Vilniaus universiteto ligoninė
SANTAROS KLINIKOS

VILNIAUS UNIVERSITETO
MEDICINOS FAKULTETAS

On the Computed Tomography Image Data to Diagnose Pancreatic Cancer Using Machine Learning

Aušra Šubonienė[1], Olga Kurasova[1], Viktor Medvedev[1], Aistė Kielaitė-Gulla[2], Artūras Samuilis[3], Raminta Lukšaitė-Lukštė[3], Džiugas Jagminas[4], Kęstutis Strupas[2], Gintautas Dzemyda[1]
1 Institute of Data Science and Digital Technologies, Vilnius University
2 Institute of Clinical Medicine, Faculty of Medicine, Vilnius University
3 Institute of Biomedical Sciences, Department of Radiology, Nuclear Medicine and Medical Physics, Faculty of Medicine, Vilnius University
4 Faculty of Medicine, Vilnius University

Institute of Data Science and Digital Technologies

Introduction

Medical imaging data, which is suitable for solving segmentation problems, are difficult to obtain due to data sensitivity issues and the effort that is required to make ground truth segmentations. In order to increase the robustness of results by including more medical images, multiple datasets are often combined. However, challenges arise when trying to combine such datasets from different sources. Populations of patients that differ by age and other conditions could affect the results. Also, there might be different approaches to segmentation and its accuracy. Experts can segment medical images as true to anatomical structures as possible, or they might include some surrounding tissues in order to speed up manual segmentation. Also, rough region boundaries can be used instead of segmentations. Lastly, there can be different diagnostic devices used, which might result in different pre-processing of images.

Different diagnostic devices and artefacts in an image

There can be different diagnostic devices used, which might result in different pre-processing of images. Spatial resolution of different CT scanners might be different, as well as image noise and artefacts. Also, different CT protocols and characteristics of the patient might result in different levels of noise.

Even after initial pre-processing, these different diagnostic devices can have unique artefacts in computer tomography images that can reduce segmentation accuracy when combining multiple datasets. Therefore additional noise removal might be needed, as well as some morphological operations to further reduce the presence of artefacts in images [1], [2].

