

Contrastive Learning and Linear Self-Attention for Robust Multi-Center Pancreas Segmentation in MRI

OPEN ACCESS

Received: 24/01/2026

Accepted: 12/02/2026

Published: 28/02/2026

Citation: Deepika M, Arockia Sahaya Sheela G (2026) Contrastive Learning and Linear Self-Attention for Robust Multi-Center Pancreas Segmentation in MRI. Indian Journal of Science and Technology 19(7): 443-454. <https://doi.org/10.17485/IJST/v19i7.104>

* **Corresponding author.**

deepika@hcctrichy.ac.in

Funding: None

Competing Interests: None

Copyright: © 2026 Deepika & Arockia Sahaya Sheela. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN

Print: 0974-6846

Electronic: 0974-5645

M Deepika^{1*}, G Arockia Sahaya Sheela²

¹ Research Scholar, Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli-2. Affiliated to Bharathidasan University, Tamil Nadu, India

² Assistant Professor, Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli-2, Affiliated to Bharathidasan University, Tamil Nadu, India

Abstract

Objectives: To develop a robust deep learning method for pancreas segmentation from multi-center MRI datasets by improving generalizability and accuracy.

Method: A novel 3D segmentation model named CTraTL-UNet was implemented. The proposed method integrates linear self-attention and contrastive learning to enhance feature representation. It addresses inter-center heterogeneity of the pancreas dataset. The model was trained and validated using a multi-center and multi-modal (T1-weighted and T2-weighted MRI) dataset. The dataset is comprising 767 cases from five institutions.

Findings: The proposed approach achieved superior segmentation performance with average Dice scores of 88.6% (T1W) and 89.2% (T2W).. The method demonstrated high reliability across varied institutional data, as confirmed by Jaccard, HD95, ASSD, and precision-recall metrics.

Novelty: This work is the first to combine linear self-attention and instance-level contrastive loss in a 3D MRI segmentation method for pancreas. It is validated on a large-scale multi-center dataset and sets a new benchmark for cross-center robustness.

Keywords: Pancreas segmentation; MRI; deep learning; linear self-attention; contrastive learning; multi-center dataset

1 Introduction

The clinical use of multi-center MRI images of the pancreas has led to accurate and automated segmentation of the pancreas; the importance of the results is shown by the fact that it can be utilized to detect early diseases, in surgical planning, and quantitative analyses. Manual annotation is labor-intensive, inter-observer variation may occur and is not feasible when dealing with large-scale multicentered studies⁽¹⁾. Besides, differences in imaging protocols across centers cause domain shifts, which dramatically deteriorate the performance of the existing segmentation models when used outside the domains of their training⁽²⁾. All these problems highlight the urgent necessity to use strong and generalizable deep learning models capable of providing consistent segmentation results on varied data sets.

A number of the recent works have tried to solve the problem of pancreas segmentation with the help of advanced deep learning architectures. Moglia *et al.*⁽³⁾ examined more than 130 pancreas segmentation studies, which found that classical CNN-based approaches reach moderate Dice scores with being sensitive to the size of the organ, and complicated anatomical boundaries. Zhang *et al.*⁽⁴⁾ proposed PanSegNet, that is, the combination of nnUNet and Transformer modules with the linear self-attention with Dice scores of over 85 percent on multi-center MRI and CT data, yet its performance reduced dramatically in case of cross-center validation. Bagci *et al.*⁽⁵⁾ also highlighted that annotated MRI data are a major bottleneck to deep-learning pipeline since little of this data is annotated and manual labeling is a costly procedure. Proietto Salanitri *et al.*⁽⁶⁾ and Uslucuk *et al.*⁽⁷⁾ have shown that 3D fully convolutional networks were able to improve the quality of spatial context modelling, but they did not have means of addressing inter-domain variations. Hussain *et al.*⁽⁸⁾ also suggested domain adaptation approaches to medical imaging, which demonstrated better generalization, but required more complex and computationally intensive models.⁽⁹⁾ Liu *et al.*⁽⁹⁾ and⁽¹⁰⁾ Podinã *et al.*⁽¹⁰⁾ used recurrent adversarial learning and reinforcement learning respectively, stating a marginal improvement in performance but with high instability in training. Recent hybrid CNN-Transformer methods^{(11), (12)} reached competitive Dice scores on internal data, however, they did not make use of cross-modal consistency (e.g., T1W vs. T2W) and did not have constraints to guarantee the feature space robustness in external validation.

