A Multi-task Model for Joint Segmentation and Classification of Pancreatic Cancer

Ching-Yuan Yu[1*], Jun Ma[1,2*], Tianhao Fu[2,3,4], and Bo Wang[1,2,5,6,7]

[1] AI Hub, University Health Network, Toronto, Canada
[2] Vector Institute Toronto, ON, Canada
[3] Villanova College. King City, ON, Canada
[4] Project Neura. Toronto, ON, Canada
[5] Peter Munk Cardiac Centre, University Health Network, Toronto, Canada
[6] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada
[7] Department of Computer Science, University of Toronto, Toronto, Canada
[*] Equal contributions
bowang@vectorinstitute.ai

Abstract. Pancreatic ductal adenocarcinoma (PDAC) is a highly aggressive malignancy with poor prognosis, underscoring the critical need for early and accurate diagnosis. We present a novel two-stage pipeline for PDAC detection from contrast-enhanced computed tomography (CECT) scans, leveraging the nnU-Net framework for segmentation and a high-resolution multi-task convolutional neural network (CNN) for joint lesion segmentation and classification. The first stage employs a low-resolution nnU-Net to segment the pancreas region, while the second stage refines segmentation and performs lesion classification. The methods achieved an AUROC of 0.9833 and an AP of 0.8011 on the PANORAMA challenge dataset.

1 Introduction

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal malignancies, characterized by its aggressive progression and poor prognosis. Early and accurate diagnosis from abdominal contrast-enhanced computed tomography (CECT) scans is critical for improving patient outcomes, as it enables timely intervention, surgical resection, and personalized treatment approaches.

Recent studies [4] have demonstrated the remarkable capability of deep learning-based algorithms in automatically and accurately detecting PDAC from CT scans. These algorithms leverage advanced neural network architectures to identify subtle patterns and features in imaging data that are often missed by human observers, offering a promising avenue for early diagnosis and improved patient outcomes. Despite these advancements, the development and comparison of such models face challenges due to the lack of benchmark platforms and the limited availability of large-scale, high-quality datasets. The PANORAMA challenge has emerged as a dedicated initiative to bridge this gap, which provides a standardized platform and a comprehensive dataset for PDAC detection.

This technical report presents our solution to the PANORAMA challenge, which ranks the first place on the public tuning leaderboard. Specifically, we introduce a two-stage pipeline for pancreas region of interest (ROI) localization followed by joint cancer segmentation and classification. The localization is based on a low-resolution nnU-Net [5] model and we further extend it to multi-task network by adding a classification head. To foster further developments of generalizable models and enable fair and transparent benchmarking across the research community, our code and model have been publicly available at https://github.com/bowang-lab/PANORAMA.

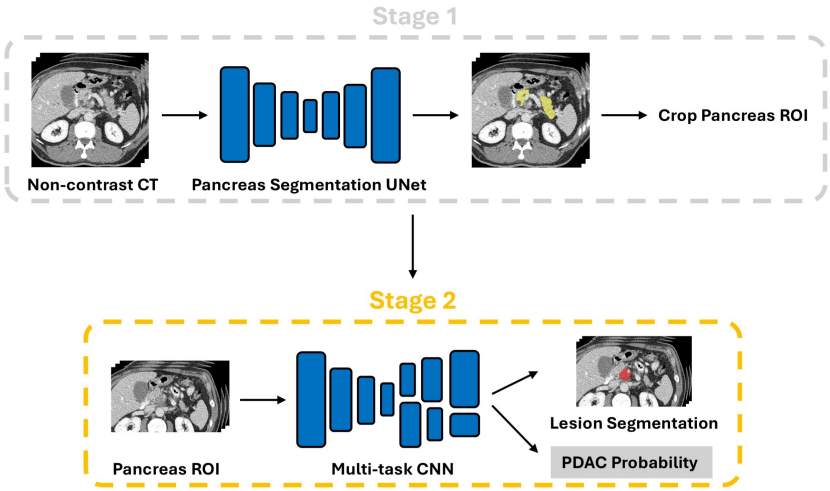

Fig. 1. Stage 1 first uses a low-resolution 3D nnU-Net to segment the pancreas from non-contrast CT scans and generate pancreas ROI. Then, Stage 2 employs a multi-task CNN for joint lesion segmentation and patient-level PDAC classification.

