DETECTION OF PANCREATIC CANCER ON CT IMAGES USING PSEUDO-LABELING METHODS

Aušra Šubonienė[1], Olga Kurasova[1], Gintautas Dzemyda[1], Viktor Medvedev[1], Aistė Gulla[2], Artūras Samuilis[3], Džiugas Jagminas[3], Kęstutis Strupas[2]

[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, Faculty of Medicine, Vilnius University

Vilnius university hospital SANTAROS KLINIKOS

Vilnius University

INTRODUCTION

- Pancreatic cancer is the **12th most frequent** cancer and the **7th highest in mortality** [1].
- Early-stage pancreatic cancer usually has **no clinical symptoms**. Although late-stage pancreatic cancer is usually symptomatic, the symptoms are not specific.
- CT scans are often used to diagnose and monitor pancreatic cancer, but early-stage pancreatic cancer is difficult to see on these images.
- The tumour may be lighter or darker than a healthy pancreas.
- The overall sensitivity of pancreatic cancer detection on CT images is 89% and the specificity is 90% (higher than on ultrasound images) [2].
- Small (< 2cm) and isoattenuating tumours may be missed on radiological examination. The sensitivity of the test may then be **as low as 58-77%** [2].

5-year survival prediction:
IA - 14%
IB - 12%
III - 3%
IV - 1%

More than half of the cases are diagnosed in stages III - IV (inoperable tumours) [3]. Therefore it is **extremely important to develop automatic pancreatic cancer detection tools for early diagnosis**.

DATA PREPARATION

The dataset consists of **61 Santaros Clinics patients** (45 cancer, 16 healthy pancreas) and **361 patients from public datasets** (281 cancer, 80 healthy pancreas) from Memorial Sloan Kettering Cancer Center dataset [4] and the TCIA dataset [5].

In order to increase the amount of data for the input of the CNN, images of the **CT scans were divided into 50x50 pixels patches**. Only patches with visible pancreas or pancreatic cancer are used. Each patch is assigned the label 0 if only healthy pancreas is visible in the patch. A patch is labeled 1 if tumor is visible in the patch. Cancerous patches might contain healthy pancreas as well.

Patients in train - validation - test split:
Santaros: 31, 10, 20
Public: 270, 30, 61

