
Xia et al. (2021)	Differentiate PDAC, non PDAC, and controls pancreas on non-contrast CTs combining an U-Net + transformer.	*non contrast CT*	1321 (450 PDACs, 394 non PDACs)	Loc : 306 (108 PDACs, 80 nonPDACs)	-**Sen** : 95.2 - **Spe** : 95.8
Alves et al. (2022)	C2F strategy using multiple surrounding anatomical structures to guide the segmentation of the tumors.	*portal CT*	242 (119 PDACs)	Ext : 363 ($\mathcal{D}_{NIH}$, $\mathcal{D}_{MSD}$)	**AUC** : - Ext : 0.889 - Ext - lesions <2cm : 0.838
Viviers et al. (2022)	Segment lesion using a 3D nnU-Net taking as input the image and the segmentation of the common bile duct, the main pancreatic duct and the pancreas.	*portal CT*	196 (99 PDACs)	- Loc : 30% of the training data - Ext : 28 cases from $\mathcal{D}_{MSD}$	**DSC** / **Sen** / **Spe** : - Loc : 31 / 100 / 90 - Ext : 31 / 99 / -
Xia et al. (2022)	C2F strategy to segment pancreatic lesions. Use of extensive data augmentation and multi-scale processing for recognizing small lesions.	*arterial + portal CT, aligned*	1592 (852 PDACs)	- Loc 1 : 513 (213 PDACs) - Loc 2 : 1421 (213 PDACs, 450 PNETs, 458 cysts) -Ext 1 : 481 ($\mathcal{D}_{NIH}$, 399 PDACs) - Portal phase only - Ext 2 : 242 (78 PDACs)	**DSC** / **Sen** / **Spe** : - Loc 1 : 65 (58-85) / 97 / 99 - Loc 2 - PNETs : 57 (25-86) / 94 / 95 - Loc 2 - Cysts : 66 (52-88) / 95 / 95 - Ext 1 : 58 (41-80) / 97 / 93 - Ext 2 : NA / 91 / 91
Qu et al. (2023b)	Use a CNN encoder and a transformer encoder to extract local and global features, and progressively fuse them in a single decoder to perform lesion segmentation.	*portal CT*	223 (8 classes, 27 PDACs)	- Loc : 90 (8 classes, 11 PDACs) - Ext 1 : 53 (8 classes, 0 PDACs) - Ext 2 : 50 (8 classes, 6 PDACs)	**DSC** : - Loc : 80.51 ± 19.62 - Ext 1 : 67.17 ± 34.58 - Ext 2 : 69.25 ± 27.29 **DR** : - All : 91.71

Table 3.4 – Summary of methods for cancer patient detection based on lesion segmentation. For each article, we report a summary of the approach, the modality used as input, the train dataset, the test sets and the evaluation measures on the test sets as reported by the authors. f CV indicates that the results were obtained doing a f fold cross-validation procedure. A local (loc) set consists of data originating from the same source as the training dataset, usually a validation or test hold-out set. On the other hand, an external set (ext) includes data from an institution different from the one that provided the training set. $\mathcal{D}_{NIH}$ and $\mathcal{D}_{MSD}$ refer to public datasets containing 80 and 281 abdominal CT scans, respectively. These datasets are detailed in Section 4. The mean Dice similarity coefficient (DSC) for the lesion segmentation, sensitivity (Sen), specificity (Spe), detection rate (DR) are given in %. When reported by the authors, standard deviations and inter-quartile ranges are indicated by ± and parenthesis, respectively. NA means the information was not reported. *AUC : Area under the curve, C2F : Coarse-to-fine, CNN : Convolutional neural network, CT : Computed tomography, $\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset, $\mathcal{D}_{NIH}$: National Institute of Health dataset, PDAC : Pancreatic ductal adenocarcinoma.*

classify a case as pathological if the number of voxels in the tumor or main pancreatic duct exceeds a certain threshold. Some models additionally apply more advanced post-processing strategy. For instance, (Xia et al., 2022) use a post-processing algorithm that evaluates uncertainty, quality, shape, and geometry of each predicted lesion to enhance specificity by eliminating false positives. Other methods perform cancer patient detection as an auxiliary task to lesion segmentation. To achieve this, a transformer is combined with a U-Net to simultaneously conduct image classification and segmentation. Xia et al. (2021) use this combination to classify non-contrast CT images into three classes : healthy patients, cancer patients, and abnormal patients without cancer. Similarly, (Zhou et al., 2023) use a model to segment lesions and classify multi-phase scans between healthy and nine different types of lesions. To better guide segmentation, the transformer takes patient metadata (gender and age) as well as feature maps from the U-Net as input.

Table 3.4 provides a concise overview of the methodology and findings presented in articles performing pancreas cancer patient detection based on lesion segmentation.

3.3.3 Feature extraction for pancreas analysis

The last focal point of this review delves into methodologies designed to extract features that can be subsequently used to provide information about the organ functional and structural properties. As elucidated in Section 3.2.2.2, features can be categorized into two types : HCR, acquired through predefined mathematical definitions, and DLR, automatically computed by CNNs.

3.3.3.1 Using hand-crafted radiomics

Numerous studies highlight the use of intermediate HCR features for diverse applications related to pancreas analysis. The most extensively studied application involves the differentiation between healthy patients and those with cancer. This differentiation is achieved through HCR features extracted from various sources, including CT scans (Chu et al., 2019a), biopsies (Momeni-Boroujeni et al., 2017), or Endoscopic Ultrasound (EUS) (Das et al., 2008; Ozkan et al., 2016; Săftoiu et al., 2008). Ozkan et al. (2016) opt to train an age-specific classifier and demonstrate that this age calibration significantly enhances diagnostic performance : they achieve accuracies of 92%, 88.5%, and 91.7% for models concerning < 40 years, 40-60 years, and > 60 years, respectively, compared to an accuracy of 85.7% for a single unified model. Zhang et al. (2020a) delve deeper into diagnostics, seeking to classify healthy and prediagnostic CT scans. The authors show that a support vector machine trained on 34 HCR achieves an AUC of 98% (95% CI : 94-98%) for identifying prediagnostic cases (median interval between examination and diagnosis : 398 days). They demonstrate that this automated method outperforms the consensus reached by two radiologists, whose classification AUC is 66% (95% CI : 46-86%) ($P < 0.001$). Other applications are also reported : from texture features extracted from EUS images, Zhu et al. (2013) distinguish cases of pancreatic cancer from chronic pancreatitis. Zhang et al. (2021a) train a classifier to determine the grade of Pancreatic Neuroendocrine Tumor (PNET) based on HCR features extracted from lesions seen in CT images. Using HCR extracted from MRI scans of patients with Pancreatic Ductal Adenocarcinoma (PDAC), Kaissis et al. (2019) train a random forest to predict whether the patient's survival will be below or above the median overall survival. These applications, summarized in Table 3.5, collectively highlight that relatively straightforward HCR features possess significant potential for addressing a range of tasks related to pancreatic cancer analysis.

3.3.3.2 Using deep-learning radiomics

The second approach to feature extraction relies on CNNs, which are directly trained on images. As explained in Section 3.2.2.2, these networks exhibit significant potential for identifying complex patterns that contribute to the analysis of pancreatic tissue images. Primarily, their application is notably prevalent in distinguishing between healthy and cancerous scans. In a study by Liu et al. (2020), image patches from portal phase CT scans were classified using a CNN. These patches, sampled within the pancreas or lesions through manual segmentation, contribute to enhancing the volume of available data. Another notable approach by Qu et al. (2023a) incorporates a contrastive learning module, constraining the network to recognize two patches from the same lesion, thus mandating the model to restore the integrity of the lesions. Additionally, Liu et al. (2019a) adopts a distinctive strategy by acknowledging that pancreatic cancer induces alterations in organ shape. They classify healthy and cancerous cases using pancreas segmentation, where a U-Net generates the pancreas segmentation, modeling the parenchyma shape, and this shape is then classified using a neural network.

Beyond cancer differentiation, Dmitriev et al. (2017) integrates portal CTs and demographic information to differentiate the four most common types of cysts. To achieve this, two classifiers operate independently : a random forest using quantitative features (age, gender, etc.), and a CNN operating on 2D slices. Another study by Corral et al. (2019) reveals that a DL-based method achieves comparable performance to radiologists in classifying Intraductal Papillary Mucinous Neoplasm (IPMN) grades from MRI images. CNNs are also applied in histological image analysis : Klimov et al. (2021) predict the risk of metastasis from histological sections of PNET, while Kriegsmann et al. (2021) automatically classify 11 structures, including cancer, in pancreatic tissue samples. Table 3.6 compiles the methods and results for pancreas analysis based on features extracted by CNNs.

Ref	Summary	Input	Train	Test	Results
Das et al. (2008)	Differentiate PDAC tissue from non-neoplastic tissue using a NN-based classification model trained on HCR.	*EUS*	319 (99 CPs, 110 PDACs)	50% of the training data	**AUC** : : 0.93
Săftoiu et al. (2008)	Differentiate benign from malignant patterns using a NN trained from hue histograms.	*EUS*	114 (11 CPs, 32 PDACs, 3 PNETs)	10f CV	**AUC** : / **Acc** : - Benign vs malignant : 0.957 / 95.31 ± 6.25 - CP vs PDAC : 0.965 / 90.01 ± 12.31
Zhu et al. (2013)	Train a SVM on HCR for the differential diagnosis of PDAC and CP.	*EUS*	388 (262 PDACs, 126 CPs)	LOO	- **Sen** : 96.25 ± 0.4460 - **Spe** : 93.38 ± 0.2076 - **Acc** : 94.2 ± 0.1749
Ozkan et al. (2016)	Differentiate cancer from non-cancer patients using age-based classifiers trained on HCR.	*EUS*	- <40 years : 43 (13 PDACs) - 40 - 60 years : 75 (27 PDACs) - >60 years : 189 (110 PDACs)	Loc : - <40 years : 25 (8 PDACs) - 40 - 60 years : 75 (14 PDACs) - >60 years : 189 (30 PDACs)	**Sen** / **Spe** / **Acc** : - <40 years : 87.5 / 94.1 / 92 - 40-60 years : 85.7 / 91.7 / 88.5 - >40 years : 93.3 / 88.9 / 91.7
Momeni-Boroujeni et al. (2017)	Distinguish between benign and malignant biopsies using a NN trained on HCR.	*FNA biopsy*	192 (118 benigns, 74 malignants)	30% of the training data	**Acc** : 100
Chu et al. (2019a)	Differentiate cancer from non-cancer patients using a RF trained on HCR.	*portal CT*	255 (130 PDACs)	Loc : 125 (60 PDACs)	- **AUC** : : 0.999 - **Sen** : 100 - **Spe** : 98.5 - **Acc** : 99.2
Kaissis et al. (2019)	Classify the survival (above vs below median overall survival) using a RF trained on HCR.	*MRI*	102 PDACs	Loc : 30 PDACs	- **AUC** : : 0.90 - **Sen** : 87 [67.3, 92.7] - **Spe** : 80 [74.0, 86.7]
Zhang et al. (2021a)	Train ML classifiers for the identification of pathological grades of PNETs based on HCR.	*arterial CT*	62 PNETs	Loc : 20 PNETs	**AUC** : : - Grade 1 vs Grade 2 : 0.82 - Grade 2 vs Grade 3 : 0.70 - Grade 1 vs Grade 3 : 0.85
Mukherjee et al. (2022)	Differentiate cancer from non-cancer patients at the prediagnostic stage using a SVM on HCR.	*portal CT*	292 (110 PDACs)	- Loc 1 : 128 (45 prediagnostic PDACs) - Loc 2 : 176 controls - Ext : $\mathcal{D}_{NIH}$	**AUC** : / **Sen** / **Spe** / **Acc** : - Loc 1 : 0.98 [0.94-0.98] / 95.5 [85.5-100.0] / 90.3 [84.3-91.5] / 92.2 [86.7-93.7] **Sen** : - Loc 2 : 92.6 - Ext : 96.2

Table 3.5 – Summary of pancreatic cancer analysis methods based on hand-crafted radiomics (HCR). For each article, we report a summary of the approach, the modality used as input, the train dataset, the test sets, and the evaluation measures on the test sets as reported by the authors. f CV and LOO indicate that the results were obtained doing a f fold cross-validation procedure and a leave-one-out procedure, respectively. A local (loc) set consists of data originating from the same source as the training dataset, usually a validation or test hold-out set. On the other hand, an external (ext) set includes data from an institution different from the one that provided the training set. $\mathcal{D}_{NIH}$ refers to a public dataset containing 80 abdominal CT scans. This dataset is detailed in Section 4. The mean sensitivity (Sen), specificity (Spe), accuracy (Acc) are given in %. When reported by the authors, standard deviations and 95% confidence intervals are indicated by ± and brackets, respectively. *AUC : Area under the curve, CP : Chronic pancreatitis, CT : Computed tomography, EUS : Endoscopic ultrasound, FNA : Fine needle aspiration, HCR : Hand-crafted radiomics, ML : Machine learning, $\mathcal{D}_{NIH}$: National Institute of Health dataset, NN : Neural network, PDAC : Pancreatic ductal adenocarcinoma, RF : Random forest, SVM : Support vector machine.*

3.3.4 Pancreatic cancer detection from a clinical perspective

We conclude this literature review with a concise overview of four research articles that focus on clinically oriented research : Cao et al. (2023); Chen et al. (2023); Korfiatis et al. (2023); Park et al. (2023). In contrast to the articles presented above where the emphasis was on methodology, the articles in this section stand out for employing simple yet extensively validated DL methods for pancreatic cancer detection. We detail the results of these articles in Table 3.7 and discuss their commonalities : objectives, straightforward yet robust methodological approaches, large training datasets, extensive validation studies across different cohorts, and comparisons with the performance of radiologists.

3.3.4.1 Objectives

The four studies presented in this section aim to detect pancreatic cancer using DL methods. Specifically, Chen et al. (2023) and Korfiatis et al. (2023) classify a patient as cancerous/non-cancerous, while Park et al. (2023) classify a patient as having a lesion (solid or cystic) or being healthy. Cao et al. (2023) seek to detect and classify the seven most common types of pancreatic lesions. While Chen et al. (2023), Korfiatis et al. (2023), and Park et al. (2023) use contrast-enhanced CT scans as input, Cao et al. (2023) develop a model operating on non-contrast CT scans. This modality is well-suited for large-scale screening as it exposes the patient to lower radiation doses and avoids the side effects of contrast agents. However, detecting

Ref	Summary	Input	Train	Test	Results
Dmitriev et al. (2017)	Train separately two classifiers : a RF trained on meta-data and a CNN trained on images. Combine the two probability vectors to classify the 4 most common cysts.	*portal CT*	134 (4 cyst classes)	10f CV	**Overall Acc** : 83.6
Corral et al. (2019)	Identify neoplasia in IMPN using CNN.	*MRI*	139 (60 high-grade dysplasia)	10f CV	**AUC** : 0.78 [0.71-0.85]
Liu et al. (2019a)	Differentiate cancer from noncancer patients using the pancreas segmentation mask.	*portal CT*	436 (136 PDACs)	4f CV	- **Sen** : 80.2 ± 0.5 - **Spe** : 90.2 ± 0.2
Liu et al. (2019c)	Differentiate cancer from noncancer patients with a CNN using the image combined with the pancreas mask generated by a first segmentation model.	*portal CT*	336 (126 PDACs)	4f CV	- **Sen** : 92 - **Spe** : 97
Liu et al. (2020)	Differentiate cancer from noncancer patients using DL on patches sampled from the pancreas and tumor.	*portal CT*	551 (295 PDACs)	- Loc 1 : 139 (75 PDACs) - Loc 2 : 189 (101 PDACs) - Ext : 364 ($\mathcal{D}_{NIH}$, $\mathcal{D}_{MSD}$)	**Sen** / **Spe** / **Acc** : - Loc 1 : 98.6 / 97.3 / 100 - Loc 2 : 98.9 / 99.0 / 98.9 - Ext : 83.2 / 79.0 / 97.6
Zhang et al. (2020b)	Detect pancreatic cancer using a faster RCNN.	*portal CT*	2650 PDACs	Loc : 240 PDACs	- **AUC** : 0.9455 - **Sen** : 83.76 - **Spe** : 91.79 - **Acc** : 90.18
Kriegsmann et al. (2021)	Identification of different anatomical tissue structures and diseases on histopathological images of pancreatic tissue specimens.	*WSI*	81 164 patches (11 classes, including 16515 PDACs)	20% of the training data	**BAcc** : - Malignant : 81.84 - Benign : 96.53
Zhou et al. (2023)	Joint segmentation and classification between controls pancreas / PDAC / 8 other lesion subtypes using a 3D U-Net and a transformer. The latter takes the feature maps of the former + additional meta-data.	*multiphase CT*	2372 (10 classes, 1088 PDACs)	Loc : 724 (10 classes, 283 PDACs)	- **Acc** : 82.9 - **BAcc** : 56.2
Qu et al. (2023a)	Three-step framework based on shape normalization, MIL and fine-tuning of the classification threshold.	*portal CT*	310 (164 PDACs)	- Loc : 5f CV - Ext : 316 (150 PDACs) - $\mathcal{D}_{MSD}$	**AUC** / **Sen** / **Spe** / **Acc** : - Loc : 0.948 ± 0.004 / 88.8 ± 1.5 / 89.1 ± 2.2 / 89.0 ± 0.8 - Ext : 0.960 ± 0.003 / 92.5 ± 2.4 / 88.7 ± 2.6 / 90.6 ± 0.6 **Sen** / **Acc** : - $\mathcal{D}_{MSD}$: 0.935 ± 0.020 / 93.5 ± 2.0

Table 3.6 – Summary of methods for pancreas classification based on CNNs. For each article, we report a summary of the approach, the modality used as input, the train dataset, the test sets and the evaluation measures on the test sets as reported by the authors. f CV indicates that the results were obtained doing a f fold cross-validation procedure. A local (loc) set consists of data originating from the same source as the training dataset, usually a validation or test hold-out set. On the other hand, an external (ext) set includes data from an institution different from the one that provided the training set. $\mathcal{D}_{NIH}$ and $\mathcal{D}_{MSD}$ refer to public datasets containing 80 and 281 abdominal CT scans, respectively. These datasets are detailed in Section 4. The mean sensitivity (Sen), specificity (Spe), accuracy (Acc), balanced accuracy (BAcc) are given in %. When reported by the authors, standard deviations and 95% confidence intervals are indicated by ± and brackets, respectively. *AUC : Area under the curve, CNN : Convolutional neural network, CT : Computed tomography, DL : Deep learning, MIL : Multiple instance learning, MRI : Magnetic resonance imaging,* $\mathcal{D}_{MSD}$ *: Medical Segmentation Decathlon dataset,* $\mathcal{D}_{NIH}$ *: National Institute of Health dataset, PDAC : Pancreatic ductal adenocarcinoma, RF : Random forest, RCNN : Region-based convolutional neural network, WSI : Whole slide imaging.*

pancreatic cancer in non-contrast images is challenging for radiologists, even experienced ones, making a diagnostic aid tool essential.

3.3.4.2 Methods

These clinical papers employ simple methodological approaches : Park et al. (2023) opt for a basic segmentation model based on a 3D nnU-Net, followed by a simple heuristic based on the number of voxels belonging to the predicted lesion. Chen et al. (2023) and Korfiatis et al. (2023) develop a classification model based on a 3D CNN. This CNN takes a two-channel ROI as input : the image and a binary segmentation mask of the pancreas. This mask is generated by a first segmentation network performing pancreas and lesion segmentation, but the resulting segmentations are combined into a binary mask to avoid negatively impacting the classification model if the first model confuses the lesion with the pancreas. Cao et al. (2023) propose a three-step methodology : pancreas localization, lesion detection, and lesion classification. First, a pancreas segmentation model generates an ROI, which is then fed into a second model performing segmentation and binary classification (normal vs. abnormal pancreas). If the classification score exceeds

a certain threshold, the ROI is passed to a third model, a U-Net combined with a transformer, which segments the lesion and classifies it into eight categories. As mentioned earlier, Cao et al. (2023) train their models on non-contrast images. Since these images are challenging for radiologists to interpret, the annotations required for model training are difficult to obtain. To overcome this challenge, annotations are performed by radiologists on the portal phase and then transferred to the non-contrast phases through registration.

3.3.4.3 Training data

The models are trained on large datasets, often exceeding 1000 samples : Cao et al. (2023), Chen et al. (2023), Korfiatis et al. (2023), and Park et al. (2023) train their methods on 852, 1023, 1776, and 3,208 patients, respectively. The proportions of pathological cases vary : 39% (Korfiatis et al., 2023), 43% (Chen et al., 2023), 59% (Park et al., 2023), 71% (Cao et al., 2023). For enhanced robustness, Chen et al. (2023) and Park et al. (2023) train multiple models following a cross-validation approach and ensemble the different models obtained during inference. After training, the models are validated on various test cohorts.

3.3.4.4 Validation of the methods

Local and external test sets In addition to testing on local data (i.e., data from the same source as the training data), the methods are evaluated on external data to assess their generalization ability. Thus, the methods of Cao et al. (2023), Chen et al. (2023), and Park et al. (2023), trained on single-center data, are validated on external datasets consisting of 589, 1473, and 5337 patients, respectively. Notably, the external dataset of Chen et al. (2023) includes data from multiple institutions in Taiwan, and the external dataset of Cao et al. (2023) aggregates data from nine centers. Korfiatis et al. (2023) gather various public datasets for external validation. We present the results on these different cohorts in Table 3.7.

Real-world cohorts While external cohorts allow assessing the model performance on unseen data, they do not evaluate the model performance in real-world scenarios. To bridge the clinical translation gap, Cao et al. (2023) and Korfiatis et al. (2023) test their methods on "real-world" cohorts, i.e., cohorts displaying low lesion prevalence or prevalence comparable to those encountered in routine clinical practice.

Korfiatis et al. (2023) bootstrap their local test cohort to simulate a cancer prevalence like that in a population with diabetes, ranging from 1 to 5%. The authors achieve an AUC of 0.97, similar to that obtained on internal data. Additionally, Cao et al. (2023) use a cohort of 16,429 patients, with a lesion prevalence of 1%, obtaining a sensitivity and a specificity of 84.6% and 99.0% for lesion detection, respectively. After retraining the model using challenging examples and excluding control cases with peri-pancreatic diseases, the updated model achieves a lesion detection and PDAC identification with a specificity of 99.9% on a new cohort of 4,110 patients wherein the lesion prevalence is 4%. Their model successfully identifies PDAC, IPMN and chronic pancreatitis with a sensitivity of 99.5%, 92.6%, and 99.0%, respectively.

Additional test cohorts Other test cohorts are assembled to further refine the analysis of the models generalization capabilities. For instance, Korfiatis et al. (2023) create a local cohort comprising 100 pre-diagnostic exams, with a median time of 475 days before clinical diagnosis. Despite being trained only on large lesions, the method detects cancers in pre-diagnostic exams with an AUC of 0.91. Similarly, Cao et al. (2023) form a cohort containing thoracic exams and demonstrate that, despite being trained on abdominal images, their model can detect lesions with an AUC of 0.979. Notably, this performance is achieved without the need for domain adaptation, signifying that the model operates effectively without specific tuning on chest CT scans.

Subgroup analysis Subgroup analyses are systematically conducted on small lesions since their detection implies early diagnosis. Cao et al. (2023) report sensitivities of 85.7% and 92.2% for the detection of PDACs < 2cm in their internal and external test cohorts, respectively. Chen et al. (2023) report sensitivities of 38% and 16% for the detection of cysts < 1cm in their internal and external test cohorts, respectively. Park

Ref	Summary	Input	Train	Test	Results
Chen et al. (2023)	Cross-validation ensembling CNNs performing pancreas and tumor segmentation and apply a heuristic to differentiate cancer from non-cancer patients.	*portal CT*	1023 (437 PDACs)	- Loc : 256 (109 PDACs) - Ext : 1473 (669 PDACs)	**AUC / Sen / Spe** : - Loc : 0.96 [0.94, 0.99] / 89.9 [82.7, 94.9] / 95.9 [91.3, 98.5] - Ext : 0.95 [0.94, 0.96] / 89.7 [87.1, 91.9] / 92.8 [90.8, 94.5]
Park et al. (2023)	Cross-validation ensembling 3D nnU-Nets performing pancreas and lesion segmentation. Then, a heuristic differentiates healthy from pathological patients (solid or cystic lesion).	*portal CT*	852 (503 with solid and cystic neoplams)	- Loc : 603 (133 with solid and cystic neoplams) - Ext : 589 (114 with solid and cystic neoplams)	**AUC / Sen / Spe / Acc** : - Loc : 0.91 [0.89, 0.94] / 90.2 [83.9, 94.7] / 85.1 [81.6, 88.2] / 86.2 [83.2, 88.9] - Ext : 0.87 [0.84, 0.89] / 83.3 [75.2, 89.7] / 82.7 [79.0, 86.0] / 82.9 [79.6, 85.8]
Korfiatis et al. (2023)	Differentiate cancer from non-cancer patients using a CNN which takes as input the image and a binary pancreas mask. The latter is generated by a first segmentation model and includes the tumor (if segmented).	*portal CT*	1776 (696 PDACs)	- Loc 1 : 1238 (409 PDACs) - Loc 2 : 234 (100 prediagnostic PDACs) - Loc real-world : Loc 1 bootstrapped 1000 times to reach a PDACs prevalence in 1-5% - Ext public : 80 from $\mathcal{D}_{NIH}$, 152 from $\mathcal{D}_{MSD}$, 42 from CPTAC	**AUC / Sen / Spe / Acc** : - Loc 1 : 0.97 [0.96, 0.98] / 88 [85, 91] / 95 [93, 96] / 92 [91, 94] - Loc 2 : 0.91 [0.86, 0.94] / 75 [67, 84] / 90 [85, 95] / 84 [79, 88] - Ext public : 0.90 [0.86, 0.95] / 88 [83, 92] / 83 [74, 90] / 86 [82, 90] **AUC / Sen / Pre / Acc** : - Loc real-world : 0.97 [0.94, 0.99] / 92 [86, 95] / 71.1 [60, 08] / 95 [94, 95]
Cao et al. (2023)	Differentiate cancer from non-cancer patients using a three-steps approach : pancreas localization (U-Net), lesion detection (U-Net based), lesion classification (U-Net + transformer). Use manual segmentations done on portal CT, aligned on non-contrast CT.	*non-contrast CT*	3208 (1431 PDACs, 839 non-PDACs)	- Loc : 902 (475 PDACs, 311 non-PDACs) - Loc chest CT : 492 (63 PDACs, 51 non-PDACs) - Loc real-world : 16 429 (44 PDACs, 135 non-PDACs) - Ext : 5337 (2737 PDACs, 932 non-PDACs)	**AUC / Sen / Spe** : - Loc : 0.996 [0.991–1.00] / 94.9 [91.4–97.8] / 100 [100–100] - Loc chest CT : 0.979 [0.962–0.993] / 86.0 [79.4–91.9] / 98.9 [97.8–100] - Loc real-world : NA / 84.6 [79.4–89.9] / 99.0 [98.9–99.2] - Ext : 0.984 [0.980–0.987] / 93.3 [92.5–94.1] / 98.8 [98.3–99.4]

Table 3.7 – Summary of clinical studies using deep learning-based methods for pancreas cancer classification. For each article, we report a summary of the approach, the modality used as input, the train dataset, the test sets and the evaluation measures on the test sets as reported by the authors. f CV indicates that the results were obtained doing a f fold cross-validation procedure. A local (loc) set consists of data originating from the same source as the training dataset, usually a validation or test hold-out set. On the other hand, an external (ext) set includes data from an institution different from the one that provided the training set. $\mathcal{D}_{NIH}$ and $\mathcal{D}_{MSD}$ refer to public datasets containing 80 and 281 abdominal CT scans, respectively. These datasets are detailed in Section 4. CPTAC refers to a public dataset containing 108 adbominal exams including 42 portal CTs (Consortium, 2018). The mean sensitivity (Sen), specificity (Spe), accuracy (Acc), precision (Pre) are given in %. When reported by the authors 95% confidence intervals are indicated by brackets. NA means the information was not reported. *AUC : Area under the curve, CNN : Convolutional neural network, CT : Computed tomography, $\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset, $\mathcal{D}_{NIH}$: National Institute of Health dataset, PDAC : Pancreatic ductal adenocarcinoma.*

et al. (2023) report sensitivities of 87.5% and 74.7% for the detection of lesions $<$ 2cm in their internal and external test cohorts, respectively. Finally, Korfiatis et al. (2023) report sensitivities of 80% and 76% for the detection of lesions $<$ 2cm in their internal and external test cohorts.

3.3.4.5 Evaluating performance against radiologists

Finally, these clinically oriented papers propose to compare the performance of their DL-based methods with the performance of radiologists. In particular, Park et al. (2023) show that their DL model exhibits similar sensitivities to radiologists for the detection of solid lesions of any size. Regarding cystic lesions, the DL model is comparable to radiologists for lesions measuring at least 1cm but shows significantly lower sensitivity for sub-centimeter cystic lesions. However, these small cysts tend to remain stable and have little clinical impact (Park et al., 2023).

From a cohort of 291 patients with surgical or biopsy pathology diagnoses, Cao et al. (2023) demonstrate that their method performs significantly better than a pool of 33 radiologists for lesion detection : the method outperforms them by +14.7% in sensitivity and +6.8% in specificity. The improvement is even more pronounced for PDAC detection : +34.1% in sensitivity and +6.3% in specificity. A second analysis on the same cohort shows that the DL method, operating on contrast-free scans, either matches or outperforms the average performance of 15 radiologists using contrast-enhanced scans.

3.4 Conclusion

In this section, we have reviewed various existing deep learning approaches for pancreas segmentation, pancreatic lesion detection, and feature extraction for pancreas analysis.

Pancreas segmentation Pancreas segmentation encounters challenges due to its complex geometry and surroundings. Its small and elongated volume, along with several nearby structures, and low contrast at its extremities, require processing both global and local information. Multiple strategies have been proposed in attempts to combine these different levels of information : 3D methods may not be inherently suitable as the pancreas occupies a small fraction of abdominal CT, but 2D methods lack the essential spatial context for pancreas localization. Various compromises have been attempted, including inter-slice regularization, 2.5D inputs, and CNNs integrating both 2D and 3D operations. Coarse-to-fine approaches (C2F), involving the division of the segmentation task into localization and fine segmentation, are the most widely used. The first step benefits from the complete spatial context, while the second step obtains information at the pancreas scale. However, these two steps are often optimized independently despite their intrinsic interdependence. Some authors tried to optimize both steps end-to-end, but computational capabilities limit this possibility. Moreover, the efficiency of C2F methods can be enhanced : the number of parameters is significantly increased due to the introduction of two models, and the extraction of the same low-level features is repeated in different stages. In response to this, methods aim to improve learning efficiency through attention modules trained to filter and amplify relevant signals. Some papers have also explored strategies to process information at different scales in a single model. For this purpose, methods have proposed varying receptive field sizes or exploiting feature maps from different stages of a U-Net. However, like C2F methods, these approaches introduce a large number of additional parameters. More recently, authors have demonstrated the benefits of coupling transformers with U-Nets to combine the global information extracted by transformers with the local information extracted by U-Nets. While all these methods have tried to tackle the complexities associated with pancreas segmentation, there has been little focus on rectifying segmentation errors that hold clinical significance. We analyzed a state-of-the-art method, the nnU-Net proposed by Isensee et al. (2021a), and observed that problematic errors occur at the extremities of the pancreas.

Consequently, this thesis introduces a novel approach to rectify these errors. The proposed method, presented in Chapter 5, leverages geometric priors to locally adjust the sensitivity of the segmentation in a model-agnostic manner. Notably, this approach succeeds in substantially improving the segmentation at the extremities of the organ, without compromising the overall segmentation quality.

Lesion detection The detection of pancreatic lesions currently remains an active area of research. The detection task typically involves an initial lesion segmentation step, facing similar challenges as parenchymal segmentation. As illustrated by Tables 3.3 and 3.1b, the DSC do not exceed 67% and 62% on the training and testing splits of the Medical Segmentation Decathlon dataset, respectively. To facilitate the segmentation, authors incorporate additional information, either through anatomical information (e.g., simultaneous pancreatic duct segmentation) or by adding imaging phases. The first strategy requires additional manual annotations, and the second strategy raises issues related to image registration. In any case, these strategies underscore the value of adding information. This additional information is particularly helpful in detecting isodense lesions or distinguishing lesions from other pathologies, such as fatty pancreas or chronic pancreatitis. This rationale aligns with the radiologists' practice of focusing on secondary signs, as described in Section 2.3.5. Therefore, in Chapter 6, we propose a lesion detection method that explicitly incorporates a secondary sign of significant importance, namely, the dilatation of the main pancreatic duct.

To achieve this, we suggest jointly segmenting the pancreas, the lesion and the main pancreatic duct. The resulting segmentations are leveraged to extract features which are subsequently used to predict the presence of a lesion and the dilatation of the main pancreatic duct. Developed on an extensive training cohort of over 2000 patients, this method has been evaluated on an external test cohort comprising 756 patients, demonstrating its robustness.

Feature extraction for pancreas analysis As seen in Section 3.3.3, most methods that extract radiomics features (HCR or DLR) from images use them for detecting patients with cancer. However, quantitative results show that segmentation-based methods are more effective for this task. Therefore, as presented in Table 3.7, many large-scale clinical studies prefer to use segmentation for cancer patient detection. Nonetheless, we posit that feature extraction can provide significant value in identifying secondary signs of pancreatic cancer, which are essential for early detection as they may manifest before the clear onset of the tumor.

In Chapter 7, we introduce two methods for detecting secondary signs of pancreatic cancer.

Starting with the detection of abnormal pancreatic shapes, we introduce a VAE-based approach to learn a normative model of healthy pancreatic shapes. This normative model is then leveraged to detect abnormal shapes. The model is learnt on a large database of 1200 healthy pancreas shapes and subsequently evaluated on a separate test set of 224 pancreas from patients with mixed conditions.

Then, we expand this work to identify four secondary signs of pancreatic cancer : abnormal shape, atrophy, senility, and fat replacement. Our method combines predefined HCR with DLR features. A non-redundancy constraint between these two types of features is used in order to extract complementary DLR features. Then, the two sets of radiomics are combined and used to detect the secondary signs. Experiments highlight the value of combining non-redundant DLR and HCR features, as evidenced by an improvement in the detection performance compared to methods that do not address redundancy or solely rely on HCR features.

Chapitre 4

Datasets

4.1 Preface

The creation of a large database stood as the primary focus during the initial months of this thesis. This phase was not only essential for this work but also aligned directly with Guerbet's strategic objective of product launch. Although less research-oriented, these activities played an indispensable role in implementing high-quality experiments, ensuring the overall success of the thesis project.

The initiation of the database work involved the formulation of working hypotheses to identify the requisite data and annotations. Then, computer development efforts were undertaken to systematically collect, clean, and organize the data. This process unfolded progressively as data were delivered by various suppliers, leading to the utilization of distinct datasets for each of the articles published throughout this thesis. The database went from zero patient exams in April 2021 to a total of over 2800 patient exams by January 2024, including the integration of two public databases into our repository.

In this section, we present the diverse datasets employed in this research project. Particular emphasis will be placed on detailing the Care Advisor for Pancreas database (refered to as $\mathcal{D}_{CAPA}$), Guerbet's proprietary database specifically designed for pancreatic research. Subsequently, we detail the process of annotations implemented for all the acquired images.

4.2 Data

The work presented in this thesis leverages a combination of public and private datasets, the latter belonging to Guerbet.

4.2.1 Public

The public datasets employed in this study include two datasets widely adopted by the scientific community for the development of Artificial Intelligence (AI) methods for pancreatic applications. We present an overview of their content, acquisition parameters, as well as the provided annotations.

4.2.1.1 National Institute of Health

Overview The National Institute of Health (NIH) clinical center provides whole abdomen Computed Tomography (CT) scans of healthy pancreases for public use (Roth et al., 2016a). This dataset is widely used by the medical imaging community for the development of pancreas segmentation algorithms (see Table 3.2). It is also commonly employed as a control cohort in problems involving the classification of cancer versus non-cancer cases (see Tables 3.4, 3.5, 3.6 and 3.7).

Content This dataset contains contrast-enhanced 3D CT scans (approximately 70 seconds after intravenous contrast injection in the portal-venous phase) of 80 subjects with morphologically normal pancreases. We refer to this dataset as $\mathcal{D}_{NIH}$. Among the 80 subjects, 17 are healthy kidney donors scanned before nephrectomy. The remaining 63 were selected by a radiologist from patients without major abdominal pathologies nor pancreatic cancer lesions. The cohort consists of 53 males and 27 females, with ages ranging from 18 to 76 years and a mean age of 46.8 ± 16.7.

Image Acquisition CT scans in this dataset were acquired using Philips and Siemens MDCT scanners with a tube voltage of 120 kVp. More information about image resolutions and pixel sizes are shown in Table 4.1.

Image Annotations Reference annotations for pancreas segmentations were performed through slice-by-slice manual segmentation by a medical student, subsequently verified and modified by an experienced radiologist.

Figure 4.1 shows a portal CT scan from the $\mathcal{D}_{NIH}$ dataset, along with the corresponding manual segmentation of the pancreas.

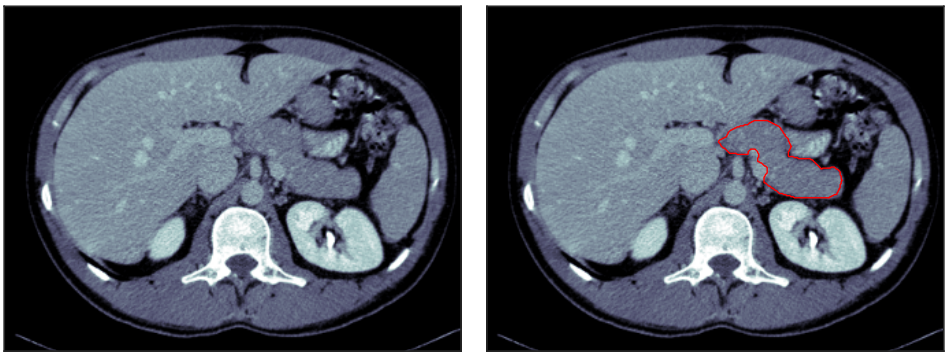


Figure 4.1 – Healthy pancreas from the $\mathcal{D}_{NIH}$ dataset (Roth et al., 2016a). Portal CT image (left) and corresponding manual segmentation (right). The pancreas is outlined in red. *CT : Computed tomography, NIH : National Institute of Health dataset.*

4.2.1.2 Medical Segmentation Decathlon

Overview As previously detailed (see Section 3.2.3.6), the Medical Segmentation Decathlon (MSD) is a renowned challenge launched to assess and compare the efficacy of medical image segmentation algorithms across diverse anatomical structures and imaging modalities (Antonelli et al., 2022). Initiated in 2018 as part of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), this challenge includes ten segmentation tasks, each associated with a distinct anatomical structure. The datasets employed in these tasks have been made openly accessible to the research community, promoting transparency, collaboration, and benchmarking. For this thesis, we used the MSD pancreas dataset, specifically intended for pancreas and pancreatic lesion segmentation and therefore widely adopted by researchers developing pancreatic segmentation algorithms and lesion detection methods (see Tables 3.2, 3.3, 3.4, 3.6, and 3.7).

Content The MSD pancreas dataset consists of 281 portal CT scans obtained from 281 distinct patients. We denote this dataset as $\mathcal{D}_{MSD}$. These images are sourced from patients undergoing resection of pancreatic masses (Intraductal Papillary Mucinous Neoplasm (IPMN), Pancreatic Neuroendocrine Tumor (PNET), or Pancreatic Ductal Adenocarcinoma (PDAC)). Yet, information about the exact lesion type for a given CT is not available.

Image Acquisition Data acquisition occurred at the Memorial Sloan Kettering Cancer Center in New York, US, and portal venous phase CT scans were obtained for each patient. Some reconstruction and acquisition parameters, as reported by Simpson et al. (2019), are shown in Table 4.1.

Image Annotations Expert abdominal radiologists manually segmented the pancreatic parenchyma and pancreatic masses (cysts or tumors) using the Scout application (Dawant et al., 2007). Segmentations were publicly available for the 281 cases of $\mathcal{D}_{MSD}$.

4.2.1.3 Discussion

While these two public datasets contribute significantly to the research community by providing manual segmentations and facilitating benchmarking, they present notable limitations.

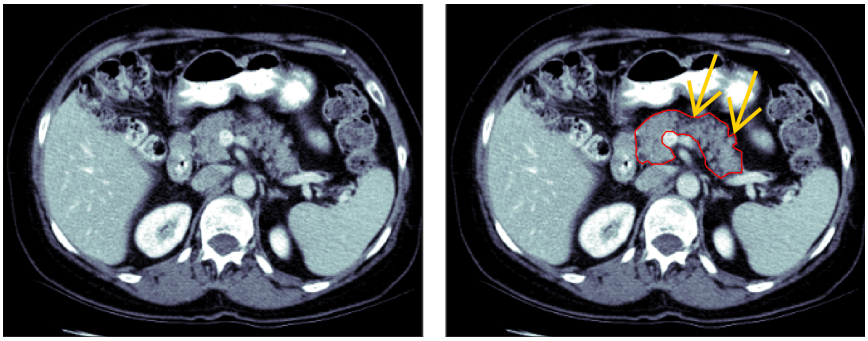
First, the total number of cases is highly insufficient for clinical applications. This raises concerns regarding potential overfitting within the Deep Learning (DL) community working on pancreas imaging, particularly in the domain of pancreas segmentation, where the majority of methods are benchmarked against the $\mathcal{D}_{NIH}$ dataset (see Table 3.2).

Secondly, both datasets lack essential information regarding patient demographics and clinical status.

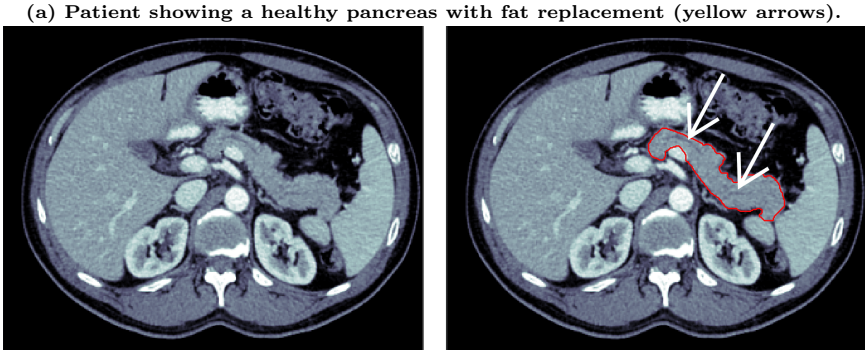
Thirdly, the control cases within the $\mathcal{D}_{NIH}$ dataset may not be fully representative of real-world pancreases, as they predominantly consist of a notably young and healthy population (mean age : 46.8 ± 16.7 years, with 20% being kidney donors). Despite some cases in the $\mathcal{D}_{NIH}$ dataset exhibiting ageing characteristics such as atrophy, fat replacement, and a visible main pancreatic duct (see Figure 4.2), these instances are limited. Moreover, even with optimal image quality, some manual segmentations in the $\mathcal{D}_{NIH}$ dataset were found to underestimate and miss portions of the pancreas (Suman et al., 2021). Ultimately, while the initial release of the $\mathcal{D}_{NIH}$ dataset in 2015 included 82 cases, an updated version in 2020 revealed that three cases originated from the same CT scan, differing only in minor cropping. Consequently, the dataset was adjusted to a total of 80 unique cases. Prior studies using this dataset may have inadvertently included these duplicates, potentially introducing bias into their findings. In this thesis, we exclusively used the revised dataset, accounting for the corrected count of 80 distinct cases.

Regarding the pathological cases within the $\mathcal{D}_{MSD}$ dataset, Suman et al. (2021) discovered that approximately 25% of them exhibited a biliary stent[1] (see Figure 4.3), and two cases showed imaging features suggestive of post-chemotherapy status. The presence of biliary stents poses a significant challenge as it

1. A biliary stent is a thin, hollow tube that is placed in the bile duct. The stent holds the duct open after the duct has been blocked or partly blocked.



(a) Patient showing a healthy pancreas with fat replacement (yellow arrows).



(b) Patient with a healthy pancreas. The main pancreatic duct is not dilated but is visible (white arrows).

Figure 4.2 – Healthy cases from $\mathcal{D}_{NIH}$ exhibiting benign characteristics. Portal CT images (left) and their corresponding manual segmentations (right). The pancreas is outlined in red. *CT : Computed tomography, $\mathcal{D}_{NIH}$: National Institute of Health dataset.*

can introduce biases in AI models (Rueckel et al., 2020) and create artifacts that obscure the delineation of tumors. Furthermore, the $\mathcal{D}_{MSD}$ dataset lacks information about the pathological status (clinical details, treatment history, histopathology) of the patients. Specifically, information about the exact pathology of pancreatic tumors in each CT scan within the $\mathcal{D}_{MSD}$ dataset is not publicly available.

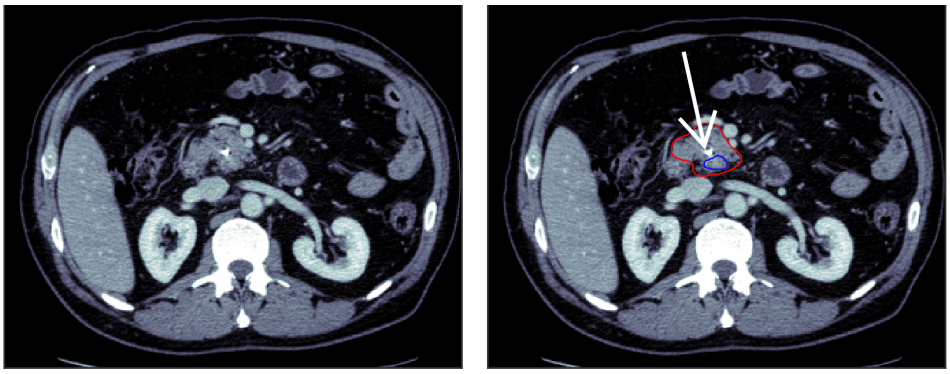


Figure 4.3 – Patient from the $\mathcal{D}_{MSD}$ dataset with a heterogeneous area (outlined in blue) in the head of the pancreas. A biliary stent is indicated by a white arrow. Portal CT image (left) and corresponding manual segmentation (right). The pancreas is outlined in red. Image from the $\mathcal{D}_{MSD}$ dataset (Antonelli et al., 2022). *CT : Computed tomography, $\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset.*

Therefore, the existing limitations in both the $\mathcal{D}_{NIH}$ and $\mathcal{D}_{MSD}$ datasets highlight the need for Guerbet to gather more comprehensive datasets for pancreatic imaging, including diverse patient demographics, detailed clinical information, and a broader spectrum of pathological cases. The acquisition of such data aims to not only enhance the robustness and generalizability of Guerbet's AI models but also contribute to a more accurate representation of the complexities associated with pancreatic diseases in real-world clinical scenarios.