2 Method

Our proposed method, based on the nnUNet framework[5], utilizes a two-stage approach to effectively detect PDAC from CECT images. This approach leverages the strengths of nnUNet's robust segmentation capabilities for coarse localization and integrates a classification head for fine-grained lesion recognition and segmentation. Figure 1 illustrates the whole pipeline for PDAC detection.

2.1 Stage 1: Low-Resolution Segmentation

Model Architecture A low-resolution nnU-Net is employed to segment the pancreas region within the CECT image. Loss Function: Dice loss is utilized to

optimize the segmentation model, encouraging accurate delineation of the pancreas. Output: The segmentation model generates a probability map indicating the likelihood of each pixel belonging to the pancreas.

2.2 Stage 2: High-Resolution Classification based on pancreas ROI

Model Architecture A high-resolution nnU-Net is employed for join segmentation and classification based on the ROIs. In particular, The segmentation head is responsible for segmenting the tumor within the ROI, which follows a similar architecture to the low-resolution segmentation network but operates on higher-resolution data to achieve more precise segmentation. The classification head uses a Feature Pyramid Network (FPN) to extract multi-scale features from the last three layers of the encoder. By fusing features across different levels, the FPN captures both local and global contextual information. The aggregated features are then passed through a fully connected layer for final classification. Figure 2 provides a detailed illustration of the Stage 2 multi-task CNN architecture.

Loss Function A combination of Dice loss and cross-entropy loss is used to train the segmentation head. To address class imbalance and prioritize the detection of PDAC cases, the classification head is optimized by wighted cross-entropy loss with a weight of 1:3 for non-PDAC and PDAC, respectively.

Stage-wise Training The network training followed a three-step approach. Initially, the segmentation head was trained with the classification head weights frozen, using a combined Dice and cross-entropy loss function. Next, the classification head was trained independently by freezing the segmentation head weights and applying a weighted cross-entropy loss for optimization. Finally, the entire network underwent end-to-end fine-tuning, where both the segmentation and classification heads were optimized simultaneously to enhance joint performance.

Post-processing The output probability map from the stage 2 model's segmentation head undergoes post-processing to generate a detection map, employing a strategy that integrates lesion candidate extraction within the report-guided annotation framework.[3].

Final Prediction To make a final prediction for a given input image, we first obtain the classification probability from the classification head. We then combine this probability with the maximum value of the detection map generated by the segmentation head. This combined score is used to make the final prediction, with a higher score indicating a higher likelihood of PDAC.