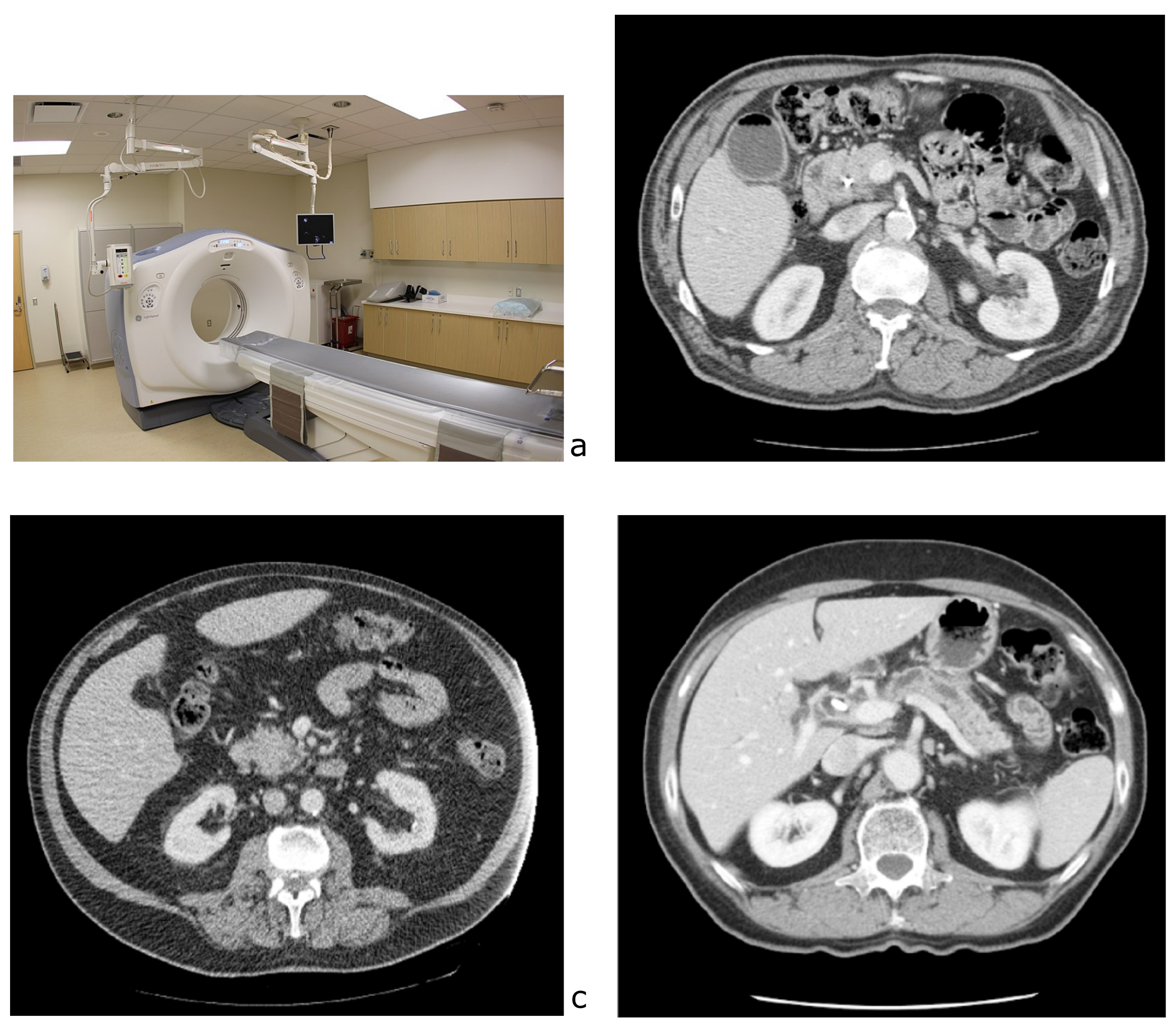

Figure 1. The presence of different imaging noise in pancreas CT images: a- CT scanner [3], b-d varying amount of noise in CT images.

Data sources

Due to data sensitivity and effort that is needed to anonymise images, there is a lack of publicly available pancreatic cancer data. Most publicly available medical data is without segmentation by experts, and images that do have some anatomical structures segmented are scarce. Although there are successful studies done on the analysis and segmentation of contrast-enhanced ultrasound (CEUS) [4] or MRI pancreas images, most of the datasets that are publicly available are computer tomography images. Currently, the largest public collections of computer tomography (CT) images of pancreatic cancer are available are the Cancer Imaging Archive (TCIA) dataset [5] and Medical Segmentation Decathlon dataset [6].

The Medical Segmentation Decathlon dataset consists of 421 portal-venous phase CT scans. Segmentations of both the pancreatic parenchyma and pancreatic mass (cyst or tumour) are provided, although done as ROI only, which makes the segmentation process quicker but less accurate. TCIA dataset consists of 82 abdominal contrast enhanced CT scans with slice thickness between 1.5−2.5 mm. Manual segmentations were done only to segment the pancreas. Here we also analyse the dataset which was acquired in the Vilnius University Hospital Santaros Klinikos. Segmentations that were provided by the experts consist of healthy pancreas, pancreatic cancer and pancreatic duct, which can be confused with pancreatic cancer by machine learning algorithms due to similar intensity of pixels. When combining publicly available datasets segmentation