	$\mathcal{D}_{NIH}$	$\mathcal{D}_{MSD}$
Country	USA	USA
Number of patients	80	281
Gender	Male : 53 Female : 27	NA
Age	46.8 ± 16.7	NA
Number of portal studies	80	281
Manufacturer	Philips, Siemens	NA
Voxel size in z (mm)	0.99 ± 0.06	2.90 ± 0.98
Voxel size in x-y (mm)	0.85 ± 0.09	0.81 ± 0.09
Tube voltage (kVp)	120	120
Scan delay (s)	≈ 70	80-85

Table 4.1 – Demographic and technical information per institution for the two public pancreas datasets, as reported by Roth et al. (2016a); Simpson et al. (2019). If not specified, data are means ± standard deviations. Square brackets indicate ranges. *$\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset, $\mathcal{D}_{NIH}$: National Institute of Health dataset, NA : Not available.*

4.2.2 Private datasets

4.2.2.1 Pancreas database

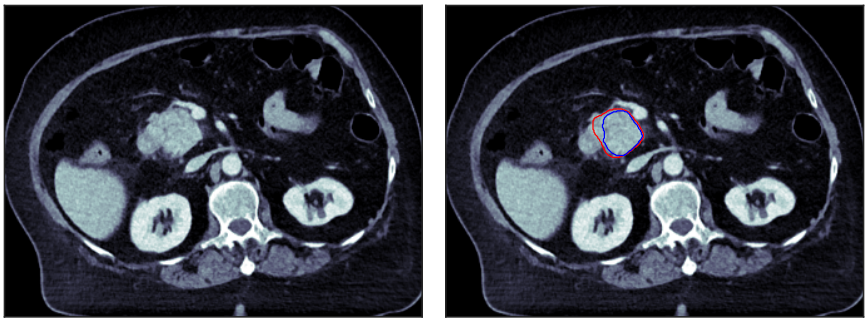
Overview The Care Advisor for Pancreas database, refered to as $\mathcal{D}_{CAPA}$, was created exclusively for Guerbet's pancreatic cancer AI project. Its collection was seamlessly integrated into the initial half of the thesis work, representing an important aspect of the research work. The creation of this database required discussions with radiologists to define the needs in terms of data and annotations, and computer development to clean and organize the data within the Guerbet database. We exclusively collected images featuring pathological pancreas conditions, with a notable emphasis on PDACs.

The $\mathcal{D}_{CAPA}$ dataset was annotated following the annotation process detailed in Section 4.3.

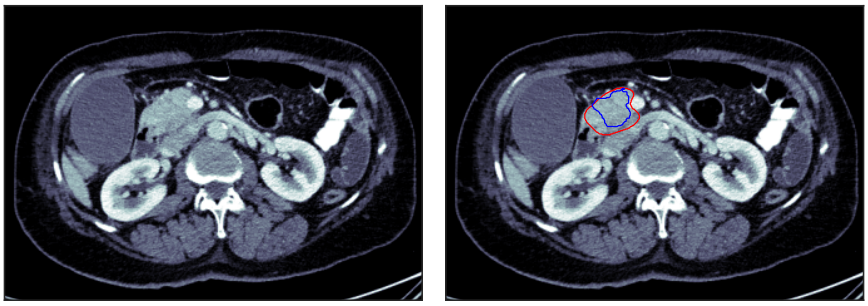
Technical specifications The primary focus of the database centers on portal phase scans, a systematic modality in abdominal examinations and on which pancreatic cancers are frequently missed (see Chapter 2), particularly when they display iso-enhancement, as depicted in Figure 4.4. Focusing on portal phase scans holds the potential to significantly decrease the mortality rate of pancreatic cancer by enhancing radiologists' sensitivity in this specific modality. Consequently, the designated modality requirements were defined as either a multi-phase exam, including at least a portal phase contrast CT scan, or a single-phase exam featuring a portal phase contrast CT scan of the abdomen. Technical specifications stipulated that images should be in DICOM format with a voxel size of less than 3mm in the z-axis.

Inclusion and exclusion criteria The formulation of inclusion and exclusion criteria was a meticulous process guided by insightful discussions with radiologists. Inclusion criteria targeted patients showing solid lesions (notably PDAC), patients with cystic pancreas (irrespective of benign or malignant nature) or patients with normal yet ageing pancreas conditions. The diagnosis of PDAC was confirmed either by examining biopsy reports or by referencing the C25 code in the *International Classification of Diseases, 10th Revision* (ICD) (DiSantostefano, 2009). Similarly, cystic lesions were identified based on their respective codes in the ICD. Figures 4.5a and 4.6a showcase examples of solid lesions, whereas Figure 4.5b illustrates a case of a cystic lesion. In addition, efforts were made to include a broad spectrum of pancreatic conditions, including cases of pancreatitis and fat replacement (see Figure 4.6). Rigorous exclusion criteria were applied, excluding studies acquired post-treatment (drug or surgery) and patients with biliary stents.

Content The creation of $\mathcal{D}_{CAPA}$ started at the beginning of the thesis and was then carried out progressively as data were delivered by the various suppliers. By January 2024, the dataset contained nearly 2500 portal CT scans sourced from five institutions, covering a spectrum of single-phase and multi-phase acquisitions (non-enhanced, arterial, portal, and delayed). Distinct sub-datasets are identified based on their originating institutions, categorized by their respective countries of origin : $\mathcal{D}_{BR,1}$ from Brazil, $\mathcal{D}_{FR,1}$ and $\mathcal{D}_{FR,2}$ from France, and $\mathcal{D}_{US,1}$ and $\mathcal{D}_{US,2}$ from the USA. The key characteristics for each institution are summarized in Table 4.2.

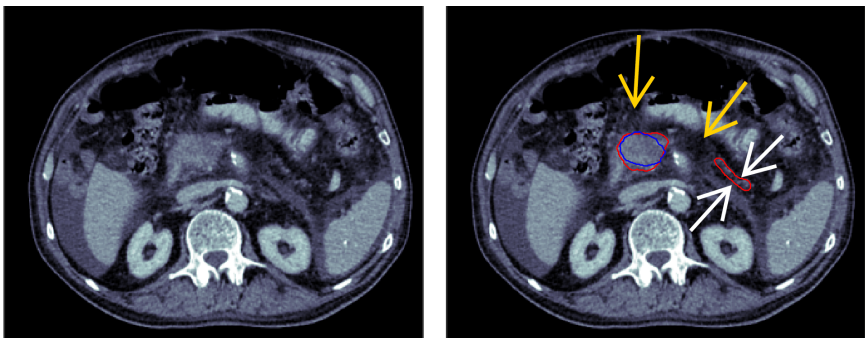


(a) 71-year-old woman with iso-enhancing PDAC.

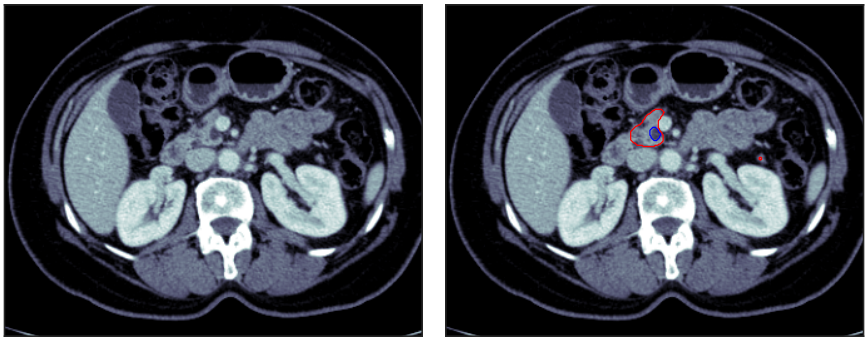


(b) 78-year-old woman with iso-enhancing PDAC.

Figure 4.4 – Images from $\mathcal{D}_{CAPA}$ dataset showing iso-enhancing PDACs. Portal CT images (left) and their corresponding manual segmentations (right). The pancreas and the lesion are outlined in red and blue, respectively. *CT : Computed tomography, $\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, PDAC : Pancreatic ductal adenocarcinoma.*

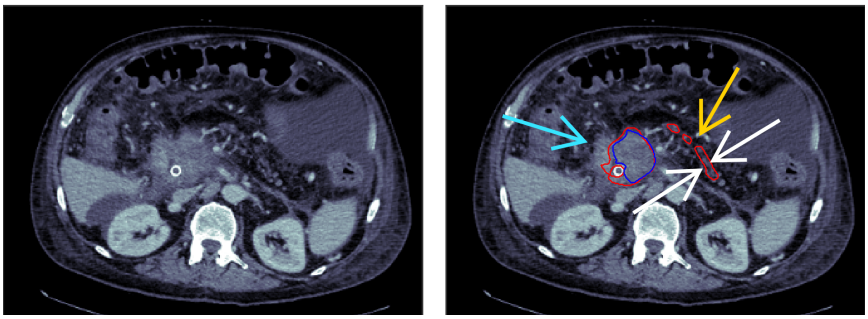


(a) 79-year-old man with hypo-enhancing PDAC and densification of peri-pancreatic fat (yellow arrows). The pancreas is atrophic (white arrows).

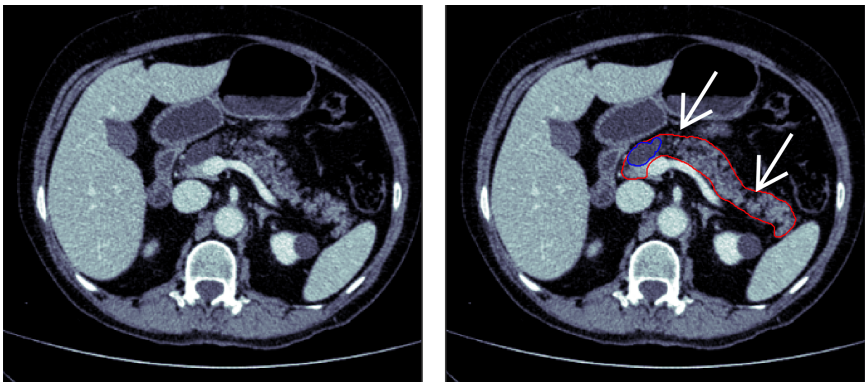


(b) 59-year-old woman with benign hypo-enhancing branch-duct IPMN.

Figure 4.5 – Solid and cystic lesions from $\mathcal{D}_{CAPA}$ dataset. Portal CT images (left) and their corresponding manual segmentations (right). The pancreas and the lesion are outlined in red and blue, respectively. *CT : Computed tomography, $\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, IPMN : Intraductal papillary mucinous neoplasm, PDAC : Pancreatic ductal adenocarcinoma.*



(a) Patient with hypo-enhancing PDAC. The pancreas shows severe atrophy (white arrows) and fat replacement (yellow arrow) in the body and tail, as well as inflammatory stranding in the head (cyan arrow) due to pancreatitis.



(b) 60-year-old woman with benign hypo-enhancing branch-duct IPMN. The pancreas exhibits fat replacement (white arrows).

Figure 4.6 – Pathological cases from $\mathcal{D}_{CAPA}$ dataset. Portal CT images (left) and their corresponding manual segmentations (right). The pancreas and the lesion are outlined in red and blue, respectively. *CT : Computed tomography, $\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, IPMN : Intraductal papillary mucinous neoplasm, PDAC : Pancreatic ductal adenocarcinoma.*

Institution	$\mathcal{D}_{BR,1}$	$\mathcal{D}_{FR,1}$	$\mathcal{D}_{FR,2}$	$\mathcal{D}_{US,1}$	$\mathcal{D}_{US,2}$
Country of origin	Brazil	France	France	USA	USA
Number of patients	1004	244	155	691	353
Gender	Male : 429 Female : 575	Male : 133 Female : 112	NA : 155	Male : 301 Female : 284 NA : 106	Male : 175 Female : 178
Age	61.77 ± 14.43	NA	NA	NA	65.65 ± 10.65
Number of portal studies	1004	244	148	699	402
Manufacturer	GE : 52 Philips : 865 Siemens : 25 Toshiba : 47 NA : 15	NA	NA	GE : 230 Philips : 133 Siemens : 236 Toshiba : 101	Canon : 1 GE : 184 Hitachi : 1 Philips : 163 Siemens : 47 Toshiba : 8
Voxel size in z (mm)	1.80 ± 0.45	1.31 ± 0.62	1.32 ± 0.25	2.52 ± 0.64	2.78 ± 1.32
Voxel size in x-y (mm)	0.81 ± 0.10	0.78 ± 0.12	0.77 ± 0.09	0.78 ± 0.10	0.76 ± 0.09
Tube voltage (kVp)	119.70 ± 6.70	NA	110.8 ± 12.62	118.26 ± 6.20	119.64 ± 4.58
X-Ray tube current (mA)	297.11 ± 133.27	NA	360.02 ± 160.01	249.70 ± 136.55	259.16 ± 136.70
Exposure (mAs)	176.81 ± 84.5	NA	140.04 ± 111.48	113.95 ± 98.31	113.33 ± 162.66

Table 4.2 – Demographic and technical information per institution for the private pancreas dataset $\mathcal{D}_{CAPA}$. Some data is missing due to regulatory constraints related to anonymization or anonymization techniques that vary by country and/or provider. Data are means ± standard deviation. *$\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, GE : General Eletric, NA : Not available.*

4.2.2.2 Cross-project databases

As the $\mathcal{D}_{CAPA}$ dataset exclusively contains pathological pancreas data, additional private datasets were used to include images of healthy pancreases. For this purpose, abdominal images collected for other Guerbet projects were exploited. Two databases were employed : the Care Advisor for Liver ($\mathcal{D}_{CALV}$) and Care Advisor for Bone Metastasis ($\mathcal{D}_{CABO}$).

$\mathcal{D}_{CALV}$ and $\mathcal{D}_{CABO}$ datasets are proprietary collections containing 3270 and 2750 multi-institutional CT

scans, respectively. These datasets contain both single-phase and multi-phase acquisitions (non-enhanced, arterial, portal, and delayed) and were annotated by expert radiologists. In particular, $\mathcal{D}_{CALV}$ includes detailed liver and tumor annotations, and $\mathcal{D}_{CABO}$ includes detailed bone lesion annotations.

In this thesis, only portal phase CT scans were used, and specific subsets of these databases were chosen. A selection process involving automated report parsing was applied, excluding any scans indicating pathological pancreas conditions. In addition, an experienced radiologist verified the selected scans to confirm the healthy status of the pancreas. The final number of scans used from these databases depended on each particular research project.

4.2.2.3 Patient consent

All patient data used in this work was obtained with explicit informed consent, and underwent a rigorous anonymization process to ensure the privacy and confidentiality of individual information in each dataset according to local regulation.

4.3 Annotations

4.3.1 Protocol

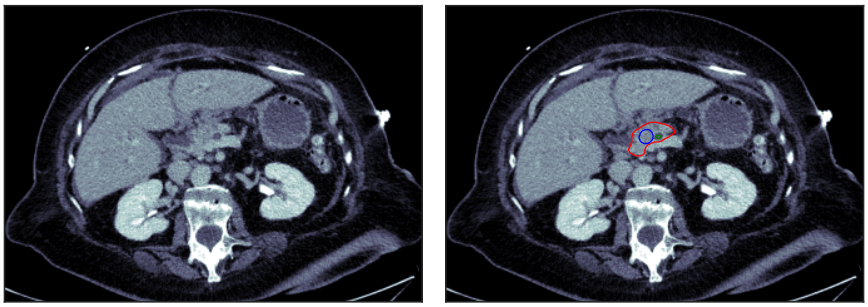
The annotation process for the pancreas was led in collaboration with radiologists to collect all pertinent information for pancreas analysis. This process was systematically applied to all images in the $\mathcal{D}_{CAPA}$ and $\mathcal{D}_{MSD}$ datasets, employing a two-level annotation approach :

- **Segmentation-based annotation** : Radiologists were asked to segment the pancreas, pancreatic lesions and the pancreatic duct when visible. In order to accelerate the segmentation process, an automatic segmentation of the parenchyma, generated by an in-house algorithm trained on public annotated data ($\mathcal{D}_{NIH}$ and $\mathcal{D}_{MSD}$), was presented to the radiologists for correction. The lesions and the main pancreatic duct were fully manually segmented. Lesion size distributions, derived from segmentations, can be visualized in Figure 4.9. Examples of segmentations are provided in Figure 4.7.
- **Question-based annotation** : Radiologists were asked to answer clinical questions related to the pancreas, potential lesions, and the pancreatic duct, based on portal image visualization. These questions were designed to replicate the cognitive processes employed by radiologists during pancreas diagnostics. Cases in which determining the lesion type proved inconclusive were denoted as unclassified. The responses to the full set of questions concerning the pancreas, lesions, and the duct can be consulted in Tables 4.3, 4.4, and 4.5, respectively.

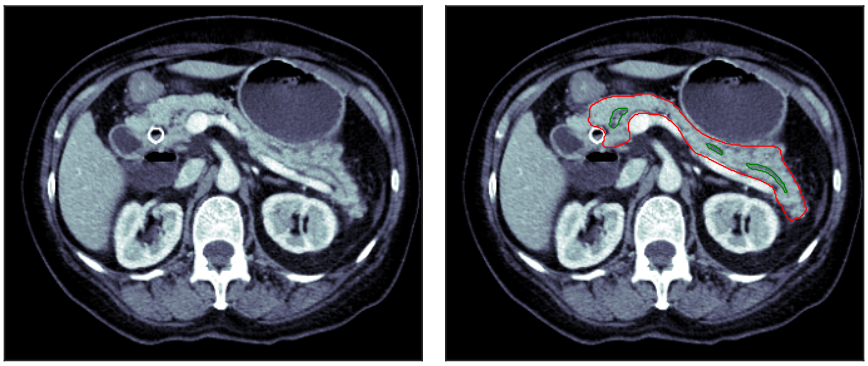
Each portal CT scan was reviewed and annotated by one in a group of 11 radiologists. In practice, the annotation was done using a custom module of the 3D Slicer software (Fedorov et al., 2012; Kikinis et al., 2013) that was developed internally. For each case, the reference 3D segmentation was generated by segmenting the pancreas, lesions, and MPD slice by slice. The customized version of 3D Slicer also allowed the radiologists to answer the clinical questions for the question-based annotation protocol. During annotation, radiologists could also consult the radiological report whenever available, especially for the precise characterization of pancreatic lesions.

4.3.2 Inter-annotator segmentation variability

To ensure a consistent annotation process across all our pathological data, $\mathcal{D}_{MSD}$ dataset was re-annotated using the same protocol, despite the availability of publicly shared manual segmentations from the Medical Segmentation Decathlon challenge. The double annotation for 281 these cases enabled us to estimate the inter-annotator variability at the dataset level. To do so, we computed for each case the Dice Similarity Coefficient (DSC) between the annotator's segmentation and the corresponding challenge segmentation. Then, we computed the mean DSC for each annotator, helping us to define a global variability between each annotator and the challenge reference segmentations for the subset of cases they reviewed.



(a) Patient showing an iso-enhancing mass in the head of the pancreas, indicated by the cutoff of the main pancreatic duct.



(b) Patient showing a severe dilatation of the main pancreatic duct.

Figure 4.7 – Portal CT images (left) and their corresponding manual segmentations (right). The pancreas, the lesion and the main pancreatic duct are outlined in red, blue, and green, respectively. Images from $\mathcal{D}_{CAPA}$ dataset. *CT : Computed tomography, $\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset.*

Number of cases	$\mathcal{D}_{BR,1}$ 1004	$\mathcal{D}_{FR,1}$ 244	$\mathcal{D}_{FR,2}$ 148	$\mathcal{D}_{US,1}$ 699	$\mathcal{D}_{US,2}$ 402	$\mathcal{D}_{MSD}$ 281
Atrophy						
Yes	446 (44)	93 (38)	51 (34)	303 (43)	167 (42)	102 (36)
No	557 (56)	151 (62)	97 (66)	394 (56)	234 (58)	179 (64)
NA	1 (0)	0 (0)	0 (0)	2 (0)	1 (0)	0 (0)
Densification						
Yes	451 (45)	84 (34)	107 (72)	336 (48)	144 (36)	109 (39)
No	553 (55)	160 (66)	41 (28)	362 (52)	257 (64)	171 (61)
NA	0 (0)	0 (0)	0 (0)	1 (0)	1 (0)	1 (0)
Fat replacement						
Yes	88 (9)	129 (53)	92 (62)	214 (31)	93 (23)	87 (31)
No	915 (91)	115 (47)	56 (38)	484 (69)	308 (77)	194 (69)
NA	1 (0)	0 (0)	0 (0)	1 (0)	1 (0)	0 (0)
Focal fat loss						
Yes	3 (0)	191 (78)	108 (73)	356 (51)	216 (54)	133 (47)
No	999 (100)	53 (22)	39 (26)	341 (49)	185 (46)	147 (52)
NA	2 (0)	0 (0)	1 (1)	2 (0)	1 (0)	1 (0)
General Shape						
Normal	520 (52)	77 (32)	20 (14)	227 (33)	216 (54)	130 (46)
Abnormal	483 (48)	167 (68)	128 (86)	471 (67)	185 (46)	151 (54)
NA	1 (0)	0 (0)	0 (0)	1 (0)	1 (0)	0 (0)
Homogeneity						
Yes	352 (35)	105 (43)	23 (16)	106 (15)	46 (11)	64 (23)
No	645 (64)	139 (57)	125 (84)	592 (85)	355 (89)	216 (77)
NA	7 (1)	0 (0)	0 (0)	1 (0)	1 (0)	1 (0)
Pancreatitis						
Yes	28 (3)	19 (8)	11 (7)	48 (7)	7 (2)	11 (4)
No	976 (97)	224 (92)	137 (93)	650 (93)	394 (98)	269 (96)
NA	0 (0)	1 (0)	0 (0)	1 (0)	1 (0)	1 (0)
Senility						
Yes	86 (9)	80 (33)	21 (14)	187 (27)	83 (21)	77 (27)
No	918 (91)	164 (67)	127 (86)	511 (73)	318 (79)	204 (73)
NA	0 (0)	0 (0)	0 (0)	1 (0)	1 (0)	0 (0)

Table 4.3 – Clinical information of the pancreatic parenchyma across diverse providers of the $\mathcal{D}_{CAPA}$ dataset. Data are number of patients with percentages in parentheses. Percentages are computed based on the number of cases for each respective provider. *$\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, NA : Not available.*

	$\mathcal{D}_{BR,1}$	$\mathcal{D}_{FR,1}$	$\mathcal{D}_{FR,2}$	$\mathcal{D}_{US,1}$	$\mathcal{D}_{US,2}$	$\mathcal{D}_{MSD}$
Visible lesion						
Yes	910	244	148	699	402	281
No	94	0	0	0	0	0
Denomination						
MD-IPMN	23 (2)	0	0	0	0	1 (0)
BD-IPMN	55 (6)	1 (0)	0	12 (2)	0	15 (5)
MCN	23 (3)	0	1 (1)	0	2 (0.5)	1 (0)
Need further examination	0 (0)	18 (7)	13 (9)	32 (5)	12 (3)	37 (13)
Other	130 (14)	4 (2)	0	6 (1)	3 (1)	1 (0)
PDAC	526⋆ (58)	103⋆ (42)	98 (66)	508 (73)	357 (89)	124 (44)
PNET	112 (12)	6 (2)	4 (3)	12 (2)	10 (2)	15 (5)
SCN	39 (4)	1 (0)	1 (1)	2 (0)	0	3 (1)
SPN	0 (0)	0	0	0	1 (0)	1 (0)
NA	2 (0)	111 (45)	31 (21)	127 (18)	17 (4)	83 (30)
Enhancement						
Heterogeneous	496 (54)	36 (15)	33 (22)	86 (12)	27 (7)	35 (12)
Hyper-enhancing	75 (8)	19 (8)	5 (3)	14 (2)	6 (1.5)	20 (7)
Hypo-enhancing	271 (30)	122 (50)	98 (66)	478 (68)	321 (80)	148 (53)
Iso-enhancing*	58 (6)	30 (12)	5 (3)	76 (11)	39 (10)	29 (10)
NA	10 (1)	37 (15)	7 (5)	45 (6)	9 (2)	49 (17)
Malignancy						
Benign	156 (17)	5 (2)	1 (1)	24 (3)	3 (0.7)	36 (13)
Doubtful	58 (6)	8 (3)	7 (5)	29 (4)	14 (3.5)	19 (7)
Malignant	689 (76)	140 (58)	131 (88)	577 (82)	376 (94)	176 (63)
NA	7 (1)	91 (37)	9 (6)	69 (10)	9 (2)	50 (18)
Margins						
Poorly defined	563 (62)	189 (78)	117 (79)	550 (78)	319 (80)	168 (60)
Well-defined	343 (38)	49 (20)	29 (20)	120 (17)	74 (18)	93 (33)
NA	4 (0.4)	6 (2)	2 (1)	29 (4)	9 (2)	20 (7)
Type						
Cystic	186 (20)	16 (7)	10 (7)	33 (5)	9 (2)	61 (22)
Solid	704 (78)	206 (85)	128 (86)	611 (87)	384 (96)	190 (68)
NA	20 (2)	22 (9)	10 (7)	55 (8)	9 (2)	30 (11)
Texture						
Heterogeneous	559 (62)	147 (61)	117 (79)	322 (46)	125 (31)	155 (55)
Homogeneous	342 (38)	81 (33)	31 (21)	345 (49)	268 (67)	106 (38)
NA	9 (1)	16 (7)	0 (0)	32 (5)	9 (2)	20 (7)

Table 4.4 – Clinical information of the pancreatic lesions across diverse providers of the $\mathcal{D}_{CAPA}$ dataset. Data are number of patients with percentages in parentheses. Percentages are computed based on the number of visible lesions for each respective provider. ⋆ indicates cases for which diagnosis was obtained via the biopsy report. * 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 main pancreatic duct or the common bile duct, allowing to confirm lesion presence. *BD-IPMN : Branch-duct intraductal papillary mucinous neoplasm, $\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, MD-IPMN : Main-duct intraductal papillary mucinous neoplasm, MCN : Mucinous cystadenoma, NA : Not available, PDAC : Pancreatic ductal adenocarcinoma, PNET : Pancreatic neuroendocrine tumor, SCN : Serous cystadenomas, SPN : Solid-pseudopapillary neoplasms.*

	$\mathcal{D}_{BR,1}$	$\mathcal{D}_{FR,1}$	$\mathcal{D}_{FR,2}$	$\mathcal{D}_{US,1}$	$\mathcal{D}_{US,2}$	$\mathcal{D}_{MSD}$
Visible duct						
Yes	576	227	125	554	328	208
No	428	17	23	145	74	73
Dilatation						
Yes	400 (69)	159 (70)	82 (66)	417 (75)	278 (85)	135 (65)
No	176 (31)	68 (30)	43 (34)	137 (25)	50 (15)	73 (35)
Stenosis						
Yes	394 (68)	176 (77)	92 (74)	434 (78)	277 (84)	144 (69)
No	182 (32)	51 (23)	33 (26)	120 (22)	51 (16)	64 (31)
Mass at the stenosis						
Yes	391 (68)	177 (78)	91 (73)	430 (78)	273 (83)	138 (66)
No	185 (32)	50 (22)	34 (27)	124 (22)	55 (17)	70 (34)

Table 4.5 – Clinical information of the main pancreatic ducts across diverse providers of the $\mathcal{D}_{CAPA}$ dataset. Data are number of patients with percentages in parentheses. Percentages are computed based on the number of visible main pancreatic ducts for each respective provider. *$\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset.*

Only annotator who reviewed more than 10 cases on this dataset were considered. The number of cases reviewed by each annotator is given in Table 4.6. We estimated that the average inter-annotator variability ranged from 0 to 0.12 DSC points, which is a reasonable variability threshold for pancreatic lesions whose contours can be hard to define.

Finally, we computed the difference between the mean DSC obtained between each annotator. While

Annotator	A2	A4	A5	A7	A8	A9
Number of cases	11	54	68	55	44	31

Table 4.6 – Number of cases reviewed by annotators on the $\mathcal{D}_{MSD}$ dataset. Results are presented only for annotators who reviewed more than 10 cases. *$\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset.*

this number is not an estimation of the inter-annotator variability on the same cases, it may be considered as a proxy measure at the dataset level. Results are illustrated in Figure 4.8 in which mean DSC differences between annotators remain mostly below 0.05, thus indicating consistency in their way to segment pancreatic lesions. While some deviation exists between the segmentations produced by the annotators, they seem to be around a threshold of variability of 0.1 points of DSC. Given the difficulty of segmenting pancreatic lesions whose margins can be poorly defined, this variability remains reasonable and should increase the model robustness rather than inducing a bias.

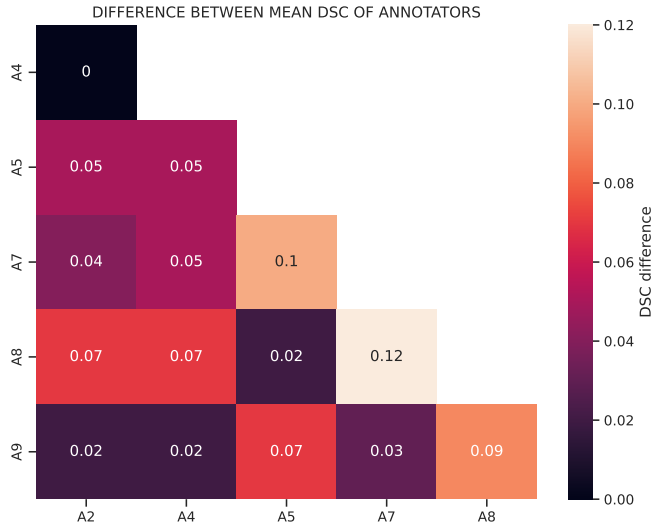


Figure 4.8 – Difference between the mean DSC of each annotator on the $\mathcal{D}_{MSD}$ dataset. For each annotator (A2, A4, A5, A7, A8, A9), the mean DSC has been computed for the cases they reviewed between their segmentation and the one provided by the Medical Segmentation Decathlon challenge organizers. *DSC : Dice similarity coefficient, $\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset.*

4.4 Discussion

Concerted efforts were undertaken to create a dataset characterized by a broad spectrum of institutions and pathologies, thereby ensuring a representative and diverse composition. However, this dataset has its limitations.

The primary limitation comes from the lack of small lesions. As illustrated in Figure 4.9, the median and interquartile range for the 2D maximum lesion diameter is $31[22-43]$ mm, with only 55 lesions falling below the subcentimetric threshold. This challenge is intricately linked to the delayed diagnosis of pancreatic cancers, wherein the infrequent detection of small lesions results in a scarcity of corresponding examinations in hospitals. Consequently, data providers face limitations in supplying such cases and tend to provide more cases of larger lesions.

A secondary limitation is attributed to the absence of a double-annotation process, a methodology that could have enabled us to estimate intra-annotator variability and identifying potential annotation errors. While we used the segmentations from the $\mathcal{D}_{MSD}$ dataset as a proxy to evaluate inter-annotator variability

at the dataset level, the inability to apply this process to all our database limits our capacity to evaluate the reliability and precision of annotations, both in terms of questions and segmentations.

Finally, the decision to employ a question-based annotation approach, while accelerating the annotation process, introduces a trade-off with interpretability. For instance, if an image is annotated as exhibiting atrophy, the precise location of the atrophy remains unmarked.

Despite these acknowledged limitations, this database stands as one of the most extensive annotated resources available for advancing AI research in the field of pancreatic cancer.

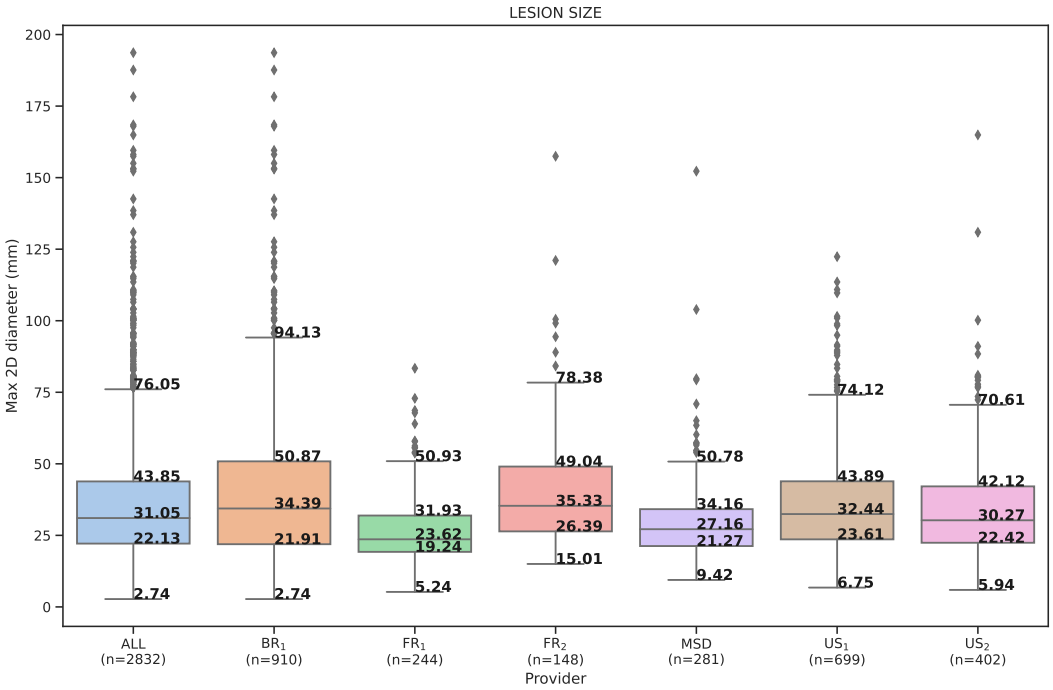


Figure 4.9 – Distribution of the maximum 2D lesion diameter in mm across diverse datasets. For each boxplot, low whisker, first quartile, median, third quartile and high whisker values are displayed. The lower whisker is at the lowest datum above $Q1 - 1.5 \times (Q3 - Q1)$, and the upper whisker at the highest datum below $Q3 + 1.5 \times (Q3 - Q1)$, where Q1 and Q3 are the first and third quartiles. Maximum lesion diameter was calculated using the maximum 2D Feret diameter (Van der Walt et al., 2014) across each axial slice.

Chapitre 5

Pancreas Segmentation

5.1 Preface

Organ segmentation is a crucial step for the development of automated decision support tools in medical analysis. Automatizing this time-consuming task holds major potential for clinical applications such as diagnosis, surgery and therapy (Yao et al., 2020b). The first objective of this thesis was to perform automatic pancreas segmentation from portal venous CT scans, which is the first line modality for abdominal imaging. To deal with the elongated shape and ambiguous contours of the pancreas that cause under-segmentation at the extremities, we propose a method that considers the pancreas specific morphology. This involves incorporating a geometrical prior into the segmentation task. This work was selected for an ***oral presentation at the IEEE International Symposium on Biomedical Imaging (ISBI) 2022*** *(Vétil et al., 2022b).*

Abstract. Deep neural networks are widely used for automated organ segmentation as they achieve promising results for clinical applications. Some organs are more challenging to delineate than others, for instance due to low contrast at their boundaries. In this chapter, we propose to improve the segmentation of elongated organs thanks to geometrical priors that can be introduced during training, using a local Tversky loss function, or at post-processing, using local thresholds. Both strategies do not introduce additional training parameters and can be easily applied to any existing network. The proposed method is evaluated on the challenging problem of pancreas segmentation. Results show that Geometrical Priors allow us to correct the systematic under-segmentation pattern of a state-of-the-art method, while preserving the overall segmentation quality.

5.2 Introduction

Automated organ segmentation is an important methodological step for the development of automated decision support tools in medical analysis. Yet, some organs are more difficult to segment than others because of ambiguous contours or elongated shapes (Kumar et al., 2019b).

Several methods have been proposed, and deep Convolutional Neural Network (CNN)s are increasingly recognized as the reference method as they achieve the best results in terms of Dice Similarity Coefficient (DSC) (Kumar et al., 2019b). Most of the recently proposed CNNs derive from the U-Net (Ronneberger et al., 2015), an encoder-decoder architecture which can operate on 2D and 3D images (see Section 3.2.3.5). Lately, Isensee et al. (2021a) introduced the nnU-Net, where the focus is put on the pre-processing, training strategy and post-processing steps rather than on the network architecture. It was ranked first in a series of biomedical segmentation challenges, among which the Medical Segmentation Decathlon challenge (detailed in Section 3.2.3.6).

While some automatic segmentation algorithms reach human-level performance (Bai et al., 2018), some organs, such as the pancreas or the colon, still constitute an open technical challenge for state-of-the-art models like the nnU-Net, as performance remains lower compared to those obtained on other abdominal organs (Yao et al., 2020b). These organs exhibit elongated shape with low-contrast at their boundaries, making the delineation of extremities difficult. Consequently, algorithms such as the nnU-Net may miss the extremities of the organ. From a clinical point of view, this error can be critical if the automatic segmentation task is part of a more complex tumor detection pipeline, since tumors at the extremities could be missed, as illustrated in Figure 5.1.A. A possible methodological approach to tackle this challenge is to perform multi-organ segmentation, delineating all the major structures close to the target one (Chu et al., 2019b; Wang et al., 2019). The costs required to build such datasets motivated the study of semi-supervised approaches (Zhou et al., 2019c). In particular, Zhou et al. (2019b) performed the segmentation of 13 abdominal structures using a loss function which embedded anatomical priors computed on unlabeled data. More precisely, the outputs of the network were enforced to match a distribution of organ sizes, learned on a small fully labeled dataset. Consequently, the quality of the prior was highly dependent on the size of the fully labeled dataset, and the prior did not take the geometry of the organs into account.

In this work, we seek to improve the automatic segmentation of elongated organs. To this end, we investigate geometry-based anatomical priors whose definitions would not depend on labeled data. We illustrate our contributions on the pancreas segmentation problem, which is a key methodological objective of this thesis as defined in Section 1.3. Besides having an elongated shape, the pancreas shows a strong ambiguity at its extremities which are highly intertwined with the duodenum, the small intestine and the spleen (see examples in Figure 5.1). As a result, automatic segmentation methods tend to underestimate the pancreas by missing its extremities. To fix these systematic under-segmentation patterns, we propose two strategies that control the sensitivity of the predictions using geometrical priors which can be introduced either during training, using a local Tversky loss function (Salehi et al., 2017), or during post-processing, using local thresholds. In order to compare with state-of-the-art methods, we evaluate our methods on the public datasets $\mathcal{D}_{NIH}$ and $\mathcal{D}_{MSD}$ (see Chapter 4) and compare it with the nnU-Net.

5.3 Method

Consider an elongated shape organ O_{el}, whose extremities e_0, e_1 are difficult to delineate. Let $\mathcal{R} = \{r_i \in \{0,1\}, i = 1...N\}$ denote its segmentation of reference, with N being the number of voxels. Let $\mathcal{P} = \{p_i \in [0,1], i = 1...N\}$ denote the probability output of a segmentation network. We define the Geometrical Prior $\mathcal{G}_{prior}$ that gives for each voxel i the desired sensitivity boost ρ_i, *i.e.*, $\mathcal{G}_{prior} = \{\rho_i \in]0,1[, i = 1...N\}$. Taking the illustrative example of pancreas segmentation, the extremities e_0 and e_1 will be referred to as the head (H) and tail (T) in the rest of this section.

This section introduces the Geometrical Priors and how to use them in practice.

Figure 5.1 – Challenging segmentation cases at the extremities of the pancreas. A : Lesion at the extremity of the tail. Portal CT scan of a 78-year-old man reveals a small iso-enhancing mass at the very end of the pancreatic tail (white arrow). Failure to analyze the pancreas in its entirety may result in overlooking the lesion. **B : Ambiguous contours with the duodenum.** Portal CT scan of a patient with hypo-enhancing lesion in the pancreatic head, displaying ambiguous contours (white arrow). **C : Ambiguous contours with the spleen.** Portal CT scan of a patient with a hypo-enhancing mass in the head (yellow arrow), where the tail exhibits similar density to the spleen (white arrow). On each image, the pancreas is outlined in red. *CT : Computed tomography.*

5.3.1 Geometrical Priors

Geometrical Priors assign a sensitivity boost to each voxel of a segmentation mask. As illustrated in Figure 5.2, the construction of one Geometrical Prior $\mathcal{G}_{prior}$ follows three steps :

1. First, a Euclidean distance map assigns to each voxel i its distance to the head, noted d_i. The distance map is normalized so that distance values are 0 and 1 at the head and the tail, respectively. Distance values for voxels outside of the pancreas, i.e. $r_i = 0$, are set to -1.
2. Secondly, a prior function f_{prior} assigns to each voxel i its sensitivity boost ρ_i according to its distance to the head d_i. The function f_{prior} is designed to give regions prone to under-segmentation a higher sensitivity boost, so as to favor their segmentation. For the pancreas, f_{prior} is built as a piece-wise linear function that emphasizes the extremities :

$$\rho_i = f_{prior}(d_i) = \begin{cases} \rho_H + \frac{\rho_* - \rho_H}{0.5} \cdot d_i & \text{if } 0 \leq d_i \leq 0.5 \\ \rho_* + \frac{\rho_T - \rho_*}{0.5} \cdot (d_i - 0.5) & \text{if } 0.5 < d_i \leq 1 \\ \rho_* & \text{if } d_i = -1 \end{cases} \tag{5.1}$$

   where d_i is the normalized distance to the head of voxel i, ρ_H, ρ_T, ρ_* are three hyper-parameters corresponding to the sensitivity boost in the head, the tail and the background, respectively. This function is designed to decrease linearly from ρ_H to ρ_* on the interval $[0, 0.5]$, to increase linearly from ρ_* to ρ_T on the interval $[0.5, 1]$, and remain constant at ρ_* otherwise. In particular, this function takes the value of ρ_H at $d_i = 0$ (i.e., at the head) and the value of ρ_T at $d_i = 1$ (i.e., at the tail). In practice, ρ_H, ρ_T, ρ_* are set so that $\rho_i \in]0, 1[$.
3. Finally, the resulting Geometrical Prior $\mathcal{G}_{prior} = \{\rho_i \in]0, 1[, i = 1...N\}$ can be used to define either a local Tversky loss function or local thresholds. This will be detailed in the subsequent sections.

5.3.2 Local Tversky loss function

A first way to use Geometrical Priors is during training, through a local version of the Tversky loss function. The Tversky loss function was introduced by Salehi et al. (2017) based on Tversky's index S, which measures the similarity between two sets $\mathcal{P}, \mathcal{R}$ in an asymmetric way (Tversky, 1977) :

$$S(\mathcal{P}, \mathcal{R}, \alpha, \beta) = \frac{|\mathcal{P} \cap \mathcal{R}|}{|\mathcal{P} \cap \mathcal{R}| + \alpha|\mathcal{P} \setminus \mathcal{R}| + \beta|\mathcal{R} \setminus \mathcal{P}|} \tag{5.2}$$

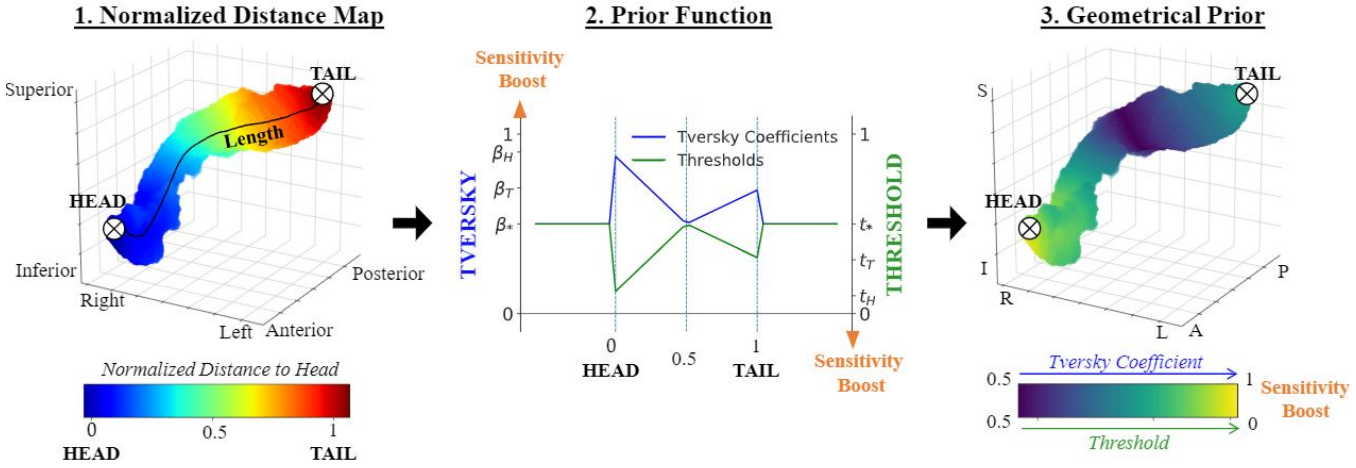


Figure 5.2 – Construction of the Geometrical Prior. Starting from a binary segmentation, a normalized distance map is computed. Then, a prior function transforms the distance map into a Geometrical Prior by assigning to each voxel a sensitivity boost. Finally, the resulting Geometrical Prior can be translated either into local Tversky coefficients or into local threshold values.

where $\alpha, \beta \geq 0$ are parameters, and $\mathcal{P} \setminus \mathcal{R}$ denotes the relative complement of $\mathcal{R}$ in $\mathcal{P}$. When applied to Boolean data, using the definition of True Positive (TP), False Positive (FP) and False Negative (FN), the Tversky index can be written as :

$$S(\mathcal{P}, \mathcal{R}, \alpha, \beta) = \frac{TP}{TP + \alpha FP + \beta FN} \tag{5.3}$$

Therefore, α and β influence the specificity and the sensibility as they control the FPs and FNs, respectively. Note that in the case of $\alpha = \beta = 0.5$, the Tversky index is equivalent to the DSC (defined in Equation 3.1). Based on this index, Salehi et al. (2017) defined the corresponding loss function :

$$T(\mathcal{P}, \mathcal{R}, \alpha, \beta) = \frac{\sum_{i=1}^{N} p_i^0 r_i^0}{\sum_{i=1}^{N} p_i^0 r_i^0 + \alpha \sum_{i=1}^{N} p_i^0 r_i^1 + \beta \sum_{i=1}^{N} p_i^1 r_i^0} \tag{5.4}$$

where p_i^0 (respectively, p_i^1) is the predicted probability of voxel i belonging to O_{el} (respectively, background). Same notations go for $r_i \in \mathcal{R}$. We propose a local version of this loss to leverage the Geometrical Priors during training :

$$T(\mathcal{P}, \mathcal{R}, \mathcal{G}_{prior}^{Tversky}) = \frac{\sum_{i=1}^{N} p_i^0 r_i^0}{\sum_{i=1}^{N} p_i^0 r_i^0 + \sum_{i=1}^{N} \alpha_i p_i^0 r_i^1 + \sum_{i=1}^{N} \beta_i p_i^1 r_i^0} \tag{5.5}$$

In contrast to Equation 5.4 that relies on global hyper-parameters (α, β), our methods uses local coefficients (α_i, β_i) that control the trade-off between the specificity and the sensitivity for voxel i. In particular, higher values of β_i penalize more under-segmentation mistakes. In practice, $\alpha_i = 1 - \beta_i$, and we use the Geometrical Priors to define the local coefficients : $\mathcal{G}_{prior}^{Tversky} = \{\beta_i = \rho_i \in]0, 1[, i = 1...N\}$. The $\mathcal{G}_{prior}^{Tversky}$ values are computed for each patient once, before training, using their reference segmentation mask.