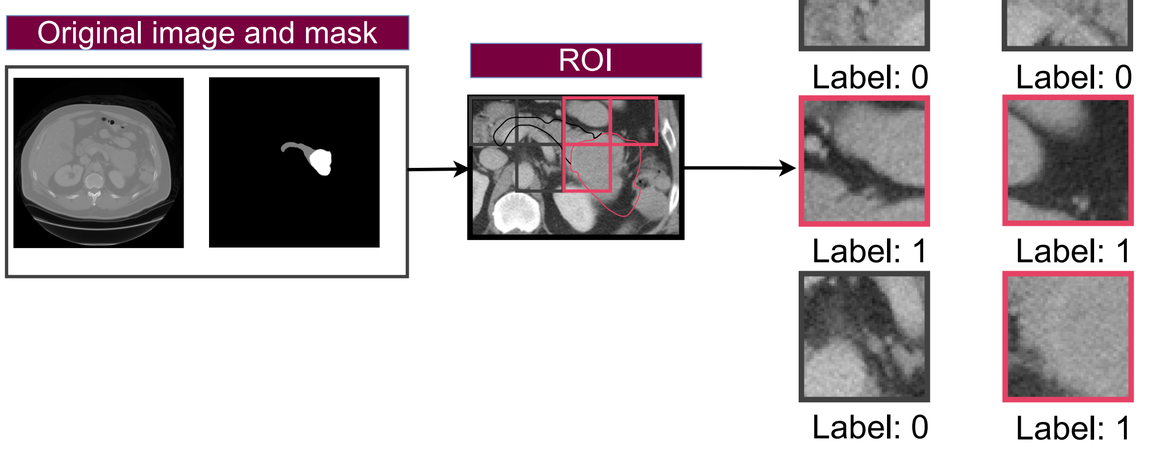

Figure 1. Image pre-processing by cropping and labelling patches

LABEL PROPAGATION

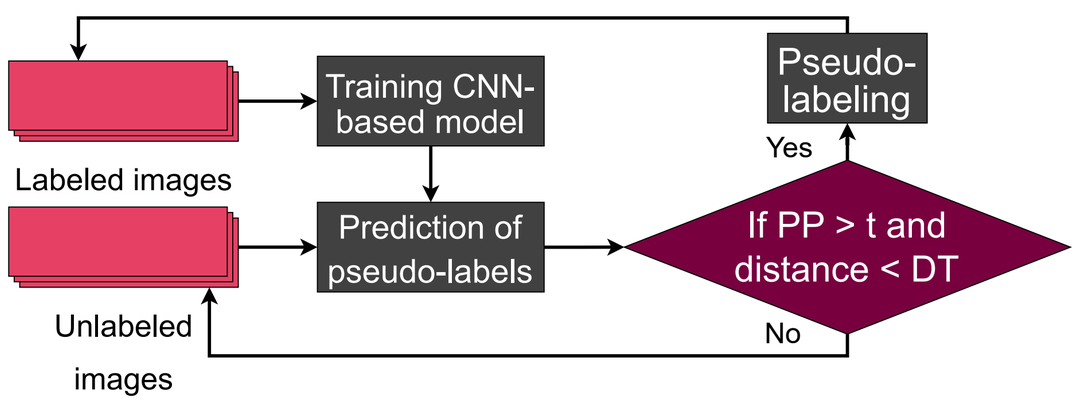

Figure 2. The process of pseudo-labeling. Patch prediction probability t = 0.9, distance threshold dist_t was calculated using median Euclidean distances to all patches in the training set.

THRESHOLD SELECTION

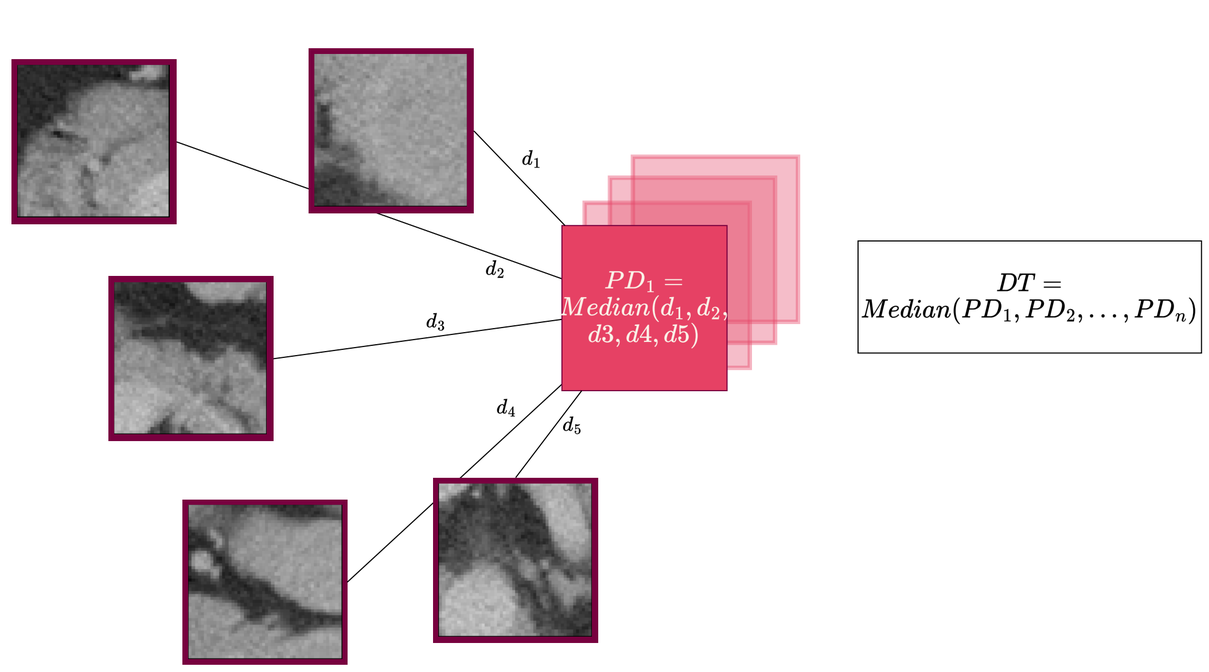

Figure 3. The process of determining optimal distance threshold. d_1 - d_5- Euclidean distances between all patches in the training set and a single patch in the test set. PD- patch distance, DT - distance threshold.

