AIMI Group 19, Panorama

1st Giedrius Mirklys
Faculty of Science
Radboud University
Nijmegen, Netherlands
giedrius.mirklys@ru.nl

2nd Ignas Golubovskis
Faculty of Science
Radboud University
Nijmegen, Netherlands
ignas.golubovskis@ru.nl

3rd Björn Westerlund
Faculty of Science
Radboud University
Nijmegen, Netherlands
bjoern.westerlund@ru.nl

4th Luka Godnič
Faculty of Science
Radboud University
Nijmegen, Netherlands
luka.godnic@ru.nl

Abstract—**Pancreatic ductal adenocarcinoma (PDAC) is among the deadliest malignancies, often eluding early detection due to its subtle presentation on contrast-enhanced CT (CECT) scans. The PANORAMA challenge provides a large-scale, multi-reader benchmark for PDAC detection, offering a unique opportunity to evaluate and enhance machine-learning–based diagnostic tools. In this project, we aimed to improve upon the winning submission of the PANORAMA competition, which leverages the nnU-Net framework for segmentation and classification. Our focus was on reducing false positives to minimize radiologist workload, primarily by incorporating the Tversky loss function, which allows fine-tuning the trade-off between false positives and false negatives. We analyzed the effect of this modification on detection accuracy and the area under the ROC curve (AUC), and further evaluated segmentation performance using Dice score metrics. This report details our methodological improvements, experimental results, and future directions for robust, generalizable PDAC detection models.**

I. INTRODUCTION

As part of course Artificial Inteligence in Medical Imaging (AIMI) [4], we were tasked with improving the baseline model of PANORAMA [5] challenge. The PANORAMA challenge is the first large-scale reader study which aimed to establish the baseline performance of radiologists at Pancreatic Ductal Adenocarcinoma (PDAC)(add reference) detection on CECT scans. PDAC) is the most common type of malignant tumor affecting the pancreas and ranks among the deadliest of all solid cancers. In the United States, it is projected that there will be approximately 67,440 new cases and 51,980 deaths from PDAC in 2025 [2]. The challenge included more 68 international readers and more than 400 cases in the hidden testing cohort. Readers study was two fold. The first task was giving a discrete diagnosis with an additional PDAC likelihood score, and the second task readers provided a point annotation of the leasion location in the PDAC cases.

II. RELATED WORK

III. MOTIVATION FOR MACHINE-LEARNING–DRIVEN PDAC DETECTION

Pancreatic ductal adenocarcinoma (PDAC) often leaves only faint radiologic footprints—e.g. minimal ductal calibre changes or subtle parenchymal texture shifts—on contrast-enhanced CT (CECT) months before a tumour becomes obvious. These patterns are easily missed by experts reviewing

thousands of abdominal scans for diverse indications. Modern deep-learning systems, however, thrive on precisely this kind of weak-signal problem: by optimising millions of parameters across 3-D feature hierarchies, they can amplify and aggregate minute cues that fall below human perceptual thresholds [7], [8].

The PANORAMA grand challenge provides the first large-scale test bed for such systems in PDAC, with the following distinguishing characteristics:

- **Data volume.** Annotated scans of $>2\,500$ cases for open development.
- **Human baseline.** 68 international radiologists supply a reader-study benchmark, enabling rigorous algorithm-vs-expert comparison.
- **Task design.** Algorithms are evaluated on (i) binary PDAC presence scores and (ii) voxel-level lesion localisation, mirroring real-world deployment needs.

State-of-the-art deep learning approaches, such as the PANDA model described by Cao et al. (2023) which utilizes components like CNNs and transformers, have reported NCCT AUCs > 0.95 for pancreatic lesion detection on single-centre cohorts. [9]. Yet external validation is rare, and most models are tuned on datasets an order of magnitude smaller than PANORAMA. Critically, these smaller datasets typically have fewer different data sources, leading to lower overall variance between samples, which will likely boost prediction scores but is unlikely to generalize well.

A. Model starting point

We chose to use the winning submission of the PANORAMA competition as starting point for out own experiments. We considered using the competition baseline but found them to be very similar, with the competition winner's repo being easier to follow. This approach uses nnUnet [1] an out of the box neural net solution that builds on the U-net architecture that is specifically tuned for biomedical use cases and contains. Additionally, it can automatically adapt preprocessing, neural net architecture, and post processing according to the data it is classifying.