5.3.3 Local thresholds

The second way to use Geometrical Priors is in post-processing, using them to define local thresholds t_i which are applied on the probability output $\mathcal{P}$, *i.e.*, voxel i is classified as belonging to the pancreas if $p_i \geq t_i$. In particular, areas with high sensitivity boost will exhibit lower thresholds in order to facilitate the

segmentation. Therefore, the local thresholds t_i are defined as $\mathcal{G}_{prior}^{threshold} = \{t_i = 1 - \rho_i \in]0, 1[\,, i = 1...N\}$. This time, the distance map $\{d_i \in \{-1\} \cup]0, 1[\,, i = 1...N\}$ is computed from the predicted segmentation obtained with a global threshold $t = 0.5$ where the head and tail are differentiated using the orientation of the images during acquisition. The final result is denoted by $\mathcal{P}_{bin}$.

Both proposed methods were based on the same rationale : lower t_i, just as larger β_i, would boost the sensitivity in desired areas, while keeping other areas unaffected. For the sake of simplicity, the sensitivity boost ρ will be used to refer to local Tversky coefficients, as $\rho_i = \beta_i$, or to local thresholds, as $\rho_i = 1 - t_i$, depending on the context.

5.4 Experiments

5.4.1 Dataset

Two public datasets were used :

(i) $\mathcal{D}_{NIH}$, containing 80 healthy subjects,

(ii) $\mathcal{D}_{MSD}$, containing 281 pathological subjects.

These datasets, detailed in Section 4.2.1, are the two prominent public datasets commonly used as benchmarks for pancreas segmentation tasks. To harmonize the datasets, $\mathcal{D}_{MSD}$ tumor masks were merged into the pancreas. The $\mathcal{D}_{NIH}$ scans, with higher z-resolution than the $\mathcal{D}_{MSD}$ scans, were resampled to halve their z-resolution. After manual review by a radiologist with 25 years of expertise in abdominal imaging, 20 cases were put aside because their reference segmentation exhibited under-segmentation of the tail or the head. The remaining 341 cases were divided into training, validation, and test sets, comprising 223, 55, and 63 cases, respectively. Within each set, 75% of the cases were sourced from the $\mathcal{D}_{MSD}$ dataset, while the remaining 25% were obtained from the $\mathcal{D}_{NIH}$ dataset.

5.4.2 Baseline

The pre-processing, network architecture, optimization and post-processing hyperparameters were automatically selected for our dataset using the nnU-Net self-configuring procedure detailed in Section 3.2.3.5. All default parameters from public implementation[1] were kept without modification in a baseline experiment.

5.4.3 Settings

To evaluate the effect of our Geometrical Prior methods, experiments were run with different combinations of sensitivity boosts (ρ_H, ρ_T, ρ_*), where each combination represented a specific prior function. Values for ρ_x varied in $[0.5, 0.7, 0.95]$, referred to as $[\rho_x^5, \rho_x^7, \rho_x^{95}]$. Values smaller than 0.5 were not explored as they would have emphasized specificity, which was unnecessary since over-segmentation was not a common occurrence.

In addition, the proposed methods were compared against the original Tversky loss function (Salehi et al., 2017), with global coefficients set to 0.7 and 0.95, chosen empirically.

5.4.4 Performance measures

The quality of the segmentation at the extremities was quantitatively evaluated with two measures. The first is the error on the length of the organ, expressed in *mm*. To compute the length, the segmentation mask was first converted into a voxel adjacency graph. In particular, n_T was the node corresponding to the tail, estimated as the closest point to the Left-Posterior-Superior corner in the abdomen, as shown in Figure 5.2. Finally, the length was estimated as the eccentricity of n_T, *i.e.*, the maximum graph distance from n_T to other nodes. This also provided the coordinates of the other extremity, the head. The signed

1. github.com/MIC-DKFZ/nnUNet

length error was then defined as *Err Length* = *Length*($\mathcal{P}_{bin}$) - *Length*($\mathcal{R}$), where $\mathcal{P}_{bin}$ and $\mathcal{R}$ refer to the predicted and reference segmentation masks, respectively. As depicted in Figure 5.3, focusing on the pancreas where extremities tend to be missed, we could estimate that negative length errors referred to under-segmentation whereas positive errors to over-segmentation.

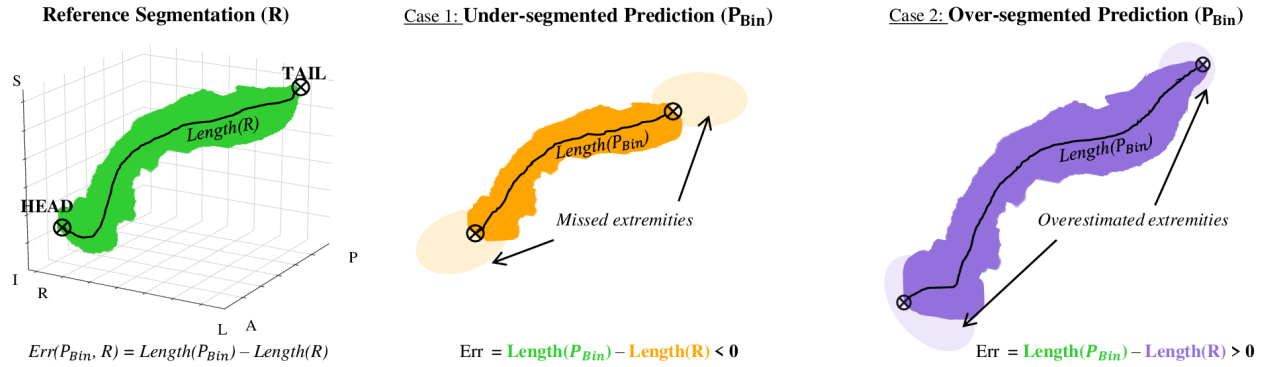


Figure 5.3 – Assessing the quality of the segmentation using the signed error on the length of the pancreas. The signed length error is defined as *Err* = *Length*($\mathcal{P}_{bin}$) - *Length*($\mathcal{R}$), where $\mathcal{P}_{bin}$ and $\mathcal{R}$ refer to the predicted and reference segmentation masks, respectively. Given that extremities are prone to being overlooked, negative errors indicate under-segmentation, while positive errors indicate over-segmentation.

The second measure is the DSC at the extremities, defined as the mean of the DSCs computed on 5% of the length of the pancreas, from each extremity. Finally, in order to ensure that the overall segmentation quality was not deteriorated, the difference of the global DSC with respect to the baseline is also reported.

5.5 Results

In total, 10 experiments, in addition to the baseline, were conducted. In a first set of experiments A-E, the baseline loss function was replaced by a local Tversky loss function. In a second set of experiments A'-E', the baseline post-processing procedure was changed for a local thresholding approach, while the baseline loss function was kept. Details and qualitative results are reported in Table 5.1.

			GLOBAL		LOCAL		
	PRIOR		ρ^{95}	ρ^{7}	$\rho^{95}_H, \rho^{95}_T, \rho^{5}_*$	$\rho^{95}_H, \rho^{7}_T, \rho^{5}_*$	$\rho^{95}_H, \rho^{7}_T, \rho^{7}_*$
TVERSKY	**Experiment**	Baseline	A	B	C	D	E
	Length Error (mm)	-3.2 (6.4)	1.2 (7.8)	-2.0 (6.5)	2.3 (7.6)	**-0.4 (8.2)**	0.2 (6.8)
	DSC at Extremities (%)	70.0 (3.4)	67.5 (2.1)	71.0 (3.4)	68.0 (1.7)	**72.5 (2.8)***	70.5 (3.2)
	δDSC (%) w.r.t. Baseline	-	-4*	0.2	-1.3*	0.1	-2*
THRESHOLD	**Experiment**	Baseline	A'	B'	C'	D'	E'
	Length Error (mm)	-3.2 (6.4)	-0.3 (6.3)	- 1.8 (7.2)	-1.4 (7.6)	-1.6 (7.3)	**-1.0 (7.7)**
	DSC at Extremities (%)	70.0 (3.4)	72.0 (2.3)*	**75.0 (2.95)***	74.0 (2.5)*	74.0 (2.7)*	74.0 (2.6)*
	δDSC (%) w.r.t. Baseline	-	-2*	0.3	0.1	-0.1	0.2

Table 5.1 – Evaluation measures obtained on the test set (65 cases). As the data are not normally distributed, we report the median instead of the mean. The interquartile range is indicated in brackets. Median Dice Similarity Coefficient (DSC) for the Baseline is 88.0 (5.0). Significant differences with baseline are indicated by *, except for the length error for which all the results are significant. Statistical significance is obtained with paired t-tests at $p = 0.05$ level. For each evaluation measure, the best result, in **bold**, is selected among the experiments that do not deteriorate significantly the DSC w.r.t. the baseline.

5.5.1 Baseline results

Baseline results were characterized by a great under-segmentation as they exhibited a negative median length error of −3.2 mm. All the experiments A-E and A'-E' managed to resolve this under-segmentation issue, as the median error was significantly shifted towards positive values.

5.5.2 Local Tversky

Experiments A-E, using global or local Tversky loss function, produced significant effects on the length error. Yet, (A) showed that global β^{95} decreased the global DSC, meaning that the overall segmentation was deteriorated. In addition, (A) also reported a drop of the DSC at the extremities. On the other hand, global β^{7}, as in (B), did not affect the quality of the segmentation but produced only slight improvement on the length error. Both (A, B) illustrated the power but also the limitations of the global Tversky loss function, that would prevent its use with strong coefficients. These drawback effects were mitigated by Geometrical Priors : C-E did not result in a significant drop in the global DSC, while they strongly corrected the length error. This mitigating capability was further illustrated in the upper row of Figure 5.4 : at the extremities, (D) (yellow) behaved as (A), trained with a strong global Tversky loss coefficient (blue), and outperformed the baseline (red). By contrast, near the body of the pancreas, where the sensitivity boost was lower, (D) mimicked the baseline and avoided the substantial over-segmentation mistakes caused by (A). In addition to reducing the length error, (D) also significantly increased the DSC at the extremities, suggesting the extremities were not only better detected, but also better delineated. Additional qualitative results are shown in Figure 5.5.

5.5.3 Local thresholds

Experiments A'-E', using local thresholds, yielded slighter improvements on the length error, with the best improvement in (A') coming at the expense of the global DSC. Thus, best results were achieved by the experiment (E'). We hypothesize that the difference in results between the two methods was due to the distribution of the probabilities outputted by the network, where values were strongly pushed towards 0 or 1. Thus, regions missed by the baseline model tended to exhibit probabilities less than 0.05, and were therefore difficult to be recovered by the post-processing method. In this regard, the local threshold technique was less efficient than the local Tversky loss function, which corrected the under-segmentation from the training. Yet, the local thresholding approach produced systematic and significant increase of the DSC at the extremities, as well as qualitative improvements illustrated in Figures 5.4 and 5.5. In particular, Figures 5.4.1', 5.4.3' and 5.5.1' show total failures of the baseline that were successfully recovered by (E').

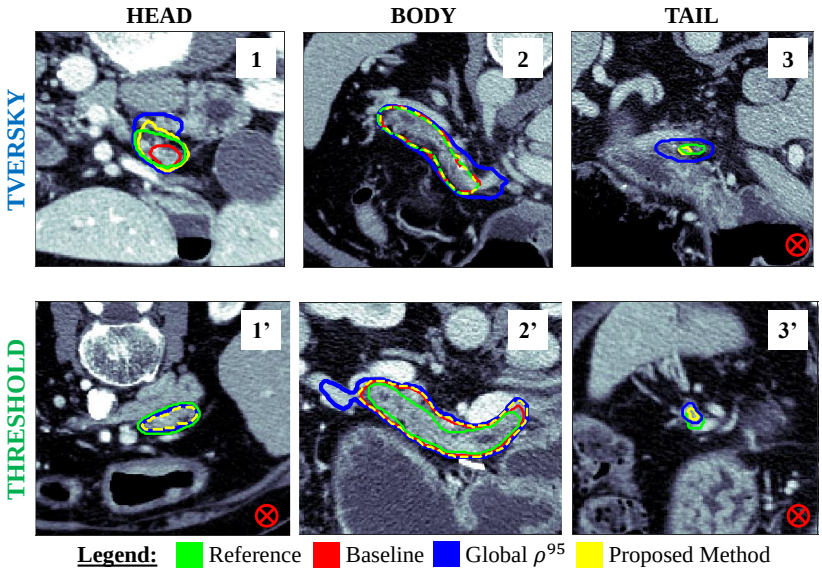


Figure 5.4 – Segmentation results obtained at the head, body and tail of the pancreas. (1/2) The first row shows local Tversky loss function experiments, where blue and yellow correspond to experiment A and D, respectively. The second row shows local thresholds experiments, where blue and yellow correspond to experiment A' and E', respectively. D and E' were selected as our *Proposed Method*, according to Table 5.1. ⊗ indicates void prediction for the baseline.

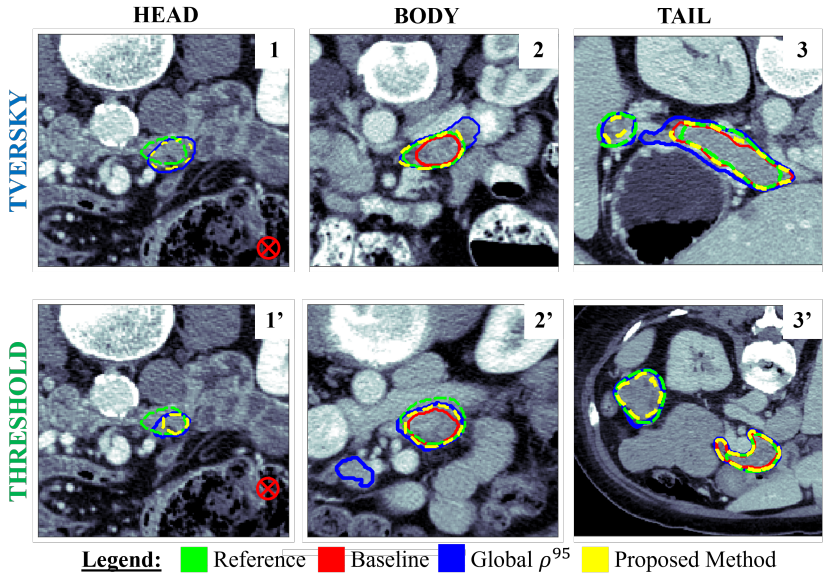


Figure 5.5 – Segmentation results obtained at the head, body and tail of the pancreas. (2/2) The first row shows local Tversky loss function experiments, where blue and yellow correspond to experiment A and D, respectively. The second row shows local thresholds experiments, where blue and yellow correspond to experiment A' and E', respectively. D and E' were selected as our *Proposed Method*, according to Table 5.1. ⊗ indicates void prediction for the baseline.

5.6 Discussion and conclusion

We proposed two methods to boost the segmentation sensitivity in extremities of elongated organs, which state-of-the-art algorithms tend to miss. Our methods rely on the computation of Geometrical Priors that assign to each voxel a sensitivity boost. This boost can be used during training, as a local Tversky coefficient, or at post-processing, as a local threshold. Quantitative results on the pancreas segmentation problem demonstrated that both proposed techniques managed to significantly increase the segmentation sensitivity at the extremities of the organ. This was supported by the improvement of both the length error and the DSC at the extremities, which were achieved without deteriorating the overall segmentation.

The strengths of our work are threefold : first, the proposed methods are interpretable thanks to the prior function that reflects the anatomical complexity of the organ. Specifically, the prior function that defines the Geometrical priors is manually crafted based on medical expertise, explicitly emphasizing the regions that are difficult to segment. Secondly, the general formulation of Geometrical Priors allows our methods to be applied to any organ with regions that are difficult to segment. Last, our methods can be easily applied to any existing network. In particular, the local thresholds strategy does not require re-training the model.

The main limitation of our work is the introduction of hyper-parameters which, although guided by anatomical knowledge, can make the search time-consuming. Thus, future work may explore the use of learnable prior functions, whose parameters would be dynamically learned during training or post-processing.

Chapitre 6

Lesion and Main Pancreatic Duct Dilatation Detection

6.1 Preface

Once built an automatic pancreas segmentation tool (see Chapter 5), we focused on radiological findings that may help radiologists identify the development of pancreatic cancer. We started by tackling the detection of pancreatic lesions and the dilatation of the main pancreatic duct, both essential indicators for pancreatic cancer. For this purpose, we constructed a predictive model leveraging a segmentation algorithm, which was developed and tested using a combination of data from public and private databases. A first version of this work was selected for an oral presentation at the ***70èmes Journées Francophones de Radiologie 2022*** *(Vétil et al., 2022), and was further extended with a* ***publication in Investigative Radiology 2023*** *(Abi Nader et al., 2023).*

Abstract. Pancreatic cancer is characterized by a low five-year survival rate of 9%, mainly due to delayed diagnosis. While early detection promises improved survival rates, it faces numerous challenges. Primarily, the initial stages of pancreatic cancer are often asymptomatic. Consequently, the incidental detection of these lesions on portal Computed Tomography (CT) scans, the primary imaging modality for abdominal examination, is crucial. However, detecting pancreatic lesions on portal CT scans presents challenges due to their often subtle or iso-enhancing appearances. Therefore, the detection of pancreatic cancer significantly benefits from the identification of secondary signs, detailed in Section 2.3.5, which can be visible before the clear appearance of the tumor. To address this, we propose a deep learning method for detecting pancreatic lesions and identifying Main Pancreatic Duct (MPD) dilatation in portal venous CT scans. The proposed method is built in three steps. First, a network segmenting the pancreas, pancreatic lesions and the MPD is trained in a five-fold cross-validation manner. Secondly, outputs of this network are 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. Thirdly, two logistic regression models are calibrated to predict lesion presence and MPD dilatation, respectively. The proposed method is trained on 2134 portal CT scans from five institutions, and its performance is evaluated on an external test cohort of 756 portal CT scans from four independent institutions. The Area Under the Curve (AUC) of the model for detecting lesion presence in a patient is 0.98 (95% Confidence Interval (CI) : [0.97,0.99]), with a reported sensitivity of 0.94 (469 of 493, 95% CI : [0.92,0.97]). Consistent performance is observed across various lesion types and characteristics. Regarding MPD dilatation detection, the model achieves an AUC of 0.97 (95% CI : [0.96,0.98]).

6.2 Introduction

As presented in Chapter 2, pancreatic cancer is a major healthcare issue on the rise (Rawla et al., 2019). Due to the increase of pancreatic cancer incidence, combined with its dramatically low five-year survival rate of 9%, the disease could become the third leading cause of cancer-related deaths by 2025 (Sung et al., 2021). The low five-year survival rate primarily stems from late-stage diagnoses. As explained in Section 2.3.3, this delay results from the lack of specific symptoms and challenges in interpreting portal venous Computed Tomography (CT) scans, which is the primary imaging modality for routine abdominal examinations. In addition, radiologists' heavy workload as well as their level of expertise and experience might further affect the interpretation of a CT scan. As a consequence, pancreatic cancer is frequently identified at advanced stages, limiting therapeutic options. Conversely, patients diagnosed with stage IA pancreatic cancer (see Section 2.3.4) are more often eligible for pancreatic resection, leading to a five-year survival rate surpassing 80% (see Figure 2.11). This underscores the critical importance of early pancreatic cancer detection.

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 (Gangi et al., 2004; Gonoi et al., 2017). As detailed in Section 2.3.5, subtle secondary signs, such as Main Pancreatic Duct (MPD) dilatation, were often visible up to one year before the cancer diagnosis. This is due to the fact that pancreatic cancer, primarily Pancreatic Ductal Adenocarcinoma (PDAC), often leads to a constriction of the MPD, resulting in dilation upstream of the lesion. Dilatation is usually defined as a duct larger than 3mm in the head and 2mm in the body and tail of the pancreas.

Deep Learning (DL) methods hold significant potential in supporting the diagnosis of pancreatic cancer by issuing alerts for patients at risk of pancreatic cancer, which could facilitate radiologists' daily practice. Our literature review identified two main categories of works : those approaching the task clinically (summarized in Table 3.7) and those focusing on methodology (summarized in Table 3.3). While clinically-oriented studies exhibited promising results, they primarily worked on lesion detection and did not leverage secondary findings, such as MPD dilatation. Moreover, despite validation experiments being performed on substantial external cohorts, these cohorts often originated from the same country as the training data, and the training data itself was frequently obtained from a single center. These factors indicate a potential lack of variability in terms of patients demographics, clinical conditions and scanners manufacturers, which raises concerns about the generalizability of these methods when applied to scanners from new medical centers. Conversely, some methodological approaches leveraged anatomical information from the MPD. For instance, Zhu et al. (2021) and Alves et al. (2022) jointly segmented lesions and the MPD, while Viviers et al. (2022) used MPD segmentation as an additional input to the portal CT scan. However, these methods used information from the MPD solely to improve lesion detection, without explicitly targeting the detection of MPD dilatation, which is an essential indicator for radiologists.

Given this context, this work presents and evaluates a DL pipeline designed to predict patients at risk of pancreatic cancer. The contribution is two-fold :

(i) A lesion segmentation algorithm is combined with explicit 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.

The entire pipeline is trained on a large multi-centric dataset, incorporating data from diverse geographic regions. Similar to the clinical works detailed in Section 3.3.4, we validate the proposed method on a large and external cohort.

6.3 Data and annotations

6.3.1 Data

Information was gathered from the nine providers outlined in Chapter 4, namely $\mathcal{D}_{BR,1}$, $\mathcal{D}_{FR,1}$, $\mathcal{D}_{FR,2}$, $\mathcal{D}_{US,1}$, $\mathcal{D}_{US,2}$, $\mathcal{D}_{MSD}$, $\mathcal{D}_{NIH}$, $\mathcal{D}_{CALV}$, and $\mathcal{D}_{CABO}$. These contributors span across Europe, the USA, and Brazil, providing a diverse geographical coverage for the dataset. Inclusion criteria were as follows : *(i)* presence of a portal venous CT scan ; *(ii)* maximum slice thickness of 3mm ; *(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 training and independent testing sets of 2134 and 756 subjects, respectively (see Figure 6.1).

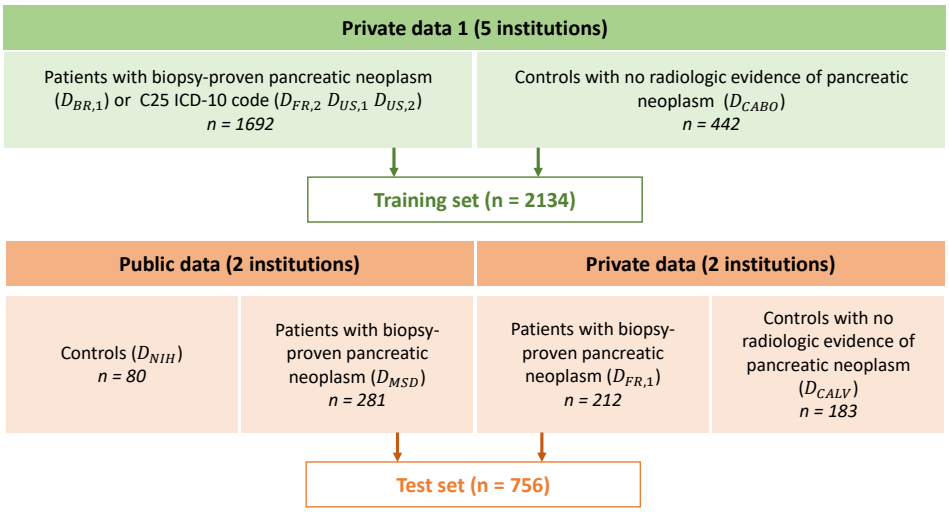


Figure 6.1 – Composition of the training and testing sets. Data in the training and testing sets come from different institutions. *ICD : International Classification of Diseases.*

The training set was composed of portal venous CT scans of 2134 patients from five providers ($\mathcal{D}_{BR,1}$, $\mathcal{D}_{FR,2}$, $\mathcal{D}_{US,1}$, $\mathcal{D}_{US,2}$, $\mathcal{D}_{CABO}$, see description in Section 4.2.2), among which 1692 had a pancreatic neoplasm. The diagnosis was either obtained via the biopsy report for 78% of the subjects (coming from $\mathcal{D}_{BR,1}$), or through the C25 code related to pancreatic neoplasms in the International Classification of Diseases (ICD-10) (DiSantostefano, 2009) for the rest of the subjects. In addition, control portal CT scans were extracted from $\mathcal{D}_{CABO}$, 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 control cases. Table 6.1 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 PDAC, 134 had Pancreatic Neuroendocrine Tumor (PNET) and 158 had unclassified solid lesion. There were also 81 subjects with Intraductal Papillary Mucinous Neoplasm (IPMN), 34 with mucinous cystic neoplasm and 42 with serous cystadenoma, 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 four institutions (see Figure 6.1). 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 ($\mathcal{D}_{MSD}$, see Section 4.2.1.2) and 80 were healthy cases ($\mathcal{D}_{NIH}$, see Section 4.2.1.1). The private dataset was composed of 212 portal venous CT scans of patients from $\mathcal{D}_{FR,1}$ (see dataset description in Section 4.2.2.1) with histopathological confirmation of pancreatic neoplasm. In addition, routine portal CT scans from $\mathcal{D}_{CALV}$ (see dataset description in Section 4.2.2.2) 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), PNET (n=48) and unclassified solid lesions (n=12), as well as IPMN (n=18) and unclassified cystic lesions (n=53) (cf. Table 6.1). Finally, 276 subjects had dilated MPD.

	Training set	**Test set**
Number of cases	2134	756
Age* (year)	64 [56, 73]	NA
Sex		NA
M	960 (45)	
F	1174 (55)	
CT manufacturer		
Philips	1113 (52)	
Siemens	168 (8)	
GE	309 (14)	NA
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)
PNET	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]
MPD		
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 6.1 – Demographic and clinical information for the different datasets. If not specified data are numbers of patients with percentages in parentheses. * Data are medians with interquartile range in square brackets. *GE : General Electric, IPMN : Intraductal papillary mucinous neoplasm, MCN : Mucinous cystic neoplasm, NA : Not available, PNET : Pancreatic neuroendocrine tumor, PDAC : Pancreatic ductal adenocarcinoma, SCN : Serous cystadenoma.*

6.3.2 Annotations

6.3.2.1 Protocol

Each portal venous CT scan was reviewed and annotated according to the protocol described in Section 4.3.1. The results of the annotation process for this dataset can be consulted in Table 6.1. In particular, it shows that 80% of the training and 65% of the testing cases exhibited pancreatic lesions, with the majority (over 85% in both splits) being solid. Notably, solid lesions were mostly PDACs (80% in train, 85% in test), whereas cystic lesions were predominantly IPMNs (38% in train, 25% in test). Lesions tended to be smaller in the test set, with a median and interquartile range of 2.4 $[2.0, 3.2]$ cm in the test set compared to 3.4 $[2.3, 4.9]$ cm in the training split. Additionally, the MPD was visible in 59% and 54% of the training and test cases, respectively. In the training set, MPD dilatation was more prevalent albeit less pronounced : 43% of the MPDs exhibited dilation, with a median and interquartile diameter of 4.5 $[3.5, 6.5]$ mm, compared to 36% in the test set with a median and interquartile diameter of 6.0 $[4.1, 8.8]$ mm. Segmentation examples can be visualized on Figure 6.2.

6.3.2.2 Inter-annotator variability

Lesion segmentation variability was investigated in Section 4.3.2. This analysis estimated that the average inter-annotator variability ranged from 0 to 0.12 Dice Similarity Coefficient (DSC) points, which

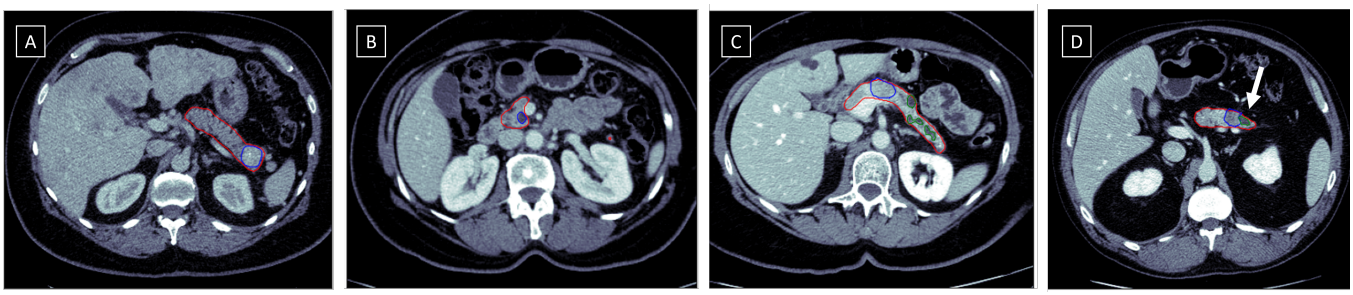


Figure 6.2 – Segmentation examples on portal CT scans. For clarity, segmentation masks are represented as contours. The pancreas, lesion, and MPD are delineated in red, blue, and green, respectively, when visible. **A :** Patient with a hyper-enhancing PNET in the tail of the pancreas. **B :** 59-year-old woman with a benign hypo-enhancing IPMN. **C :** Patient with a hypo-enhancing PDAC following the dilatation of the MPD. **D :** Patient with a hypo-enhancing PDAC characterized by MPD cutoff (white arrow). Images A and B are sourced from the $\mathcal{D}_{CAPA}$ dataset and are used from training, while images C and D are sourced from the $\mathcal{D}_{MSD}$ dataset and are used for testing. Details about the $\mathcal{D}_{CAPA}$ and $\mathcal{D}_{MSD}$ datasets are provided in Chapter 4. *$\mathcal{D}_{CAPA}$: Care Advisor for Pancreas dataset, CT : Computed tomography, IPMN : Intraductal papillary mucinous neoplasm, $\mathcal{D}_{MSD}$: Medical Segmentation Decathlon dataset, MPD : Main pancreatic duct, PDAC : Pancreatic ductal adenocarcinoma, PNET : Pancreatic neuroendocrine tumor.*

is a reasonable variability threshold for pancreatic lesions whose contours can be hard to define.

6.4 Detection method

6.4.1 Segmentation

A segmentation model is trained to segment the pancreas, the pancreatic lesion (if any) and the MPD (if visible). A 3D nnU-Net (Isensee et al., 2021a) is selected for its established robustness, as discussed in Section 3.2.3.5. This model is trained in a five-fold cross-validation setting.

6.4.1.1 Image preprocessing

Image preprocessing is done according to the nnU-Net automatic preprocessing pipeline, which involves intensity normalization and resampling.

Intensity normalization Based on the 2134 training scans, the mean μ_{HU} and standard deviation σ_{HU} intensities are computed : $\mu_{HU} = 68.7$ Hounsfield Unit (HU) and $\sigma_{HU} = 51.0$ HU, respectively. They are subsequently used to perform a z-normalization on each training and testing case, based on the following formula :

$$i' = \frac{i - \mu_{HU}}{\sigma_{HU}}$$

where i', i are the new and original intensity values of a given voxel, respectively.

Resampling Based on the 2134 training scans, the median voxel size for each spatial dimension is computed. In this case, a median voxel size of $(0.79 \times 0.79 \times 2)$ mm^3 is obtained along the x, y and z axes, respectively. During training, each input image is resampled to match this target spacing using a 3rd order B-spline interpolation. This preprocessing procedure has demonstrated effectiveness across various biomedical image segmentation tasks, including the segmentation of small lesions (Isensee et al., 2021a).

6.4.1.2 Architecture

The network architecture follows a standard 3D U-Net with a topology automatically inferred through a set of heuristics defined within the nnU-Net framework. The precise network topology is shown in Figure 6.3.

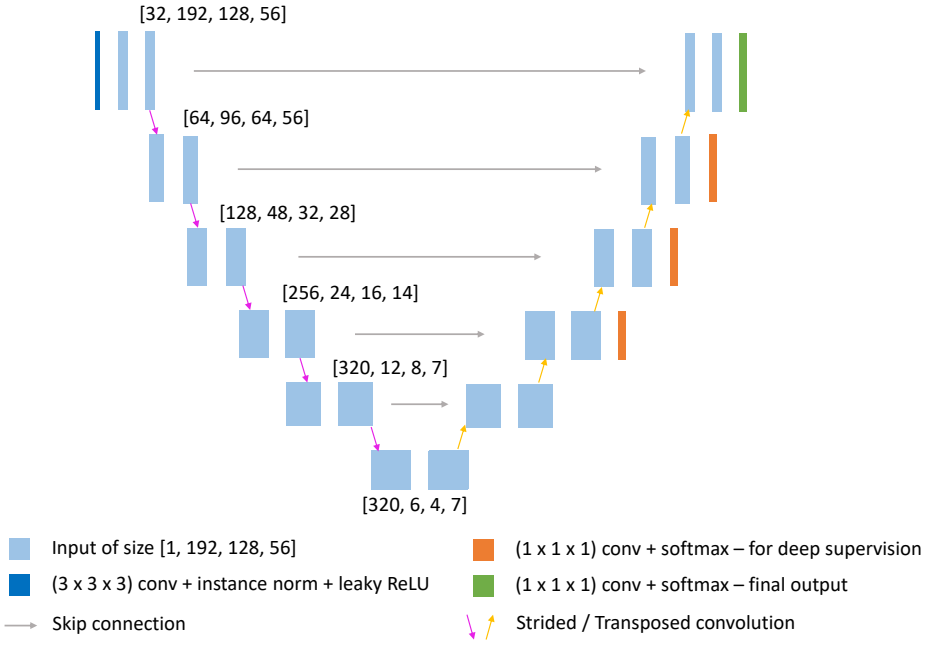


Figure 6.3 – Architecture of the U-Net segmentation network used to segment the pancreas, lesions, and the main pancreatic duct. The U-Net architecture uses two blocks per resolution step in both the encoder and decoder modules, with each block comprising a convolutional layer followed by instance normalization and a leaky ReLU activation function. Instance normalization is preferred over batch normalization due to its superior performance with small batch sizes (Kickingereder et al., 2019). Downsampling is achieved through strided convolutions, while upsampling is carried out using transposed convolutions. To balance performance and memory consumption, the initial number of feature maps is set to 32 and doubled (or halved) with each downsampling (or upsampling) operation. The network adopts deep supervision by incorporating additional auxiliary loss functions in the decoder, applied to all resolutions except the two lowest ones. This strategy enables gradients to penetrate deeper into the network, facilitating the training of all layers effectively.

6.4.1.3 Training parameters

Once image preprocessing is performed, the nnU-Net looks for the best compromise to maximize both the batch size and patch size given the available GPU budget (a Tesla V100 16 GB). In this case, the obtained values are a batch size of 2 and patches of size ($192 \times 128 \times 56$) voxels along the x, y, and z axes, respectively.

Regarding the loss function, the learning rate, and the optimizer, the default configuration of the 3D nnU-Net is kept. The loss function consisted in the sum of the Cross-Entropy and Soft Dice loss functions. The optimizer is Stochastic Gradient Descent with an initial learning rate of 0.01 and polynomial decay.

Data augmentation is used during training, following the standard transformations proposed in the nnU-Net framework : translation, rotation, scaling, mirroring, gamma augmentation. Detailed implementation and parameterization is provided in the original nnU-Net paper (Isensee et al., 2021a).

6.4.1.4 Inference

In inference, a new given image undergoes the image preprocessing procedure described in Section 6.4.1.1, for which parameters were defined on the training samples. Then, when applied on a new image, the trained nnU-Net generates a segmentation of the pancreas, pancreatic lesion and MPD, as well as a lesion probability map (output from the final softmax layer) assigning to each voxel a probability to be a lesion (see Figure 6.4).

6.4.2 Feature extraction

Using the output of the nnU-Net, features are extracted as follows :

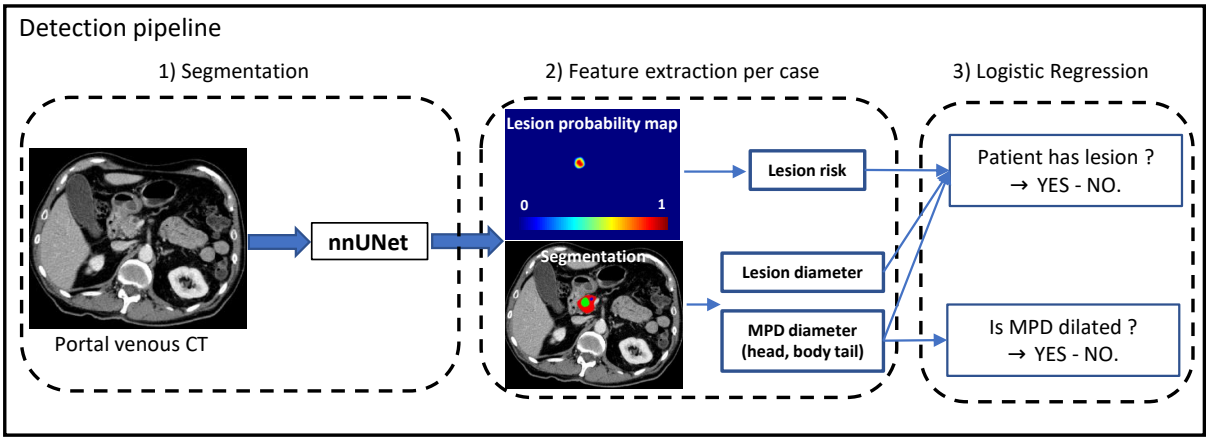


Figure 6.4 – Illustration of the detection pipeline. The algorithm is composed of three steps : 1) Based on a portal venous CT scan, the nnU-Net generates a probability map as well as a segmentation of the pancreas (red), lesions (green) and MPD (blue). 2) Lesion and MPD features are extracted. 3) Lesion presence and MPD dilatation are predicted using two logistic regression models. *CT : Computed Tomography, MPD : Main pancreatic duct.*

- A lesion risk between 0 and 1, computed using the 3D probability map generated by the final softmax layer of the nnU-Net. The computation is done as follows :
  1. First, predicted lesions that form a connected component out of the predicted pancreas segmentation mask are removed ;
  2. Second, the resulting predicted probability map for the lesion segmentation is binarized using a threshold at 0.01. From this binary lesion segmentation map, each lesion candidate (i.e., lesions forming a connected component) is given a lesion probability equal to the average predicted probability of the voxels within the connected component.
  3. The lesion candidate is discarded if the lesion probability is below 0.05 or if the number of voxels in the lesion is inferior to 10.
- The maximum diameter of the lesion segmented by the nnU-Net, computed as the maximum 2D axial Feret diameter (Van der Walt et al., 2014) of the lesion segmented by the nnU-Net. If no lesion is segmented, the maximum diameter is set to 0 ;
- The maximum 2D MPD diameter in the head, body, and tail of the pancreas. Computation is carried out using the IMEA (Kroell, 2021) library. Details on how the head, body, and tail of the pancreas are identified are provided in Appendix A.1. The maximum value of the MPD diameter in the pancreas is also computed by taking the maximum diameter between head, body, and tail. For each region, the MPD diameter is set to 0 if there is no predicted segmentation.

6.4.3 Logistic regression

We propose 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 nnU-Net is applied on the validation set of each fold, leading to a total of 2134 3D probability maps and segmentations, from which features are extracted. Two logistic regression models are trained using the Scikit-Learn library (Pedregosa et al., 2011). To reduce the potential impact of class imbalance, the two logistic regression models are regularized by weighting the binary cross-entropy loss function with the inverse class frequencies :

$$\text{Weighted Binary Cross-Entropy Loss} = -\frac{1}{N}\sum_{i=1}^{N}[w_0 \times y_i \times \log(p_i) + w_1 \times (1-y_i) \times \log(1-p_i)]$$

where N represents the number of samples in the dataset, y_i is the true label (either 0 or 1) for the i-th sample, and $p_i \in]0,1[$ is the predicted probability that the i-th sample belongs to class 1. Additionally,

$w_0 = \frac{N}{2 \times N_0}$ and $w_1 = \frac{N}{2 \times N_1}$ are inverse-class frequencies, where N_0 and N_1 are the number of samples with true label 0 and 1, respectively.

Prediction of lesion presence The first logistic regression predicts lesion presence based on three features : 1) lesion risk ; 2) lesion diameter ; 3) maximum MPD diameter in the pancreas. These features are extracted for all candidate lesions of each case. Candidate lesions matching a reference lesion (DSC greater than 0.1) are considered as true positives, while candidate lesions that do not match a reference lesion are considered as false positives. Based on these target classes the logistic regression can be calibrated on the training set.

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

6.4.4 Evaluation pipeline

Once the two logistic regressions are trained, evaluation of a test subject comprises three steps :

1. Generation of the lesion probability map and segmentation of the pancreas, lesion, and MPD using the nnU-Net.
2. Extraction of features. As the focus is on lesion detection at the patient level, only the lesion with the highest lesion risk among all candidate lesions is considered.
3. Application of the two previously trained logistic regression models to predict lesion presence and MPD dilatation (see Figure 6.4).

6.4.5 Statistical analysis

To evaluate performance, the Receiver Operating Characteristic (ROC) curve is generated, illustrating the trade-off between sensitivity and false positive rate across various thresholds of the predicted probability from the logistic regression. The Area Under the Curve (AUC) is measured, as well as sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) at the operating point maximizing the balanced accuracy.

As the primary focus is on achieving accurate cancer detection in patients, the evaluation is conducted at the case level : in the case of lesion detection, sensitivity is 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. Similarly, in the case of MPD dilatation detection, sensitivity is defined as the ratio of the number of patients correctly detected with a dilatation by the model to the total number of patients with a dilatation. The computation of the other evaluation measures is performed accordingly. Details on how each metric is computed are provided in Appendix A.2 and A.3. Bootstrap sampling is used to provide median values and 95% Confidence Interval (CI) for AUC, sensitivity, specificity, PPV and NPV.

Finally, the segmentation performance is evaluated by calculating the DSC and the Normalized Surface Dice (NSD) scores, both defined in Section 3.2.3.4. As the NSD allows for a tolerance error between the reference and predicted segmentations (Ostmeier et al., 2022), it 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 is 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) are computed. The FPR is defined for each class as the number of wrongly segmented voxels to the total number of segmented voxels. The FNR is computed accordingly.

6.5 Results

6.5.1 Detecting patients with pancreatic neoplasms

The model performance on the test set is reported in Figure 6.5 and Table 6.2. 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 distributions of the evaluation metrics is provided in Appendix A.4.

In order to compare with other methods, the model performance on public data (defined in Figure 6.1) is also reported in Table 6.2. Evaluation measures provided similar values than the ones obtained on the whole test set. The model was also evaluated on subjects with specific lesions characteristics and types. Further examination was conducted on five particularly significant subgroups :

- patients with lesions less than 2cm in diameter, corresponding to stage IA tumors (refer to Section 2.3.4), thereby indicating early detection ;
- patients with isodense lesions, defined as lesions with a density similar to adjacent pancreatic parenchyma, often accompanied by secondary signs like abrupt cutoffs of the MPD or common bile duct, making their identification challenging on portal CT scans (Elbanna et al., 2020) ;
- patients with PDAC, constituting 80% of pancreatic cancers and typically appearing hypodense (see Section 2.3) ;
- patients with PNET, representing 10% of pancreatic cancers and typically appearing hyperdense (see Section 2.3) ;
- patients with IPMN, cystic lesions with the potential to become malignant, thereby requiring early and appropriate diagnosis (refer to Section 2.4.1).

Performance obtained on these subsets is reported in Table 6.2. 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 PNET 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.

	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 $\leq$ 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 types					
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]
PNET (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 6.2 – 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. * 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 main pancreatic duct or the common bile duct, allowing to confirm lesion presence. *AUC : Area under the curve, D : diameter, IPMN : Intraductal papillary mucinous neoplasm, PNET : Pancreatic neuroendocrine tumor, NPV : Negative predictive value, PDAC : Pancreatic ductal adenocarcinoma, PPV : Positive predictive value.*

6.5.2 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 (see Section 6.4.2). 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 6.3 reports the sensitivity of these three models on the test set, as well as on the subsets previously

defined in Section 6.5.1. 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 $\leq$ 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 types			
PDAC (418/681)	0.94 [0.91, 0.97]	0.91 [0.88, 0.95]	0.89 [0.85, 0.94]
PNET (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 6.3 – 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, IPMN : Intraductal papillary mucinous neoplasm, PNET : Pancreatic neuroendocrine tumor, PDAC : Pancreatic ductal adenocarcinoma.*

6.5.3 MPD dilatation detection performance

Regarding MPD dilatation performance, the model of Section 6.4.3 was also evaluated on the test set. Results are reported in Table 6.4. 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 A.4.