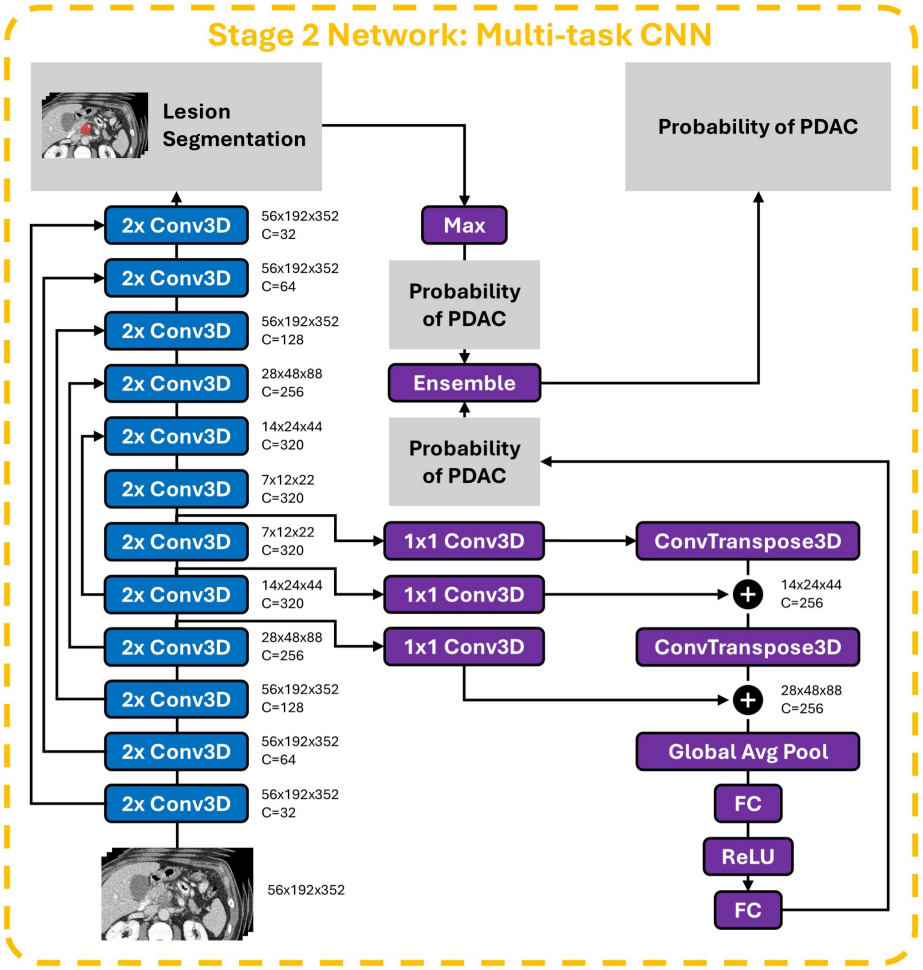

Fig. 2. The network contains a shared encoder for feature extraction from the pancreas ROI and two heads for join lesion segmentation and classification.

3 Experiments

3.1 Dataset and evaluation measures

The PANORAMA study dataset[2] is a comprehensive resource for AI-driven pancreatic cancer diagnostics, consisting of approximately 3,000 cases from centers in the Netherlands, Sweden, and Norway. The dataset includes contrast-enhanced CT (CECT) scans in the portal-venous phase, with cases excluding patients who have undergone prior pancreatic cancer treatments or have positive PDAC histopathology findings. It features a public training and development set of 2,238 cases, including 676 positive PDAC cases—482 with manual lesion annotations by trained investigators using ITK-SNAP v3.80 [6] under expert supervision, and 194 with algorithm-generated annotations validated by Alves et al.[1]. Segmentation of key anatomical structures (e.g., pancreas parenchyma, ducts, veins, arteries, and bile ducts) is provided, generated by the same validated algorithm. The dataset also includes a hidden validation cohort of 100 cases for model tuning via a public leaderboard and a hidden testing cohort of over 400 cases to benchmark AI and radiologists and validate hypotheses.

Evaluation Description The performance of AI algorithms is evaluated through a combination of patient-level and lesion-level metrics. Patient-level diagnosis performance is measured using the Area Under the Receiver Operating Characteristic (AUROC), while lesion-level detection performance is evaluated using the Average Precision (AP) metric. The overall ranking score for each algorithm is calculated as the average of its ranks across these two metrics:

$$\text{Overall Ranking Score} = \frac{\text{Rank(AP)} + \text{Rank(AUROC)}}{2}$$

Criteria for Lesion Detection Success The criteria define conditions under which predicted lesions are classified as true positives. For AI-detected 3D lesions, the following rules are applied:

- **True Positives:** Predicted lesions must achieve an overlap of at least 0.15 IoU with the ground truth annotations.
- **False Positives:** Predictions that fail to meet the minimum overlap or have no overlap at all are classified as false positives, irrespective of their size or location.
- **Ambiguous Cases:** When multiple predictions surpass the overlap threshold (IoU ≥ 0.15) for the same ground truth lesion, only the prediction with the highest overlap is treated as a true positive. Any other overlapping predictions are ignored but are not counted as false positives, addressing split-merge scenarios.

3.2 Implementation details

Environment settings The development environments and requirements are presented in Table 1.

Table 1. Development environments and requirements.

System	Ubuntu 20.04.5 LTS
CPU	AMD Ryzen Threadripper PRO 5995WX 64-Cores
RAM	16×4GB; 2.67MT/s
GPU (number and type)	NVIDIA RTX A6000
CUDA version	12.4
Programming language	Python 3.9.20
Deep learning framework	torch 2.5.0, torchvision 0.20.0

Table 2. Training protocols for stage 1 model.

Network initialization	
Batch size	2
Patch size	128×192×192
Total epochs	1000
Optimizer	Adam
Initial learning rate (lr)	0.01
Lr decay schedule	Reduced by factor of 0.1 every 75 epochs
Training time	16 hours
Loss function	Dice loss with Cross Entropy

Table 3. Training protocols for stage 2 model.