Differences in populations of patients and cancer types

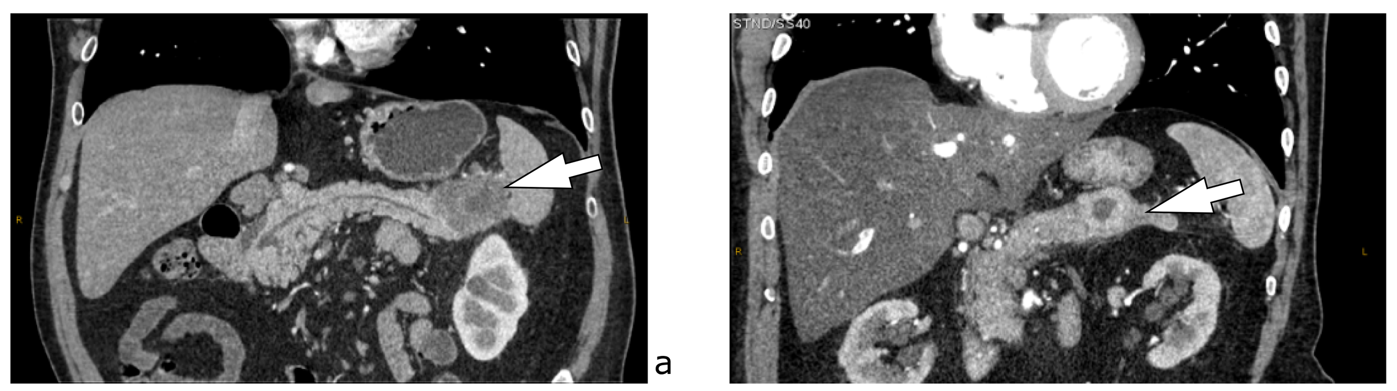

Figure 2. The differences between the types of pancreatic cancer: a- ductal adenocarcinoma, b- neuroendocrine tumor with central cystic degeneration

Due to the limited availability of medical images and pancreatic cancer being more common later in life, it is difficult to achieve equal coverage of all age categories of pancreatic cancer patients. This can be partially solved by increasing the dataset size of categories by retrospective analysis of computer tomography images provided by the medical institution, that have been filtered by the desired conditions. Also, different types of pancreatic cancer can be visualised differently in CT images. The detection of pancreatic cancer becomes more difficult when multiple cancer types are used for training and testing. Tumor pixels might have either higher or lower pixel intensity values than healthy pancreas, which complicates classification of tissue into healthy and cancerous. Neuroendocrine tumor appears brighter than healthy pancreas tissue in CT images, while ductal adenocarcinoma has darker pixel values. This creates problems if traditional morphological operations-based methods or radiomics features are used on their own or as part of a fusion method for pancreatic cancer detection.

Issues in the quality of the segmentation

Manual segmentations of computer tomography images can be prepared using different approaches to segmentation and its accuracy. Experts can segment medical images as true to anatomical structures as possible, or they might include some surrounding tissues in order to speed up manual segmentation. This creates inconsistencies in segmentation as part of the training data might have surrounding tissue segmented as pancreas as ground truth. This might reduce the stability of results of segmentation and classification. Ground truth segmentation masks can be created either manually by radiologists or by using available segmentation software, such as 3D Slicer. While computer-aided segmentation might save some time at initial segmentation, its results might be not accurate enough [7].

Also, rough region boundaries can be used instead of segmentations. Data pre-processing might be done by removing fat tissue based on the values of Hounsfield units during the process of manual segmentation. This results in holes and irregular edges of segmented regions. This creates additional problems when trying to unify segmentations across multiple datasets and might reduce the accuracy of segmentations when using machine learning. One possible solution might be dilation followed by erosion in 3D images in order to fill the holes and smooth the edges of segmented regions [8].

Different segmentation standards might also be an issue. For example, some segmentations pancreatic duct can be marked as part of healthy pancreas tissue. This might reduce the accuracy of classification, since pancreatic duct might have similar pixel intensity to tumor pixels.