	AUC	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Test set (493/756)	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 6.4 – Evaluation metrics obtained by the logistic regression predicting main pancreatic duct dilatation. 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, NPV : Negative predictive value, PPV : Positive predictive value.*

6.5.4 Segmentation performance

The segmentations predicted by the nnU-Net of Section 6.4.1 were evaluated both quantitatively and qualitatively. The DSC between the reference and the nnU-Net segmentation maps is reported in Table 6.5 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 measures were also measured on public data only in order to compare our segmentation network to other works. Qualitative examples of model segmentations are presented in Figure 6.6, where red, green, and blue voxels indicate the predicted pancreas, lesion, and MPD regions, respectively. The top row displays a small PNET tumor located at the very end of the pancreatic tail and correctly segmented by the model. The second and third rows depict correct predicted segmentations of IPMN and PDAC lesions measuring 13mm and 18mm, respectively. The last row illustrates a dilated MPD, with a diameter of 4mm, correctly segmented by the model.

On the whole test set, the mean DSC was 0.91 ($\pm$ 0.06), 0.69 ($\pm$ 0.34) and 0.58 ($\pm$ 0.37) for the pancreas, lesions, and MPD, respectively. The mean NSD score was 0.71 ($\pm$ 0.39) and 0.77 ($\pm$ 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 A.5.

	Test set (n=756)			Public data (n=361)		
	Pancreas	**Lesions**	**Main Pancreatic Duct**	**Pancreas**	**Lesions**	**Main Pancreatic Duct**
DSC	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 6.5 – Dice similarity coefficient (DSC), NSD score, FPR, and FNR measured on the whole test set and public data. The DSC, 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. *DSC : Dice similarity coefficient, FPR : False positive rate, FNR : False negative rate, NSD : Normalized surface Dice.*

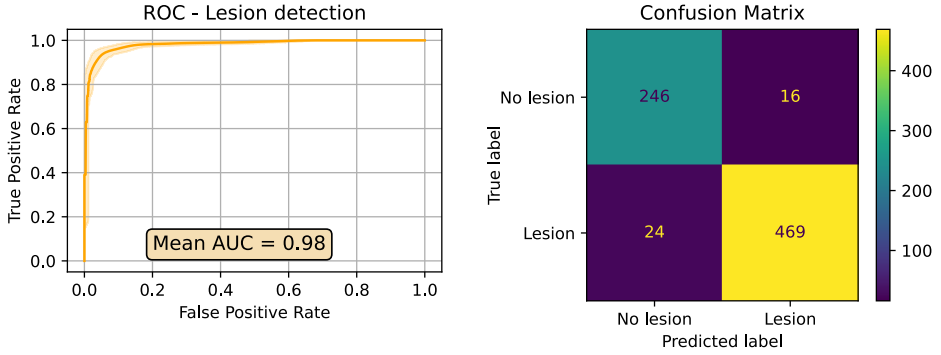


Figure 6.5 – Quantitative results. Left : ROC curve of the logistic regression model predicting lesion presence. The central line represents the average curve and shadowed areas the 95% confidence interval. Right : Confusion matrix of the model obtained at the operating point maximizing the balanced accuracy. *ROC : Receiver operating characteristic, AUC : Area under the curve.*

6.6 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 that 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.

Subgroup analysis for the detection of patients with pancreatic neoplasms 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 PNET (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 6.2 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 control cases. 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.

Comparison with other works on public data The results on public data can be compared to competing DL models that used this database to test their approach. Alves et al. (2022) reported an AUC of 0.91

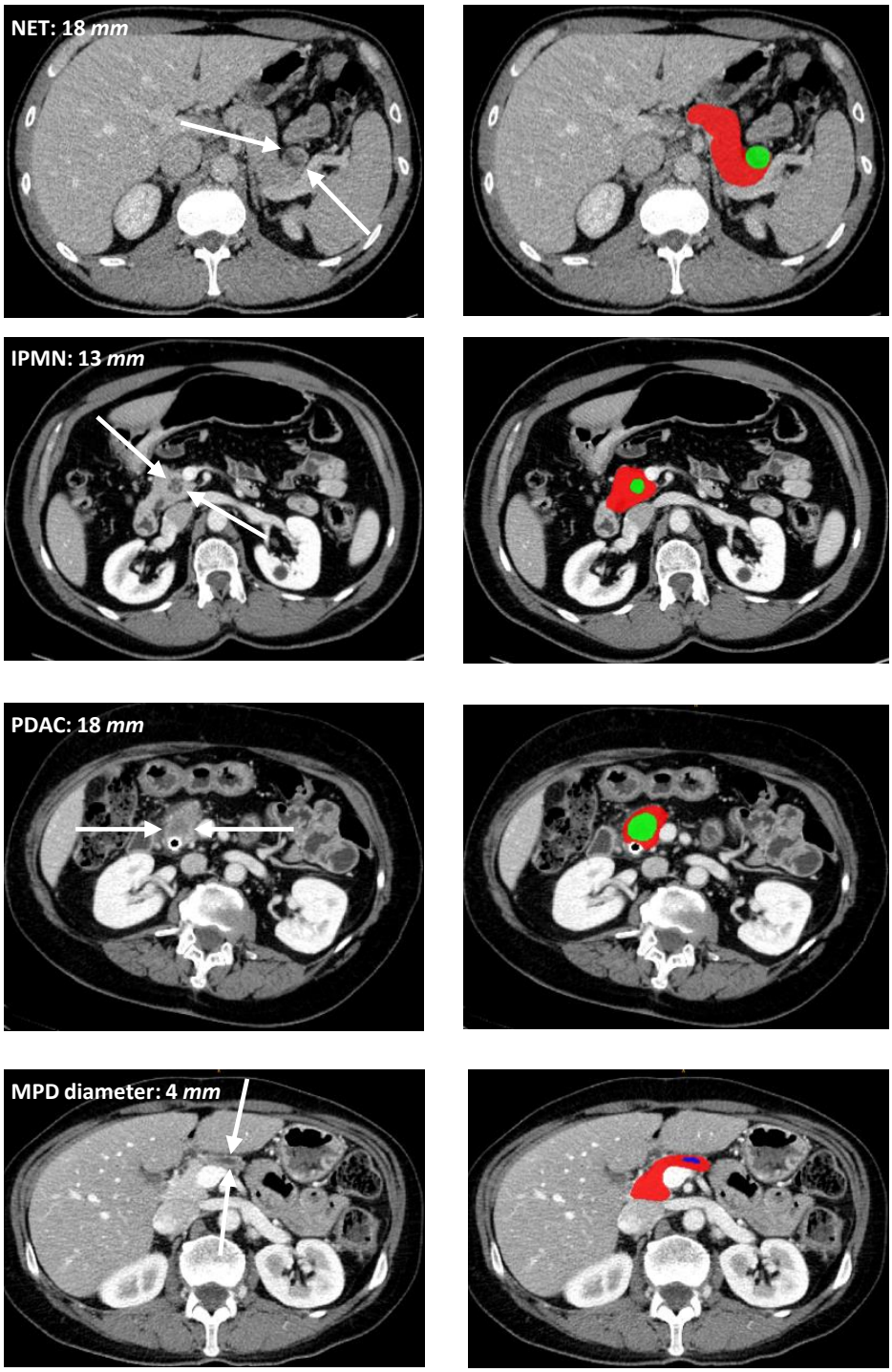


Figure 6.6 – Qualitative results. **Left column :** Axial portal venous CT slices of four patients. White arrows indicate lesions location for the first three rows and the MPD for the last row. **Right column :** Segmentation of pancreas (red), lesions (green) and MPD (blue) output by our model. *CT : Computed tomography, IPMN : Intraductal papillary mucinous neoplasm, MPD : Main pancreatic duct, NET : Pancreatic neuroendocrine tumor, PDAC : Pancreatic ductal adenocarcinoma.*

and Liu et al. (2020) 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 (Chen et al., 2023; Park et al., 2023), thus allowing for 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, PNET, 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 two-fold : *(i)* it does not require to modify the U-Net 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 10%). 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 (Yoon et al., 2011). Improvement of sensitivity for PDAC detection (plus 5%) when using the lesion and MPD diameter was also observed. Regarding pancreas PNET, a milder effect was observed (plus 2% sensitivity). However, as PNETs are not linked to MPD dilatation, effect on sensitivity in patient with PNET 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.

MPD dilatation detection performance 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 6.4.3). An AUC of 0.97 (95% CI : [0.96,0.98]) was reported. We highlight that even if other DL methods allow segmenting the MPD (Viviers et al., 2022; Xia et al., 2022), 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 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 DSC on the same datasets as in this study. However, three DL models which reported a mean DSC of 0.87 on their test set were found in (Zhou et al., 2019a; Zhu et al., 2019b, 2021). The segmentation network presented in this work obtained a similar result on public data and a higher mean DSC on the whole test set. As for lesion segmentation, a mean DSC of 0.63 was obtained on public data. This is a 9% improvement over the nnU-Net trained by Isensee et al. (2021a) who reported a DSC of 0.54 on this dataset. Finally, we could not find DSCs obtained by other DL approaches for the MPD. However, given the small size of the MPD and compared to the DSC 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 DSC values remained below 0.7 for lesions and the MPD. The reason for this is two-fold : *(i)* for false negative and false positive cases, the associated DSC will be 0, thus strongly deteriorating the mean value. We show in Appendix A.5 that using the median instead of the mean is more robust to these extreme cases. A median DSC 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 DSC strongly penalizes segmentation errors in this scenario. Using the NSD better highlights the predicted segmentation quality for these structures. Compared to the DSC, a gain of 0.08 and 0.13 points was observed for the NSD in the case of lesions and the MPD, respectively.

Limitations There are 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 6.2 and 6.3. Secondly, 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

PNETs (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. Thirdly, the reference label for MPD dilatation was derived from visual assessments by radiologists rather than relying on a quantitative measurement of the MPD diameter.

6.7 Conclusion

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 pancreatic 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.

Chapitre 7

Towards early diagnosis : detection of secondary signs

7.1 Preface

Following the detection of pancreatic lesions, our focus shifted towards identifying secondary signs of pancreatic cancer. We initiated this exploration with the identification of abnormal pancreatic shapes, presenting a data-driven approach to learn shape distributions from large databases of healthy organs. Leveraging this healthy shape distribution, we achieved the detection of abnormal pancreatic shapes in both zero-shot and few-shot settings. This methodology, presented in Section 7.2, was patented and led to a ***poster communication at Medical Image Computing and Computer Assisted Interventions (MICCAI) 2022.*** *Then, we expanded this work with the detection of four secondary signs of pancreatic cancer using radiomics. Specifically, we tackled the challenge of extracting both hand-cratfed and non-redundant deep learning radiomics to predict four secondary signs of pancreatic cancer : abnormal shape, atrophy, senility, and fat replacement. This work, presented in Section 7.3, was selected for an* ***oral presentation at Cancer Prevention through Early Detection (CaPTion), a workshop of MICCAI 2023****, and won the best presentation award.*

7.2 Learning shape distributions from large databases of healthy organs : applications to zero-shot and few-shot abnormal pancreas detection

Abstract. We propose a scalable and data-driven approach to learn shape distributions from large databases of healthy organs. To do so, volumetric segmentation masks are embedded into a common probabilistic shape space that is learned with a variational auto-encoding network. The resulting latent shape representations are leveraged to derive zero-shot and few-shot methods for abnormal shape detection. The proposed distribution learning approach is illustrated on a large database of 1200 healthy pancreas shapes. Downstream qualitative and quantitative experiments are conducted on a separate test set of 224 pancreas from patients with mixed conditions. The abnormal pancreas detection area under the curve reached up to 65.41% in the zero-shot configuration, and 78.97% in the few-shot configuration with as few as 15 abnormal examples, outperforming a baseline approach based on the sole volume.

7.2.1 Introduction

Anatomical alterations of organs such as the brain or the pancreas may be informative of functional impairments. For instance, hippocampal atrophy and duct dilatation are well-known markers of Alzheimer's disease and Pancreatic Ductal Adenocarcinoma (PDAC), respectively (Fox et al., 1996; Liu et al., 2019b). In these examples, quantifying anatomical differences bears therefore a great potential for determining the patient's clinical status, anticipating its future progression or regression, and supporting the treatment planning.

Since the seminal work of Thompson (1917), the computational anatomy literature proposed several Variational Autoencoder (VAE) approaches, which embed geometrical shapes into metric spaces where notions of distance and difference can be defined and quantified (Beg et al., 2005; Christensen et al., 1996; Kendall, 1984). Taking advantage of these representations, statistical shape models were then proposed to perform group analyses of shape collections. In particular, atlas models (Pennec, 2006) learn geometrical distributions in terms of an "average" representative shape and associated variability, generalizing the Euclidean mean-variance analysis. In medical imaging, learning atlases from healthy examples allows for the definition of normative models for anatomical structures or organs, such as brain Magnetic Resonance Imaging (MRI)s or subcortical regions segmented from neuroimaging data (Gori et al., 2017; Zhang et al., 2013), thus providing a natural framework for the detection of abnormal anatomies.

In practice, leveraging an atlas model to compute the likelihood of a given shape to belong to the underlying distribution either requires to identify landmarks (Cootes et al., 1995), or to solve a registration problem (Bône et al., 2018). To circumvent the computational cost of this shape embedding operation, Yang et al. (2017) proposed to train an encoder network to predict registration parameters from image pairs. Dalca et al. (2018); Krebs et al. (2019) built on this idea and used the VAE of Kingma and Welling (2014) to learn the embedding space jointly with the atlas model, instead of relying on predetermined parametrization strategies. However, the structure of the decoding network remained constrained by hyperparameter-rich topological assumptions, enforced via costly smoothing and numerical integration operators from a computational point of view.

Alternative approaches proposed to drop topological hypotheses by relying on variations of the Autoencoder (AE) or its variational counterparts (Kingma and Welling, 2014) to learn normative models that are subsequently used to perform Anomaly Detection (AD). These methods compress and reconstruct images of healthy subjects to capture a normative model of organs (Baur et al., 2021; Zimmerer et al., 2019). Yet, they are usually applied on the raw imaging data, thus they entail the risk of extracting features related to the intensity distribution of a dataset which are not necessarily specific to the organ anatomy. Therefore, regularization constraints (Baur et al., 2021; Chen and Konukoglu, 2022) are introduced to improve the detection performance compared to the vanilla AE. To further reduce the overfitting risk, these methods artificially increase the dataset size by working on 2D slices.

Given this context, we propose a VAE-based method to learn a normative model of organ shape that can subsequently be used to detect anomalies, thus bridging the gap between VAE and AD models. Although VAE methods with explicit modeling constraints proved effective to learn relevant shape spaces from relatively small collections of high-dimensional data, we propose to further reduce the set of underlying hypotheses and leave the decoding network unconstrained in its architecture. With the objective to learn normative shape models from collections of healthy organs, we argue that sufficiently large databases of relevant medical images can be constructed by pooling together different data sources, see (Dufumier et al., 2021) for instance. To reduce the risk of overfitting and focus on the anatomy of organs, the VAE is learned from 3D binary segmentation masks and is coupled with a shape-preserving data augmentation strategy consisting of translations, rotations and scalings. An approach to study and visualize group differences is also proposed.

Section 7.2.2 details the proposed method, which is then illustrated on a pancreas shape problem in Section 7.2.3. Section 7.2.4 discusses the results and concludes.

7.2.2 Methods

7.2.2.1 Modeling organ shape

We consider an image acquired via a standard imaging technique. For a given organ in the image, its anatomy can be represented by a binary segmentation mask $\mathbf{X} = \{x_i, i = 1...d\}$ with $x_i \in \{0, 1\}$ and d the number of voxels in the image. We are interested in studying the shape of this organ, and assume that it is characterized by a set of underlying properties that can be extracted from the segmentation mask. Therefore, we hypothesize the following generative process for the segmentation mask :

$$p_\theta(\mathbf{X} \mid \mathbf{z}) = \prod_{i=1}^{d} f_\theta(\mathbf{z})_i^{x_i} (1 - f_\theta(\mathbf{z})_i)^{1-x_i} \tag{7.1}$$

where $0^0 = 1$ by convention, and $\mathbf{z}$ is a latent variable generated from a prior distribution $p(\mathbf{z})$. This latent variable provides a low-dimensional representation of the segmentation mask embedding its main shape features. The function f_θ is a non-linear function mapping $\mathbf{z}$ to a predicted probabilistic segmentation mask.

We are interested in inferring the parameters θ of the generative process, as well as approximating the posterior distribution of the latent variable $\mathbf{z}$ given a segmentation mask $\mathbf{X}$. We rely on the VAE framework (Kingma and Welling, 2014), detailed in Section 3.2.2.5, to estimate the model parameters. Hence, we assume that $p(\mathbf{z})$ is a multivariate Gaussian with zero mean and identity covariance. We also introduce the approximate posterior distribution $q_\phi(\mathbf{z} \mid \mathbf{X})$ parameterized by ϕ, and optimize a lower bound $\mathcal{L}$ of the marginal log-likelihood, which can be written for the segmentation mask $\mathbf{X}^p$ of a subject p as :

$$\mathcal{L} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{X}^p)}[\log p_\theta(\mathbf{X}^p \mid \mathbf{z})] - KL[q_\phi(\mathbf{z} \mid \mathbf{X}^p) \mid p(\mathbf{z})], \tag{7.2}$$

where $q_\phi(\mathbf{z} \mid \mathbf{X}^p)$ follows a Gaussian distribution $\mathcal{N}(\mu_\phi(\mathbf{X}^p), \sigma^2_\phi(\mathbf{X}^p)\mathbf{I})$ with $\mathbf{I}$ the identity matrix, and KL is the Kullback-Leibler divergence.

To capture shape features, we use a convolutional network and adopt the U-Net (Ronneberger et al., 2015) encoder-decoder architecture, detailed in Section 3.2.3.5. However, to align with the VAE framework, skip connections are omitted, as depicted in Figure 7.1. In practice, the number of convolutional layers and the convolutional blocks are automatically inferred thanks to the nnU-Net self-configuring procedure (Isensee et al., 2021a), detailed in Section 3.2.3.5. Due to this encoder-decoder architecture, the segmentation masks are progressively down-sampled to obtain low-resolution feature maps which are mapped through a linear transformation to the latent variable $\mathbf{z}$. The latent code is subsequently decoded by a symmetric path to reconstruct the original masks.

The model was trained on 800 epochs. We used the Stochastic Gradient Descent optimizer with an initial learning rate of 10^{-4} following a polynomial decay. To avoid memory issues and mimic larger batch sizes, better suited for VAEs (Kingma and Welling, 2014), we used a batch size of 8 with a gradient accumulation step of 5.

7.2.2.2 Anomaly detection

We propose to learn a normative model of organ shapes by applying the VAE framework previously presented on the segmentation masks of a large cohort of N healthy patients, allowing the model to capture a low-dimensional embedding characteristic of a normal organ anatomy. In addition, we use a data augmentation procedure consisting of random translations, rotations and scalings, in order to be invariant to these transformations and force the network to extract shape features. Based on this learned model, we propose two approaches to perform AD by leveraging the latent representation of normal organ shapes.

7.2.2.3 Zero-shot learning method

After training, the recognition model $q_\phi(\mathbf{z} \mid \mathbf{X})$ can be used to project the segmentation maps $\mathbf{X}^p$ of the cohort of healthy subjects and obtain an empirical distribution of normal shapes in the latent space. We

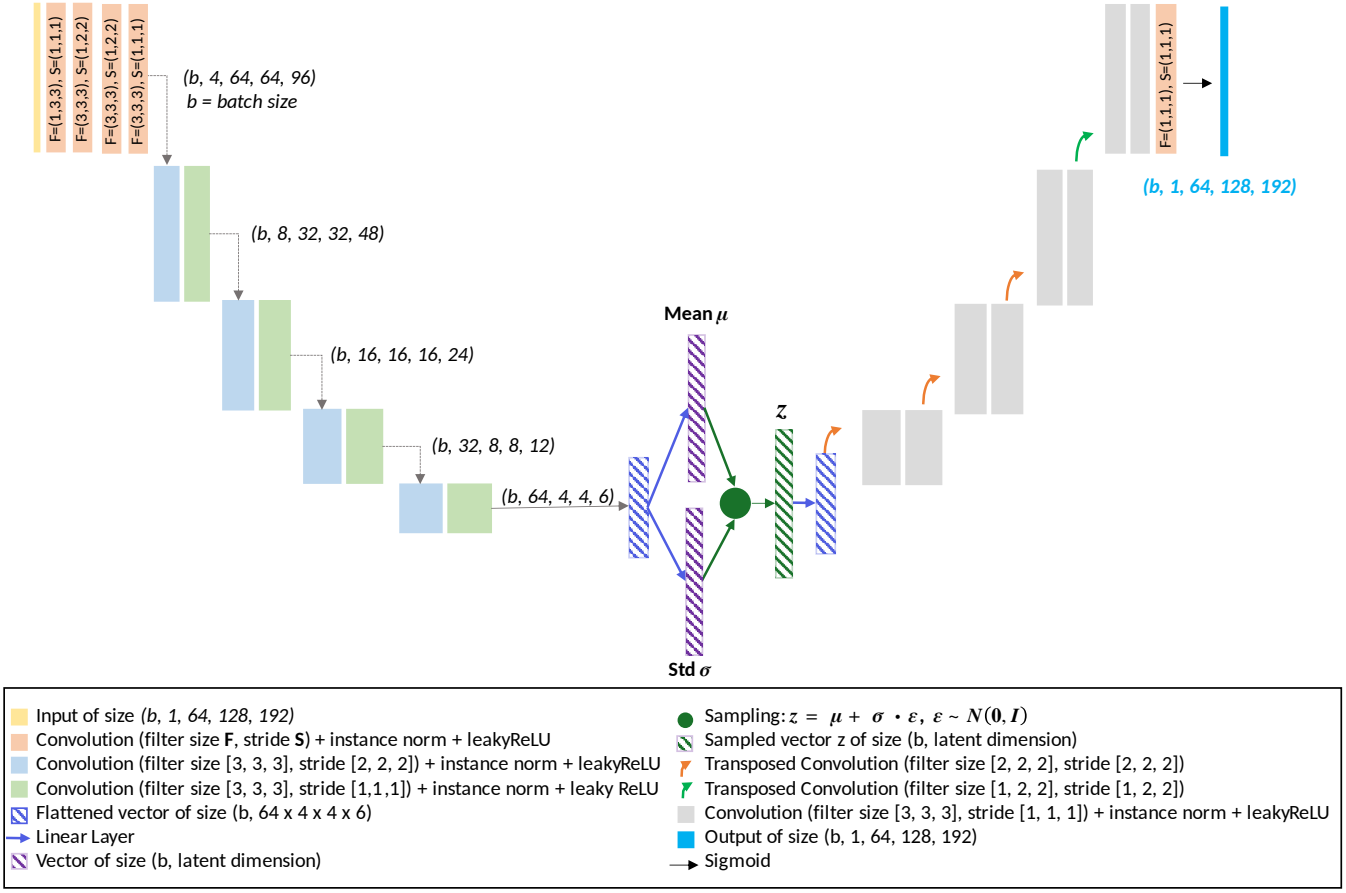


Figure 7.1 – Architecture of the proposed variational autoencoder.

rely on this low-dimensional distribution of normality to detect abnormal shapes. To do so, we compute the mean of the healthy subjects projection, and define abnormality through the L_2 distance to this mean latent representation.

7.2.2.4 Few-shot learning method

Another approach is to classify normal and abnormal shapes based on their low-dimensional representations. In practice, we project the segmentation maps from a set of healthy and pathological subjects in the latent space using the recognition model $q_\phi(\mathbf{z} \mid \mathbf{X})$. Therefore, we obtain for all these subjects a set of low-dimensional organ shape features that we can use to learn any type of classifier (*e.g.,* linear Support Vector Machine (SVM)).

7.2.2.5 Studying organ shapes differences

Our framework can also be used to study organ differences between groups. Let us consider a set of healthy and pathological subjects, as well as their segmentation masks. Based on the recognition model $q_\phi(\mathbf{z} \mid \mathbf{X})$, we can compute the average of the subject's latent projection for each group, denoted by $\mathbf{z}_{normal}$ and $\mathbf{z}_{abnormal}$, respectively. We consider the line of equation $(1-t) \times \mathbf{z}_{normal} + t \times \mathbf{z}_{abnormal}$ with $t \in \mathbb{R}$. When moving along this line with increasing values of t, we progress from a healthy mean latent shape representation to a pathological one, and can reconstruct the corresponding segmentation mask using the probabilistic decoder $p_\theta(\mathbf{X} \mid \mathbf{z})$.

The comprehensive framework, along with its three applications, is depicted in Figure 7.2.

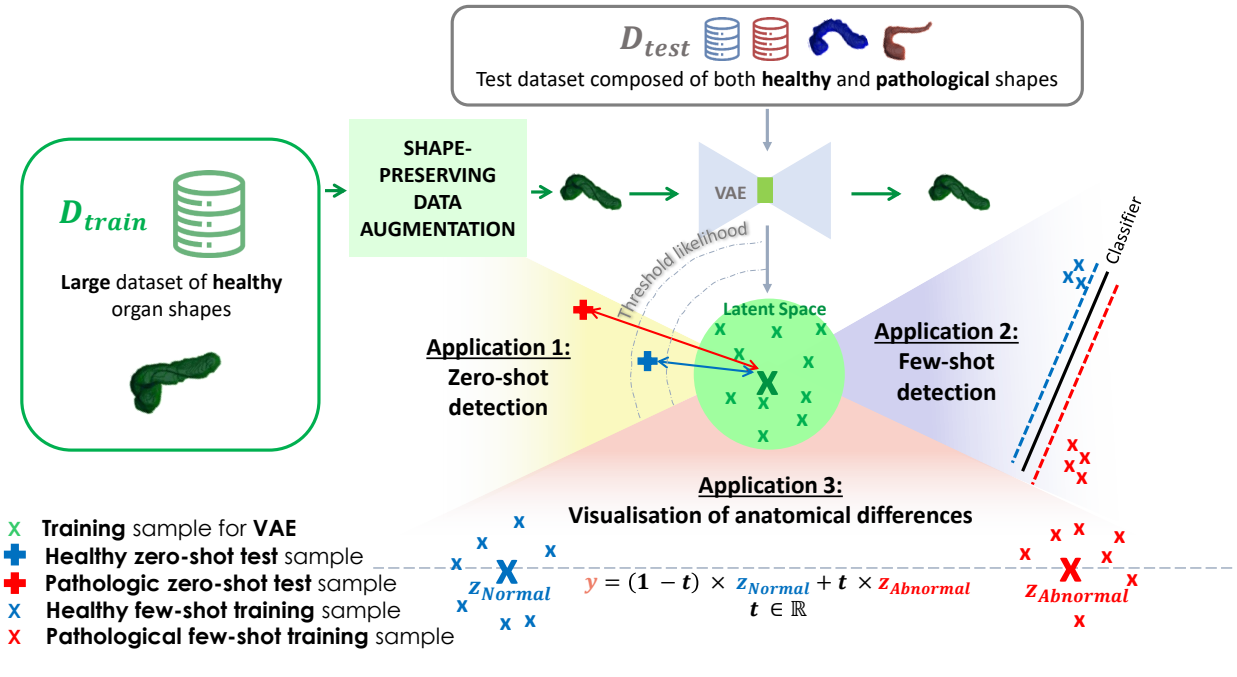


Figure 7.2 – Framework of the proposed method. A variational autoencoder (VAE) is trained on many 3D segmentation masks representing healthy pancreas. After training, we assume the model has captured a low-dimensional embedding, characteristic of a normal shape. By projecting the training samples, we get an empirical distribution of normal shapes in the latent space (green circle). Based on this model, we can perform anomaly detection in two ways : firstly, in a zero-shot setting. To do so, we compute the mean of the projections of the healthy training samples. Then, we project the test shapes in the latent space, and we define abnormality through the L_2 distance to this mean latent representation. Secondly, in a few-shot setting, with very few samples. We train a classifier to classify normal and abnormal shapes based on their low-dimensional representations. Lastly, our framework can also be used to visualize shape differences between groups. For example, for a set of healthy and pathological subjects, we can compute the average of the healthy and pathological latent projection : $\mathbf{z}_{normal}$ and $\mathbf{z}_{abnormal}$. Then, we consider the line going through these two points, of equation $y = (1-t) \times \mathbf{z}_{normal} + t \times \mathbf{z}_{abnormal}$ with $t \in \mathbb{R}$. When moving along this line, we progress from a healthy latent shape representation to a pathological one. We can generate the shapes corresponding to each point on this line.

7.2.3 Experiments

In this section, we applied our method in the case of the pancreas. A normative model of pancreas shape was learned on a large cohort of healthy subjects, and was then leveraged for AD on an independent test cohort. Several configurations were proposed to assess the model performance, including the impact of the number of training subjects and of the latent space dimensionality on the AD performance. Detection with the few-shot learning method was performed using SVM. Finally, we showed how the proposed framework can be used to visualize differences between the healthy and pathological pancreas.

7.2.3.1 Training

The training dataset $\mathcal{D}^{Train}$ was created from $\mathcal{D}_{CALV}$, our private cohort containing 3270 abdominal Portal Computed Tomography (CT) scans of patients with potential liver cancer (see Section 4.2.2.2). To ensure the healthy condition and shape of the pancreas, several exclusion criteria were applied :

— First, all patients with metastatic cancer or lesions reported as “unevaluable” were excluded.

— Secondly, in order to avoid overly large liver lesions, patients with lesion volumes exceeding the 99^{th} percentile were also excluded.

— Thirdly, cases in which lesions protruded from the liver were also removed from the study.

— Finally, the pancreas segmentation masks were automatically generated for the remaining patients, and only the cases for which the mask consisted of one single connected component were retained, this last criterion acting as a quality control on the segmentation masks.

In the end, 1200 cases were retained to build the healthy pancreas database for training. To explore the influence of the number of samples seen during training, subsets $\mathcal{D}_N$ with a growing number of subjects were created (see Table 7.1). For each $\mathcal{D}_N$, 80% and 20% of the samples were used for training and validation, respectively. Splitting was done such that the pancreas volume distribution was balanced across the splits.

Name	N_{tot}	N_{train}	N_{val}	**Liver lesion**			**Other abdominal pathology**
				Benign	Non suspicious	Suspicious	
$\mathcal{D}_{300}$	300	240	60	✓	✗	✗	✗
$\mathcal{D}_{600}$	600	480	120	✓	✓	✗	✗
$\mathcal{D}_{900}$	900	720	180	✓	✓	✓	✗
$\mathcal{D}_{1200}$	1200	960	240	✓	✓	✓	✓

Table 7.1 – Individual splits of the different training sets $\mathcal{D}_N$, as well as their corresponding inclusion criteria.

7.2.3.2 Testing

The test database $\mathcal{D}^{Test}$ was obtained by combining two datasets : *(i)* 144 cases from $\mathcal{D}_{BR,1}$, diagnosed with pancreatic cancer, and for whom the pancreas shape was evaluated as abnormal by an expert radiologist ; *(ii)* $\mathcal{D}_{NIH}$, containing 80 CT scans of patients with healthy pancreas and for whom the assumption of normal pancreas shape held. $\mathcal{D}_{BR,1}$ and $\mathcal{D}_{NIH}$ are detailed in Chapter 4. Examples of normal and abnormal shapes can be seen in Figure 7.3. Centers, machines and protocols differed among $\mathcal{D}^{Train}$, NIH, and $\mathcal{D}_{BR,1}$.

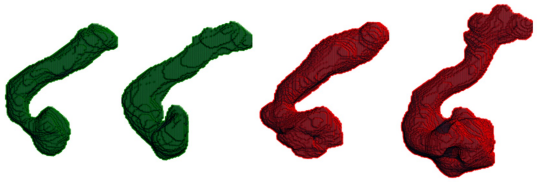


Figure 7.3 – Examples of normal and abnormal pancreas shapes from $\mathcal{D}^{Test}$. Green and red figures are examples taken from $\mathcal{D}_{NIH}$ and $\mathcal{D}_{BR,1}$, respectively. $\mathcal{D}_{BR,1}$ and $\mathcal{D}_{NIH}$ are detailed in Chapter 4.

7.2.3.3 Preprocessing

The first step consisted in obtaining the pancreas segmentation masks. For the public dataset $\mathcal{D}_{NIH}$, we used the reference pancreas segmentation masks publicly available. For $\mathcal{D}^{Train}_{normal}$ and $\mathcal{D}_{BR,1}$, the masks were obtained semi-automatically using an in-house segmentation algorithm derived from the nnU-Net, and validated by a radiologist with 25 years of expertise in abdominal imaging. Finally, all the masks were resampled to $1 \times 1 \times 2$ mm^3 in (x, y, z) directions, and centered in a volume of size $192 \times 128 \times 64$ voxels.

7.2.3.4 Zero-shot AD

We trained our model on the different datasets $\mathcal{D}^{Train}_N$, with a growing number of latent dimensions L ranging from 16 to 1024 (denoted by $L_{16}...L_{1024}$). For each experiment, we applied the zero-shot AD procedure, as previously explained, on $\mathcal{D}^{Test}$. We report the Area Under the Curve (AUC), in %, in Table 7.2. Increasing the dimension of the latent space L improved the classification performance on each dataset $\mathcal{D}_N$. Moreover, for each dataset size the best result was consistently obtained when L was set at the maximum value L_{1024}. We also observed that the effect of the latent space dimension on the performance seemed to attenuate as the dataset size increased. Indeed, we observed that when going from L_{16} to L_{1024}, the mean AUC for $\mathcal{D}_{300}$, $\mathcal{D}_{600}$, $\mathcal{D}_{900}$, $\mathcal{D}_{1200}$ improved by 10.7, 5.9, 4.0 and 3.1 points, respectively. Regarding the effect of the database size, we observed that increasing the training set size seemed to globally improve the AUC scores. For instance, going from $\mathcal{D}_{300}$ to $\mathcal{D}_{600}$ increased the classification performance for all the experiments, particularly for L_{16} which gained 9.3 points. This beneficial effect of both larger training sets and latent dimension was also observed on the Dice Similarity Coefficient (DSC) between the original and

reconstructed segmentation masks, as shown in Table 7.3. Thus, for the following experiments, we chose the model trained on $\mathcal{D}_{1200}$ with a latent dimension L_{1024} as it gave the best results in terms of AUC and DSCs.

	L_{16}	L_{64}	L_{256}	L_{1024}
$\mathcal{D}_{300}$	51.51 ±0.37	59.08 ±0.37	62.16 ±0.37	62.17 ±0.37
$\mathcal{D}_{600}$	59.24 ±0.38	60.97 ±0.36	**64.32** ±0.36	65.11 ±0.36
$\mathcal{D}_{900}$	60.77 ±0.37	**62.64** ±0.37	64.04 ±0.36	64.81 ±0.36
$\mathcal{D}_{1200}$	**62.28** ±0.36	61.74 ±0.37	62.58 ±0.37	**65.41** ±0.36

Table 7.2 – Results for zero-shot AD. For each experiment, corresponding to a specific training size $\mathcal{D}$ and latent space dimension L, we report the mean and standard deviation of AUC scores (in %) obtained by bootstrapping with 10000 repetitions. Best results by line are underlined and by column are in **bold**. *AD : Anomaly detection, AUC : Area under the curve.*

	L_{16}	L_{64}	L_{256}	L_{1024}
$\mathcal{D}_{300}$	73.7 ±7.2	79.9 ±5.7	80.9 ±7.4	81.5 ±7.4
$\mathcal{D}_{600}$	75.5 ±7.5	83.4 ±5.4	85.0 ±5.8	83.9 ±5.9
$\mathcal{D}_{900}$	**78.1 ±5.9**	85.0 ±5.8	86.1 ±5.4	85.4 ±5.8
$\mathcal{D}_{1200}$	77.3 ±6.4	**85.9 ±5.4**	**86.2 ±5.3**	**88.5 ±4.4**

Table 7.3 – Results for segmentation masks reconstruction. For each experiment, corresponding to a specific training size and latent space dimension, we report the mean and standard deviation of Dice Similarity Coefficients obtained by bootstrapping with 10000 repetitions. Best results by line are underlined and by column are in **bold**.

To visualize the separation between normal and abnormal shapes, we projected each subject from $\mathcal{D}^{test}$ using the recognition model $q_\phi(\mathbf{z}|\mathbf{X})$. Based on the subjects' latent representation, we applied three dimensionality reduction techniques, namely principal components analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE) and Isomap. Results are displayed in Figure 7.4, on which each point represents the latent projection of a test subject reduced on a 2D plane. We observed that, independently of the projection technique, normal and abnormal shapes tended to be separated in two different clusters.

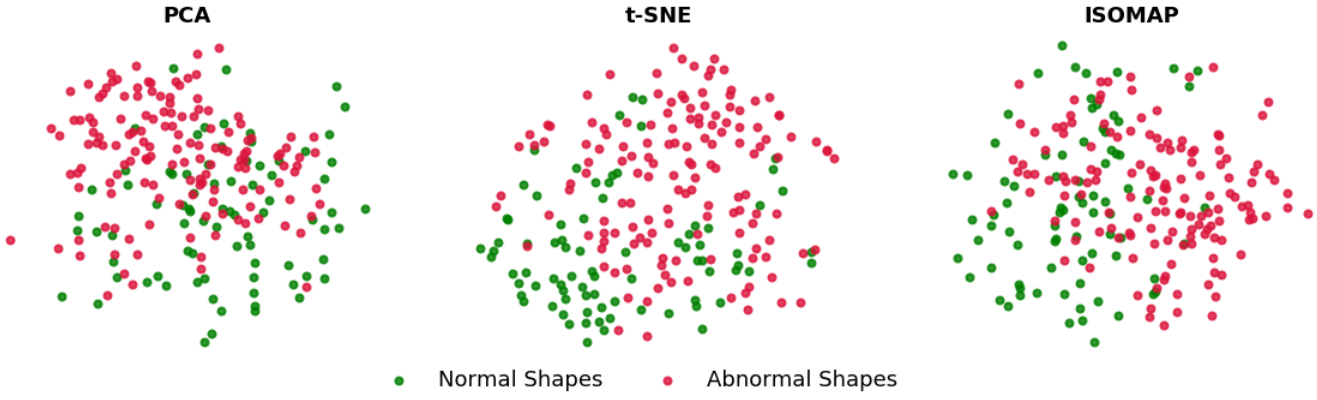


Figure 7.4 – 2D reduction of the latent representation of the test subjects $\mathcal{D}^{Test}$. The 80 samples from $\mathcal{D}_{NIH}$ are in green, and the 144 samples from $\mathcal{D}_{BR,1}$ are in red. *PCA : Principal components analysis, t-SNE : t-distributed stochastic neighbor embedding.*

7.2.3.5 Few-shot AD

We trained a linear SVM classifier on the latent representation of $\mathcal{D}^{Test}$ with stratified k-fold cross-validation. We varied the number k of folds to test the performance of the classifier depending on the train/test samples ratio. Experiments ranged from a 0.05 train/test ratio to a leave-one-out cross-validation and are presented in Table 7.4. We noticed that using only 8 healthy and 15 abnormal training samples increased the performance to 78.9%. We also observed that the AUC scores and the balanced accuracy

increased with the number of training samples, reaching a maximum of 91.1% and 83.2% respectively in the leave-one-out configuration.

Train/Test ratio	**0.05**	**0.11**	**0.25**	**1**	**223**
Number of training samples	*12 (8)*	*23 (15)*	*45 (29)*	*112 (72)*	*223 (144)*
AUC	66.02 ±0.02	78.97 ±0.03	81.87 ±0.07	86.95 ±0.23	91.18 ±0.19
Balanced Accuracy	67.78 ±0.03	70.88 ±0.05	73.96 ±0.10	75.49 ±0.37	83.26 ±0.34

Table 7.4 – Results for few-shot AD. For each experiment, we indicate the number of training samples, as well as the number of abnormal samples (in brackets). We report the means and standard deviations for AUC (in %) and balanced accuracy (in %), obtained by bootstrapping with 10000 repetitions. *AD : Anomaly detection, AUC : Area under the curve.*

7.2.3.6 Comparison with a baseline method

We compared our approach with a baseline method classifying shapes based on their volume. We applied this method on $\mathcal{D}^{Test}$ with bootstrap sampling and obtained an average AUC of 51% with a 95% confidence interval of $[49.9; 51.7]$, below the maximum AUC scores of 65.4% and 91.1% previously reported in the zero-shot and few-shot cases, respectively.

7.2.3.7 Comparison with other SSM methods

We also compared the proposed method with two state-of-the-art methods : Active Shape Models (ASM) (Cootes et al., 1995) and Large Deformation Diffeomorphic Metric Mapping (LDDMM) using the Deformetrica software (Bône et al., 2018). Concerning ASM, we computed the signed distance map of the pancreas 3D contours of each subject. For LDDMM, we estimated a Bayesian Atlas (Gori et al., 2017) parametrized by 576 control points. In both cases, we performed a PCA on the shape-encoding parameters to obtain a latent vector of dimension 1024 for each subject. We compared them with our method also using a latent dimension $L = 1024$. All methods were trained on $\mathcal{D}_{1200}$. Results, reported in Table 7.5, show that our method outperforms ASM and LDMM when the number of training samples is small (Train/Test ratio ≤ 0.05).

Configuration	**Zero-shot**	**Few-shot**		
Train/Test ratio	**0**	**0.05**	**1**	**223**
Ours	**65.41±0.36**	**66.02±0.02**	86.95±0.23	91.18±0.19
LDDMM	54.6±0.36	58.68±0.02	**89.43±0.19**	**95.41±0.11**
ASM	58.42±0.36	61.1±0.02	82.64±0.26	93.79±0.14

Table 7.5 – Few-shot and zero-shot AD results. For each experiment, we report the mean and standard deviation for AUC (in %), obtained by bootstrapping with 10000 repetitions. Best results by column are in **bold**. *AD : Anomaly detection, ASM : Active shape model, AUC : Area under the curve, LDDMM : Large deformation diffeomorphic metric mapping.*

7.2.3.8 Studying pancreas shapes differences between groups

To model differences in the pancreas shape between healthy and pathological groups, we applied the procedure presented in Section 7.2.2.5 on the subjects from $\mathcal{D}^{Test}_{normal}$ and $\mathcal{D}^{Test}_{abnormal}$. Figure 7.5 shows the pancreas shapes obtained for different values of t. When going from a healthy towards a pathological latent representation, we observed a shrinkage of the shape in the body for the generated pancreas.

7.2.4 Discussion and conclusion

We presented a method based on a VAE to learn a normative model of organ shape. We hypothesized that such a model could be learned from large databases of healthy subjects. The method was applied in the case of the pancreas, for which morphological changes can be a marker of disease.

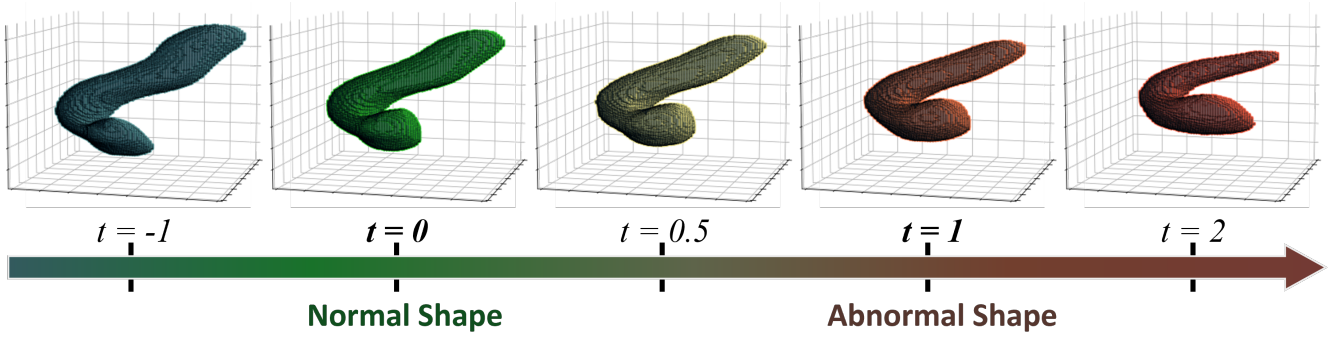


Figure 7.5 – Generated pancreas shapes. Pancreas shapes generated by decoding latent representations lying on the line of equation$(1-t) \times \mathbf{z}_{normal} + t \times \mathbf{z}_{abnormal}$.

We empirically observed that large training sets and latent dimensions were beneficial to the model in terms of AD performance. Our results also demonstrated that the model captured features that distinguished between normal and abnormal shapes in the latent space, as illustrated in Figure 7.4. From a quantitative point of view, we observed in the zero-shot case - *i.e.* without supervision - that the best model obtained an AUC score of $65.41 \pm 0.36\%$, which significantly outperformed a naive model classifying shapes based on their volume. In the few-shot experiments, we obtained a mean AUC score of 77.4% by training a SVM with only 8 healthy subjects and 15 pathological subjects. These findings highlight the discriminating properties of the latent normative model of pancreas shape estimated by our model. Moreover, classification performance reached up to 91.1% AUC and 83.2% balanced accuracy when training the classifier on 223 samples in a leave-one-out fashion. These results are in line with Liu et al. (2019b), where the authors reported a balanced accuracy of 85.2% on their private dataset. Yet, our approach differs from theirs by its paradigm. Instead of training a supervised model for joint shape representation and classification, we propose to learn a normative model of shape. The advantage of this approach is that it does not require different types of patients to be trained but solely a database of healthy subjects. Moreover, it can be used in an unsupervised manner (cf. zero-shot) or with few labeled data (cf. few-shot), with good performance in both cases.

Finally, we also showed that our framework could be used to study and visualize the morphological differences between the organ shape of different clinical groups, based on an exploration of the latent space. The anatomical changes observed in Figure 7.5 seemed to concur with clinical evidence as the shrinkage suggests partial parenchymal atrophy (Yamao et al., 2020). This hypothesis would require further medical evaluation, and could be the subject of a proper clinical validation.