Network initialization	
Batch size	2
Patch size	56×192×352
Total epochs	1000
Optimizer	Adam
Initial learning rate (lr)	0.0003
Lr decay schedule	Reduced by factor of 0.1 every 75 epochs
Training time	36 hours
Loss function (segmentation)	Dice loss with Cross Entropy
Loss function (classification)	Weighted Cross Entropy

4 Results and discussion

The proposed method was evaluated on a hidden validation cohort to ensure unbiased performance assessment. Using the classification head alone, the model achieved an AUROC (Area Under the Receiver Operating Characteristic) of 0.9671 and an Average Precision (AP) of 0.8011.

To further enhance predictive performance, we employed an ensemble strategy that combines the output from the maximum of the detection map and the classification head. This approach led to a notable improvement, achieving an AUROC of 0.9833, while maintaining an AP of 0.8011.

The superior AUROC obtained with the ensemble highlights its effectiveness in integrating complementary information from the detection map and classification head. This integration significantly enhances the model's discrimination ability between PDAC and non-PDAC cases.

These results, validated on a hidden cohort, demonstrate the potential of our approach for accurate and reliable diagnosis of pancreatic ductal adenocarcinoma, warranting further clinical evaluation.

Table 4. Performance on hidden validation cohort

Algorithm	AUROC	AP
nnUNet + FPN	0.9671	0.8011
nnUNet + FPN + ensemble	0.9833	0.8011

5 Conclusion

We proposed a two-stage pipeline for detecting pancreatic ductal adenocarcinoma (PDAC) in contrast-enhanced CT images, combining nnU-Net-based segmentation and a classification model for enhanced precision. Evaluated on a hidden validation cohort, our method achieved an AUROC of 0.9833 and an AP of 0.8011, with the ensemble strategy significantly improving diagnostic accuracy. These results highlight the approach's potential for reliable PDAC diagnosis, warranting further clinical validation.

References

1. Alves, N., Schuurmans, M., Litjens, G., Bosma, J.S., Hermans, J., Huisman, H.: Fully automatic deep learning framework for pancreatic ductal adenocarcinoma detection on computed tomography. Cancers **14**(2) (2022) 5
2. Alves, N., Schuurmans, M., Rutkowski, D., Yakar, D., Haldorsen, I., Liedenbaum, M., Molven, A., Vendittelli, P., Litjens, G., Hermans, J., Huisman, H.: The panorama study protocol: Pancreatic cancer diagnosis - radiologists meet ai (2024) 5

3. Bosma, J.S., Saha, A., Hosseinzadeh, M., Slootweg, I., de Rooij, M., Huisman, H.: Semisupervised learning with report-guided pseudo labels for deep learning–based prostate cancer detection using biparametric mri. Radiology: Artificial Intelligence **5**(5), e230031 (2023) 3
4. Cao, K., Xia, Y., Yao, J., Han, X., Lambert, L., Zhang, T., Tang, W., Jin, G., Jiang, H., Fang, X., et al.: Large-scale pancreatic cancer detection via non-contrast ct and deep learning. Nature medicine **29**(12), 3033–3043 (2023) 1
5. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021) 2
6. Yushkevich, P.A., Gao, Y., Gerig, G.: Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 3342–3345 (2016) 5

Table 5. Checklist Table. Please fill out this checklist table in the answer column.

Requirements	Answer
A meaningful title	Yes/No
The number of authors (≤6)	Number
Author affiliations and ORCID	Yes/No
Corresponding author email is presented	Yes/No
Validation scores are presented in the abstract	Yes/No
Introduction includes at least three parts: background, related work, and motivation	Yes/No
A pipeline/network figure is provided	Figure number
Pre-processing	Page number
Strategies to use the partial label	Page number
Strategies to use the unlabeled images.	Page number
Strategies to improve model inference	Page number
Post-processing	Page number
The dataset and evaluation metric section are presented	Page number
Environment setting table is provided	Table number
Training protocol table is provided	Table number
Ablation study	Page number
Efficiency evaluation results are provided	Table number
Visualized segmentation example is provided	Figure number
Limitation and future work are presented	Yes/No
Reference format is consistent.	Yes/No