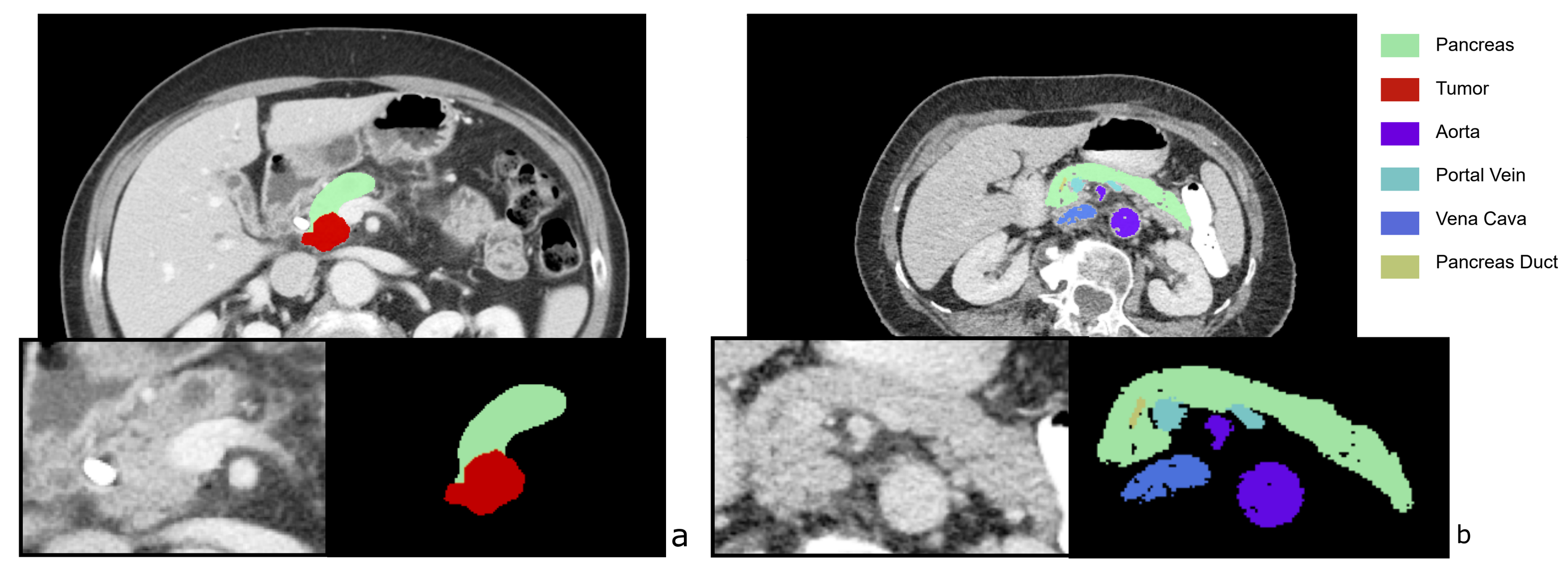

Figure 3. Differences in the quality of Pancreatic cancer segmentation a- only rough region of interest segmented, b- full segmentation with surrounding objects

References

[1] Arabi, H., & Zaidi, H. (2021). Deep learning–based metal artefact reduction in PET/CT imaging. European radiology, 1-13.
[2] Mouton, A., Megherbi, N., Van Slambrouck, K., Nuyts, J., & Breckon, T. P. (2013). An experimental survey of metal artefact reduction in computed tomography. Journal of X-ray Science and Technology, 21(2), 193-226.
[3] https://commons.wikimedia.org/wiki/File:UPMCEast_CTscan.jpg
[4] Kielaitė-Gulla, A., Samuilis, A., Raišutis, R., Dzemyda, G., & Strupas, K. (2021). The concept of AI-based algorithm: analysis of CEUS images and HSPs for identification of early parenchymal changes in severe acute pancreatitis. Informatica, 32(2), 305-3019.
[5] Holger R. Roth, Amal Farag, Evrim B. Turkbey, Le Lu, Jiamin Liu, and Ronald M. Summers. (2016). Data From Pancreas-CT. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU
[6] Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., van Ginneken, B. and Bilello, M. & Cardoso, M. J. (2021). The Medical Segmentation Decathlon. arXiv preprint arXiv:2106.05735.
[7] Pocė, I., Arsenjeva, J., Kielaitė-Gulla, A., Samuilis, A., Strupas, K., & Dzemyda, G. (2021). Pancreas segmentation in CT images: state of the art in clinical practice. Baltic journal of modern computing, 9(1), 25-34.
[8] Liu, K. L., Wu, T., Chen, P. T., Tsai, Y. M., Roth, H., Wu, M. S., Liao, W.C. & Wang, W. (2020). 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.