One limitation of this study lies in the experimental setup, particularly in the normal *vs.* abnormal shape classification experiments where normal and abnormal shapes are sourced from distinct databases. While the use of binary inputs significantly minimizes bias related to the data origin, it is important to investigate potential implications further. Additionally, as our proposed method is grounded in a generative approach, there is interest in exploring discriminative approaches, specifically contrastive learning methods. The details of these limitations and preliminary results are elaborated in Appendix B.

7.3 Detection of Secondary Signs of Pancreatic Cancer using Non-Redundant Combination of Hand-Crafted and Deep Learning Radiomics

Building on the work on abnormal pancreatic shapes, we further studied the identification of additional secondary signs of pancreatic cancer using radiomics.

Abstract. We address the problem of learning Deep-Learning Radiomics (DLR) that are not redundant with Hand-Crafted Radiomics (HCR). To do so, we extract DLR features using a variational autoencoder while enforcing their independence with HCR features by minimizing their mutual information. The resulting DLR features can be combined with hand-crafted ones and leveraged by a classifier to predict signs of cancer. We illustrate our method on four secondary signs of pancreatic cancer and validate it on a large independent test set. Our results highlight the value of combining non-redundant DLR and HCR features, as evidenced by an improvement in the area under the curve compared to baseline methods that do not address redundancy or solely rely on HCR features.

7.3.1 Introduction

Computational methods in medical imaging hold the potential to support radiologists in the early diagnosis of cancer, either by detecting small-size abnormal neoplasms (Litjens et al., 2017), or even earlier in the disease course by recognizing indirect signs of malignancy. Such signs are usually subtle and organ-dependent, thus requiring a time-consuming and demanding clinical assessment. For example, in the case of pancreatic cancer, radiologists analyze the overall shape of the organ, check for fat replacement and note whether the pancreas shows atrophy and/or senile characteristics (Khoury et al., 2017; Matsuda, 2019; Miura et al., 2020). The identification of cancerous signs using automated tools can be based on radiomics, which are descriptors of texture and shape of a medical image, computed based on spatial relationships between voxels and their intensity distribution (Kumar et al., 2012; Lambin et al., 2012). Radiomics can be divided into two categories : (i) Hand-Crafted Radiomics (HCR), which are based on predefined mathematical formulas (Kumar et al., 2012; Lambin et al., 2012) ; (ii) Deep-Learning Radiomics (DLR), estimated using deep neural networks (Kumar et al., 2015; Shafiee et al., 2017), which may unveil additional complex relationships between voxels. HCR are generally extracted by open-source frameworks such as pyradiomics (Van Griethuysen et al., 2017). While such tools facilitate the standardization of the HCR, they only provide a limited number of predefined features. On the other hand, DLR features are typically extracted using either discriminative or generative models. Discriminative models frequently rely on one or multiple simple Convolutional Neural Network (CNN)s (Antropova et al., 2017; Chen et al., 2016; Huynh et al., 2016; Lao et al., 2017; Paul et al., 2016). To prevent overfitting, some methods extract DLR by using pretrained models trained on large datasets like ImageNet (Antropova et al., 2017; Huynh et al., 2016; Paul et al., 2016). The deep neural networks commonly employed for computing these DLR features consist of multiple layers, with each layer producing potential features as its output. As a result, the choice of the layers to retain varies, with each method employing different heuristics to identify them (Huynh et al., 2016; Paul et al., 2016). In the realm of generative models, autoencoder networks are widely used (Afshar et al., 2019). Autoencoders encode an image in a latent vector that is subsequently used to reconstruct the original image. This latent vector is considered to encapsulate the most descriptive features of the input image, making it a natural choice for representing the DLR (Kumar et al., 2015; Ravì et al., 2016).

The two types of radiomics are complementary : the computation of DLR is data-driven, which ensures that the extracted features are adapted to a specific problem or type of data. On the other hand, the predefined and generic definitions of HCR may make them less adapted for a given specific task, but favors generalization and interpretability. Therefore, it has been recently proposed to combine HCR with DLR, arguing that this approach could result in an improved feature set for predictive or prognostic

models (Afshar et al., 2019). As depicted in Figure 7.6, the literature reports two main approaches to perform this combination : decision-level methods that train separate classifiers on DLR and HCR before aggregating their predictions (Antropova et al., 2017; Huynh et al., 2016; Liu et al., 2017), and feature-level methods that concatenate the two types of radiomics in a single feature vector which is then leveraged by a classifier (Chen et al., 2016; Lao et al., 2017; Paul et al., 2016). These approaches extract HCR and DLR features independently, without guaranteeing complementarity between the two sets of features. As a result, the extracted DLR may be highly redundant with the HCR, limiting the value of their combination.

Given this context, we propose to extract DLR features that will complement the information already contained in the HCR. Our contributions are two-fold :

— A deep learning method, based on the Variational Autoencoder (VAE) framework (Kingma and Welling, 2014), that extracts non-redundant DLR features with respect to a predetermined set of HCR. This is achieved by minimizing the mutual information between the two types of radiomics during the training of the VAE. The resulting HCR and DLR features are leveraged to predict secondary signs of cancer.

— Validation of the proposed approach in the case of pancreatic cancer, using 2319 training and 1094 test subjects collected from 9 medical institutions with a split performed at the institution level. This is all the more important as most combination approaches have been solely evaluated in a cross-validation setting on mono-centric data (Antropova et al., 2017; Huynh et al., 2016; Liu et al., 2017).

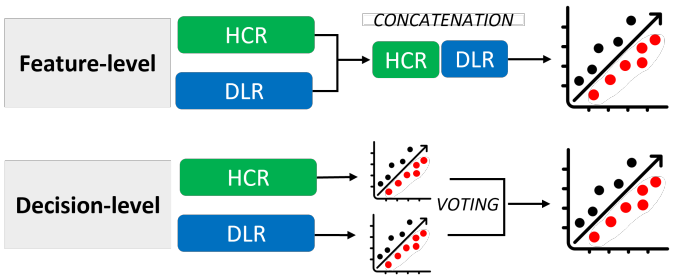


Figure 7.6 – Existing methods for combining HCR and DLR. Combinations at the feature-level concatenate the two types of radiomics in a single feature vector which is then leveraged by a classifier, while combinations at the decision-level train separate classifiers on HCR and DLR before aggregating their predictions. *DLR : Deep learning radiomics, HCR : Hand-crafted radiomics.*

7.3.2 Method

Our method, illustrated in Figure 7.7, relies on a generative model that recreates a 3D input image from the concatenation of HCR and DLR features. Feature extraction is done analytically for the HCR and through a VAE encoder for the DLR. Independence between the features is encouraged through the minimization of their mutual information, which is estimated by a discriminator relying on the density-ratio trick (Kim and Mnih, 2018). Finally, the resulting features are given to a classifier for cancer marker prediction.

7.3.2.1 Generative framework

Let $x \in \mathbb{R}^V$ be a 3D image acquired via a standard imaging technique, and $y \in \{0,1\}^V$ the corresponding binary segmentation mask of a given organ, with V the number of voxels. In order to focus on a specific organ and facilitate the extraction of specific features, we work on the masked image $x^* = x \times y$. We postulate the existence of a generative model enabling us to create an image x^* from a low-dimensional representation space $[h, d]$ where $h \in \mathrm{R}^{N_h}$ and $d \in \mathrm{R}^{N_d}$ represent the HCR and DLR features with N_h and N_d being the number of hand-crafted and deep features, respectively. Assuming that x^* follows an independent and identically distributed Gaussian distribution, and that f_θ is a non-linear function mapping

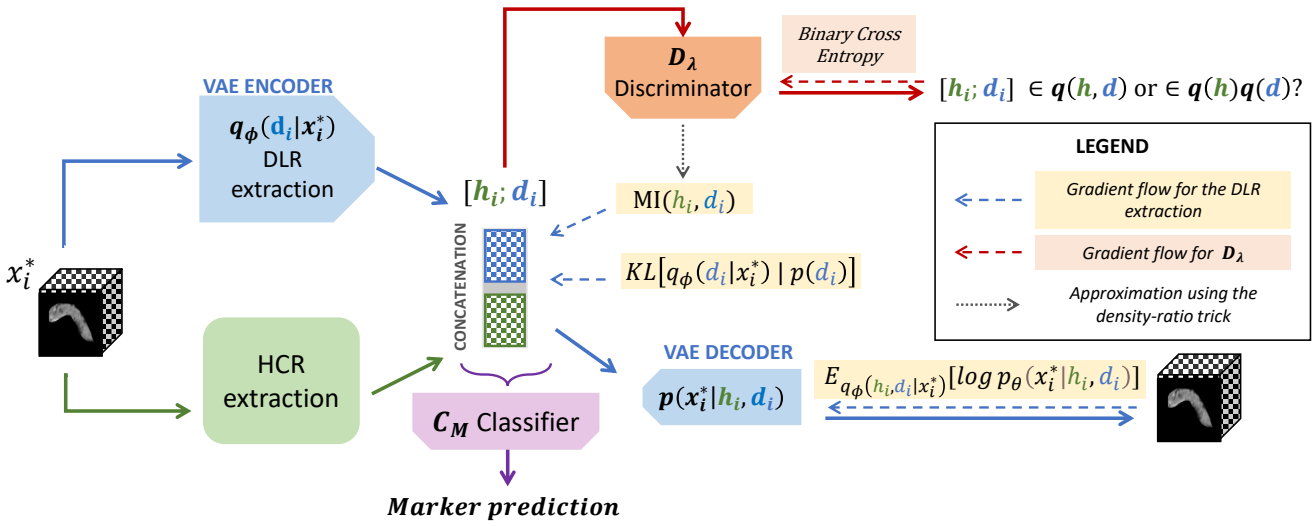


Figure 7.7 – Overview of our method. Starting from a masked image, HCR are calculated analytically, while DLR are extracted by the encoder of a VAE. These two types of radiomics are subsequently combined and given to the decoder for image reconstruction. The independence of HCR and DLR is enforced by the minimization of the MI. The latter is approximated by the density-ratio trick (Kim and Mnih, 2018), involving a discriminator $\mathcal{D}_\lambda$. Following the training of the VAE, a classifier $\mathcal{C}_M$ can be trained using both the HCR and DLR features to predict a specific marker of interest. *DLR : Deep learning radiomics, HCR : Hand-crafted radiomics, MI : Mutual information, VAE : Variational autoencoder.*

the concatenation of vectors $[h, d]$ to the masked image x^*, we hypothesize the following generative process :

$$p_\theta(x^* \mid y, h, d) = \prod_{v=1/y_v=1}^{V} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x_v^* - f_\theta([h,d])_v)^2}{2\sigma^2} \tag{7.3}$$

7.3.2.2 HCR and DLR features computation

We place ourselves within the VAE framework (Kingma and Welling, 2014), detailed in Section 3.2.3.5, and assume that $p(d)$ follows a Gaussian distribution with zero mean and identity covariance. HCR features are calculated analytically, while DLR features are computed by introducing the approximate posterior distribution $q_\phi(d \mid x^*)$. We hypothesize $q_\phi(d \mid x^*) \sim \mathcal{N}(\mu_\phi(x^*), \sigma_\phi^2(x^*)\mathbf{I})$, and maximize a lower bound of the marginal log-likelihood $\log p_\theta(x^* \mid y)$. We obtain the following loss function :

$$\mathcal{L}_{\text{VAE}} = -\mathbb{E}_{q_\phi(d|x^*)}[\log(p_\theta(x^* \mid y, h, d))] + KL[q_\phi(d \mid x^*) \mid p(d)] \tag{7.4}$$

where KL refers to the Kullback-Leibler divergence.

7.3.2.3 Mutual Information Minimization

To promote the independence between HCR and DLR features, we propose to minimize their Mutual Information (MI), expressed here as $KL[q(h,d) \mid q(h)q(d)]$, where $q(h,d)$ represents the joint distribution of the DLR and HCR features, and $q(h)q(d)$ the product of their marginal distributions. These terms involve mixtures with a large number of components, making them intractable. Moreover, obtaining the direct Monte Carlo estimate necessitates processing the entire dataset in a single pass. Thus, we sample from these distributions to compute the MI : to sample from $q(h,d)$, we randomly choose an image x_i^*, extract its HCR features h_i as well as its DLR features d_i using the VAE encoder, and concatenate them. Samples from $q(h)q(d)$ are obtained by concatenating vectors h_k and d_j with $k \neq j$. Finally, to compute the MI, we need to compute the density-ratio between $q(h,d)$ and $q(h)q(d)$. To do so, we resort to the density-ratio trick (Kim and Mnih, 2018), which consists in introducing a discriminator $\mathcal{D}_\lambda([h,d])$ able to discriminate

between samples from $q(h,d)$ and samples from $q(h)q(d)$. Thus, we obtain :

$$KL[q(h,d) \mid q(h)q(d)] = \mathbb{E}_{q(h,d)}\left[\log \frac{q(h,d)}{q(h)q(d)}\right] \approx \sum_i \text{ReLU}\left(\left[\log \frac{\mathcal{D}_\lambda(h_i,d_i)}{1-\mathcal{D}_\lambda(h_i,d_i)}\right]\right). \tag{7.5}$$

where the ReLU function forces the estimate of the MI to be positive, which prevents from back-propagating wrong estimates of the density-ratio.

7.3.2.4 Estimating the Mutual Information

The MI is estimated following the density-ratio trick (Kim and Mnih, 2018) which requires to train a discriminator $\mathcal{D}_\lambda$ predicting whether concatenated radiomics vectors $[h,d]$ come from $q(h,d)$ or $q(h)q(d)$. Samples for training $\mathcal{D}_\lambda$ are obtained following the procedure shown in Figure 7.8. In practice, $\mathcal{D}_\lambda$ is modeled as a 2-layer Multi Layer Perceptron with ReLu activation, which is trained by minimizing a binary cross-entropy loss term. Once the discriminator is trained, the MI between HCR and DLR features can be approximated as follows :

$$\text{MI}(h,d) = \mathbb{E}_{q(h,d)}\left[\log \frac{q(h,d)}{q(h)q(d)}\right] \approx \sum_i \text{ReLU}\left(\left[\log \frac{\mathcal{D}_\lambda(h_i,d_i)}{1-\mathcal{D}_\lambda(h_i,d_i)}\right]\right). \tag{7.6}$$

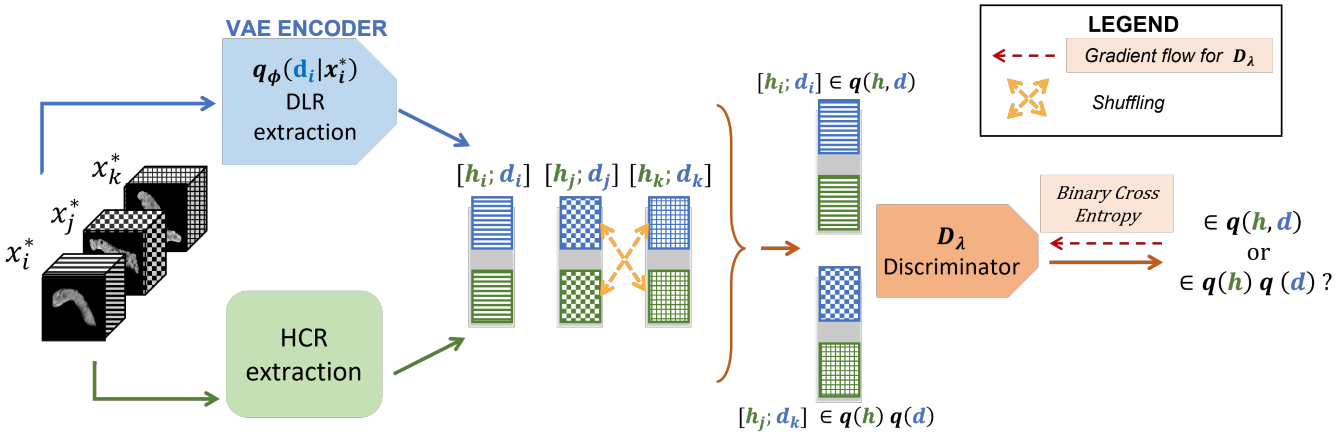


Figure 7.8 – Training the discriminator $\mathcal{D}_\lambda$. Given three different input images x_i^*, x_j^* and x_k^*, the corresponding HCR and DLR features are computed : h_j, h_j, h_k and d_i, d_j, d_k. Samples from $q(h,d)$ are obtained by concatenating features of a same image (h_i and d_i for instance), while samples from $q(h)q(d)$ are obtained by concatenating h_k and d_j with $k \neq j$. *DLR : Deep learning radiomics, HCR : Hand-crafted radiomics.*

7.3.2.5 Optimization

The final loss function is :

$$\mathcal{L} = \mathcal{L}_{\text{VAE}} + \kappa KL[q(h,d) \mid q(h)q(d)] \tag{7.7}$$

This loss function is composed of two terms : the left-hand term, which is the common VAE loss function and promotes the reconstruction of the masked image while regularizing the approximate posterior distribution ; and the right-hand term which minimizes the MI between $q(h,d)$ and $q(h)q(d)$, and enforces the extraction of DLR features which are not redundant with HCR features. The importance of the MI in the loss function is weighted by κ, which we empirically set to 1 according to the results shown in Table 7.6. To ensure that the density-ratio is well-estimated, as explained in (Kim and Mnih, 2018), we opt for an alternate optimization scheme between the VAE model and the discriminator $\mathcal{D}_\lambda$: every 5 epochs, we freeze the optimization of the VAE, train the discriminator for 150 epochs, and continue the optimization of the VAE model.

	$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 1$	$\kappa = 10$
General Shape	70.44±0.07	70.01±0.07	70.07±0.07	**71.03±0.07**
Atrophy	80.82±0.05	81.43±0.06	**82.57±0.06**	80.77±0.06
Fat Replacement	69.52±0.08	70.5±0.07	**71.05±0.07**	68.65±0.08
Senility	**73.14±0.08**	72.36±0.08	72.44±0.07	72.38±0.08

Table 7.6 – Cancer marker prediction scores for different values of κ. For each experiment, we report the means and standard deviations of the area under the curve (in %) obtained by bootstrapping with 10000 repetitions. For each line, best result is in bold.

7.3.2.6 Early cancer markers prediction

Once the VAE model is trained, DLR can be extracted and leveraged to predict cancer markers. We propose to train, for each marker of interest, a classifier $\mathcal{C}_M$ based on the concatenation of HCR and DLR extracted by our model. Unlike VAE training, which is unsupervised and task-agnostic, $\mathcal{C}_M$ training is supervised and specific to a cancer marker.

7.3.3 Experiments

We illustrate our method on the pancreas, for which we aim to predict four secondary signs of abnormality that manifest prior to the onset of visible lesions :

(i) *Abnormal shape :* Changes in the shape of the pancreas can be associated with pancreatic cancer as the tumor growth can lead to various structural changes in the pancreas (Liu et al., 2019b; Vétil et al., 2022a) ;

(ii) *Atrophy :* Pancreatic atrophy may signal pancreatic cancer (Miura et al., 2020) and can indicate small isodense lesions (Yamao et al., 2020) ;

(iii) *Fat replacement :* Fat replacement is characterized by the accumulation of fat within the pancreas and is associated with various metabolic diseases, pancreatitis, pancreatic cancer, and precancer (Khoury et al., 2017; Majumder et al., 2017; Miura et al., 2020). While this mainly modifies the texture, severe fat replacement can also affect the shape by inducing lobulated margins ;

(iv) *Senility :* Anatomical changes in the pancreas, such as pancreatic atrophy, fatty replacement and fibrosis have been documented in elderly individuals and increase the susceptibility of individuals to pancreatic cancer (Khoury et al., 2017; Matsuda, 2019).

These early signs are illustrated in Figure 7.9, and are further detailed in Section 2.9.

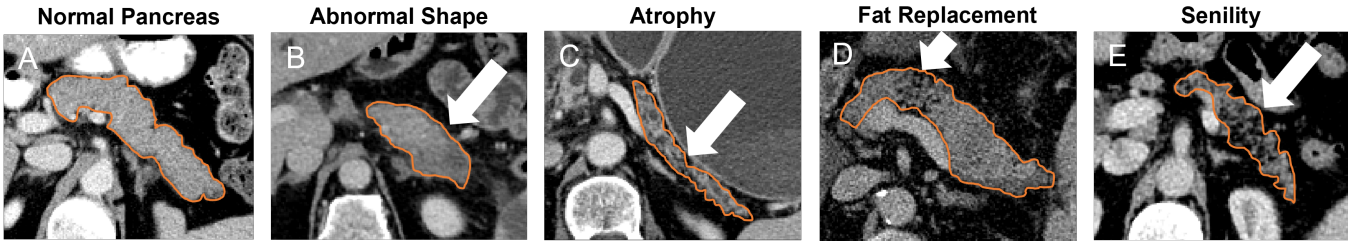


Figure 7.9 – Portal CT scans showing secondary signs of pancreatic cancer. Pancreas are delineated in orange. (A) shows a normal pancreas. White arrows indicate an abnormal enlarged tail (B), a parenchymal atrophy (C), fat replacement in the neck of the pancreas (D) and senile characteristics (E). *CT : Computed tomography.*

7.3.3.1 Dataset

Data were obtained from our public and private cohorts and split at the institution-level into two independent datasets $\mathcal{D}^{Train}$ and $\mathcal{D}^{Test}$. $\mathcal{D}^{Train}$ contained 2319 portal Computed Tomography (CT) scans from $\mathcal{D}_{BR,1}$, $\mathcal{D}_{FR,1}$, $\mathcal{D}_{FR,2}$, $\mathcal{D}_{US,2}$, $\mathcal{D}_{CALV}$, and $\mathcal{D}_{CABO}$, while $\mathcal{D}^{Test}$ contained 1095 abdominal portal CT scans from $\mathcal{D}_{US,1}$, $\mathcal{D}_{NIH}$, and $\mathcal{D}_{MSD}$. The reference labels regarding the secondary signs previously

described were obtained based on the assessment of the CT scan by a pool of 11 radiologists. Details about the datasets and the annotation process are given in Chapter 4. At the time of this work, reference labels were available for 676 cases of $\mathcal{D}^{Train}$ and all the subjects from $\mathcal{D}^{Test}$.

7.3.3.2 Preprocessing

For all the subjects, pancreas segmentation masks were obtained using a segmentation model derived from the nnU-Net (Isensee et al. (2021a), detailed in Section 3.2.3.5) and manually reviewed by radiologists. The CT images and corresponding masks were resampled to $1 \times 1 \times 2$ mm^3 in the (x, y, z) directions, and centered in a volume of size $192 \times 128 \times 64$ voxels. Images intensities were clipped to the $[0.5, 99.5]$ percentiles and standardized based on the percentiles, mean and standard deviation of the pancreas intensities in $\mathcal{D}^{Train}$.

7.3.3.3 Extracting HCR and DLR

32 HCR features were extracted using the pyradiomics library (Van Griethuysen et al., 2017), focusing exclusively on shape and first-order intensity features (see Appendix C for the comprehensive list). Complementary DLR features were extracted using the VAE model of Section 7.3.2. The architecture followed the U-Net (Ronneberger et al. (2015), detailed in Section 3.2.3.5) encoder-decoder scheme without skip connections to align with the VAE framework. The number of convolutional layers and the convolutional blocks were automatically inferred thanks to the nnU-Net self-configuring procedure (Isensee et al., 2021a), resulting in $1,110,240$ trainable parameters. The VAE was trained on $\mathcal{D}^{Train}$ for 1000 epochs with a batch size of size 32. Every five epochs, the VAE was frozen and the discriminator $\mathcal{D}_\lambda$ was trained for 150 epochs with a batch size equal to the total training dataset. The VAE and $\mathcal{D}_\lambda$ were optimized using two independent Adam optimizers with a learning rate of 10^{-3}. The dimension of DLR features d was set to 32, resulting in a final latent space dimension for the VAE of 64. Data augmentation consisting of rotation and cropping was applied during training.

7.3.3.4 Predicting early cancer markers

For each marker, a logistic regression was trained based on the concatenation of HCR and DLR features extracted from the subjects in $\mathcal{D}^{Train}$ for whom reference labels were available. The logistic regression was regularized using $\mathcal{L}_2$ penalty, with a default regularization coefficient of 1. Final predictions for $\mathcal{D}^{Test}$ were derived by ensembling models obtained through a four-fold cross-validation setup.

7.3.4 Results

7.3.4.1 Quantitative results

To demonstrate the usefulness of extracting DLR with MI minimization, two VAEs were trained. Both followed the same procedure (detailed in Figure 7.7) but differed only in the presence or absence of the MI minimization term in their loss function. Then, several logistic regression models with different inputs were trained in order to assess the effect of combining HCR and DLR features. In total, the following experiments were run :

— **HCR only :** H_{32} **and** H_{64}. These two experiments use the 32 basic HCR features described in Appendix C, and H_{64} uses a further 32 HCR gray-level features calculated by the pyradiomics library (Van Griethuysen et al., 2017) and selected by recursive feature elimination.

— **DLR only :** $\mathrm{D}_{32}^{\mathrm{MI}}$ **and** D_{32}. 32 DLR features extracted by a VAE with and without MI minimization, respectively ;

— **HCR + DLR :** $\mathrm{HD}_{64}^{\mathrm{MI}}$ **and** HD_{64}. 32 basic HCR features + 32 DLR features extracted by a VAE with and without MI minimization, respectively.

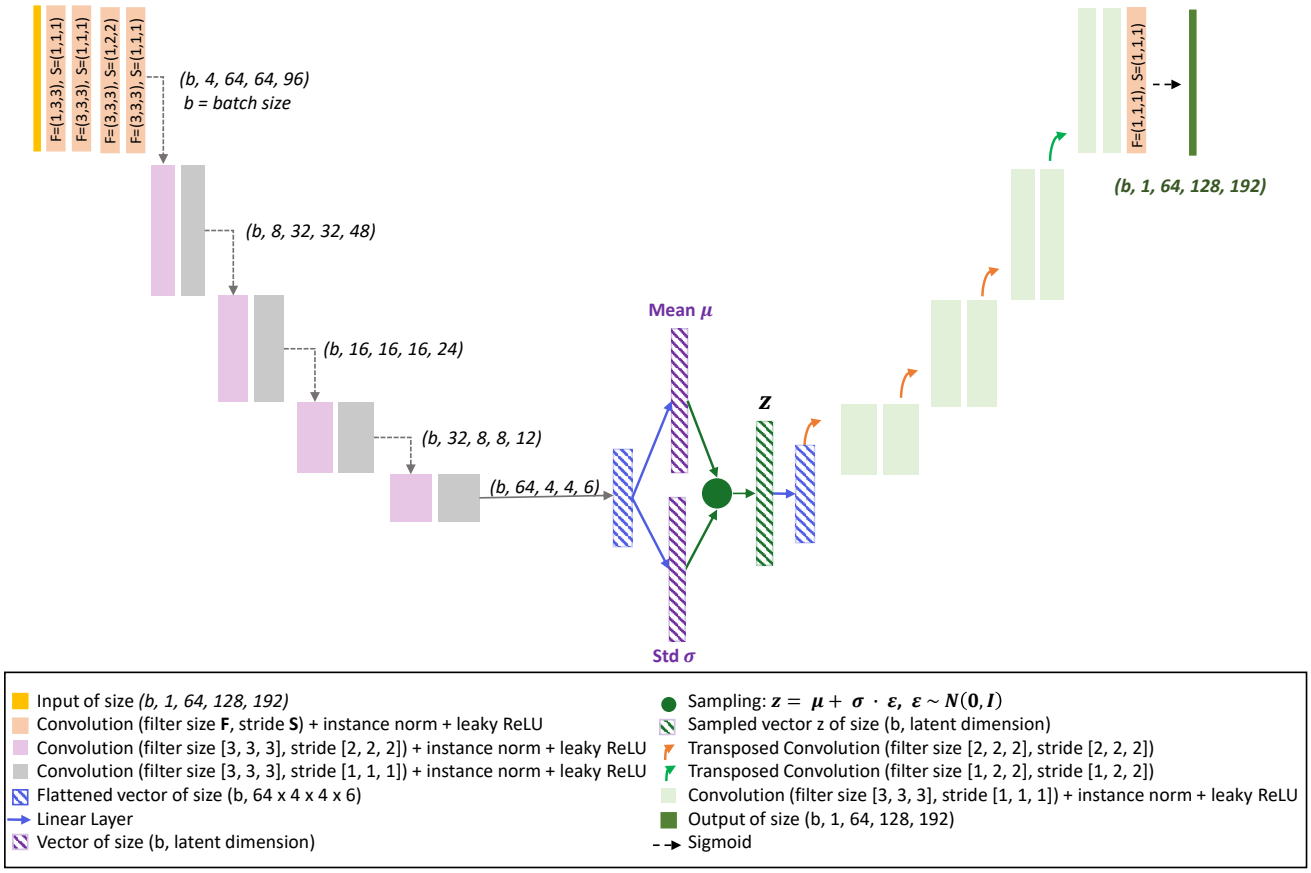


Figure 7.10 – Architecture of the proposed variational autoencoder.

Thus, the logistic regressions of H_{32}, D_{32} and D_{32}^{MI} used vectors of size 32, while those of H_{64}, HD_{64} and HD_{64}^{MI} used vectors of size 64. Prediction results for each of the four cancer markers are presented in Table 7.7.

	HCR only		**DLR only**		**HCR + DLR**	
	H_{32}	H_{64}	D_{32}	D_{32}^{MI}	HD_{64}	HD_{64}^{MI}
Abnormal Shape	68.38±0.07	68.11±0.07	67.66±0.07	**72.41±0.07**	71.2±0.07	70.07±0.07
Atrophy	81.05±0.06	81.57±0.05	74.08±0.07	79.08±0.06	80.82±0.06	**82.57±0.06**
Fat Replacement	70.55±0.07	69.78±0.08	65.96±0.08	65.74±0.07	69.28±0.08	**71.05±0.07**
Senility	71.63±0.08	70.21±0.08	70.18±0.07	69.1±0.08	72.28±0.08	**72.44±0.07**
δ w.r.t H_{32}	-	-0.48±0.07	-3.43±0.07	-1.32±0.07	0.49±0.07	**1.13±0.07**

Table 7.7 – Pancreatic cancer marker prediction. For each experiment, we report the means and standard deviations of the AUC (in %) obtained by bootstrapping with 10000 repetitions. For each line, first and second best results are in bold and underlined, respectively. The last row shows the difference in AUC compared with H_{32}, averaged over the different markers. *AUC : Area under the curve, DLR : Deep learning radiomics, HCR : Hand-crafted radiomics.*

The comparison between H_{32} and H_{64} showed that adding 32 gray-level HCR features was not beneficial as results were similar, or even decreased : for instance, for senility, the Area Under the Curve (AUC) went from 71.63% (H_{32}) to 70.21% (H_{64}). On average, the AUC of H_{64} lost -0.48 points compared with H_{32}. These experiments demonstrated the power of the 32 basic HCR features, and the need to find complementary features that would add value.

Then, for almost all markers, H_{32} outperformed D_{32} and D_{32}^{MI}, meaning that no VAE, whether trained with or without MI minimization, managed to automatically extract 32 DLR features as informative as the 32 basic HCR features used by H_{32}. For texture-related markers, such as fat replacement and senility,

MI minimization did not produce clear differences. On the other hand, on shape-related markers, the DLR features learned by D_{32}^{MI} were shown to be more relevant than those learned by D_{32} with a basic VAE. Thus, on average, DLR features were better when extracted by a VAE trained with MI minimization, but still proved less informative than HCR features.

Finally, experiments HD_{64} and HD_{64}^{MI} showed that combining the two types of radiomics is beneficial since the average AUC gained 0.49 (HD_{64}) and 1.13% (HD_{64}^{MI}) compared to H_{32}. Yet, results demonstrated that minimizing the redundancy produced the best results compared with all other approaches. Indeed, in HD_{64}, adding 32 DLR features produced variable results depending on the markers : compared to H_{32}, the AUC increased by a maximum of 2.82% for abnormal shape prediction, and dropped by a maximum of 1.27% for predicting fat replacement. On the other hand, HD_{64}^{MI} outperformed H_{32} on all prediction problems, meaning that the non-redundant DLR features systematically provided useful information.

7.3.4.2 Influence of the latent space

To explore the influence of the latent space dimension on the prediction performance, we replicated the HD_{64}^{MI} experiment with increasing size L of the latent space, and reported prediction results in Table 7.8. Table 7.8 shows that increasing the latent space size resulted in lower classification performance. Specifically, a latent space size of 32 provided the most relevant DLR features.

	$L = 32$	$L = 64$	$L = 256$	$L = 512$	$L = 1024$	$L = 2048$
Abnormal Shape	**70.07±0.07**	69.02±0.07	68.87±0.07	69.91±0.07	69.33±0.07	68.68±0.07
Atrophy	82.57±0.06	82.28±0.05	81.77±0.06	**82.68±0.05**	80.9±0.06	80.21±0.06
Fat Replacement	**71.05±0.07**	70.91±0.07	70.23±0.08	70.45±0.08	69.55±0.07	68.96±0.08
Senility	**72.44±0.07**	72.02±0.07	70.38±0.08	71.65±0.08	72.03±0.07	69.6±0.08

Table 7.8 – Pancreatic cancer marker prediction with varying latent space size. For each experiment, a VAE with MI minimization and latent space size L was trained. Predictions were obtained after training logistic regressions on 32 basic hand-crafted radiomics features + L deep learning radiomics features extracted by a VAE with MI minimization. We report the means and standard deviations of the AUC (in %) obtained on the test set by bootstrapping with 10000 repetitions. For each line, first best results are in bold. *AUC : Area under the curve, MI : Mutual information, VAE : Variational autoencoder.*

7.3.4.3 Qualitative results

To visualize the effect of the extracted DLR features, we looked at the absolute value of the logistic regression weights for D_{32} and D_{32}^{MI} in two ways. In Figure 7.11-A, the absolute value of these coefficients are displayed. The higher the absolute value of the coefficient, the higher its importance in the logistic regression prediction. When the MI was not minimized, HCR features had stronger importance than DLR ones. On the other hand, when we encouraged the independence between the two types of features through MI minimization, the contribution of DLR features to the prediction increased. Figure 7.11-B shows the number of DLR features among the k features with highest importance, for increasing values of k. HD_{64}^{MI} and HD_{64} are shown in blue and orange, respectively. In addition, two extreme scenarios are shown : one where the logistic regression is predominantly influenced by the DLR features (in green), and another one where the logistic regression is primarily driven by the HCR features (in red). We can see that the blue curve approached the green curve, meaning that DLR features from HD_{64}^{MI} contributed more to the outcome prediction. When the MI was not minimized, DLR features had less influence on the predictions as the orange curve approached the scenario in which DLR would be ignored.

7.3.4.4 Reconstruction performance

To explore the reconstruction performance of the VAE, we computed the average L_2 error per voxel between the original test images and their corresponding reconstructions. Upon applying nnU-Net automatic intensity normalization procedure (Isensee et al., 2021a), voxel intensities were observed to range from -3 to 2.3. Specifically, we employed a VAE with a latent space dimension of $L = 32$ and MI minimization during training. The resulting reconstruction error was found to be $(4.4 \pm 1.4) \times 10^{-3}$, which was comparable

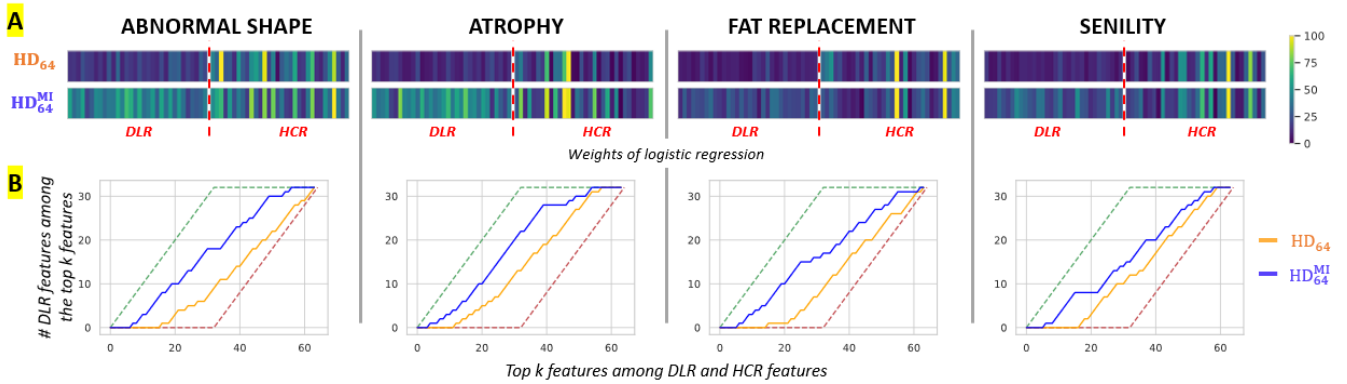


Figure 7.11 – Qualitative assessment of the DLR and HCR features through the coefficients of the logistic regressions. A : Absolute value of the coefficients of the logistic regressions. We plot, for each logistic regression corresponding to one marker, the absolute value of the coefficient for each of the 64 features. The first 32 features corresponded to DLR, while the 32 remaining features corresponded to HCR. **B : Number of DLR features among the top k features.** Dashed lines represent the extreme scenarios in which all 32 DLR are more informative than all 32 HCR (green), or all 32 HCR are more informative than all 32 DLR (red). *DLR : Deep learning radiomics, HCR : Hand-crafted radiomics.*

to the L_2 error obtained from a VAE trained without MI minimization, amounting to $(4.1 \pm 1.4) \times 10^{-3}$. These observations suggest that the introduction of MI minimization did not significantly impact the quality of the reconstructed images, neither resulting in deterioration nor improvement. Additionally, Table 7.9 further explores the relationship between reconstruction performance and latent space sizes, demonstrating that increasing the latent space size did not have a discernible effect on the quality of the reconstructions.

	$L = 32$	$L = 64$	$L = 256$	$L = 512$	$L = 1024$	$L = 2048$
L_2 error $\times 10^3$	4.4±1.4	4.4±1.4	4.4±1.4	4.4±1.4	4.3±1.4	4.3±1.5

Table 7.9 – Reconstruction performance with varying latent space sizes. For each experiment, a variational autoencoder with mutual information minimization and latent space size L was trained. We report the L_2 error per voxel between the original image and its reconstruction, with voxel intensities varying in $[-3, 2.3]$.

7.3.5 Discussion and conclusion

We presented a method to learn DLR features that are not redundant with HCR ones. The method was based on the well-known VAE framework (Kingma and Welling, 2014) that extracted DLR features from masked images in an unsupervised manner. The complementarity between the two types of radiomics features was enforced by minimizing their MI, and the resulting features were used to train classifiers predicting different cancer markers. Experiments in the case of four secondary signs of pancreatic cancer indicated that our method increased prediction performance with respect to two state-of-the-art approaches. These findings suggest that our approach holds potential to improve patient survival outcomes. Qualitative results confirmed the advantages of minimizing the MI during training, as it resulted in the generation of DLR features that were complementary to HCR features and more prominently used for marker prediction. These results were obtained on a large and independent test set, which is particularly important as radiomics models require robust validation strategies to ensure their generalization and reproducibility when applied to new datasets (Aerts et al., 2014). With this in mind, it might be interesting to further encourage this feature efficiency by imposing independence between the DLR features themselves. Another research avenue could be to simplify the proposed pipeline by developing an end-to-end network capable of performing both feature extraction and classification tasks within a unified framework. Achieving this objective would necessitate the simultaneous training of the feature extractor and multiple sub-networks for each classification task. However, this approach might pose challenges in terms of training complexity, particularly due to the presence of substantial class imbalances across the various classification tasks. Alternatively, another possibility is to train an end-to-end CNN. Although more direct in nature, this approach would entail the training of a separate CNN for each question, which could be computationally heavier

compared to the calibration of a logistic regression based on a single feature extractor, as suggested in our current work. Future studies should also address the interpretability of the extracted DLR features, as this aspect was not covered in the present work.

Chapitre 8

Conclusion and Perspectives

8.1 Preface

This last chapter aims to summarize the contributions made in this thesis and to outline potential avenues for future research in this area. The future directions are categorized into clinical, technical, and industrial perspectives.

8.2 Main contributions

The purpose of this thesis was to propose Artificial Intelligence (AI)-based methods designed to assist radiologists in identifying pancreatic cancer on portal Computed Tomography (CT) scans. The primary objectives included the systematic identification of pancreatic lesions to prevent missed or delayed diagnoses, and the pursuit of early diagnosis by detecting secondary signs that could be visible before the clear appearance of pancreatic lesions. These objectives aimed to facilitate timely diagnosis, increase eligibility for surgical intervention—currently the primary therapeutic option—and ultimately enhance the patients' survival.

The pursuit of these objectives began with the creation of a database enabling the development of efficient and robust AI methods for pancreatic cancer analysis. Extensive effort was put into creating $\mathcal{D}_{CAPA}$, a large and multi-centric database with diverse pathologies, detailed in Chapter 4. This resulted in the establishment of one of the largest annotated databases for pancreatic cancer research, currently containing over $2,800$ portal CT scans. This database is planned for continuous expansion to serve Guerbet's research and development needs.

Then, an important methodological step was the development of an automatic pancreas segmentation algorithm on portal CT scans. While pancreas segmentation was not an objective *per se*, achieving precise segmentation was crucial as it provided a Region of Interest (ROI) for addressing more complex tasks. Segmentation of the pancreas posed challenges due to its elongated shape and low contrast at its extremities. In particular, ensuring accurate segmentation at these extremities was crucial as lesions could be situated within these areas. The proposed method, detailed in Chapter 5, locally adjusted the sensitivity of the segmentation by incorporating geometric priors. While the method was showcased on the pancreas, its general formulation renders it applicable to any existing network and any organ presenting similar segmentation challenges. The results demonstrated that geometrical priors effectively corrected the missed extremity errors made by a state-of-the-art method, while also preserving the overall quality of segmentation. However, this approach was constrained by the introduction of hyper-parameters, which, although guided by anatomical knowledge, can lead to time-consuming search processes.

With these fundamental steps addressed, the thesis objectives shifted towards pancreatic cancer detection on portal CT scans.

Chapter 6 presented a method allowing for the identification of pancreatic lesions and Main Pancreatic Duct (MPD) dilatation, crucial indicators of pancreatic cancer. As these two signs rely on clearly identifiable structures (the lesion and the MPD), the proposed method started with the segmentation of the pancreas, the lesion and the MPD. Then, features were extracted from the predicted segmentations and leveraged to predict the presence of a lesion and the dilatation of the MPD. For this matter, a state-of-the-art segmentation model was trained on 2134 patients from multi-centric databases covering multiple geographical regions. The model was thoroughly evaluated on 756 patients from independent cohorts, reaching an Area Under the Curve (AUC) of 98% for the detection of lesions, with a sensitivity of 94%. Subgroup analyses demonstrated similar sensitivities across different lesion types and characteristics. Regarding the detection of MPD dilatation, the AUC was 97%. This method represents the first AI-driven approach for detecting MPD dilatation, and demonstrated that enhancing lesion detection is achievable through the incorporation of explicit features derived from the MPD. However, this method focused solely on detection and did not address the task of characterization, namely identifying the lesion type or determining the malignant nature of the MPD dilatation. Additionally, while the method underwent evaluation on a large external cohort, this validation cohort lacked a sufficient number of cases with isodense or small lesions, which are the lesions frequently overlooked during diagnosis. These limitations will be elaborated upon in the subsequent section focusing on perspectives.

Continuing towards early diagnosis, Chapter 7 aimed at detecting secondary signs of pancreatic cancer for which only image-level labels were available. Different strategies were explored.

The first approach proposed to address the detection of abnormal pancreatic shapes. To this end, a method leveraged large databases of healthy pancreases in order to learn a normative model of healthy shapes, hence facilitating the identification of abnormal shapes. Volumetric segmentation masks were embedded into a common probabilistic shape space using a Variational Autoencoder (VAE). The resulting

latent shape representations were leveraged to derive zero-shot and few-shot methods for abnormal shape detection. Trained on 1200 healthy shapes and tested on 224 shapes with mixed conditions, the abnormal pancreas detection AUC reached up to 65.41% in the zero-shot configuration, and 78.97% in the few-shot configuration with as few as 15 abnormal examples, outperforming baseline approaches. Besides anomaly detection, the resulting method could be used to study and visualize the morphological differences between the organ shapes of different clinical groups. In addition to the suggested generative approach, alternative discriminative methods could be explored in order to build similar normative models. Particularly, contrastive learning methods could be employed, as elaborated in the following Section 8.3.2.1 on perspectives.

The second approach went further and aimed to detect several secondary signs of pancreatic cancer using radiomic features. To this end, a novel method closing the gap between Deep-Learning Radiomics (DLR) and Hand-Crafted Radiomics (HCR) was proposed : DLR features were extracted using a VAE, while their independence with HCR features was enforced by minimizing their mutual information. The resulting DLR features were combined with the HCR ones, and leveraged by a classifier to predict four secondary signs of pancreatic cancer : abnormal shape, atrophy, senility and fat replacement. Trained on 2319 cases and tested on 1094 subjects from independent institutions, this method demonstrated improved diagnostic accuracy compared to baseline methods that did not address independence or solely relied on HCR features. In addition, qualitative results confirmed that the generated DLR were complementary to HCR features and more prominently used for the subsequent prediction. Despite these quantitative and qualitative assessments, the interpretability of these DLR needs to be addressed to improve the clinical utility of this method.

Both approaches proved efficient, requiring training only a single VAE, and the resulting latent space was leveraged for multiple downstream tasks. Like the segmentation method proposed in Chapter 5, these approaches were illustrated in the context of pancreatic cases but hold broader applicability. Yet, their common limitation arises from the use of annotations that were obtained through visual assessment, without quantitative measures. As a result, significant inter- and intra-annotator variability in the annotations can complicate the training and validation of these methods. Further discussion of this topic will be presented in Section 8.3.1.1.

8.3 Perspectives

We present here several general research directions related to the thesis topic. These research avenues are categorized into clinical, technical, and industrial domains. The clinical section proposes several research topics that may be pertinent from a clinical perspective. Technical perspectives emphasize three primary areas, including the use of contrastive learning methods, exploration of the use of the arterial phase, and development of multi-modality algorithms. Lastly, the industrial perspectives underscore essential considerations for the integration of the proposed AI-based methods into clinical practice.