RESULTS

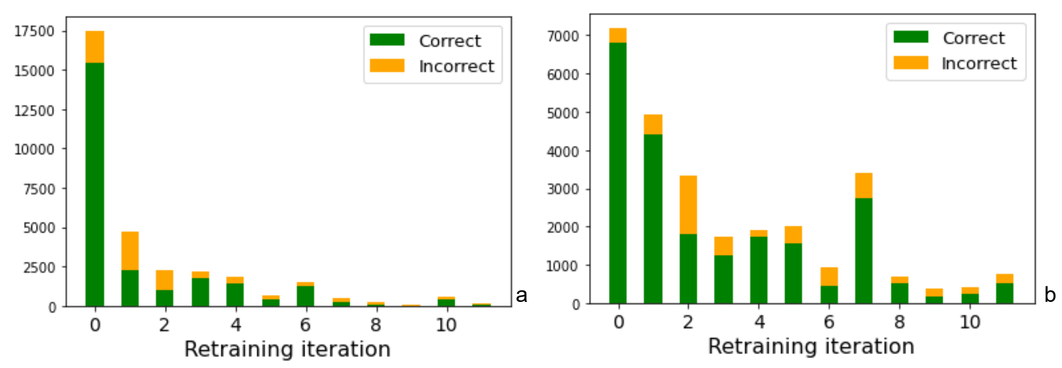

Figure 4. Number of added labels after retraining iterations, a- using only probabilities for threshold, b- using probabilities and distances for threshold

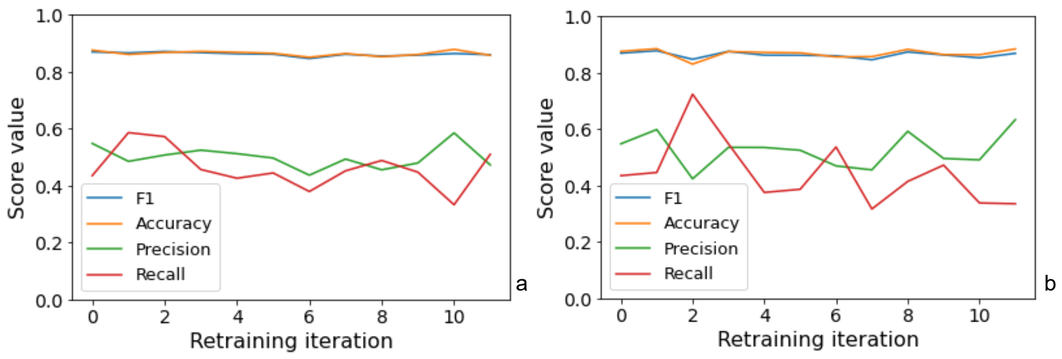

Figure 5. Prediction results after each retraining iteration using validation set, a- using only probabilities for threshold, b- using probabilities and distances for threshold

Total patches added:

- Using probabilities: 32061
- Using probabilities and distances: 27638

Correctly labeled patches:

- Using probabilities: 24446, 76.2%
- Using probabilities and distances: 22170, 80.2%

Adding calculations of distances to pseudo-labeling threshold selection resulted in **bigger proportion of patches labeled correctly**. Also, the increase of training dataset was more steady.
Using pseudo-labeling method the results of classification evaluation metrics remained approximately the same. Therefore adding pseudo-labeled patches **did not affect the classification results negatively**.

This study was approved by the Institutional Ethics Committee of Vilnius University (protocol code Nr. 158200-17-941-455.

[1] Globoscan 2020: https://gco.iarc.fr/today/data/factsheets/cancers/13-Pancreas-fact-sheet.pdf
[2] Rawla P, Sunkara T, Gaduputi V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. World J Oncol. 2019 Feb;10(1):10-27. doi: 10.14740/wjon1166. Epub 2019 Feb 26. PMID: 30834048; PMCID: PMC6396775.
[3] Elbanna, K.Y., Jang, HJ. & Kim, T.K. Imaging diagnosis and staging of pancreatic ductal adenocarcinoma: a comprehensive review. Insights Imaging 11, 58 (2020). https://doi.org/10.1186/s13244-020-00861-y
[4] H. R. Roth, A. Farag, E. B. T. Turkbey, L. Lu, J. Liu, and R. M. Summers, "Data from Pancreas-CT. The Cancer Imaging Archive," 2016, doi: 10.7937/K9/TCIA.2016.tNB1kqBU.
[5] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," J. Digit. Imaging, vol. 26, no. 6, pp. 1045–1057, Dec. 2013, doi: 10.1007/s10278-013-9622-7.