The full inference pipeline consists of 4 steps: down-sampling, low-res prediction, masking and high-res prediction. First images are loaded and the voxels are down-sampled according to a set spacing, main to save. Then the nnU-net makes a mask prediction, aiming to identify the rough area

of the pancreas. This is very important as the pancreas is quite small and the CECT scans used as input cover the whole body of the patients. Standardized down-sampling is important in order to save computational cost without losing predictive ability. The mask is represented as a bounding box and in the next step, this bounding box is applied to the original image. Then a second prediction is made, this time at full resolution but only using the area within the bounding box. This prediction produces two outputs, the full scan of the patient with lesion candidate likelihoods applied to each voxel as well as a single number representing the probability that a patient has a pancreatic lesion.

$$TI = \frac{TP}{TP + \alpha \cdot FN + \beta \cdot FP}$$

IV. IMPROVING THE BASELINE

When considering how to improve the baseline, we explored various potential approaches and evaluated their feasibility. We quickly determined that the most effective objective was to shift the AUC curve to the left. This shift reduces the false positive (FP) rate, thereby decreasing the workload for radiologists and physicians, who must review all cases flagged as positive.

One of the approaches we explored was training the network using the Tversky loss, a loss function commonly used in medical image analysis. It measures the dissimilarity between two sets—typically the predicted segmentation and the ground truth—and is a generalization of the Jaccard index. [3] By tuning the parameters α and β, you can control how the loss penalizes different types of segmentation errors. For instance, setting $\alpha > \beta$ places greater emphasis on reducing false positives, whereas $\beta > \alpha$ focuses more on minimizing false negatives. This flexibility allows Tversky loss to fine-tune the trade-off between precision and recall, making it particularly useful in scenarios where the cost of false positives and false negatives is not equal. In our case, where minimizing false positives and maximizing precision is crucial, Tversky loss provides an effective way to bias the model accordingly.

There are a few more things we need to describe in the report:

- checking the data: balance and differences in test and training data - besides the loss, dice score curves of both baseline and tversky, maybe some outputs of the filters

- some FP/FN analysis.

V. RESULTS

TODO

VI. CONCLUSIONS

TODO

This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.

Index Terms—component, formatting, style, styling, insert

REFERENCES

[1] Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z

[2] National Cancer Institute, "SEER Cancer Stat Facts: Pancreatic Cancer," https://seer.cancer.gov/statfacts/html/pancreas.html, accessed May 15, 2025.

[3] Saba, M., "Tversky Loss," *Medium*, https://medium.com/@saba99/tversky-loss-902f5f8cc35f, accessed May 15, 2025.

[4] Radboud University, "Course: NWI-IMC037 – Artificial Intelligence and Medical Imaging," Osiris Course Catalogue, https://ru.osiris-student.nl/onderwijscatalogus/extern/cursus?taal=en&cursuscode=NWI-IMC037&collegejaar=2024, accessed May 15, 2025.

[5] PANORAMA Grand Challenge, "PANORAMA: AI Grand Challenge for Pancreatic Cancer Detection," https://panorama.grand-challenge.org/, accessed May 15, 2025.

[6] The Radiology Assistant, "Pancreas - Carcinoma," https://radiologyassistant.nl/abdomen/pancreas/pancreas-carcinoma-1, accessed May 15, 2025.

[7] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.

[8] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, A. Chesus, D. C. Corrado, B. Darzi, *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, 2020, doi: 10.1038/s41586-019-1799-6.

[9] K. Cao, Y. Xia, J. Yao, H. Xu, L. Lambert, T. Zhang, W. Tang, G. Jin, J. Hui, F. Xu, I. Nogues, X. Li, W. Guo, Y. Wang, W. Fang, M. Qiu, Y. Hou, T. Kovarnik, M. Vocka, Y. Lu, Y. Chen, X. Chen, Z. Liu, Z. Chen, C. Xie, R. Zhang, H. Lu, G. Hager, A. L. Yuille, L. Lu, C. Shao, Y. Shi, T. Liang, L. Zhang, and J. Lu, "Large-scale pancreatic cancer detection via non-contrast CT and deep learning," *Nature Medicine*, vol. 29, no. 12, pp. 3033–3043, 2023, doi: 10.1038/s41591-023-02640-w.