8.3.1 Clinical perspectives

8.3.1.1 Further study of secondary signs

The detection of secondary signs faces limitations due to the challenges and constraints associated with annotations. As detailed in Chapter 4, all secondary signs were annotated using image-level labels, lacking insight into which radiological feature indicated their absence or presence (with the exception of MPD dilatation, for which manual segmentation of the MPD provided anatomical information). Therefore, since these labels were assigned based on visual assessment, they lack quantitative measures (again excluding MPD dilatation, which relies on the MPD diameter) and are subject to significant inter and intra-annotator variability. This variability in the image-level annotation process could not be measured, as each case was annotated by a single radiologist. Given this context, several avenues for improvement can be considered. First, conducting an annotation process involving multiple experts to obtain consensus labels could enhance the reliability of the annotations. However, due to the associated costs, this approach may be feasible only for a subset of cases. Secondly, the accurate pancreas segmentations at hand could be used in clinical

studies to derive automated measurements. These measurements could help to establish guidelines for characterizing atrophy or fatty infiltration, for example.

8.3.1.2 Towards more precise diagnostic

Characterization represents a pivotal stage in refining the accurate and timely diagnosis of pancreatic cancer. This characterization can concern lesions and secondary signs.

Lesion characterization The first step towards a more precise diagnosis consists in characterizing the lesions. While lesion detection is necessary to prevent missing cancerous lesions, the critical aspect is determining the precise subtype of each lesion. This determination is essential not only for ensuring appropriate patient care and follow-up but also for generating meaningful output from a computer-aided detection system. The significance of differentiation becomes evident in the varying levels of concern associated with different subtypes : for example, the alert level for Pancreatic Ductal Adenocarcinoma (PDAC) should not be equivalent to that for a benign serous cystadenoma. Accurately diagnosing pancreatic lesions poses a significant challenge for radiologists, especially considering the complexity introduced by various pancreatic cancer mimics. Indeed, conditions like pancreatitis, Pancreatic Neuroendocrine Tumor (PNET)s, solid pseudopapillary neoplasms, metastases, and lymphoma can mimic PDAC, complicating the differentiation process due to shared imaging features. Consequently, radiologists frequently resort to biopsies, despite their notable limitations including insufficient sensitivity (84%) and a relatively high morbidity rate (3%) (Zins, 2023). Even in cases where biopsies are negative for malignancy, the recommendation for resection persists due to sampling error concerns (Santo and Bar-Yishay, 2017). Consequently, there is a tendency towards excessive resections (Fitzgerald et al., 2003; Sachs et al., 2009). These challenges highlight the complexity of achieving a precise diagnosis and underscore the importance of automated methods to enhance diagnostic confidence (Barat et al., 2021). Therefore, addressing the automated characterization of lesions emerges as a crucial step in developing a clinically applicable method. Initially, lesion characterization could be conducted at broader levels (solid vs. cystic, or benign vs. malignant), although the ultimate goal would be the development of a method capable of identifying precise lesion subtypes.

Secondary signs characterization Similarly, there is a need to delve deeper into the characterization of secondary signs. While the current focus is on their detection, future work should shift towards their differentiation or quantification. For instance, quantifying the degree of fatty infiltration (Hoogenboom et al., 2021; Kim et al., 2014), assessing the malignant nature of the MPD dilatation (Chen et al., 2020b; Kim et al., 2017), and characterizing parenchymal atrophy (Yamao et al., 2020) could provide additional insights. This characterization is crucial for distinguishing between pathological and ageing pancreases, which may exhibit similar features due to physiological evolutions (Löhr et al., 2018).

To tackle these two research topics aiming at refining diagnostics, a crucial step involves augmenting the existing database to include a broader spectrum of cases. As the database currently contains approximately 60% and 80% of PDACs and solid lesions, respectively (see Table 4.4), efforts should be directed towards incorporating more diverse cases in order to distinguish between various lesion subtypes and potential mimics. Similarly, examples of both benign and malignant cases for a given secondary sign should be added. However, even with the addition of these cases, the database may still exhibit imbalance across certain classes. To address this imbalance, few-shot learning methods, as discussed in Chapter 7, can be employed, enabling models to be trained with minimal examples of a class. Moreover, to train models that could characterize different lesions or signs, pertinent features could be extracted using techniques such as radiomics or contrastive learning (elaborated further below, in the technical perspectives Section 8.3.2.1).

8.3.1.3 Clinical validation

To further validate the proposed methods for early diagnosis, an essential step entails conducting a clinical study. This necessitates the establishment of a specific validation database comprising distinct cohorts, including :

— **Patients with small lesions** : The inclusion of such cases is imperative for validating the effectiveness of the methods in detecting early-stage lesions. The current Guerbet's dataset inadequately represents this cohort, as only 55 out of 2917 cases (1.9%) feature subcentrimetric lesions (see Section 4.4).

— **Longitudinal cohorts** : Incorporating follow-up examinations with pre-diagnostic images would be essential to assess how early the cancer can be reliably detected.

— **Real-life cohorts** : Accessing cohorts with a low prevalence of pancreatic cancer ($< 1\%$), as done by Cao et al. (2023) and Korfiatis et al. (2023), is critical for affirming the practical applicability of the proposed methods within routine clinical contexts.

The collection of these cohorts presents significant challenges due to the specific nature of the cases required. Nonetheless, it is indispensable for extensive validation of the proposed algorithms in detecting early lesions.

8.3.1.4 Leveraging AI for pancreatic surgery

Ultimately, AI could be integrated into surgical workflows, serving both preoperative and postoperative purposes (Schlanger et al., 2022). Potential applications encompass the assessment of tumor resectability (Barat et al., 2021) and the prediction of postoperative complications based on imaging features. Achieving this requires the development of a precise segmentation method able to identify various organs, vessels, and arteries involved in pancreatic surgery. Moreover, the incorporation of multi-modal data could highly benefit models predicting postoperative complications. In particular, clinical data such as age, sex or diagnosis can significantly influence the prognostic. Integrating such diverse data sources into predictive models has the potential to greatly augment the decision-making process (Schlanger et al., 2022). Multi-modal models are further discussed in the technical perspectives (see Section 8.3.2.3).

8.3.2 Technical perspectives

8.3.2.1 Contrastive learning for improved detection of secondary signs

As detailed in Chapter 7, the two methods introduced for detecting secondary signs relied on features extracted by a VAE. These features served as powerful descriptors used to determine the presence or absence of a specific radiological sign within an image. Consequently, enhancing the feature extraction process emerges as a crucial strategy for improving secondary sign detection. Recent studies have highlighted the remarkable ability of self-supervised contrastive methods in extracting highly informative features, underscoring their potential for improving the detection of secondary signs (Chaitanya et al., 2020). Self-supervised contrastive methods aim to learn image representations without relying on annotations, by contrasting positive and negative pairs (Liu et al., 2021a). They operate under the assumption that similar instances (i.e., positive pairs) should be closer together in a learned embedding space, while dissimilar instances (i.e., negative pairs) should be further apart. The pairs are generated using data augmentation transforms : a positive pair would consist of different augmentations of a same instance, while a negative pair would consist of two different instances. These transformations are carefully designed to encourage the model to capture meaningful features and similarities in the data (Ruppli et al., 2022). Self-supervised contrastive learning has demonstrated impressive results in various tasks, including image classification, object detection, and image generation, across both natural and medical images (Chaitanya et al., 2020; Liu et al., 2021a).

Preliminary experiments using contrastive methods have been conducted to predict abnormal pancreatic shapes (see Appendix B). However, the exploration of their potential applicability to other secondary signs, particularly those associated with texture, merits further investigation. Moreover, recent advances have integrated weak labels into the contrastive framework, offering promising avenues for learning from noisy image-level labels (Sarfati et al., 2023). This approach could be a technical solution to the inter and intra-annotator variability discussed above (see Section 8.3.1.1). Finally, the application of contrastive learning methods could also be considered for lesion characterization, as feature descriptors could be used to classify lesions types (Li et al., 2021a).

8.3.2.2 Use of the arterial phase for improved lesion detection

While the primary focus of this thesis centered on portal phase CT scans, it is acknowledged that this modality may not be the most optimal for pancreas imaging. Indeed, during the portal phase (typically 60 seconds after contrast injection), the contrast enhancement between normal pancreas, pancreatic tumors, and surrounding critical vascular structures is suboptimal (Boland et al., 1999). By contrast, the arterial phase (typically 20-35 seconds after contrast injection) or the pancreatic phase (typically 40-45 seconds after contrast injection) are considered more suitable, allowing for maximum differentiation between normal parenchyma and potential tumors. For hypodense tumors such as PDACs, the contrast is most pronounced in the pancreatic phase, whereas they may appear iso-dense (and barely visible) in the portal phase (Quencer et al., 2013) as illustrated in Figure 8.1.

Therefore, several works addressing pancreatic lesion detection develop algorithms that use both arterial and portal phases as input (Xia et al., 2020; Zhou et al., 2019a; Zhu et al., 2021). However, their direct application is unsuitable for clinical use cases since they require both modalities in inference. In real-world practice, patients without suspected pancreatic cancer typically undergo a portal CT examination. Given this clinical context, the following question arises : *is it possible to use both portal and arterial phases during training, in order to improve lesion detection performance on portal phase during inference ?* This approach, where the arterial phase serves as additional input during training but not during inference, seeks to enhance training by leveraging the distinct information present in each phase. We hypothesize that this approach could significantly improve lesion detection and potentially extend to lesion characterization.

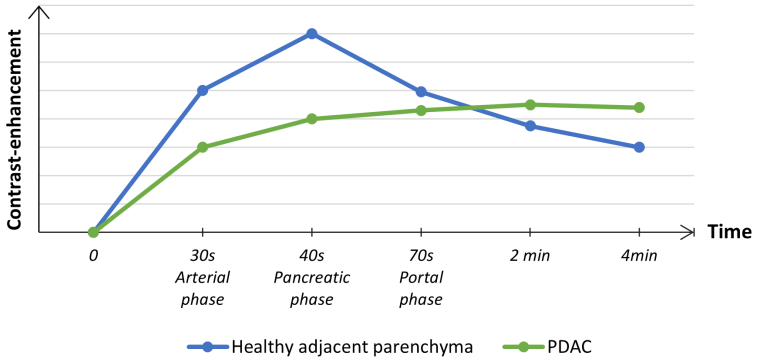


Figure 8.1 – Schematic diagram showing CT contrast enhancement over time for a PDAC mass and healthy parenchyma. The enhancement pattern of the PDAC mass is characterized by a slow and gradual increase, reaching a plateau, whereas healthy parenchyma exhibits a rapid increase followed by a gradual decrease in enhancement. Maximal contrast between the two structures is reached in the pancreatic phase. Adapted from Zhang et al. (2017). *CT : Computed tomography, PDAC : Pancreatic ductal adenocarcinoma.*

Before exploring strategies to tackle this research topic, preliminary experiments are essential to confirm how useful the arterial phase is for detecting lesions. The next section introduces the dataset created for these experiments and shares the results obtained. We then detail a few research avenues to address this topic.

Dataset 1281 exams containing both arterial and portal CT phases are retrieved from $\mathcal{D}_{CAPA}$, Guerbet's private database, to form $\mathcal{D}^{Train}_{PortArt}$. All the cases have segmentation masks for the pancreas, pancreatic lesion (if any) and MPD (if visible). These masks are manually drawn by radiologists on the portal phase. For each case, the arterial phase is registered on the portal phase using the Simple Elastix library (Marstal et al., 2016). In particular, a rigid registration process is executed, employing the mean square loss function optimized only within the pancreas region (determined as the union of the three reference segmentation masks). $\mathcal{D}^{Train}_{PortArt}$ is split into four folds, stratified based on both the originating institution and the lesion size.

Experiments In order to assess the contribution of the arterial phase for the detection of pancreatic lesion, a first set of experiments is run. It consists in training automatic segmentation models (nnUNet by Isensee et al. (2021a), described in Section 3.2.3.5) with varying input :

- **T1** : Portal CT scans,
- **T2** : Arterial CT scans,
- **T3** : Concatenation of Portal and Arterial CT scans.

Each model is trained to segment the pancreas, the lesion, and the MPD. As arterial images are rigidly registered on the corresponding portal images, the reference segmentation masks used to train each model are the portal phase masks. For each experiment, data preprocessing is automatically done according to the nnU-Net procedure (described in Section 3.2.3.5). Each model is trained in a four-fold cross-validation procedure for 1000 epochs.

To evaluate the models at the patient level (lesion detected or not), each case is assigned a lesion risk score. To this end, a procedure similar to the one in Section 6.4.2 is proposed :

1. First, predicted lesions that form a connected component out of the predicted pancreas segmentation mask are removed ;
2. Secondly, the predicted lesion probability map is binarized using a threshold at 0.01. Then, each lesion candidate (i.e., lesion forming a connected component) is given a lesion probability equal to the average voxel probability within the connected component ;
3. Thirdly, the candidate lesion is discarded if the lesion probability is below 0.05 or if the number of voxels in the lesion is inferior to 10.

Finally, the lesion risk by patient corresponds to the maximum lesion probability among all the candidate lesions.

Results Table 8.1 shows the results obtained on $\mathcal{D}_{PortArt}^{Train}$ by cross-validation. In addition, subgroup analysis were conducted by analyzing results on isodense tumors, PNET and small lesions (with lesion diameter inferior to 10 or 20 mm).

Validation set	# of cases (# with lesions)	**T1 : Portal**	**T2 : Arterial**	**T3 : [Portal, Arterial]**
All	1281 (1240)	91.28 ±1.91	89.62 ±2.19	**92.12 ± 1.75**
Isodense	136 (95)	82.94 ±3.89	81.01 ±4.14	**84.79 ±3.56**
PNET	133 (92)	86.04 ±3.4	82.93 ±3.64	**88.76 ±3.02**
Small (<10mm)	74 (33)	56.06 ±6.99	58.62 ±6.65	**60.63 ±6.88**
Small (<20mm)	243 (202)	77.92 ±3.92	75.98 ±4.27	**80.93 ±3.61**

Table 8.1 – Sub-group analysis obtained on $\mathcal{D}_{PortArt}^{Train}$ by cross-validation on the four folds. Means ± standard deviations of the AUC (in %), obtained by bootstrapping with 1000 repetitions, are reported. Each column represents one experiment, with its input type written in brackets. Each row represents one subgroup, with best results per row in bold. *AUC : Area under the curve, PNET : Pancreatic neuroendocrine tumor.*

As shown in Table 8.1, detection performance went from 91.28% to 89.62% when going from a model using portal phase images (T1) to a model using arterial phase images (T2). This result, which disagrees with the initial hypothesis that arterial phases are better for lesion detection, can be explained by imperfect registration and by arterial acquisition times that are not optimal for the observation of pancreatic lesions.

The best results were obtained by T3, which concatenates both phases as input, as the AUC reaches 92.12% (vs 91.28% on the baseline experiment T1). This suggests that the arterial phase brings additional information that is useful for the lesion segmentation when used jointly with the portal phase. Sub-group analysis results confirm the usefulness of using both phases as the AUC gains 1.85% on isodense tumors and 2.72% on PNETs compared to a model using solely the portal phase. AUC improvements are even higher on small lesions, with gains of 4.57% and 3.01% on lesions with diameter inferior to 10 and 20 mm, respectively.

Possible strategies As indicated by the preceding results, concatenating the arterial phase yields a tangible improvement in lesion detection, particularly for isodense or small lesions. As the arterial phase is not always available in clinical routine, different strategies can be considered to deal with this constraint.

A first strategy consists in building a feature space which is modality-invariant, as in HeMIS (Havaei et al., 2016) where a feature extractor is trained for each modality. Then, feature maps are merged by computing statistics (mean, variance) and these merged feature maps produce the final output of the network. As these statistics do not depend on the number of maps being concatenated, the network can work with any number of input modalities. However, when only one modality is available as input (which would be our case), these methods are less performing than a model trained exclusively on that specific modality (Hu et al., 2020a).

A second strategy consists in leveraging knowledge distillation methods (Hinton et al., 2015). These methods typically involve a small student network learning to mimic a teacher network, which is usually bigger and pre-trained. In Hu et al. (2020a), distillation is combined with the concept of Privileged Information (Vapnik et al., 2015), where the teacher network contains additional information for each training instance. Combination of both paradigms allows for knowledge transfer between a multi-modal teacher network and a single-modal student network, and the method is illustrated on the BraTS 2018 dataset (BRATS; Hu et al., 2020a). Several variations have been proposed with deep mutual learning (Zhang et al., 2018), multi-view learning (Blum and Mitchell, 1998) and cooperative learning (Batra and Parikh, 2017), where a pool of networks learn the same task simultaneously, but using inputs from different domains.

A third strategy consists in using a generative network that will generate the missing modality in inference. This research avenue was explored in Appendix D, with initial attempts made to synthetically generate the arterial phase using a Generative Adversarial Model (GAN). As our problem involved paired data (i.e., both portal and arterial phases from the same acquisition), the state-of-the-art Pix2Pix model by Isola et al. (2017) was employed. The details of the Pix2Pix model, including its principle and architecture, are provided in Appendix D.1. Several experiments were conducted, and both qualitative and quantitative results are outlined in Appendix D. However, none of the outcomes proved conclusive, and various challenges were encountered. First, the generated images exhibited either jagged edges or blurriness, maybe due to imperfect registration. Cyclic GANs offer a potential solution to this issue : while they are commonly applied in unpaired data scenarios, recent research by Wolterink et al. (2017) has highlighted their effectiveness in paired data situations, particularly in addressing the problem of blurred images as they do not rely on imperfect data pairing (in this case, registration). Secondly, we struggled to find an appropriate evaluation metric allowing to determine whether the synthetically generated arterial phase provided additional information that could enhance a lesion segmentation model. In practice, this evaluation involves training a GAN to generate synthetic arterial phase images, followed by training a segmentation model using both the true portal image and the generated arterial image as inputs. The performance of lesion detection can then be assessed. Given this context, conducting a hyper-parameter search to identify the optimal generated images for lesion segmentation proves to be costly in terms of time and computational resources, rendering this approach inefficient. Alternatively, the GAN could integrate a segmentation task by directly introducing an additional Dice score loss function. This approach would prompt the GAN to generate synthetic images that are directly useful for the downstream task of lesion segmentation, avoiding the need for separate training steps. This concept aligns with the perspectives of Velikova et al. (2022) and La Barbera et al. (2022), asserting that generating perfect images may not be imperative when focusing on a downstream task.

8.3.2.3 Enhancing diagnosis with multi-modal data

Using the arterial phase represents a step towards harnessing the available imaging modalities ; however, this potential can be further extended by leveraging all accessible data, particularly clinical data and electronic health records (EHR). Indeed, the diagnosis of a specific pancreatic lesion is influenced by various factors, including demographic, clinical, or biological data. Therefore, AI models should be constructed as hybrid models that incorporate not only imaging features but also a comprehensive range of non-imaging information (Barat et al., 2021). Such models have gained research interest, particularly for disease diagnosis

and predicting different outcomes (e.g. mortality or survival prediction, treatment outcome prediction), and a meta-review by Mohsen et al. (2022) found that multi-modal models combining EHR and medical imaging data generally outperform single modality models for the same task. Many of these models employ early fusion techniques, where each modality (imaging or non-imaging) is encoded in a vector given to a single machine learning algorithm for training. Typically, a non-redundant combination of DLR and HCR, as proposed in Chapter 7.2, could be employed to encode the images into vectors. Moreover, large language models (LLM) could be leveraged to integrate the information contained in the EHR (Deng and Lin, 2022; Thirunavukarasu et al., 2023). These models have recently demonstrated proficiency within the medical field, with different applications ranging from determining the optimal imaging modality given specific clinical presentations (Nazario-Johnson et al., 2023) to predicting differential diagnoses based on free text describing imaging patterns (Kottlors et al., 2023). Thus, coupling LLMs with imaging features extracted by deep learning models could be of major interest for pancreatic cancer analysis, although LLMs are data- and computationally-intensive.

8.3.3 Industrial perspectives

Given that this thesis is part of an industrial project initiated by Guerbet, it is important to discuss some perspectives related to the industrial development of an AI method for pancreatic cancer detection. As with many AI applications in medical imaging, transitioning from research to clinical use raises several questions :

— **Displaying Results :** The various works conducted in this thesis, along with the clinical perspectives mentioned earlier, address numerous diagnostic questions. However, presenting the results of all these analyses might burden radiologists, impeding their workflow efficiency. For example, while segmentations may be valuable for interpretability, their display may add complexity to the diagnostic workflow. Therefore, a balance between conciseness and interpretability must be found. For instance, the outputs of AI algorithms could be combined to provide a binary output (alert or not), allowing physicians to re-examine a case and display the segmentation if necessary. However, defining the alert level remains an open question : should an alert be exclusively issued for PDACs or for any type of abnormality, including pancreatitis and atrophy ? For this purpose, alert levels based on clinical data may be beneficial. These considerations are essential and require thorough testing to assess user experience qualitatively.

— **Performance Expectations :** It is necessary to establish the performance level at which a tool becomes genuinely useful for assisting radiologists in clinical practice. An under-performing tool may not be used or could lead to diagnostic errors. Since global performance involves a trade-off between sensitivity and specificity, the expected performance for each of these measures needs to be distinguished. In this regard, Chu et al. (2023) conducted a survey of 161 radiologists to assess their expectations for the clinical use of an AI tool for pancreatic cancer detection. Their study indicates that the majority of respondents prioritized sensitivity over specificity. Specifically, the most preferred cutoff for sensitivity and specificity in an AI system was 99% sensitivity and 75% specificity. The same study reports that, according to these 161 respondents, the minimum lesion size that an AI tool should detect is 5 mm. However, these parameters could be user-dependent and determined beforehand by each hospital or radiologist. These considerations must also be examined through an ethical lens, and the consequences of prediction errors (whether false positives or false negatives) must be carefully considered.

— **Hardware Requirements :** The industrial deployment of such tools also requires technical considerations, given that DL algorithms demand significant computational resources. These computations could be executed using dedicated hardware installed within the medical institution or through cloud-based computing, each option presenting distinct challenges in terms of installation and maintenance. These considerations are critical and necessitate careful planning regarding equipment installation and staff training.

— **Regulatory Aspects :** In addition to these questions, regulatory issues, particularly concerning the clinical studies necessary for the market approval of such algorithms, must be addressed.

Thus, bridging the gap between research and application in clinical practice is a sensitive step, and numerous challenges must be addressed by multidisciplinary teams to cover various clinical, technical, ethical, regulatory, administrative, and user-related aspects. Despite these challenges, 50% of radiologists surveyed by Chu et al. (2023) intend to integrate AI tools into their clinical practice within the next five years, highlighting practitioners' interest in these diagnostic aids.

8.4 Conclusion

This thesis led to the development of various AI-based methods aimed at detecting pancreatic cancer on portal CT scans. It started with the development of a pancreas segmentation algorithm, and then addressed the detection of several signs of pancreatic cancer, including lesions and secondary radiological signs which are crucial for early diagnosis. These methodologies were presented in national conferences (GRETSI, JFR, IABM), international conferences (ISBI, MICCAI), and journals (Investigative Radiology). Notably, one of these methods has been patented, and two others have received awards, including the Best Oral Presentation Awards at the CaPTion workshop, MICCAI 2023.

Our vision for the future research directions in AI-based pancreatic cancer detection includes a focus on precise lesion characterization, and the creation of expansive databases with specialized validation cohorts. These advances have the potential to significantly improve diagnostic accuracy and lead to enhanced patients outcomes. From a technical standpoint, the incorporation of multi-modality approaches appears to be of paramount importance in achieving these objectives.

Appendices

Annexe A

Lesion and Main Pancreatic Duct Dilatation Detection

We present supplementary results and additional details related to Chapter 6.

A.1 Regional measurement of the main pancreatic duct diameter

The segmentation network presented in Chapter 6 generates a segmentation of the pancreas. Based on this segmentation, the pancreas can be sub-segmented into head, body, and tail. To do so, the extremities of the tail and head of the pancreas are identified, and the centerline going through the pancreas and connecting these two points is estimated. The centerline is divided into tail, body, and head according to the following heuristic : the tail represents 25% of the total length, the body represents the following 50%, and the head the last 25% of the total length. An illustration of the pancreas sub-segmentation is shown in Figure A.1.

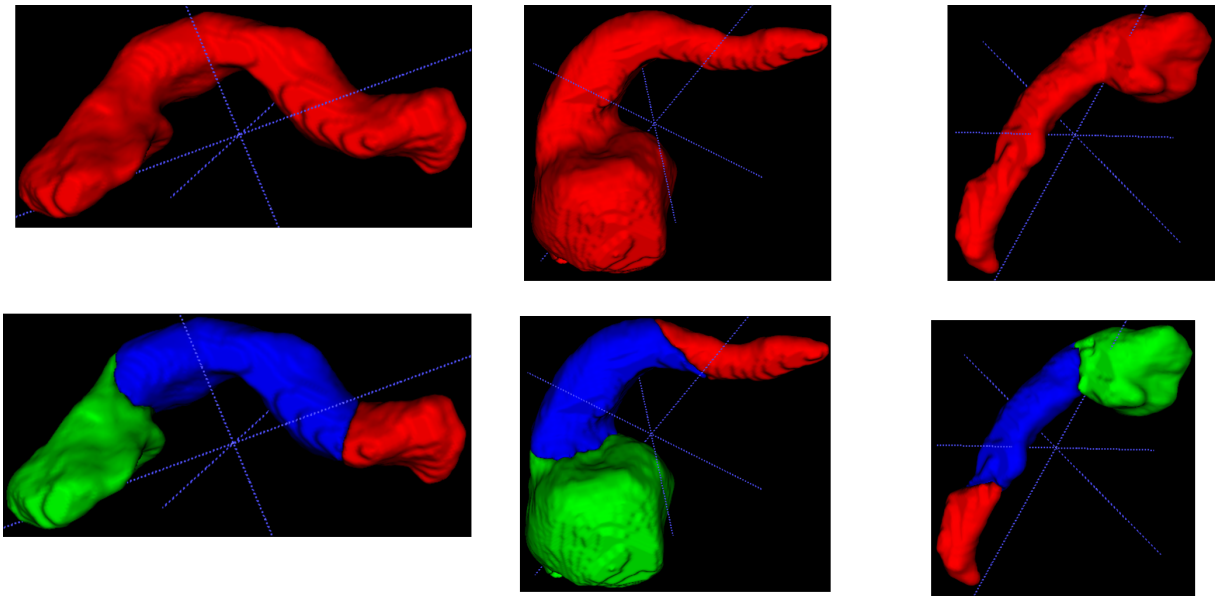


Figure A.1 – Example of the sub-segmentation of the pancreas into head (green), body (blue), and tail (red).

Finally, to measure the Main Pancreatic Duct (MPD) diameter in the head, body, and tail, each MPD voxel is assigned the region of its closest centerline point. Then, the MPD is also sub-segmented into three parts and its size can be measured regionally using the IMEA library (Kroell, 2021) as explained in Section 6.4.2.

A.2 Measuring lesion detection performance

We detail how the evaluation measures of interest were computed for the logistic regression model predicting lesion presence.

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. Specificity was computed as the ratio of healthy subjects correctly identified by the model to the total number of healthy subjects. Positive Predictive Value (PPV) was defined as the ratio of the number of patients with a pancreatic lesion correctly detected by the model to the total number of patients for which the model detected a lesion. Negative Predictive Value (NPV) was defined as the ratio of healthy subjects correctly found by the model to the total number of healthy subjects predicted by the model.

A.3 Measuring main pancreatic duct dilatation detection performance

We detail how the evaluation measures of interest were computed for the logistic regression model predicting MPD dilatation.

Sensitivity was computed as the ratio of the number of subjects with a dilated MPD detected by the model to the total number of subjects with a dilated MPD. Specificity was defined as the ratio of the number of subjects with a normal MPD identified by the model to the total number of subjects with a normal MPD. PPV was defined as the number of subjects with a dilated MPD correctly detected by the model to the total number of patients with a dilated MPD found by the model. NPV was computed as the ratio of subjects without a dilated MPD correctly identified by the model to the total number of subjects without a dilated MPD predicted by the model.

A.4 Visualizing the distribution of the evaluation measures

Regarding Tables 6.1, 6.3 and 6.4 in Chapter 6, it is important to clarify that the Area Under the Curve (AUC), sensitivity, specificity, PPV, and NPV, were computed for the whole test set. Median values and confidence intervals were calculated thanks to bootstrap sampling, which allows estimating the sampling distribution of all the evaluated measures by using random sampling with replacement. We provide in Figure A.2 a boxplot of the evaluation measures for both lesion detection and MPD dilatation detection on the whole test set, thus giving a visualization of Tables 6.2 and 6.4.

A.5 Analyzing segmentation performance of lesions and the MPD

It can be observed in Figure A.3 that Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD) median values are higher than mean values for both lesions and the MPD. Regarding the DSC, a median value of 0.83 is obtained compared to a mean value of 0.69. In the case of the MPD, a median value of 0.67 is obtained compared to a mean value of 0.58. This can be explained by the fact that for false negative and false positive cases, the associated DSC will be 0, thus strongly deteriorating the mean value. Another reason that can explain the apparently low segmentation performance is due to the choice of the evaluation measure. The DSC computes the overlap between the predicted and reference segmentations and penalizes every wrongly detected voxel. When used on large and well-defined anatomical structures such as the pancreas, this evaluation measure is adapted to evaluate segmentation performance. However, pancreatic lesions can have poor margins and be small (63 of lesions in the test set had a diameter lower than 30mm), while the MPD is a thin structure which can be hard to accurately delineate for the annotators. Moreover, these structures usually contain much less voxels than the pancreas. Therefore, the DSC can strongly penalize the evaluation of segmentation performance on these structures, while the segmentations generated by the model are acceptable. This is why we used the NSD, which allows for a tolerance of 2mm along the three spatial dimensions to measure segmentation performance. In Figure A.3, the median

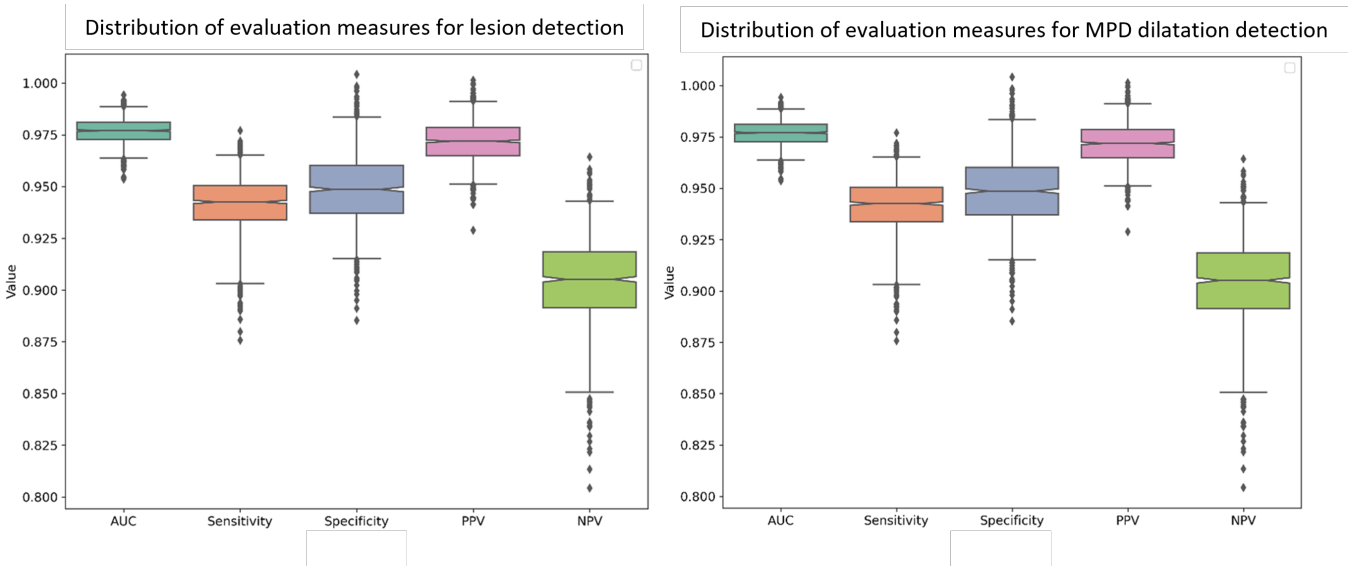


Figure A.2 – Boxplot of the AUC, sensitivity, specificity, PPV, and NPV obtained on the test set by the two logistic regression models. Left : Lesion presence detection. Right : MPD dilatation detection. *AUC : Area under the curve, MPD : Main pancreatic duct, NPV : Negative predictive value, PPV : Positive predictive value.*

NSD is 0.97 and 0.91 for lesions and the MPD, respectively. This result shows that the model generated segmentation which were overall correctly overlapping the reference segmentations for both lesions and the MPD.

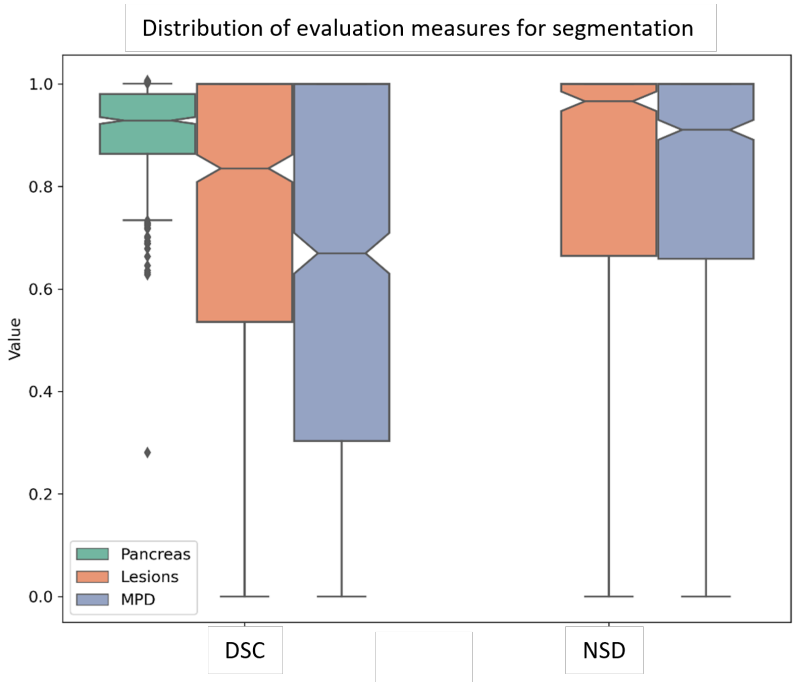


Figure A.3 – Boxplot of the DSC and NSD obtained for the pancreas, lesions, and the MPD on each case of the test set. The NSD is computed for the lesions and the MPD only, with a tolerance error of 2 mm along each spatial dimension. *DSC : Dice similarity coefficient, MPD : Main pancreatic duct, NSD : Normalized surface Dice.*

Annexe B

Learning shape distributions from large databases of healthy organs

We present supplementary results and additional details related to Section 7.2.

B.1 Improving the experimental part

A limitation of the work presented in Section 7.2 lies in the experimental setup. Indeed, for the normal *vs.* abnormal shape classification experiments, all normal shapes come from patients in database $\mathcal{D}_{NIH}$, while all abnormal shapes come from database $\mathcal{D}_{BR,1}$. Although the bias on the origin of the data is very strongly limited by the use of binary inputs, it is worth studying its potential implications. To do this, a first straightforward solution is to add a third database containing normal shapes. The expected behavior of the model would be to accurately differentiate abnormal shapes from all normal shapes, irrespective of their source institution. However, it is not expected to distinguish among normal shapes based on their origin.

B.2 Contrastive learning for shape analysis

Contrastive learning falls within the field of self-supervised learning, where image representations are learned without annotations. More specifically, contrastive methods aim to learn such representations by contrasting positive against negative pairs. For instance, in SimCLR (Chen et al., 2020a), distance between representations of positive pairs is reduced while distance between representations of negative pairs is increased. Key ingredients for training contrastive models are heavy data augmentations and careful choice of the negative samples, which is usually done with large batch sizes or memory banks.

Recent works explored non-contrastive paradigms, where the loss functions do not include any contrastive terms. By consequence, such models are trained without negative samples and do not require large batch sizes. While asymmetry was introduced by Grill et al. (2020) and Chen and He (2021) in the framework to prevent collapse, trivial solutions are avoided by construction of its loss function in (Zbontar et al., 2021), making the latter method more principled.

The direct transposition of our problem into the SimCLR framework (Chen et al., 2020a) would lead us to learn an embedding space in which positive shape pairs (*i.e.*, one single shape which would have undergone two random isometric transformations[1]) stay close to each other, while negative shape pairs are far apart. In our setup, where the model is learned exclusively from healthy shapes, this formulation would make two different shapes distant in the latent space, though both healthy. Actually, our working hypothesis, which is the availability of a large database of healthy shapes for training the model, does not allow for the definition of negative pairs since all the training examples are healthy. This motivated us to

1. A shape-preserving transformation in the plane or in space.

explore the use of non-contrastive frameworks, such as in (Zbontar et al., 2021), for learning a distribution of normal shapes. This is illustrated in Figure B.1. Besides of its formulation which seems more adapted to our problem, Zbontar et al. (2021) claim that their model is robust to small batch sizes, which is our case as we work with 3D inputs.

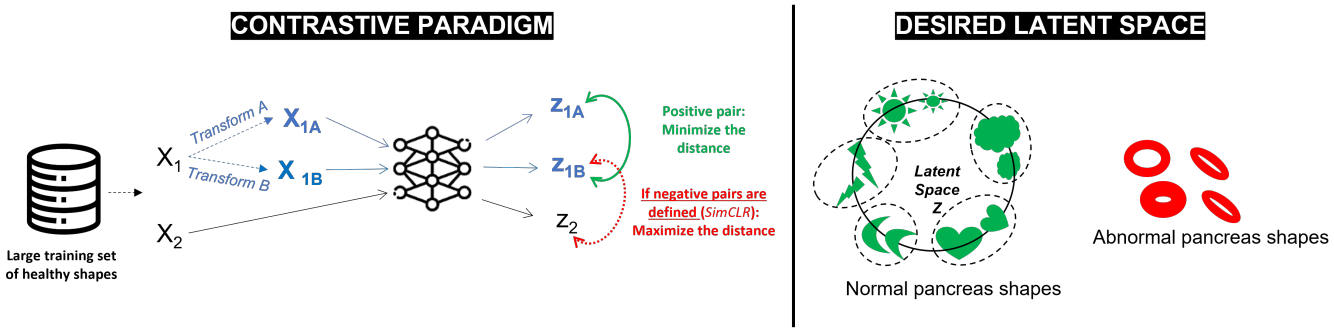


Figure B.1 – Left : Illustration of the contrastive paradigm on our problem. Two random isometric transformations are applied to a single shape, and the distance between their latent representations (which constitute a positive pair) is minimized. If the contrastive framework explicitly defines negative pairs, as it is the case in SimCLR (Chen et al., 2020a), then the objective function also maximizes the distance between the latent representations of negative pairs. **Right : Design of the desired latent space for our problem.** Green and red shapes represent normal and abnormal shapes, respectively. As shown with the green pairs, training the model on healthy samples should make it able to encode healthy shapes properly. As pathologic pancreas exhibit different shapes, the model should naturally encode them further away.

B.3 Preliminary results

We display here some preliminary results with the new experimental set-up described above. This time, we tested the zero-shot and the few-shot classification with three databases. For the sake of clarity, we adopt the following notations :

— D^{Test}_{normal} : the 80 cases from $\mathcal{D}_{NIH}$, for whom the assumption of normal pancreas shape holds ;

— $D^{Test}_{normal,2}$: 80 Computed Tomography (CT) scans from $\mathcal{D}_{CABO}$ who did not exhibit pancreatic lesions, and for whom the assumption of normal pancreas shape was done ;

— $D^{Test}_{abnormal}$: 144 cases from $\mathcal{D}_{BR,1}$, diagnosed with pancreatic cancer, and for whom the pancreas shape was evaluated as abnormal by an expert radiologist.

$\mathcal{D}_{NIH}$, $\mathcal{D}_{CABO}$ and $\mathcal{D}_{BR,1}$ are described in Chapter 4. For all the scans, segmentation masks of the pancreas were automatically generated by the same in-house algorithm. The expected behavior of the model would be to differentiate abnormal shapes from all normal shapes, i.e. differentiate D^{Test}_{normal} from $D^{Test}_{abnormal}$, and $D^{Test}_{normal,2}$ from $D^{Test}_{abnormal}$. However, it should not to be able to distinguish D^{Test}_{normal} from $D^{Test}_{normal,2}$.

We compared our Variational Autoencoder (VAE)-based method, presented in Chapter 7.2, with the following methods : ASM (Cootes et al., 1995), LDDMM using the Deformetrica software (Bône et al., 2018), SimCLR (Chen et al., 2020a) and Barlow Twins (BT) (Zbontar et al., 2021). Concerning ASM and Large Deformation Diffeomorphic Metric Mapping (LDDMM), we used the same procedure as detailed in Section 7.2.3, and we performed a PCA on the shape-encoding parameters to obtain a latent vector of dimension 256 for each subject. The three other models under comparison (VAE, SimCLR, BT) also had a latent dimension $L = 256$, and all methods were trained on the 600 subjets.

Results are reported in Table B.1. Tables B.1aand B.1b underline the superiority of contrastive methods, specifically when the number of training samples is small. Yet, there is no clear superiority between SimCLR and BT, and further experiments need to be carried out. Except for LDDMM, all models perform better at distinguishing D^{Test}_{normal} from $D^{Test}_{abnormal}$ than $D^{Test}_{normal,2}$ from $D^{Test}_{abnormal}$, which can be easily explained by the origin of the databases : D^{Test}_{normal} are young, very healthy organ donors while patients from $D^{Test}_{normal,2}$ are more diverse and may exhibit some abdominal pathologies. In Table B.1c, the AUC is

expected to be low as the classification between D^{Test}_{normal} and $D^{Test}_{normal,2}$ should not work. Although classification performance is significantly lower than in the other experiments, the AUCs still increase with the number of training examples, suggesting that the models are able to capture some features related to the origin of the database. The reasons for these results shoud be investigated. De-biasing techniques could be explored, such as in (Tartaglione et al., 2021) where a regularization strategy disentangles the bias and the relevant information during training.

CONFIGURATION	Zero-shot	Few-shot		
	-	CV-20	CV-10	LOO-CV
# of train/test samples	*0/224*	*12/222*	*112/112*	*223/1*
LDDMM	63.51±0.34	75.95±0.02	80.06±0.03	89.5±0.2
ASM	63.22±0.34	60.33±0.02	63.1±0.04	84.13±0.26
VAE (Ours)	71.18±0.28	76.13±0.02	80.66±0.03	89.34±0.21
SimCLR	**80.15±0.28**	80.36±0.01	84.51±0.03	**93.12±0.16**
BT	76.83±0.29	**89.46±0.01**	**90.43±0.02**	92.26±0.18

(a) Few-shot and zero-shot AD results for D^{Test}_{normal} vs $D^{Test}_{abnormal}$

CONFIGURATION	Zero-shot	Few-shot		
	-	CV-20	CV-10	LOO-CV
# of train/test samples	*0/224*	*12/222*	*112/112*	*223/1*
LDDMM	72.67±0.33	73.07±0.02	78.28±0.03	**90.56±0.19**
VAE (Ours)	68.00±0.3	72.88±0.02	77.92±0.03	89.09±0.21
SimCLR	**76.91±0.3**	73.74±0.02	79.32±0.03	89.76±0.22
BT	74.96±0.3	**83.92±0.01**	**84.05±0.03**	88.44±0.24

(b) Few-shot and zero-shot AD results for $D^{Test}_{normal,2}$ vs $D^{Test}_{abnormal}$

CONFIGURATION	Zero-shot	Few-shot		
	-	CV-20	CV-10	LOO-CV
# of train/test samples	*0/224*	*12/222*	*112/112*	*223/1*
LDDMM	39.44±0.37	60.0±0.02	64.76±0.04	73.78±0.33
ASM	**49.01±0.39**	**52.91±0.02**	**54.72±0.04**	**58.39±0.37**
VAE (Ours)	51.03±0.3	57.77±0.02	62.45±0.04	73.06±0.32
SimCLR	57.97±0.4	57.31±0.02	61.1±0.04	71.51±0.34
BT	52.05±0.4	61.09±0.02	65.57±0.04	73.01±0.34

(c) Few-shot and zero-shot AD results for D^{Test}_{normal} vs $D^{Test}_{normal,2}$

Table B.1 – For each experiment, we report the mean and standard deviation for AUC (in %), obtained by bootstrapping with 10000 repetitions. Best results (in this experiment, AUC closest to 50%) by column are in **bold**. *AD : Anomaly detection, ASM : Active shape models, AUC : Area under the curve, BT : Barlow Twins, CV : Cross-validation, LDDMM : Large deformation diffeomorphic metric mapping, LOO : Leave-one-out, VAE : Variational autoencoder.*

Annexe C

Early Detection of Pancreatic Cancer using Non-Redundant Combination of Hand-Crafted and Deep Learning Radiomics

The 32 hand-crafted radiomics used in Chapter 7.3 were the followings :

— **14 shape features** describing the size and shape of the pancreas
  — Mesh Volume
  — Voxel Volume
  — Surface Area
  — Surface Area to Volume ratio
  — Sphericity
  — Maximum 3D diameter
  — Maximum 2D diameter in the axial plane
  — Maximum 2D diameter in the coronal plane
  — Maximum 2D diameter in the sagittal plane
  — Major Axis Length
  — Minor Axis Length
  — Least Axis Length
  — Elongation
  — Flatness

— **18 first-order intensity features** describing the intensities distribution within the organ
  — Energy
  — Total Energy
  — Entropy
  — Minimum
  — 10^{th} percentile
  — 90^{th} percentile
  — Maximum
  — Mean
  — Median
  — Interquartile Range

— Range
— Mean Absolute Deviation
— Robust Mean Absolute Deviation
— Root Mean Squared
— Skewness
— Kurtosis
— Variance
— Uniformity

These features were extracted using the pyradiomics library (Van Griethuysen et al., 2017). More details about each feature can be found in the online documentation[1].

1. https ://pyradiomics.readthedocs.io/en/latest/features.html

Annexe D

Improving lesion detection using the arterial phase

As shown in Table 8.1 in Section 8.3.2.2, the use of arterial phase brings a real gain for lesion detection, especially when lesions are isodense or small. As this arterial phase is rarely available in clinical routine, we begin by looking at synthetic generation of arterial phase using a Generative Adversarial Model (GAN). Since we have paired training data (*i.e.*, we have both portal and arterial phase from the same acquisition), we rely on the state-of-the-art Pix2Pix model (Isola et al., 2017).

D.1 Pix2Pix

Pix2Pix model (Isola et al., 2017) is a start-of-the-art conditional GAN whose architecture is depicted in Figure D.1. A conditional GAN is conditioned on an input image : the generator is constrained, via an additional term in the loss function, to generate an output image which should be near the paired ground-truth image. The principle of the Pix2Pix model is as follows :

1. First, starting from the real portal I^{real}_{portal}, the generator G, which is a U-Net (Ronneberger et al., 2015), generates a fake arterial $I^{fake}_{arterial}$.
2. Then, the discriminator D is trained :
   (a) It takes, as input, the concatenation of the real paired images $I^{real}_{pair} = [I^{real}_{portal}, I^{real}_{arterial}]$. D is a PatchGAN discriminator which outputs a vector P_{real} of size $[b, 1, 5, 14, 22]$, where b is the batch size and 5, 14, and 22 depend on the input size and the discriminator architecture. This vector represents $5 \times 14 \times 22 = 1540$ patches of size $70 \times 70 \times 70$, according to the receptive field of D. P_{real} assigns a probability of each of the 1540 patches for being fake, and is used to compute $\mathcal{L}_{real} = MSE(P_{real}, \mathbb{1}_{[b,1,5,14,22]})$ where MSE stands for the Mean Square Error function and $\mathbb{1}_{[b,1,5,14,22]}$ represents a size $[b, 1, 5, 14, 22]$ vector of 1s.
   (b) Then, D takes as input the concatenation of the fake paired images $I^{fake}_{pair} = [I^{real}_{portal}, I^{fake}_{arterial}]$, and outputs a vector P_{fake} that gives the probability for each of the 1540 patches of I^{fake}_{pair} to be fake. We compute $\mathcal{L}_{fake} = MSE(P_{fake}, \mathbb{0}_{[b,1,5,14,22]})$ where $\mathbb{0}_{[b,1,5,14,22]}$ represents a size $[b, 1, 5, 14, 22]$ vector of 0s.
   (c) D is optimized with the following loss function : $\mathcal{L}_D = 0.5 \times (\mathcal{L}_{fake} + \mathcal{L}_{real})$.
3. Secondly, the generator G is updated with the objective to fool D (so to minimize $\mathcal{L}_{fake}$) and to minimize the L_1 distance between $I^{fake}_{arterial}$ and $I^{real}_{arterial}$. Therefore, the parameters of G are optimized with the following loss function : $\mathcal{L}_G = \mathcal{L}_{fake} + \lambda \times L_1(I^{fake}_{arterial}, I^{real}_{arterial})$, where λ is a hyperparameter set to 100 in Pix2Pix original implementation.
4. D and G are updated in an alternate fashion until convergence.

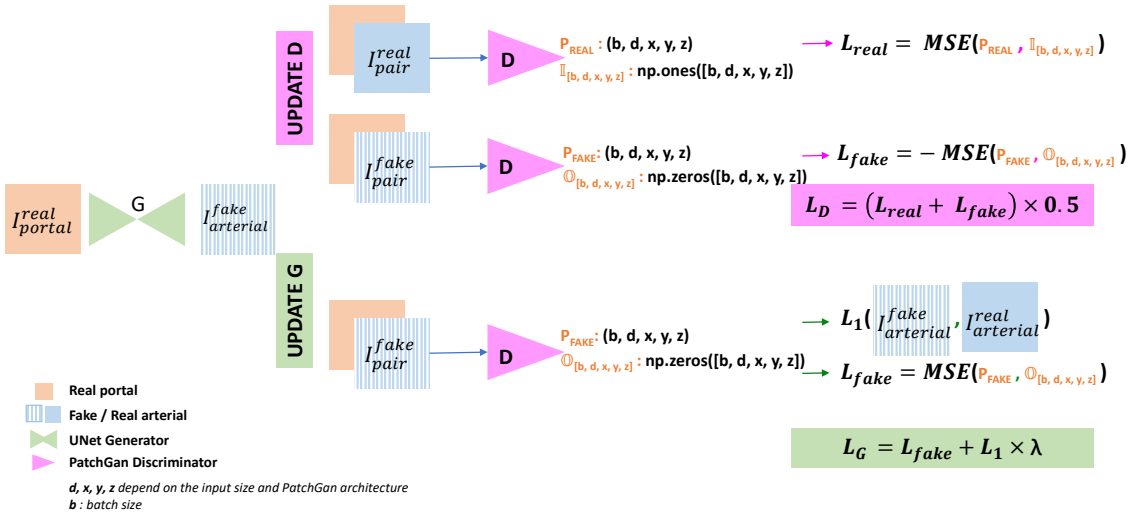


Figure D.1 – Overview of Pix2Pix (Isola et al., 2017). A generator G generates a fake arterial image $I^{fake}_{arterial}$ from a portal image input I^{real}_{portal}. Then, a PatchGan discriminator D is trained to distinguish real pairs ($I^{real}_{pair} = [I^{real}_{portal}, I^{real}_{arterial}]$) and fake pairs ($I^{fake}_{pair} = [I^{real}_{portal}, I^{fake}_{arterial}]$). Then, G is trained to fool D and to minimize the L_1 error between $I^{fake}_{arterial}$ and $I^{real}_{arterial}$. *MSE : Mean square error.*

D.2 Arterial phase images generation experiments

D.2.1 Description

A first series of experiments aims to synthesize arterial phase images from portal phase images using a Pix2Pix model. The following experiments are run :

- **A :** Baseline experiment, consisting of the original implementation [1] of Pix2Pix but adapted for 3D inputs. Use of a PatchGAN discriminator with receptive field $70 \times 70 \times 70$. Use of a generator with U-Net architecture. $\lambda = 100$.
- **B, C, D :** Trying several values for λ : 0.1, 1, 10.
- **E :** PatchGAN discriminator with a bigger receptive field ($127 \times 127 \times 127$).
- **F :** PatchGAN discriminator with a smaller receptive field ($25 \times 31 \times 46$).
- **G :** Making G generate $\delta = I^{real}_{arterial} - I^{real}_{portal}$.
- **H :** Masking the input portal image by the reference pancreas region mask — determined as the union of the three reference segmentation masks (pancreas, lesion, MPD).
- **I :** Masking the MSE loss function by the reference pancreas region mask.
- **J :** Conditions of experiments G + I
- **K :** Conditions of experiments G + I + F

Each experiment is run on one same fold of the dataset $\mathcal{D}^{Train}_{PortalArt}$ introduced in Section 8.3.2.2, resulting in 954 and 327 samples for training and validation, respectively. Arterial images are rigidly registered on the portal images, as described in Section 8.3.2.2. Data pre-processing follows the nnU-Net procedure (see Section 3.2.3.5) except for intensity normalization, as a normalization between -1 and 1 is chosen instead. Each model is trained for 1000 epochs.

D.2.2 Results

To assess the quality of the generated arterial phase, we perform qualitative and quantitative analysis of the different experiments listed above.

1. https ://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

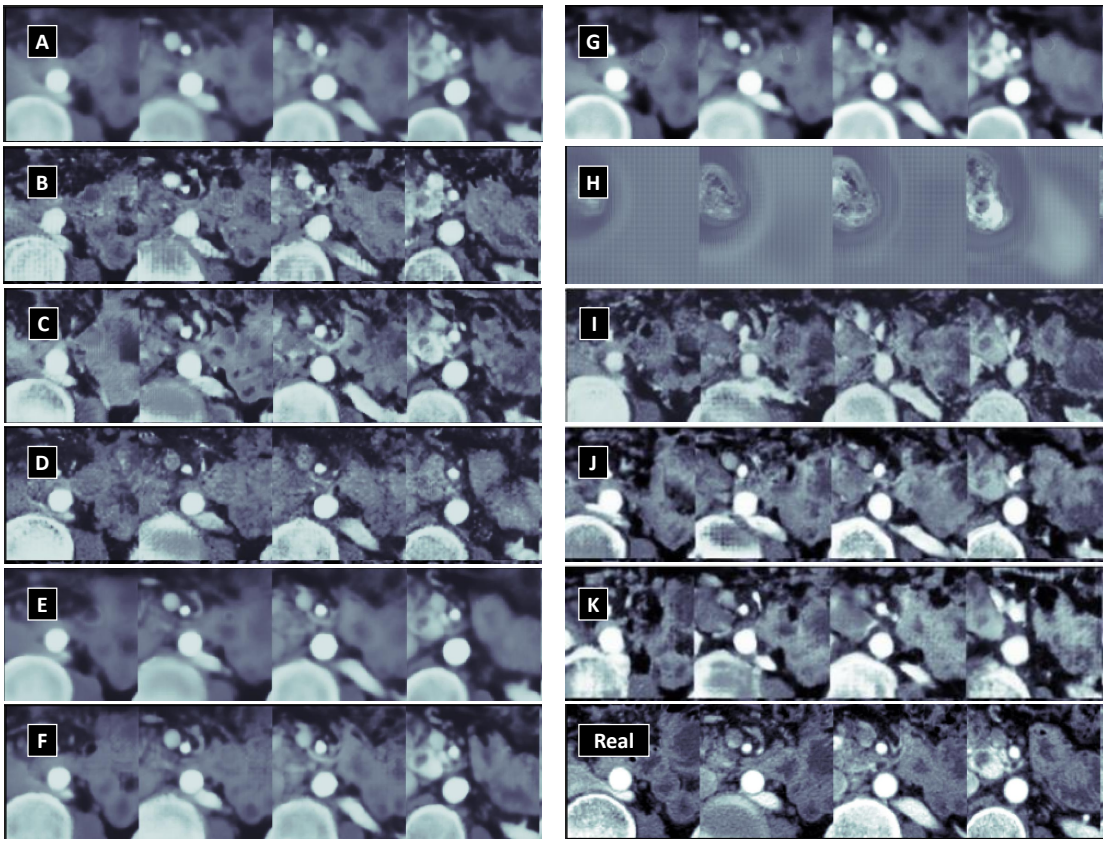


Figure D.2 – Qualitative results of the different conditional GANs trained. We plot the fake arterial phase of a validation sample generated by different models. *A, B, C, D, E, F, G, H, I, J, K : Experiments described in Section D.2.1. Last row, right : Real arterial phase. GAN : Generative adversarial network.*

Qualitative results In Figure D.2, we display the synthetic arterial phase for a validation sample. Results show that the images generated by the baseline experiment (A) tend to be blurry. Reducing the value of λ (experiments B, C, D) or reducing the field of view of the discriminator (experiment F) reduced the blur but rendered images with jagged edges. By masking the input images by the pancreas mask (experiment H), the images were sharper but less realistic.

Quantitative results In Table D.1, we show the L_1 error between the true and fake arterial images, computed over all the validation samples. We also compute the ratio of the mean intensity within the lesion to the mean intensity within the pancreas, computed across all validation samples. This metric is intended to identify which generated images exhibit the highest contrast between the lesion and the surrounding pancreatic tissue. Thus, ratios far from 1 suggest that the images generated present high contrasts and may therefore be more useful for lesion detection, as explored in Section D.3.

The quantitative results do not reflect the qualitative results, and no experiment clearly stands out. We nevertheless choose GAN K for further experiments as it demonstrated the lowest L_1 error in the pancreas, the furthest lesion/parenchyma ratio from 1, and satisfactory qualitative results.

D.3 Lesion segmentation experiments

D.3.1 Description

A second series of experiments aims to assess the value of these generated images for the final objective, i.e. lesion detection.

	A - Baseline	B	C	D	E	F
L_1	0.23 ± 0.05	0.28 ± 0.05	0.26 ± 0.05	0.29 ± 0.05	0.23 ± 0.05	0.23 ± 0.05
L_1 in the pancreas	0.16 ± 0.06	0.21 ± 0.06	0.19 ± 0.06	0.22 ± 0.08	0.16 ± 0.06	0.16 ± 0.06
Ratio lesion / pancreas	0.92 ± 0.49	0.87 ± 0.38	0.93 ± 0.42	0.85 ± 0.42	0.91 ± 0.44	0.94 ± 0.48

	A - Baseline	G	H	I	J	K
L_1	0.23 ± 0.05	0.21 ± 0.05	0.45 ± 0.08	0.28 ± 0.05	0.27 ± 0.05	0.27 ± 0.05
L_1 in the pancreas	0.16 ± 0.06	0.16 ± 0.06	0.19 ± 0.06	0.17 ± 0.06	0.17 ± 0.06	0.16 ± 0.06
Ratio lesion / pancreas	0.92 ± 0.49	0.85 ± 0.48	0.85 ± 0.37	0.94 ± 0.40	0.88 ± 0.43	0.84 ± 0.43

Table D.1 – Quantitative results of the different conditional GANs trained. The first and second rows show the L_1 error between the fake and the real arterial images in the whole image and in the pancreas only, respectively. The third row shows the ratio of the mean intensity within the lesion to the mean intensity within the pancreas. Results are values computed over all the validation samples. *A, B, C, D, E, F, G, H, I, J, K : experiments described in Section D.2.1. GAN : Generative adversarial network.*

To this end, GAN K is run in a four-fold cross-validation manner in order to get, for all the 1281 samples of the database $\mathcal{D}^{Train}_{PortArt}$, a generated arterial phase obtained by validation. Then, these generated arterial phases are used as input for segmentation models trained to segment the pancreas, the lesion and the MPD. In particular, the following experiments are carried out, extending the analysis of Table 8.1 :

- **T1** : using portal CT scans as intput,
- **T2** : using arterial CT scans as input,
- **T3** : concatenation of portal and arterial CT scans as input,
- **T4** : concatenation of portal phase CT and arterial phase CT generated by GAN K as input.

Experiment T4 is performed to evaluate the improvement in lesion detection performance brought by the arterial images generated by GAN K. These experiments are carried out on $\mathcal{D}^{Train}_{PortalArt}$. Data preprocessing is automatically done according to the nnU-Net procedure. Each model is trained in a four-fold cross-validation way for 1000 epochs.

To evaluate the models at the patient level (lesion detected or not), each case is assigned a lesion risk score following the procedure detailed in Section 8.3.2.2. Table D.2 shows the results obtained on $\mathcal{D}^{Train}_{PortArt}$ by cross-validation on the whole dataset. Subgroup analyses are conducted by analyzing results on isodense tumors, Pancreatic Neuroendocrine Tumor (PNET) and small lesions (with lesion diameter inferior to 10 or 20 mm). Detection results are shown in Table D.2.

D.3.2 Results

Validation set	# of cases (# with lesions)	T1 : Portal	T2 : Arterial	T3 : [Portal, Arterial]	T4 : [Portal, A^{GAN}_H]
All	1281 (1240)	91.28 ±1.91	89.62 ±2.19	**92.12 ± 1.75**	90.93 ±1.84
Isodense	136 (95)	82.94 ±3.89	81.01 ±4.14	**84.79 ±3.56**	83.24 ±3.77
PNET	133 (92)	86.04 ±3.4	82.93 ±3.64	**88.76 ±3.02**	82.33 ±3.54
Small (<10mm)	74 (33)	56.06 ±6.99	58.62 ±6.65	**60.63 ±6.88**	55.07 ±6.97
Small (<20mm)	243 (202)	77.92 ±3.92	75.98 ±4.27	**80.93 ±3.61**	76.79 ±4.01

Table D.2 – Lesion detection results obtained on $\mathcal{D}^{Train}_{PortArt}$ by cross-validation on the four folds. A^{GAN}_K refers to arterial images generated by GAN K. Means ± standard deviations of the AUC (in %), obtained by bootstrapping with 1000 repetitions, are reported. Each column represents one experiment, with its input type written in brackets. Each row represents one subgroup, with best results per row in bold. *AUC : Area under the curve, GAN : Generative adversarial network, PNET : Pancreatic neuroendocrine tumor.*

The results shown in Table D.2 reveal that the AUC of experiment T4 is lower than that of experiment T3, meaning that the arterial images generated by GAN K fail to substitute real arterial images when used in concatenation with portal images. Moreover, experiment T4 even achieves lower detection performance than experiment T1 (using portal image input only), meaning that adding as input these synthetic arterial images is not beneficial, if not detrimental.

D.4 Conclusion

Preliminary experiments attempted to synthetically generate the arterial phase using a state-of-the-art Pix2Pix GAN. The generated arterial images were then concatenated to portal images and used as input to train a lesion segmentation model.

Despite several experiments, this work was limited by several factors. The generated images exhibited either jagged edges or blurriness. A hyper-parameter search was carried out on the GANs to obtain realistic arterial images, and one GAN experiment was selected for further investigation based on qualitative and quantitative results. The generated images were used to train a lesion segmentation model. However, analyses of the resulting model showed that the arterial images generated by the GAN did not bring any performance gain compared to baseline experiments. This means that the evaluation measures used did not allow to determine whether the synthetic images provided additional information that could enhance a lesion segmentation model, and perhaps another GAN among the different ones tested would have been more interesting. Identifying the optimal generated images for lesion segmentation proves to be costly in terms of time and computational resources, as it involves training a GAN to generate synthetic arterial phase images, followed by training a segmentation model using both the true portal image and the generated arterial image as inputs. These problems and their potential solutions are discussed further in the main body of the manuscript (see Section 8.3.2.2).

Bibliographie

L. Abel, J. Wasserthal, T. Weikert, A.W. Sauter, I. Nesic, M. Obradovic, S. Yang, S. Manneck, C. Glessgen, J.M. Ospel, et al. Automated detection of pancreatic cystic lesions on CT using deep learning. *Diagnostics*, 11(5) :901, 2021.

C. Abi Nader, R. Vétil, L.K. Wood, M.M. Rohé, A. Bône, H. Karteszi, and M.P. Vullierme. Automatic detection of pancreatic lesions and main pancreatic duct dilatation on portal venous CT scans using deep learning. *Investigative Radiology*, 2023.

H.J.W.L. Aerts, E.R. Velazquez, R.T.H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5(1) :4006, 2014.

P. Afshar, A. Mohammadi, K.N. Plataniotis, A. Oikonomou, and H. Benali. From handcrafted to deep-learning-based cancer radiomics : challenges and opportunities. *IEEE Signal Processing Magazine*, 36 (4) :132–160, 2019.

S.S. Ahn, M.J. Kim, J.Y. Choi, H.S. Hong, Y.E. Chung, and J.S. Lim. Indicative findings of pancreatic cancer in prediagnostic CT. *European Radiology*, 19 :2448–2455, 2009.

M.M. Al-Hawary, I.R. Francis, S.T. Chari, E.K. Fishman, D.M. Hough, D.S. Lu, M. Macari, A.J. Megibow, F.H. Miller, K.J. Mortele, et al. Pancreatic ductal adenocarcinoma radiology reporting template : consensus statement of the society of abdominal radiology and the american pancreatic association. *Radiology*, 270(1) :248–260, 2014.

S. Albawi, T.A. Mohammed, and S. Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology*, pages 1–6. IEEE, 2017.

N. Alves, M. Schuurmans, G. Litjens, J.S. Bosma, J. Hermans, and H. Huisman. Fully automatic deep learning framework for pancreatic ductal adenocarcinoma detection on computed tomography. *Cancers*, 14(2) :376, 2022.

American Cancer Society. Cancer facts and figures, 2014. `https://www.cancer.org/research/cancer-facts-sta-tistics/all-cancer-facts-figures/cancer-facts-figures-2014.html`, 2014. Accessed 05 October 2018.

M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B.A. Landman, G. Litjens, B. Menze, O. Ronneberger, and R.M. et al. Summers. The medical segmentation decathlon. *Nature Communications*, 13(1) :4128, 2022.

N. Antropova, B.Q. Huynh, and M.L. Giger. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Medical Physics*, 44(10) :5162–5171, 2017.

L.D. Arnold, A.V. Patel, Y. Yan, E.J. Jacobs, M.J. Thun, E.E. Calle, and G.A. Colditz. Are racial disparities in pancreatic cancer explained by smoking and overweight/obesity ? *Cancer Epidemiology, Biomarkers & Prevention*, 18(9) :2397–2405, 2009.

S. Asgari Taghanaki, K. Abhishek, J.P. Cohen, J. Cohen-Adad, and G. Hamarneh. Deep semantic segmentation of natural and medical images : a review. *Artificial Intelligence Review*, 54 :137–178, 2021.

W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, G. Vaillant, A.M. Lee, N. Aung, E. Lukaschuk, M.M. Sanghvi, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1) :1–12, 2018.

U.K. Ballehaninna and R.S. Chamberlain. The clinical utility of serum CA19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma : An evidence based appraisal. *Journal of Gastrointestinal Oncology*, 3(2) :105, 2012.

I.M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach. Comparison of deep learning approaches for multi-label chest X-ray classification. *Scientific Reports*, 9(1) :6381, 2019.

D. Bank, N. Koenigstein, and R. Giryes. Autoencoders. *Machine Learning for Data Science Handbook : Data Mining and Knowledge Discovery Handbook*, pages 353–374, 2023.

M. Barat, G. Chassagnon, A. Dohan, S. Gaujoux, R. Coriat, C. Hoeffel, C. Cassinotto, and P. Soyer. Artificial intelligence : A critical review of current applications in pancreatic imaging. *Japanese Journal of Radiology*, 39 :514–523, 2021.

T. Batra and D. Parikh. Cooperative learning with visual attributes. *arXiv preprint arXiv :1705.05512*, 2017.

C. Baur, S. Denner, B. Wiestler, N. Navab, and S. Albarqouni. Autoencoders for unsupervised anomaly segmentation in brain MR images : a comparative study. *Medical Image Analysis*, 69 :101952, 2021.

M.F. Beg, M.I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2) :139–157, 2005.

C. Bettegowda, M. Sausen, R.J. Leary, I. Kinde, Y. Wang, N. Agrawal, B.R. Bartlett, H. Wang, B. Luber, R.M. Alani, et al. Detection of circulating tumor DNA in early-and late-stage human malignancies. *Science Translational Medicine*, 6(224) :224ra24–224ra24, 2014.

G.J. Betts, P. Desaix, E. Johnson, O. Korol, D. Kruse, B. Poe, et al. Human anatomy and physiology. *OpenStax College*, 2013.

P. Bilic, P. Christ, H.B. Li, E. Vorontsov, A. Ben-Cohen, G. Kaissis, A. Szeskin, C. Jacobs, G.E.H. Mamani, G. Chartrand, et al. The liver tumor segmentation benchmark (LiTS). *Medical Image Analysis*, 84 : 102680, 2023.

A.L. Blackford, M.I. Canto, A.P. Klein, R.H. Hruban, and M. Goggins. Recent trends in the incidence and survival of stage 1A pancreatic cancer : a surveillance, epidemiology, and end results analysis. *Journal of the National Cancer Institute*, 112(11) :1162–1169, 2020.

A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Eleventh annual Conference on Computational Learning Theory*, pages 92–100, 1998.

D.E. Bockman. Anatomy of the pancreas. *The Exocrine Pancreas : Biology, Pathobiology and Disease. Raven Press, New York*, pages 1–8, 1993.

G.W. Boland, M.E. O'malley, M. Saez, C. Fernandez-del Castillo, A.L. Warshaw, and P.R. Mueller. Pancreatic-phase versus portal vein-phase helical CT of the pancreas : optimal temporal window for evaluation of pancreatic adenocarcinoma. *American Journal of Roentgenology*, 172(3) :605–608, 1999.

A. Bône, M. Louis, B. Martin, and D. Stanley. Deformetrica 4 : An open-source software for statistical shape analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–13. Springer, 2018.

P. Borghei, F. Sokhandon, A. Shirkhoda, and D.E. Morgan. Anomalies, anatomic variants, and sources of diagnostic pitfalls in pancreatic imaging. *Radiology*, 266(1) :28–36, 2013.

C. Bosetti, P. Bertuccio, E. Negri, C. La Vecchia, M.P. Zeegers, and P. Boffetta. Pancreatic cancer : overview of descriptive epidemiology. *Molecular Carcinogenesis*, 51(1) :3–13, 2012.

L. Bottou. Large-scale machine learning with stochastic gradient descent. In *19th International Conference on Computational Statistics*, pages 177–186. Springer, 2010.

BRATS. Brain tumor segmentation challenge. http://www.braintumorsegmentation.org/. Accessed February 2024.

F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, and A. Jemal. Global cancer statistics 2018 : GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA : A Cancer Journal for Clinicians*, 68(6) :394–424, 2018.

D.D. Brennan, G.A. Zamboni, V.D. Raptopoulos, and J.B. Kruskal. Comprehensive preoperative assessment of pancreatic adenocarcinoma with 64-section volumetric CT. *Radiographics*, 27(6) :1653–1666, 2007.

K.A. Brune, B. Lau, E. Palmisano, M. Canto, M.G. Goggins, R.H. Hruban, and A.P. Klein. Importance of age of onset in pancreatic cancer kindreds. *Journal of the National Cancer Institute*, 102(2) :119–126, 2010.

J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang. Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function. *arXiv preprint arXiv :1707.04912*, 2017.

L. Cai, J. Gao, and D. Zhao. A review of the application of deep learning in medical image classification and segmentation. *Annals of Translational Medicine*, 8(11), 2020.

M.I. Canto, R.H. Hruban, E.K. Fishman, I.R. Kamel, R. Schulick, Z. Zhang, M. Topazian, N. Takahashi, J. Fletcher, G. Petersen, et al. Frequent detection of pancreatic lesions in asymptomatic high-risk individuals. *Gastroenterology*, 142(4) :796–804, 2012.

M.I. Canto, F. Harinck, R.H. Hruban, G.J. Offerhaus, J.W. Poley, I. Kamel, Y. Nio, R.S. Schulick, C. Bassi, I. Kluijt, et al. International cancer of the pancreas screening (CAPS) consortium summit on the management of patients with increased risk for familial pancreatic cancer. *Gut*, 62(3) :339–347, 2013.

H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang. Swin-unet : Unet-like pure transformer for medical image segmentation. In *European Conference on Computer Vision*, pages 205–218. Springer, 2022.

K. Cao, Y. Xia, J. Yao, X. Han, L. Lambert, T. Zhang, W. Tang, G. Jin, H. Jiang, X. Fang, et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nature Medicine*, pages 1–11, 2023.

K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in Neural Information Processing Systems*, 33 :12546–12558, 2020.

B. Chauffert, F. Mornex, F. al Bonnetain, P. Rougier, C. Mariette, O. Bouché, J.F. Bosset, T. Aparicio, L. Mineur, A. Azzedine, et al. Phase III trial comparing intensive induction chemoradiotherapy (60 gy, infusional 5-FU and intermittent cisplatin) followed by maintenance gemcitabine with gemcitabine alone for locally advanced unresectable pancreatic cancer. definitive results of the 2000–01 FFCD/SFRO study. *Annals of Oncology*, 19(9) :1592–1599, 2008.

L. Chen and L. Wan. CTU-Net : automatic pancreas segmentation using a channel-wise transformer and 3D U-Net. *The Visual Computer*, 39(11) :5229–5243, 2023.

P.T. Chen, T. Wu, P. Wang, D. Chang, K.L. Liu, M.S. Wu, H.R. Roth, P.C. Lee, W.C. Liao, and W. Wang. Pancreatic cancer detection on CT scans with deep learning : a nationwide population-based study. *Radiology*, 306(1) :172–182, 2023.

S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, and J.Z. Cheng. Automatic scoring of multiple semantic attributes with multi-task feature leverage : a study on pulmonary nodules in CT images. *IEEE Transactions on Medical Imaging*, 36(3) :802–814, 2016.

T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020a.

W. Chen, R.K. Butler, Y. Zhou, R.A. Parker, C.Y. Jeon, and B.U. Wu. Prediction of pancreatic cancer based on imaging features in patients with duct abnormalities. *Pancreas*, 49(3) :413, 2020b.

X. Chen and K. He. Exploring simple siamese representation learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15750–15758, 2021.

X. Chen and E. Konukoglu. Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders. In *Medical Imaging with Deep Learning*, 2022.

X. Chen, Z. Chen, J. Li, Y-D. Zhang, X. Lin, and X. Qian. Model-driven deep learning method for pancreatic cancer segmentation based on spiral-transformation. *IEEE Transactions on Medical Imaging*, 41(1) :75–87, 2021.

G.E. Christensen, R.D. Rabbitt, and M.I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10) :1435–1447, 1996.

L.C. Chu, S. Park, S. Kawamoto, D.F. Fouladi, S. Shayesteh, E.S. Zinreich, J.S. Graves, K.M. Horton, R.H. Hruban, A.L. Yuille, et al. Utility of CT radiomics features in differentiation of pancreatic ductal adenocarcinoma from normal pancreatic tissue. *American Journal of Roentgenology*, 213(2) :349–357, 2019a.

L.C. Chu, S. Park, S. Kawamoto, Y. Wang, Y. Zhou, W. Shen, Z. Zhu, Y. Xia, L. Xie, F. Liu, et al. Application of deep learning to pancreatic cancer detection : lessons learned from our initial experience. *Journal of the American College of Radiology*, 16(9) :1338–1342, 2019b.

L.C. Chu, T. Ahmed, A. Blanco, A. Javed, E.M. Weisberg, S. Kawamoto, R.H. Hruban, K.W. Kinzler, B. Vogelstein, and E.K. Fishman. Radiologists' expectations of artificial intelligence in pancreatic cancer imaging : How good is good enough ? *Journal of Computer Assisted Tomography*, 47(6) :845–849, 2023.

T. Conroy, P. Hammel, M. Hebbar, M. Ben Abdelghani, A. C. Wei, J.L. Raoul, L. Choné, E. Francois, P. Artru, J.J. Biagi, et al. FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. *New England Journal of Medicine*, 379(25) :2395–2406, 2018.

National Cancer Institute Clinical Proteomic Tumor Analysis Consortium. Radiology data from the clinical proteomic tumor analysis consortium pancreatic ductal adenocarcinoma [CPTAC-PDA], 2018. The Cancer Imaging Archive.

T.F. Cootes, C.J. Taylor, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1) :38–50, 1995.

J.E. Corral, S. Hussein, P. Kandel, C.W. Bolan, U. Bagci, and M.B. Wallace. Deep learning to classify intraductal papillary mucinous neoplasms using magnetic resonance imaging. *Pancreas*, 48(6) :805–810, 2019.

M.I. Costache, C.A. Costache, C.I. Dumitrescu, A.A. Tica, M. Popescu, E.A. Baluta, A.C. Anghel, A. Saftoiu, and D. Dumitrescu. Which is the best imaging method in pancreatic adenocarcinoma diagnosis and staging-CT, MRI or EUS ? *Current Health Sciences Journal*, 43(2) :132, 2017.

A.V. Dalca, G. Balakrishnan, J. Guttag, and M.R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 729–738. Springer, 2018.

A. Das, C.C. Nguyen, F. Li, and B. Li. Digital image analysis of EUS images accurately differentiates pancreatic cancer from chronic pancreatitis and normal tissue. *Gastrointestinal Endoscopy*, 67(6) :861–867, 2008.

B.M. Dawant, R. Li, B. Lennon, and S. Li. Semi-automatic segmentation of the liver and its evaluation on the MICCAI 2007 grand challenge data set. *3D Segmentation in The Clinic : A Grand Challenge*, pages 215–221, 2007.

M. Del Chiaro, C. Verbeke, R. Salvia, G. Klöppel, J. Werner, C. McKay, H. Friess, R. Manfredi, E. Van Cutsem, M. Löhr, et al. European experts consensus statement on cystic tumours of the pancreas. *Digestive and Liver Disease*, 45(9) :703–711, 2013.

M. Del Chiaro, R. Segersvärd, M. Lohr, and C. Verbeke. Early detection and prevention of pancreatic cancer : is it really possible today ? *World journal of Gastroenterology*, 20(34) :12118, 2014.

J. Deng and Y. Lin. The benefits and challenges of ChatGPT : An overview. *Frontiers in Computing and Intelligent Systems*, 2(2) :81–83, 2022.

L.R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3) :297–302, 1945.

I. Dimitriou, A. Katsourakis, E. Nikolaidou, and G. Noussios. The main anatomical variations of the pancreatic duct system : review of the literature and its importance in surgical practice. *Journal of Clinical Medicine Research*, 10(5) :370, 2018.

J. DiSantostefano. International classification of diseases 10th revision (ICD-10). *Journal for Nurse Practitioners*, 5(1) :56–57, 2009.

A. Djuric-Stefanovic, D. Masulovic, J. Kostic, K. Randjic, and D. Saranovic. CT volumetry of normal pancreas : correlation with the pancreatic diameters measurable by the cross-sectional imaging, and relationship with the gender, age, and body constitution. *Surgical and Radiologic Anatomy*, 34 :811–817, 2012.

K. Dmitriev, A.E. Kaufman, A.A. Javed, R.H. Hruban, E.K. Fishman, A.M. Lennon, and J.H. Saltz. Classification of pancreatic cysts in computed tomography images using a random forest and convolutional neural network ensemble. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 3, pages 150–158, Quebec City, QC, Canada, 2017. Springer.

R.O. Dogan, H. Dogan, C. Bayrak, and T. Kayikcioglu. A two-phase approach using mask R-CNN and 3D U-Net for high-accuracy automatic segmentation of pancreas in CT imaging. *Computer Methods and Programs in Biomedicine*, 207 :106141, 2021.

A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words : Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2020.

B. Dufumier, P. Gori, J. Victor, A. Grigis, M. Wessa, P. Brambilla, P. Favre, M. Polosan, C. Mcdonald, C.M. Piguet, et al. Contrastive learning with continuous proxy meta-data for 3D MRI classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 58–68. Springer, 2021.

K.Y. Elbanna, H.J. Jang, and T.K. Kim. Imaging diagnosis and staging of pancreatic ductal adenocarcinoma : a comprehensive review. *Insights Into Imaging*, 11(1) :1–13, 2020.

M. Ezzati, S.J. Henley, A.D. Lopez, and M.J. Thun. Role of smoking in global and regional cancer epidemiology : current patterns and data needs. *International Journal of Cancer*, 116(6) :963–971, 2005.

C. Fang, G. Li, C. Pan, Y. Li, and Y. Yu. Globally guided progressive fusion network for 3D pancreas segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 2, pages 210–218, Shenzhen, China, 2019. Springer.

H. Fang, Y. Fang, and X. Yang. Multi-organ segmentation network with adversarial performance validator. *arXiv preprint arXiv :2204.07850*, 2022.

K. Fang, B. He, L. Liu, H. Hu, C. Fang, X. Huang, and F. Jia. UMRFormer-net : a three-dimensional u-shaped pancreas segmentation method based on a double-layer bridged transformer network. *Quantitative Imaging in Medicine and Surgery*, 13(3) :1619, 2023.

A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9) :1323–1341, 2012.

T.L. Fitzgerald, A.J. Smith, M. Ryan, M. Atri, F.C. Wright, C.H.L. Law, and S.S. Hanna. Surgical treatment of incidentally identified pancreatic masses. *Canadian Journal of Surgery*, 46(6) :413, 2003.

N.C. Fox, E.K. Warrington, P.A. Freeborough, P. Hartikainen, A.M. Kennedy, J.M. Stevens, and M.N. Rossor. Presymptomatic hippocampal atrophy in Alzheimer's disease : a longitudinal MRI study. *Brain*, 119(6) :2001–2007, 1996.

M. Fu, W. Wu, X. Hong, Q. Liu, J. Jiang, Y. Ou, Y. Zhao, and X. Gong. Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images. *BMC Systems Biology*, 12 :119–127, 2018.

H. Fujita. AI-based computer-aided diagnosis (AI-CAD) : the latest review to read first. *Radiological Physics and Technology*, 13(1) :6–19, 2020.

S. Gangi, J.G. Fletcher, M.A. Nathan, J.A. Christensen, W.S. Harmsen, B.S. Crownhart, and S.T. Chari. Time interval between abnormalities seen on CT and the clinical diagnosis of pancreatic cancer : retrospective review of CT scans obtained before diagnosis. *American Journal of Roentgenology*, 182(4) : 897–903, 2004.

P. Ghiorzo. Genetic predisposition to pancreatic cancer. *World Journal of Gastroenterology*, 20(31) :10778, 2014.

Globocan. Globocan 2018. https://www.uicc.org/news/global-cancer-data-globocan-2018, 2018. Accessed February 2024.

W. Gonoi, T.Y. Hayashi, H. Okuma, M. Akahane, Y. Nakai, S. Mizuno, R. Tateishi, H. Isayama, K. Koike, and K. Ohtomo. Development of pancreatic cancer is predictable well in advance using contrast-enhanced CT : a case-cohort study. *European Radiology*, 27(12) :4941–4950, 2017.

M. Goodman, J.K. Willmann, and R.B. Jeffrey. Incidentally discovered solid pancreatic masses : imaging and clinical observations. *Abdominal Imaging*, 37 :91–97, 2012.

P. Gori, O. Colliot, L. Marrakchi-Kacem, Y. Worbe, C. Poupon, A. Hartmann, N. Ayache, and S. Durrleman. A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes. *Medical Image Analysis*, 35 :458–474, 2017.

J.B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, 33 :21271–21284, 2020.

Z. Guo, L. Zhang, L. Lu, M. Bagheri, R.M. Summers, M. Sonka, and J. Yao. Deep LOGISMOS : Deep learning graph-based 3D segmentation of pancreatic tumors on CT scans. In *IEEE 15th International Symposium on Biomedical Imaging*, pages 1230–1233. IEEE, 2018.

A.E. Hanbidge et al. Cancer of the pancreas : the best image for early detection-ct, MRI, PET or us ? *Canadian Journal of Gastroenterology and Hepatology*, 16 :101–105, 2002.

A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, and D. Xu. Unetr : Transformers for 3D medical image segmentation. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 574–584, 2022.

M. Havaei, N. Guizard, N. Chapados, and Y. Bengio. Hemis : Hetero-modal image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 469–477. Springer, 2016.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *IEEE international conference on computer vision*, pages 2961–2969, 2017.

M.P. Heinrich and O. Oktay. BRIEFnet : Deep pancreas segmentation using binary sparse convolutions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 329–337. Springer, 2017.

G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *Journal of Computational Science*, 14(7) :38–39, 2015.

S.A. Hoogenboom, C.W. Bolan, A. Chuprin, M.T. Raimondo, J.E. van Hooft, M.B. Wallace, and M. Raimondo. Pancreatic steatosis on computed tomography is an early imaging feature of pre-diagnostic pancreatic cancer : A preliminary study in overweight patients. *Pancreatology*, 21(2) :428–433, 2021.

M. Hu, M. Maillard, Y. Zhang, T. Ciceri, G. La Barbera, I. Bloch, and P. Gori. Knowledge distillation from multi-modal to mono-modal segmentation networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 772–781. Springer, 2020a.

P. Hu, X. Li, Y. Tian, T. Tang, T. Zhou, X. Bai, S. Zhu, T. Liang, and J. Li. Automatic pancreas segmentation in CT images with distance-based saliency-aware DenseASPP network. *IEEE Journal of Biomedical and Health Informatics*, 25(5) :1601–1611, 2020b.

H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.W. Chen, and J. Wu. Unet 3+ : A full-scale connected unet for medical image segmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1055–1059. IEEE, 2020.

M. Huang, C. Huang, J. Yuan, and D. Kong. A semiautomated deep learning approach for pancreas segmentation. *Journal of Healthcare Engineering*, 2021, 2021.

B.Q. Huynh, H. Li, and M.L. Giger. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3) :034501–034501, 2016.

M. Ilic and I. Ilic. Epidemiology of pancreatic cancer. *World Journal of Gastroenterology*, 22(44) :9694, 2016.

F. Isensee, P.F. Jaeger, S.A.A. Kohl, J. Petersen, and K.H. Maier-Hein. nnU-Net : a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2) :203–211, 2021a.

F. Isensee, P.F. Jaeger, S.A.A. Kohl, J. Petersen, and K.H. Maier-Hein. nnU-Net : a self-configuring method for deep learning-based biomedical image segmentation. `https://github.com/MIC-DKFZ/nnUNet`, 2021b. GitHub repository of the nnU-Net, acessed 19 January 2023.

P. Isola, J.Y. Zhu, T. Zhou, and A.A. Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017.

Y. Jiang, L. Chen, H. Zhang, and X. Xiao. Breast cancer histopathological image classification using convolutional neural networks with small se-resnet module. *PLOS One*, 14(3) :e0214587, 2019.

John Hopkins Medicine. John Hopkins Medicine website. https://www.hopkinsmedicine.org/health/conditions-and-diseases/pancreatic-cancer/pancreatic-cancer-types, a. Accessed February 2024.

John Hopkins Medicine. John Hopkins Medicine website. https://www.hopkinsmedicine.org/health/conditions-and-diseases/pancreatic-cancer/pancreatic-cancer-screening, b. Accessed February 2024.

G. Kaissis, S. Ziegelmayer, F. Lohöfer, H. Algül, M. Eiber, W. Weichert, R. Schmid, H. Friess, E. Rummeny, D. Ankerst, et al. A machine learning model for the prediction of survival and tumor subtype in pancreatic ductal adenocarcinoma from preoperative diffusion-weighted imaging. *European Radiology Experimental*, 3(1) :1–9, 2019.

J.D. Kang, S.E. Clarke, and A.F. Costa. Factors associated with missed and misinterpreted cases of pancreatic ductal adenocarcinoma. *European Radiology*, 31(4) :2422–2432, 2021.

Y. Kanno, S. Koshita, T. Ogawa, H. Kusunose, K. Masu, T. Sakai, K. Yonamine, Y. Kawakami, Y. Fujii, K. Miyamoto, et al. Predictive value of localized stenosis of the main pancreatic duct for early detection of pancreatic cancer. *Clinical Endoscopy*, 52(6) :588–597, 2019.

B.S. Kelly, C. Judge, S.M. Bollard, S.M. Clifford, G.M. Healy, A. Aziz, P. Mathur, S. Islam, K.W. Yeom, A. Lawlor, et al. Radiology artificial intelligence : a systematic review and evaluation of methods (RAISE). *European Radiology*, 32(11) :7998–8007, 2022.

D.G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2) :81–121, 1984.

T. Khoury, A.W. Asombang, T.M. Berzin, J. Cohen, D.K. Pleskow, and M. Mizrahi. The clinical implications of fatty pancreas : a concise review. *Digestive Diseases and Sciences*, 62 :2658–2667, 2017.

P. Kickingereder, F. Isensee, I. Tursunova, J. Petersen, U. Neuberger, D. Bonekamp, G. Brugnara, M. Schell, T. Kessler, M. Foltyn, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks : a multicentre, retrospective study. *The Lancet Oncology*, 20 (5) :728–740, 2019.

R. Kikinis, S.D. Pieper, and K.G. Vosburgh. 3D slicer : a platform for subject-specific image analysis, visualization, and clinical support. In *Intraoperative Imaging and Image-Guided Therapy*, pages 277–289. Springer, 2013.

H. Kim and A. Mnih. Disentangling by factorising. In *International Conference on Machine Learning*, pages 2649–2658. PMLR, 2018.

J.H. Kim, S.H. Park, E.S. Yu, M.-H. Kim, J. Kim, J.H. Byun, S.S. Lee, H.J. Hwang, J.-Y. Hwang, S.S. Lee, et al. Visually isoattenuating pancreatic adenocarcinoma at dynamic-enhanced CT : frequency, clinical and pathologic characteristics, and diagnosis at imaging examinations. *Radiology*, 257(1) :87–96, 2010.

S.W. Kim, S.H. Kim, D.H. Lee, S.M. Lee, Y.S. Kim, J.Y. Jang, and J.K. Han. Isolated main pancreatic duct dilatation : CT differentiation between benign and malignant causes. *American Journal of Roentgenology*, 209(5) :1046–1055, 2017.

S.Y. Kim, H. Kim, J.Y. Cho, S. Lim, K. Cha, K.H. Lee, Y.H. Kim, J.H. Kim, Y.S. Yoon, H.S. Han, et al. Quantitative assessment of pancreatic fat by using unenhanced CT : pathologic correlation and clinical implications. *Radiology*, 271(1) :104–112, 2014.

D.P. Kingma and J. Ba. Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*, 2014.

D.P. Kingma and M. Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations 2014, Banff, AB, Canada*, 2014.

KiTS. Kidney and kidney tumor segmentation challenge. https://kits-challenge.org/kits23/. Accessed February 2024.

S. Klimov, Y. Xue, A. Gertych, R.P. Graham, Y. Jiang, S. Bhattarai, S.J. Pandol, E.A. Rakha, M.D. Reid, and R. Aneja. Predicting metastasis risk in pancreatic neuroendocrine tumors using deep learning image analysis. *Frontiers in Oncology*, 10 :593211, 2021.

P. Korfiatis, G. Suman, N.G. Patnam, K.H. Trivedi, A. Karbhari, S. Mukherjee, C. Cook, J.R. Klug, A. Patra, H. Khasawneh, et al. Automated artificial intelligence model trained on a large data set can detect pancreas cancer on diagnostic computed tomography scans as well as visually occult preinvasive cancer on prediagnostic computed tomography scans. *Gastroenterology*, 165(6) :1533–1546, 2023.

J. Kottlors, G. Bratke, P. Rauen, C. Kabbasch, T. Persigehl, M. Schlamann, and S. Lennartz. Feasibility of differential diagnosis based on imaging patterns using a large language model. *Radiology*, 308(1) : e231167, 2023.

T. Kozu, K. Suda, and F. Toki. Pancreatic development and anatomical variation. *Gastrointestinal Endoscopy Clinics of North America*, 5(1) :1–30, 1995.

J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi. Learning a probabilistic model for diffeomorphic registration. *IEEE Transactions on Medical Imaging*, 38(9) :2165–2176, 2019.

G.J. Krejs. Pancreatic cancer : epidemiology and risk factors. *Digestive Diseases*, 28(2) :355–358, 2010.

M. Kriegsmann, K. Kriegsmann, G. Steinbuss, C. Zgorzelski, A. Kraft, and M.M. Gaida. Deep learning in pancreatic tissue : Identification of anatomical structures, pancreatic intraepithelial neoplasia, and ductal adenocarcinoma. *International Journal of Molecular Sciences*, 22(10) :5385, 2021.

N. Kroell. imea : A python package for extracting 2D and 3D shape measurements from images. *Journal of Open Source Software*, 6(60) :3091, 2021.

D. Kumar, A. Wong, and D.A. Clausi. Lung nodule classification using deep features in CT images. In *Conference on Computer and Robot Vision*, pages 133–138. IEEE, 2015.

H. Kumar, S.V. DeSouza, and M.S. Petrov. Automated pancreas segmentation from computed tomography and magnetic resonance images : A systematic review. *Computer Methods and Programs in Biomedicine*, 178 :319–328, 2019a.

H. Kumar, S.V. DeSouza, and M.S. Petrov. Automated pancreas segmentation from computed tomography and magnetic resonance images : A systematic review. *Computer Methods and Programs in Biomedicine*, 178 :319–328, 2019b.

V. Kumar, Y. Gu, S. Basu, A. Berglund, S.A. Eschrich, M.B. Schabath, K. Forster, H.J.W.L. Aerts, A. Dekker, D. Fenstermacher, et al. Radiomics : the process and the challenges. *Magnetic Resonance Imaging*, 30(9) :1234–1248, 2012.

G. La Barbera, H. Boussaid, F. Maso, S. Sarnacki, L. Rouet, P. Gori, and I. Bloch. Anatomically constrained CT image translation for heterogeneous blood vessel segmentation. In *The 33rd British Machine Vision Conference*, 2022.

P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G.P.M. Van Stiphout, P. Granton, C.M.L. Zegers, R. Gillies, R. Boellard, A. Dekker, et al. Radiomics : extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4) :441–446, 2012.

J. Lao, Y. Chen, Z.C. Li, Q. Li, J. Zhang, J. Liu, and G. Zhai. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Scientific Reports*, 7(1) :10353, 2017.

S.C. Larsson and A. Wolk. Red and processed meat consumption and risk of pancreatic cancer : meta-analysis of prospective studies. *British Journal of Cancer*, 106(3) :603–607, 2012.

Y. LeCun, Y. Bengio, et al. Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361(10) :1995, 1995.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553) :436–444, 2015.

E.S. Lee and J.M. Lee. Imaging diagnosis of pancreatic cancer : a state-of-the-art review. *World Journal of Gastroenterology*, 20(24) :7864, 2014.

K.S. Lee, A. Sekhar, N.M. Rofsky, and I. Pedrosa. Prevalence of incidental pancreatic cysts in the adult population on mr imaging. *Official journal of the American College of Gastroenterology— ACG*, 105(9) : 2079–2084, 2010.

P.J. Lee and G.I. Papachristou. New insights into acute pancreatitis. *Nature Reviews Gastroenterology & Hepatology*, 16(8) :479–496, 2019.

P. Lévy, V. Rebours, M. Barthet, and L. Buscail. Un plaidoyer pour la pancréatologie ! *Hépato-Gastro & Oncologie Digestive*, 24(4) :307–318, 2017.

B. Li, Y. Li, and K.W Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2021a.

H. Li, J. Li, X. Lin, and X. Qian. A model-driven stack-based fully convolutional network for pancreas segmentation. In *5th International Conference on Communication, Image and Signal Processing*, pages 288–293. IEEE, 2020.

J. Li, X. Lin, H. Che, H. Li, and X. Qian. Pancreas segmentation with probabilistic map guided bi-directional recurrent UNet. *Physics in Medicine & Biology*, 66(11) :115010, 2021b.

J. Li, L. Qi, Q. Chen, Y.D. Zhang, and X. Qian. A dual meta-learning framework based on idle data for enhancing segmentation of pancreatic cancer. *Medical Image Analysis*, 78 :102342, 2022.

M. Li, F. Lian, C. Wang, and S. Guo. Accurate pancreas segmentation using multi-level pyramidal pooling residual U-Net with adversarial mechanism. *BMC Medical Imaging*, 21(1) :1–8, 2021c.

Q. Li, X. Liu, Y. He, D. Li, and J. Xue. Temperature guided network for 3D joint segmentation of the pancreas and tumors. *Neural Networks*, 157 :387–403, 2023.

G. Lippi and C. Mattiuzzi. The global burden of pancreatic cancer. *Archives of Medical Science*, 16(1), 2020.

G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, and C.I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42 :60–88, 2017.

F. Liu, L. Xie, Y. Xia, E. Fishman, and A. Yuille. Joint shape representation and classification for detecting PDAC. In *International Workshop on Machine Learning in Medical Imaging*, pages 212–220. Springer, 2019a.

F. Liu, L. Xie, Y. Xia, E. Fishman, and A. Yuille. Joint shape representation and classification for detecting PDAC. In *International Workshop on Machine Learning in Medical Imaging*, pages 212–220. Springer, 2019b.

F. Liu, Y. Zhou, E. Fishman, and A. Yuille. Fusionnet : Incorporating shape and texture for abnormality detection in 3D abdominal ct scans. In *Machine Learning in Medical Imaging (MLMI) Workshop*, volume 10, pages 221–229, Shenzhen, China, 2019c. Springer.

K.L. Liu, T. Wu, P.T. Chen, Y.M. Tsai, H. Roth, M.S. Wu, W.C. Liao, and W. Wang. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue : a retrospective study with cross-racial external validation. *The Lancet Digital Health*, 2(6) :e303–e313, 2020.

S. Liu, Y. Xie, A. Jirapatnakul, and A.P. Reeves. Pulmonary nodule classification in lung cancer screening with three-dimensional convolutional neural networks. *Journal of Medical Imaging*, 4(4) :041308–041308, 2017.

S. Liu, X. Yuan, R. Hu, S. Liang, S. Feng, Y. Ai, and Y. Zhang. Automatic pancreas segmentation via coarse location and ensemble learning. *IEEE Access*, 8 :2906–2914, 2019d.

X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang. Self-supervised learning : Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 35(1) :857–876, 2021a.

Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer : Hierarchical vision transformer using shifted windows. In *IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021b.

P.J. Loehrer Sr, Y. Feng, H. Cardenes, L. Wagner, J. M. Brell, D. Cella, P. Flynn, R.K. Ramanathan, C.H. Crane, S. R. Alberts, et al. Gemcitabine alone versus gemcitabine plus radiotherapy in patients with locally advanced pancreatic cancer : an eastern cooperative oncology group trial. *Journal of Clinical Oncology*, 29(31) :4105, 2011.

J-M. Löhr, N. Panic, M. Vujasinovic, and C.S. Verbeke. The ageing pancreas : a systematic review of the evidence and analysis of the consequences. *Journal of Internal Medicine*, 283(5) :446–460, 2018.

J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

C. Lu, C. F. Xu, X.Y. Wan, H.T. Zhu, C.H. Yu, and Y.M. Li. Screening for pancreatic cancer in familial high-risk individuals : A systematic review. *World Journal of Gastroenterology*, 21(28) :8678, 2015.

L. Lu, L. Jian, J. Luo, and B. Xiao. Pancreatic segmentation via ringed residual U-Net. *IEEE Access*, 7 : 172871–172878, 2019.

J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A.L Martel. Loss odyssey in medical image segmentation. *Medical Image Analysis*, 71 :102035, 2021.

S. Macdonald, U. Macleod, N.C. Campbell, D. Weller, and E. Mitchell. Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. *British Journal of Cancer*, 94(9) :1272–1280, 2006.

S. Majumder, N.A. Philip, N. Takahashi, M.J. Levy, V.P. Singh, and S.T. Chari. Fatty pancreas : should we be concerned ? *Pancreas*, 46(10) :1251, 2017.

M. Malvezzi, G. Carioli, P. Bertuccio, T. Rosso, P. Boffetta, F. Levi, C. La Vecchia, and E. Negri. European cancer mortality predictions for the year 2016 with focus on leukaemias. *Annals of Oncology*, 27(4) : 725–731, 2016.

Y. Man, Y. Huang, J. Feng, X. Li, and F. Wu. Deep q learning driven CT pancreas segmentation with geometry-aware U-Net. *IEEE Transactions on Medical Imaging*, 38(8) :1971–1980, 2019.

D.V. Mann, R. Edwards, S. Ho, W.Y. Lau, and G. Glazer. Elevated tumour marker CA19-9 : clinical interpretation and influence of obstructive jaundice. *European Journal of Surgical Oncology*, 26(5) : 474–479, 2000.

K. Marstal, F. Berendsen, M. Staring, and S. Klein. Simpleelastix : A user-friendly, multi-lingual library for medical image registration. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 134–142, 2016.

L. Marti-Bonmati, L. Cerdá-Alberich, A. Pérez-Girbés, R. Díaz Beveridge, E. Montalvá Orón, J. Pérez Rojas, and A. Alberich-Bayarri. Pancreatic cancer, radiomics and artificial intelligence. *The British Journal of Radiology*, 95(1137) :20220072, 2022.

Y. Matsuda. Age-related morphological changes in the pancreas and their association with pancreatic carcinogenesis. *Pathology International*, 69(8) :450–462, 2019.

L.M. Mazur, P.R. Mosaly, L.M. Hoyle, E.L. Jones, B.S. Chera, and L.B. Marks. Relating physician's workload with errors during radiation therapy planning. *Practical Radiation Oncology*, 4(2) :71–75, 2014.

S. Midha, S. Chawla, and P.K. Garg. Modifiable and non-modifiable risk factors for pancreatic cancer : A review. *Cancer Letters*, 381(1) :269–277, 2016.

K.W. Millikan, D.J. Deziel, J.C. Silverstein, T.M. Kanjo, J.D. Christein, A. Doolas, and R.A. Prinz. Prognostic factors associated with resectable adenocarcinoma of the head of the pancreas. *The American Surgeon*, 65(7) :618–624, 1999.

R.M.S. Mitchell, M.F. Byrne, and J. Baillie. Pancreatitis. *The Lancet*, 361(9367) :1447–1455, 2003.

S. Miura, K. Kume, K. Kikuta, S. Hamada, T. Takikawa, N. Yoshida, S. Hongo, Y. Tanaka, R. Matsumoto, T. Sano, et al. Focal parenchymal atrophy and fat replacement are clues for early diagnosis of pancreatic cancer with abnormalities of the main pancreatic duct. *The Tohoku Journal of Experimental Medicine*, 252(1) :63–71, 2020.

J.D. Mizrahi, R. Surana, J.W. Valle, and R.T. Shroff. Pancreatic cancer. *The Lancet*, 395(10242) :2008–2020, 2020.

J. Mo, L. Zhang, Y. Wang, and H. Huang. Iterative 3D feature enhancement network for pancreas segmentation from CT images. *Neural Computing and Applications*, 32 :12535–12546, 2020.

F. Mohsen, H. Ali, N. El Hajj, and Z. Shah. Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Scientific Reports*, 12(1) :17981, 2022.

A. Momeni-Boroujeni, E. Yousefi, and J. Somma. Computer-assisted cytologic diagnosis in pancreatic FNA : an application of neural networks to image analysis. *Cancer Cytopathology*, 125(12) :926–933, 2017.

S. Mukherjee, A. Patra, H. Khasawneh, P. Korfiatis, N. Rajamohan, G. Suman, S. Majumder, A. Panda, M.P. Johnson, N.B. Larson, et al. Radiomics-based machine learning models can detect pancreatic cancer on prediagnostic cts at a substantial lead time prior to clinical diagnosis. *Gastroenterology*, pages S0016–5085, 2022.

S. Naudin, V. Viallon, D. Hashim, H. Freisling, M. Jenab, E. Weiderpass, F. Perrier, F. McKenzie, H.B. Bueno-de Mesquita, and A. et al. Olsen. Healthy lifestyle and the risk of pancreatic cancer in the EPIC study. *European Journal of Epidemiology*, 35 :975–986, 2020.

L. Nazario-Johnson, H.A. Zaki, and G.A. Tung. Use of large language models to predict neuroimaging. *Journal of the American College of Radiology*, 20(10) :1004–1009, 2023.

J.P. Neoptolemos, J.A. Dunn, D.D. Stocken, J. Almond, K. Link, H. Beger, C. Bassi, M. Falconi, P. Pederzoli, C. Dervenis, et al. Adjuvant chemoradiotherapy and chemotherapy in resectable pancreatic cancer : a randomised controlled trial. *The Lancet*, 358(9293) :1576–1585, 2001.

C. Niederau and J.H. Grendell. Diagnosis of pancreatic carcinoma : imaging techniques and tumor markers. *Pancreas*, 7(1) :66–86, 1992.

S. Nikolov, S. Blackwell, A. Zverovitch, R. Mendes, M. Livne, J. De Fauw, Y. Patel, C. Meyer, H. Askham, B. Romera-Paredes, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. *arXiv preprint arXiv :1809.04430*, 2018.

M. Nishio, S. Noguchi, and K. Fujimoto. Automatic pancreas segmentation using coarse-scaled 2D model of deep learning : usefulness of data augmentation and deep U-Net. *Applied Sciences*, 10(10) :3360, 2020.

H. Okasha, S. Elkholy, R. El-Sayed, M.N. Wifi, M. El-Nady, W. El-Nabawi, W.A. El-Dayem, M.I. Radwan, A. Farag, Y. El-Sherif, et al. Real time endoscopic ultrasound elastography and strain ratio in the diagnosis of solid pancreatic lesions. *World Journal of Gastroenterology*, 23(32) :5962, 2017.

O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al. Attention U-Net : Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*, 2022.

M. Orellana-Donoso, D. Milos-Brandenberg, A. Benavente-Urtubia, J. Guerra-Loyola, A. Bruna-Mejias, P. Nova-Baeza, Á. Becerra-Farfán, W. Sepulveda-Loyola, R.M. Luque-Bernal, and J.J. Valenzuela-Fuenzalida. Incidence and clinical implications of anatomical variations in the pancreas and its ductal system : A systematic review and meta-analysis. *Life*, 13(8) :1710, 2023.

S. Ostmeier, B. Axelrod, J. Bertels, F. Isensee, M.G. Lansberg, S. Christensen, G.W. Albers, L-J. Li, and J.J. Heit. Evaluation of medical image segmentation models for uncertain, small or empty reference annotations. *arXiv preprint arXiv :2209.13008*, 2022.

D.K. Owens, K.W. Davidson, A.H. Krist, M.J. Barry, M. Cabana, A.B. Caughey, S.J. Curry, C.A. Doubeni, J.W. Epling, M. Kubik, et al. Screening for pancreatic cancer : US preventive services task force reaffirmation recommendation statement. *Journal of the American Medical Association*, 322(5) :438–444, 2019.

M. Ozkan, M. Cakiroglu, O. Kocaman, M. Kurt, B. Yilmaz, G. Can, U. Korkmaz, E. Dandil, and Z. Eksi. Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. *Endoscopic Ultrasound*, 5(2) :101, 2016.

P. Paluszkiewicz, K. Smolińska, I. Debińska, and W.A. Turski. Main dietary compounds and pancreatic cancer risk. the quantitative analysis of case-control and cohort studies. *Cancer Epidemiology*, 36(1) : 60–67, 2012.

Nicolas Parent. Imagerie du Pancreas. `http://memoires.scd.univ-tours.fr/Medecine/Theses/2012_Medecine_ParentNicolas/web/html/index.html`, 2012. Accessed November 2023.

H.J. Park, K. Shin, M.W. You, S.G. Kyung, S.Y. Kim, S.H. Park, J.H. Byun, N. Kim, and H.J. Kim. Deep learning-based detection of solid and cystic pancreatic neoplasms at contrast-enhanced CT. *Radiology*, 306(1) :140–149, 2023.

R. Paul, S.H. Hawkins, Y. Balagurunathan, M. Schabath, R.J. Gillies, L.O. Hall, and D.B. Goldgof. Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. *Tomography*, 2(4) :388–395, 2016.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn : Machine learning in python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.

X. Pennec. Intrinsic statistics on Riemannian manifolds : Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1) :127–154, 2006.

S.P. Pereira, L. Oldfield, A. Ney, P.A. Hart, M.G. Keane, S.J. Pandol, D. Li, W. Greenhalf, C.Y. Jeon, E.J. Koay, et al. Early detection of pancreatic cancer. *The Lancet Gastroenterology & Hepatology*, 5(7) : 698–710, 2020.

N.L. Pernick, F.H. Sarkar, P.A. Philip, P. Arlauskas, A.F. Shields, V.K. Vaitkevicius, M.C. Dugan, and N.V. Adsay. Clinicopathologic analysis of pancreatic adenocarcinoma in african americans and caucasians. *Pancreas*, 26(1) :28–32, 2003.

P.A. Philip, J. Lacy, F. Portales, A. Sobrero, R. Pazo-Cid, J.L. M. Mozo, E.J. Kim, S. Dowden, A. Zakari, C. Borg, et al. Nab-paclitaxel plus gemcitabine in patients with locally advanced pancreatic cancer (LAPACT) : a multicentre, open-label phase 2 study. *The lancet Gastroenterology & Hepatology*, 5(3) : 285–294, 2020.

F. Proietto Salanitri, G. Bellitto, I. Irmakci, S. Palazzo, U. Bagci, and C. Spampinato. Hierarchical 3D feature learning forpancreas segmentation. In *Machine Learning in Medical Imaging (MLMI) Workshop*, volume 12, pages 238–247, Strasbourg, France, 2021. Springer.

C. Qiu, Z. Liu, Y. Song, J. Yin, K. Han, Y. Zhu, Y. Liu, and V.S. Sheng. RTUNet : Residual transformer unet specifically for pancreas segmentation. *Biomedical Signal Processing and Control*, 79 :104173, 2023.

J. Qu, X. Wei, and X. Qian. Generalized pancreatic cancer diagnosis via multiple instance learning and anatomically-guided shape normalization. *Medical Image Analysis*, 86 :102774, 2023a.

T. Qu, X. Li, X. Wang, W. Deng, L. Mao, M. He, X. Li, Y. Wang, Z. Liu, L. Zhang, et al. Transformer guided progressive fusion network for 3D pancreas and pancreatic mass segmentation. *Medical Image Analysis*, 86 :102801, 2023b.

K. Quencer, A. Kambadakone, D. Sahani, and A.S.R. Guimaraes. Imaging of the pancreas : part 1. *Applied Radiology*, 42(09) :14–20, 2013.

Radiopaedia. Radiopaedia. https://radiopaedia.org/cases/normal-pancreas-ultrasound?lang=us, a. Accessed November 2023.

Radiopaedia. Radiopaedia. https://radiopaedia.org/, b. Accessed November 2023.

F. Rahman, M. Cotterchio, S.P. Cleary, and S. Gallinger. Association between alcohol consumption and pancreatic cancer risk : a case-control study. *PLOS One*, 10(4) :e0124489, 2015.

S. Raimondi, A.B. Lowenfels, A.M. Morselli-Labate, P. Maisonneuve, and R. Pezzilli. Pancreatic cancer in chronic pancreatitis ; aetiology, incidence, and early detection. *Best Practice & Research Clinical Gastroenterology*, 24(3) :349–358, 2010.

D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang. Deep learning for health informatics. *Journal of Biomedical and Health Informatics*, 21(1) :4–21, 2016.

P. Rawla, T. Sunkara, and V. Gaduputi. Epidemiology of pancreatic cancer : global trends, etiology and risk factors. *World Journal of Oncology*, 10(1) :10, 2019.

T.J. Re, A. Lemke, M. Klauss, F.B. Laun, D. Simon, K. Grünberg, S. Delorme, L. Grenacher, R. Manfredi, and R. P. et al. Mucelli. Enhancing pancreatic adenocarcinoma delineation in diffusion derived intravoxel incoherent motion f-maps through automatic vessel and duct segmentation. *Magnetic Resonance in Medicine*, 66(5) :1327–1332, 2011.

O. Ronneberger, P. Fischer, and T. Brox. U-Net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

H.F. Roth, A. Farag, E.B. Turkbey, L. Lu, J. Liu, and R.M. Summers. Data from pancreas-CT. The Cancer Imaging Archive, 2016a.

H.R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E.B. Turkbey, and R.M. Summers. Deeporgan : Multi-level deep convolutional networks for automated pancreas segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 556–564. Springer, 2015.

H.R. Roth, L. Lu, A. Farag, A. Sohn, and R.M. Summers. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016b.

J. Rueckel, L. Trappmann, B. Schachtner, P. Wesp, B.F. Hoppe, N. Fink, J. Ricke, J. Dinkel, M. Ingrisch, and B.O. Sabel. Impact of confounding thoracic tubes and pleural dehiscence extent on artificial intelligence pneumothorax detection in chest radiographs. *Investigative Radiology*, 55(12) :792–798, 2020.

C. Ruppli, P. Gori, R. Ardon, and I. Bloch. Optimizing transformations for contrastive learning in a differentiable framework. In *Medical Image Learning with Limited and Noisy Data (MILLanD) Workshop*, pages 96–105, Singapore, 2022. Springer.

T. Sachs, W.B. Pratt, M.P. Callery, and C.M. Vollmer. The incidental asymptomatic pancreatic lesion : nuisance or threat ? *Journal of Gastrointestinal Surgery*, 13 :405–415, 2009.

A. Săftoiu, P. Vilmann, F. Gorunescu, D.I. Gheonea, M. Gorunescu, T. Ciurea, G.L. Popescu, A. Iordache, H. Hassan, and S. Iordache. Neural network analysis of dynamic sequences of EUS elastography used for the differential diagnosis of chronic pancreatitis and pancreatic cancer. *Gastrointestinal Endoscopy*, 68(6) :1086–1094, 2008.

S.S.M. Salehi, D. Erdogmus, and A. Gholipour. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *International Workshop on Machine Learning in Medical Imaging*, pages 379–387. Springer, 2017.

E. Santo and I. Bar-Yishay. Pancreatic solid incidentalomas. *Endoscopic Ultrasound*, 6(Suppl 3) :S99, 2017.

E. Sarfati, A. Bône, M.M. Rohé, P. Gori, and I. Bloch. Weakly-supervised positional contrastive learning : application to cirrhosis classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 227–237. Springer, 2023.

D. Sarwinda, R.H. Paradisa, A. Bustamam, and P. Anggia. Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer. *Procedia Computer Science*, 179 : 423–431, 2021.

A. Sauvanet, S. Gaujoux, B. Blanc, A. Couvelard, S. Dokmak, M.P. Vullierme, P. Ruszniewski, J. Belghiti, and P. Lévy. Parenchyma-sparing pancreatectomy for presumed noninvasive intraductal papillary mucinous neoplasms of the pancreas. *Annals of Surgery*, 260(2) :364–371, 2014.

C. Scapicchio, M. Gabelloni, A. Barucci, D. Cioni, L. Saba, and E. Neri. A deep look into radiomics. *La Radiologia Medica*, 126(10) :1296–1311, 2021.

D. Schlanger, F. Graur, C. Popa, E. Moiș, and N. Al Hajjar. The role of artificial intelligence in pancreatic surgery : A systematic review. *Updates in Surgery*, 74(2) :417–429, 2022.

M. Schmidt-Hansen, S. Berendse, and W. Hamilton. Symptoms of pancreatic cancer in primary care : a systematic review. *Pancreas*, 45(6) :814–818, 2016.

S. Seidlitz, J. Sellner, J. Odenthal, B. Özdemir, A. Studier-Fischer, S. Knödler, L. Ayala, T.J Adler, H.G Kenngott, M. Tizabi, et al. Robust deep learning-based semantic organ segmentation in hyperspectral images. *Medical Image Analysis*, 80 :102488, 2022.

M.J. Shafiee, A.G. Chung, F. Khalvati, M.A. Haider, and A. Wong. Discovery radiomics via evolutionary deep radiomic sequencer discovery for pathologically proven lung cancer detection. *Journal of Medical Imaging*, 4(4) :041305–041305, 2017.

N. Sharma and L.M. Aggarwal. Automated medical image segmentation techniques. *Journal of Medical Physics/Association of Medical Physicists of India*, 35(1) :3, 2010.

A. Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D : Nonlinear Phenomena*, 404 :132306, 2020.

S.V. Shrikhande, S.G. Barreto, M. Goel, and S. Arya. Multimodality imaging of pancreatic ductal adenocarcinoma : a review of the literature. *The Official Journal of the International Hepato Pancreato Biliary Association*, 14(10) :658–668, 2012.

A.L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B.A. Landman, G. Litjens, B. Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv :1902.09063*, 2019.

K. Sirinukunwattana, J.P.W. Pluim, H. Chen, X. Qi, P.-A. Heng, Y.B. Guo, L.Y. Wang, B.J. Matuszewski, E. Bruni, and U. et al. Sanchez. Gland segmentation in colon histology images : The glas challenge contest. *Medical Image Analysis*, 35 :489–502, 2017.

J.A. Soto, B.C. Lucey, and J.W. Stuhlfaut. Pancreas divisum : depiction with multi-detector row CT. *Radiology*, 235(2) :503–508, 2005.

R.Z. Stolzenberg-Solomon, C. Schairer, S. Moore, A. Hollenbeck, and D.T. Silverman. Lifetime adiposity and risk of pancreatic cancer in the NIH-AARP diet and health study cohort. *American Journal of Clinical Nutrition*, 98(4) :1057–1065, 2013.

G. Suman, A. Patra, P. Korfiatis, S. Majumder, S.K. Kang, S. Mukherjee, C. Cook, J.R. Klug, N.G. Patnam, K.H. Trivedi, et al. Quality gaps in public pancreas imaging datasets : Implications & challenges for AI applications. *Gastroenterology*, 160(6) :1881–1895, 2021.

H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. Global cancer statistics 2020 : GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA : A Cancer Journal for Clinicians*, 71(3) :209–249, 2021.

Synapse Multi Organ dataset. Multi-atlas labeling beyond the cranial vault - workshop and challenge. `https://www.synapse.org//#!Synapse:syn3193805/wiki/217789`, 2013. Accessed December 2023.

M. Takahashi, M. Hori, R. Ishigamori, M. Mutoh, T. Imai, and H. Nakagama. Fatty pancreas : A possible risk factor for pancreatic cancer in animals and humans. *Cancer science*, 109(10) :3013–3023, 2018.

T. Tamada, K. Ito, N. Kanomata, T. Sone, A. Kanki, A. Higaki, M. Hayashida, and A. Yamamoto. Pancreatic adenocarcinomas without secondary signs on multiphasic multidetector CT : association with clinical and histopathologic features. *European Radiology*, 26 :646–655, 2016.

S. Tanaka, A. Nakaizumi, T. Ioka, O. Oshikawa, H. Uehara, M. Nakao, K. Yamamoto, O. Ishikawa, H. Ohigashi, and T. Kitamra. Main pancreatic duct dilatation : a sign of high risk for pancreatic cancer. *Japanese Journal of Clinical Oncology*, 32(10) :407–411, 2002.

S. Tanaka, M. Nakao, T. Ioka, R. Takakura, Y. Takano, H. Tsukuma, H. Uehara, R. Suzuki, and J. Fukuda. Slight dilatation of the main pancreatic duct and presence of pancreatic cysts as predictive signs of pancreatic cancer : a prospective study. *Radiology*, 254(3) :965–972, 2010.

E. Tartaglione, C.A. Barbano, and M. Grangetto. EnD : Entangling and disentangling deep representations for bias correction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13508–13517, 2021.

A.J. Thirunavukarasu, D.S.J. Ting, K. Elangovan, L. Gutierrez, T.F. Tan, and D.S.W. Ting. Large language models in medicine. *Nature Medicine*, 29(8) :1930–1940, 2023.

D.W. Thompson. *On Growth and Form*. Cambridge Univ. Press, 1917.

J.R. Treadwell, H.M. Zafar, M.D. Mitchell, K. Tipton, U. Teitelbaum, and J. Jue. Imaging tests for the diagnosis and staging of pancreatic adenocarcinoma : a meta-analysis. *Pancreas*, 45(6) :789–795, 2016.

A. Turečková, T. Tureček, Z. Komínková Oplatková, and A. Rodríguez-Sánchez. Improving CT image tumor segmentation through deep supervision and attentional gates. *Frontiers in Robotics and AI*, 7 : 106, 2020.

A. Tversky. Features of similarity. *Psychological Review*, 84(4) :327, 1977.

B. Vachiranubhap, Y.H. Kim, N.C. Balci, and R.C. Semelka. Magnetic resonance imaging of adenocarcinoma of the pancreas. *Topics in Magnetic Resonance Imaging*, 20(1) :3–9, 2009.

C. Valls, E. Andía, A. Sanchez, J. Fabregat, O. Pozuelo, J.C. Quintero, R. Serrano, F. Garcia-Borobia, and R. Jorba. Dual-phase helical CT of pancreatic adenocarcinoma. *American Journal of Roentgenology*, 178(4) :821–826, 2002.

S. Van der Walt, J.L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner, N. Yager, E. Gouillart, and T. Yu. scikit-image : image processing in python. *Journal of Life & Environmental Sciences*, 2 : e453, 2014.

J.J.M. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H.J.W.L. Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, 77(21) :e104–e107, 2017.

V. Vapnik, R. Izmailov, et al. Learning using privileged information : similarity control and knowledge transfer. *Journal of Machine Learning Research*, 16(1) :2023–2049, 2015.

H.F.A. Vasen, B. Boekestijn, I.S. Ibrahim, A. Inderson, B.A. Bonsing, W.H. de V. tot Nederveen, S. Feshtali, M.N. Wasser, et al. Dilatation of the main pancreatic duct as first manifestation of small pancreatic ductal adenocarcinomas detected in a hereditary pancreatic cancer surveillance program. *The Official Journal of the International Hepato Pancreato Biliary Association*, 21(10) :1371–1375, 2019.

Y. Velikova, W. Simson, M. Salehi, M.F. Azampour, P. Paprottka, and N. Navab. CACTUSS : Common anatomical CT-US space for US examinations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 492–501. Springer, 2022.

E. Versteijne, M. Suker, K. Groothuis, J.M. Akkermans-Vogelaar, M.G. Besselink, B.A. Bonsing, J. Buijsen, O. R. Busch, G.J.M. Creemers, R.M. van Dam, et al. Preoperative chemoradiotherapy versus immediate surgery for resectable and borderline resectable pancreatic cancer : results of the dutch randomized phase III PREOPANC trial. *Journal of Clinical Oncology*, 38(16) :1763, 2020.

R. Vétil, C. Abi-Nader, A. Bône, M.-P. Vullierme, M.-M. Rohé, P. Gori, and I. Bloch. Learning shape distributions from large databases of healthy organs : applications to zero-shot and few-shot abnormal pancreas detection. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 2, pages 464–473, Singapore, 2022a. Springer.

R. Vétil, A. Bône, M.-P. Vullierme, M.-M. Rohé, P. Gori, and I. Bloch. Improving the automatic segmentation of elongated organs using geometrical priors. In *IEEE 19th International Symposium on Biomedical Imaging*, pages 1–4, Kolkata, India, 2022b. IEEE.

R. Vétıl, C. Abi-Nader, A. Bône, M.-P. Vullierme, M.-M. Rohé, P. Gori, and I. Bloch. Détection automatique de lésions pancréatiques par apprentissage profond. Séance de communications orales : l'Intelligence Artificielle en Imagerie oncologique et cardiaque, 10 2022. 70èmes Journées Francophones de Radiologie.

C.G.A. Viviers, M. Ramaekers, P.H.N. de With, D. Mavroeidis, J. Nederend, M. Luyer, and F. van der Sommen. Improved pancreatic tumor detection by utilizing clinically-relevant secondary features. In *Cancer Prevention through Early Detection (CaPTion) Workshop*, pages 139–148, Singapore, 2022. Springer.

F.M. Walter, K. Mills, S.C. Mendonça, G.A. Abel, B. Basu, N. Carroll, S. Ballard, J. Lancaster, W. Hamilton, and G.P. et al. Rubin. Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study) : a prospective cohort study. *The Lancet Gastroenterology & Hepatology*, 1(4) :298–306, 2016.

R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A.K. Nandi. Medical image segmentation using deep learning : A survey. *IET Image Processing*, 16(5) :1243–1267, 2022.

Y. Wang, Y. Zhou, W. Shen, S. Park, E.K. Fishman, and A.L. Yuille. Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. *Medical Image Analysis*, 55 :88–102, 2019.

Y. Wang, G. Gong, D. Kong, Q. Li, J. Dai, H. Zhang, J. Qu, X. Liu, and J. Xue. Pancreas segmentation using a dual-input v-mesh network. *Medical Image Analysis*, 69 :101958, 2021.

D.C. Whitcomb. Mechanisms of disease : advances in understanding the mechanisms leading to chronic pancreatitis. *Nature Clinical Practice Gastroenterology & Hepatology*, 1(1) :46–52, 2004.

J.M. Wolterink, A.M. Dinkla, M.H.F. Savenije, P.R. Seevinck, C.A.T. van den Berg, and I. Išgum. Deep MR to CT synthesis using unpaired data. In *Simulation and Synthesis in Medical Imaging (SASHIMI) Workshop,*, volume 2, pages 14–23, Québec City, QC, Canada, 2017. Springer.

Y. Xia, Q. Yu, W. Shen, Y. Zhou, E.K. Fishman, and A.L. Yuille. Detecting pancreatic ductal adenocarcinoma in multi-phase CT scans via alignment ensemble. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 285–295. Springer, 2020.

Y. Xia, J. Yao, L. Lu, L. Huang, G. Xie, J. Xiao, A. Yuille, K. Cao, and L. Zhang. Effective pancreatic cancer screening on non-contrast CT scans via anatomy-aware transformers. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 259–269. Springer, 2021.

Y. Xia, Q. Yu, L. Chu, S. Kawamoto, S. Park, F. Liu, J. Chen, Z. Zhu, B. Li, Z. Zhou, et al. The felix project : deep networks to detect pancreatic neoplasms. *medRxiv*, pages 2022–09, 2022.

X. Xiao, S. Lian, Z. Luo, and S. Li. Weighted res-unet for high-quality retina vessel segmentation. In *9th International Conference on Information Technology in Medicine and Education*, pages 327–331. IEEE, 2018.

K. Yamao, M. Takenaka, R. Ishikawa, A. Okamoto, T. Yamazaki, A. Nakai, S. Omoto, K. Kamata, K. Minaga, I. Matsumoto, et al. Partial pancreatic parenchymal atrophy is a new specific finding to diagnose small pancreatic cancer ($\leq$ 10 mm) including carcinoma in situ : comparison with localized benign main pancreatic duct stenosis patients. *Diagnostics*, 10(7) :445, 2020.

Y. Yan and D. Zhang. Multi-scale u-like network with attention mechanism for automatic pancreas segmentation. *PLOS One*, 16(5) :e0252287, 2021.

X. Yang, R. Kwitt, M. Styner, and M. Niethammer. Quicksilver : Fast predictive image registration - A deep learning approach. *NeuroImage*, 158 :378–396, 2017.

X. Yao, Y. Song, and Z. Liu. Advances on pancreas segmentation : a review. *Multimedia Tools and Applications*, 79 :6799–6821, 2020a.

X. Yao, Y. Song, and Z. Liu. Advances on pancreas segmentation : a review. *Multimedia Tools and Applications*, 79(9) :6799–6821, 2020b.

J. Yoon, K.G. Kim, Y.J. Kim, S. Lim, Y-H. Park, D. Kim, H-T. Kang, and D-H. Lee. Distribution and characteristics of pancreatic volume using computed tomography volumetry. *Healthcare Informatics Research*, 26(4) :321–327, 2020.

S.H. Yoon, J.M. Lee, J.Y. Cho, K.B. Lee, J.E. Kim, S.K. Moon, S.J. Kim, J.H. Baek, S.H. Kim, S.H. Kim, et al. Small ($\leq$ 20 mm) pancreatic adenocarcinomas : analysis of enhancement patterns and secondary signs with multiphasic multidetector CT. *Radiology*, 259(2) :442–452, 2011.

J. Yu, M.A. Turner, A.S. Fulcher, and R.A. Halvorsen. Congenital anomalies and normal variants of the pancreaticobiliary tract and the pancreas in adults : part 2, pancreatic duct and pancreas. *American Journal of Roentgenology*, 187(6) :1544–1553, 2006.

Q. Yu, L. Xie, Y. Wang, Y. Zhou, E.K. Fishman, and A.L. Yuille. Recurrent saliency transformation network : Incorporating multi-stage visual cues for small organ segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8280–8289, 2018.

Q. Yu, D. Yang, H. Roth, Y. Bai, Y. Zhang, A.L. Yuille, and D. Xu. C2fnas : Coarse-to-fine neural architecture search for 3D medical image segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4126–4135, 2020.

J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny. Barlow twins : Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pages 12310–12320. PMLR, 2021.

L. Zhang, Y. Shi, J. Yao, Y. Bian, K. Cao, D. Jin, J. Xiao, and L. Lu. Robust pancreatic ductal adenocarcinoma segmentation with multi-institutional multi-phase partially-annotated CT scans. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 4, pages 491–500, Lima, Peru, 2020a. Springer.

M. Zhang, N. Singh, and P.T. Fletcher. Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In *International Conference on Information Processing in Medical Imaging*, pages 37–48. Springer, 2013.

T. Zhang, Y. Zhang, X. Liu, H. Xu, C. Chen, X. Zhou, Y. Liu, and X. Ma. Application of radiomics analysis based on CT combined with machine learning in diagnostic of pancreatic neuroendocrine tumors patient's pathological grades. *Frontiers in Oncology*, 10 :521831, 2021a.

T.T. Zhang, L. Wang, H.H. Liu, C.Y. Zhang, X.M. Li, J.P. Lu, and D.B. Wang. Differentiation of pancreatic carcinoma and mass-forming focal pancreatitis : qualitative and quantitative assessment by dynamic contrast-enhanced MRI combined with diffusion-weighted imaging. *Oncotarget*, 8(1) :1744, 2017.

X.-M. Zhang, D.G. Mitchell, M. Dohke, G.A. Holland, and L. Parker. Pancreatic cysts : depiction on single-shot fast spin-echo mr images. *Radiology*, 223(2) :547–553, 2002.

Y. Zhang, T. Xiang, T.M. Hospedales, and H. Lu. Deep mutual learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4320–4328, 2018.

Y. Zhang, J. Wu, Y. Liu, Y. Chen, W. Chen, E.X. Wu, C. Li, and X. Tang. A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set. *Medical Image Analysis*, 68 : 101884, 2021b.

Z. Zhang and U. Bagci. Dynamic linear transformer for 3D biomedical image segmentation. In *International Workshop on Machine Learning in Medical Imaging*, pages 171–180. Springer, 2022.

Z. Zhang, S. Li, Z. Wang, and Y. Lu. A novel and efficient tumor detection framework for pancreatic cancer via CT images. In *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pages 1160–1164. IEEE, 2020b.

N. Zhao, N. Tong, D. Ruan, and K. Sheng. Fully automated pancreas segmentation with two-stage 3D convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 201–209. Springer, 2019.

B. Zhou, Y. Xia, J. Yao, L. Lu, J. Zhou, C. Liu, J.S. Duncan, and L. Zhang. Meta-information-aware dual-path transformer for differential diagnosis of multi-type pancreatic lesions in multi-phase CT. In *International Conference on Information Processing in Medical Imaging*, pages 119–131. Springer, 2023.

Y. Zhou, L. Xie, E.K. Fishman, and A.L. Yuille. Deep supervision for pancreatic cyst segmentation in abdominal CT scans. In *International conference on Medical Image Computing and Computer-Assisted Intervention*, pages 222–230. Springer, 2017a.

Y. Zhou, L. Xie, W. Shen, Y. Wang, E.K. Fishman, and A.L. Yuille. A fixed-point model for pancreas segmentation in abdominal CT scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 693–701. Springer, 2017b.

Y. Zhou, Y. Li, Z. Zhang, Y. Wang, A. Wang, E.K. Fishman, A.L. Yuille, and S. Park. Hyper-pairing network for multi-phase pancreatic ductal adenocarcinoma segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 155–163. Springer, 2019a.

Y. Zhou, Z. Li, S. Bai, C. Wang, X. Chen, M. Han, E. Fish man, and A.L. Yuille. Prior-aware neural network for partially-supervised multi-organ segmentation. In *IEEE/CVF International Conference on Computer Vision*, pages 10672–10681, 2019b.

Y. Zhou, Y. Wang, P. Tang, S. Bai, W. Shen, E. Fishman, and A. Yuille. Semi-supervised 3D abdominal multi-organ segmentation via deep multi-planar co-training. In *IEEE Winter Conference on Applications of Computer Vision*, pages 121–140. IEEE, 2019c.

M. Zhu, C. Xu, J. Yu, Y. Wu, C. Li, M. Zhang, Z. Jin, and Z. Li. Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images : a diagnostic test. *PLOS One*, 8(5) :e63820, 2013.

Z. Zhu, Y. Xia, W. Shen, E.K. Fishman, and A.L. Yuille. A 3D coarse-to-fine framework for automatic pancreas segmentation. *arXiv preprint arXiv :1712.00201*, 2 :2, 2017.

Z. Zhu, C. Liu, D. Yang, A. Yuille, and D. Xu. V-NAS : Neural architecture search for volumetric medical image segmentation. In *International Conference on 3D Vision*, pages 240–248. IEEE, 2019a.

Z. Zhu, Y. Xia, L. Xie, E.K. Fishman, and A.L. Yuille. Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–12. Springer, 2019b.

Z. Zhu, Y. Lu, W. Shen, E.K. Fishman, and A.L. Yuille. Segmentation for classification of screening pancreatic neuroendocrine tumors. In *IEEE/CVF International Conference on Computer Vision*, pages 3402–3408, 2021.

D. Zimmerer, F. Isensee, J. Petersen, S. Kohl, and K. Maier-Hein. Unsupervised anomaly localization using variational auto-encoders. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 289–297. Springer, 2019.

M. Zins. Invited commentary : Failures in imaging diagnosis of pancreatic cancer : Can we avoid them and how ? *RadioGraphics*, 43(11) :e230217, 2023.

ECOLE DOCTORALE

Titre : Méthodes d'Intelligence Artificielle pour l'Aide au Diagnostic des Maladies du Pancréas en Radiologie

Mots clés : Apprentissage Automatique, Imagerie Médicale, Pancréas, Cancer

Résumé : Avec l'augmentation de son incidence et son taux de survie à cinq ans (9%), le cancer du pancréas pourrait devenir la troisième cause de décès par cancer d'ici à 2025. Ces chiffres sont notamment dus aux diagnostics tardifs, limitant les options thérapeutiques. Cette thèse vise à assister les radiologues dans le diagnostic du cancer du pancréas sur des images scanner grâce à des outils d'intelligence artificielle (IA) qui faciliteraient un diagnostic précoce. Pour atteindre ces objectifs, trois pistes de recherche ont été explorées. Premièrement, une méthode de segmentation automatique du pancréas a été développée. Le pancréas présentant une forme allongée et des extrémités subtiles, la méthode proposée utilise des informations géométriques pour ajuster localement la sensibilité de la segmentation. Deuxièmement, une méthode réalise la détection des lésions et de la dilatation du canal pancréatique principal (CPP), deux signes cruciaux du cancer du pancréas. La méthode proposée commence par segmenter le pancréas, les lésions et le CPP. Ensuite, des caractéristiques quantitatives sont extraites des segmentations prédites puis utilisées pour prédire la présence d'une lésion et la dilatation du CPP. La robustesse de la méthode est démontrée sur une base externe de 756 patients. Dernièrement, afin de permettre un diagnostic précoce, deux approches sont proposées pour détecter des signes secondaires. La première utilise un grand nombre de masques de segmentation de pancréas sains pour apprendre un modèle normatif des formes du pancréas. Ce modèle est ensuite exploité pour détecter des formes anormales, en utilisant des méthodes de détection d'anomalies avec peu ou pas d'exemples d'entraînement. La seconde approche s'appuie sur deux types de radiomiques : les radiomiques profonds (RP), extraits par des réseaux de neurones profonds, et les radiomiques manuels (RM), calculés à partir de formules prédéfinies. La méthode extrait des RP non redondants par rapport à un ensemble prédéterminé de RM afin de compléter l'information déjà contenue. Les résultats montrent que cette méthode détecte efficacement quatre signes secondaires : la forme anormale, l'atrophie, l'infiltration de graisse et la sénilité. Pour élaborer ces méthodes, une base de données de 2800 examens a été constituée, ce qui en fait l'une des plus importantes pour la recherche en IA sur le cancer du pancréas.

Title : Artificial Intelligence Methods to Assist the Diagnosis of Pancreatic Diseases in Radiology

Keywords : Machine Learning, Medical Imaging, Pancreas, Cancer

Abstract : With its increasing incidence and its five-year survival rate (9%), pancreatic cancer could become the third leading cause of cancer-related deaths by 2025. These figures are primarily attributed to late diagnoses, which limit therapeutic options. This thesis aims to assist radiologists in diagnosing pancreatic cancer through artificial intelligence (AI) tools that would facilitate early diagnosis. Several methods have been developed. First, a method for the automatic segmentation of the pancreas on portal CT scans was developed. To deal with the specific anatomy of the pancreas, which is characterized by an elongated shape and subtle extremities easily missed, the proposed method relied on local sensitivity adjustments using geometrical priors. Then, the thesis tackled the detection of pancreatic lesions and main pancreatic duct (MPD) dilatation, both crucial indicators of pancreatic cancer. The proposed method started with the segmentation of the pancreas, the lesion and the MPD. Then, quantitative features were extracted from the segmentations and leveraged to predict the presence of a lesion and the dilatation of the MPD. The method was evaluated on an external test cohort comprising hundreds of patients. Continuing towards early diagnosis, two strategies were explored to detect secondary signs of pancreatic cancer. The first approach leveraged large databases of healthy pancreases to learn a normative model of healthy pancreatic shapes, facilitating the identification of anomalies. To this end, volumetric segmentation masks were embedded into a common probabilistic shape space, enabling zero-shot and few-shot abnormal shape detection. The second approach leveraged two types of radiomics: deep learning radiomics (DLR), extracted by deep neural networks, and hand-crafted radiomics (HCR), derived from predefined formulas. The proposed method sought to extract non-redundant DLR that would complement the information contained in the HCR. Results showed that this method effectively detected four secondary signs of pancreatic cancer: abnormal shape, atrophy, senility, and fat replacement. To develop these methods, a database of 2800 examinations has been created, making it one of the largest for AI research on pancreatic cancer.

Institut Polytechnique de Paris
91120 Palaiseau, France