Although these improvements have been made, the pancreas segmentation in a wide variety of clinical contexts is still a major challenge. The performance of most of the available approaches significantly deteriorates when used on any external multi-center data, and thus they can be characterized as sensitive to domain changes and poor extrapolation to test distributions outside of the training distributions^{(4), (6), (8)}. Moreover, although attention models based on transformers have been shown to better capture long-range dependencies, their computation complexity, which is quadratic, limits their scalability to large-resolution 3D MRI volumes, and cannot be used in real-time or resource-constrained settings^{(4), (11)}. Also, existing segmentation techniques do not explicitly use cross-modal feature alignment strategy, because they rarely incorporate instance level similarity restrictions between T1W and T2W modalities, resulting in poor features representation and performance variance across different imaging protocols^{(5), (9)}.

The lack of simultaneous global context modeling and cross-modal feature alignment in existing pancreas segmentation methods. Current transformer-based models suffer from quadratic complexity and domain sensitivity, while CNN-based methods lack robustness to inter-center variability. To overcome this challenge, this study proposes a new 3D pancreas segmentation architecture CTraTL-UNet, with Linear Self-Attention (LSA) to model global context effectively and Contrastive Representation-Aware Training Loss (CTraTL) to learn features in a cross-center and cross-modal manner. The proposed method seeks to provide the state-of-the-art performance in large-scale and multi-center MRI datasets combined with computational efficiency and feature space alignment, as well as guaranteeing cross-domain generalization and stability.

2 Methodology

2.1 Dataset

In this work, the massive publicly available multi-center pancreatic MRI data consisting of 385 T1-weighted (T1W) and 382 T2-weighted (T2W) MRI scans, as proposed in Bagci *et al.*⁽⁵⁾, was used. The data set is based on five dissimilar imaging centers with dissimilar scanning protocol, different hardware setup, and location-specific acquisition characteristic. This variability guarantees that segmentation algorithms are normally assessed with realistic multi-center conditions that reduce the chances of overfitting to a single source of image.

All MRI volumes were provided with high-quality manual annotations of the boundaries of the pancreas, which was organized and checked by professional radiologists under the standard annotation guidelines. The information therefore presents a sound ground truth to supervised learning activities. All the volumes were subjected to a standardized preprocessing pipeline prior to training, to make the inter site comparative. This as well involved resampling to a constant voxel spacing and constant spatial dimensions ($96 \times 96 \times 32$). It will be configured to resolve the resolution differences between imaging centers. Additionally, the intensity normalization was done to compensate scanner coupled differences in image contrast and brightness and as such produced standard intensity distributions across the dataset.

The last dataset was randomly divided into training, validation and testing divisions and there was an equal representation of all the centers to avoid site specific bias. It is important to note that T1W and T2W images of the same patient were put in the same split to prevent leakage of information at the level of patient. This form of strict dataset management ensures the strong assessment and cross-modality comparisons.

The suggested CTraTL-UNet architecture in Figure 1 is constructed on the basis of the 3D DynUNet backbone, further improved by two key innovations: (1) a Linear Self-Attention (LSA) module that can effectively model global context; (2) a Contrastive Representation-Aware Training Loss (CTraTL) that enables effective multi-center learning of features. The LSA

module substitutes the quadratic complexity of the traditional attention with a linear one, which allows processing volumetric MRI efficiently. This process elicits long-range interaction between spatial locations at a cost that is not too high. The CTraTL component makes sure that embeddings of similar anatomical structures across modalities (T1W, T2W) are drawn nearer in the latent space; simultaneously those embeddings of distinct anatomical types are distanced and enhances inter-domain invariance.

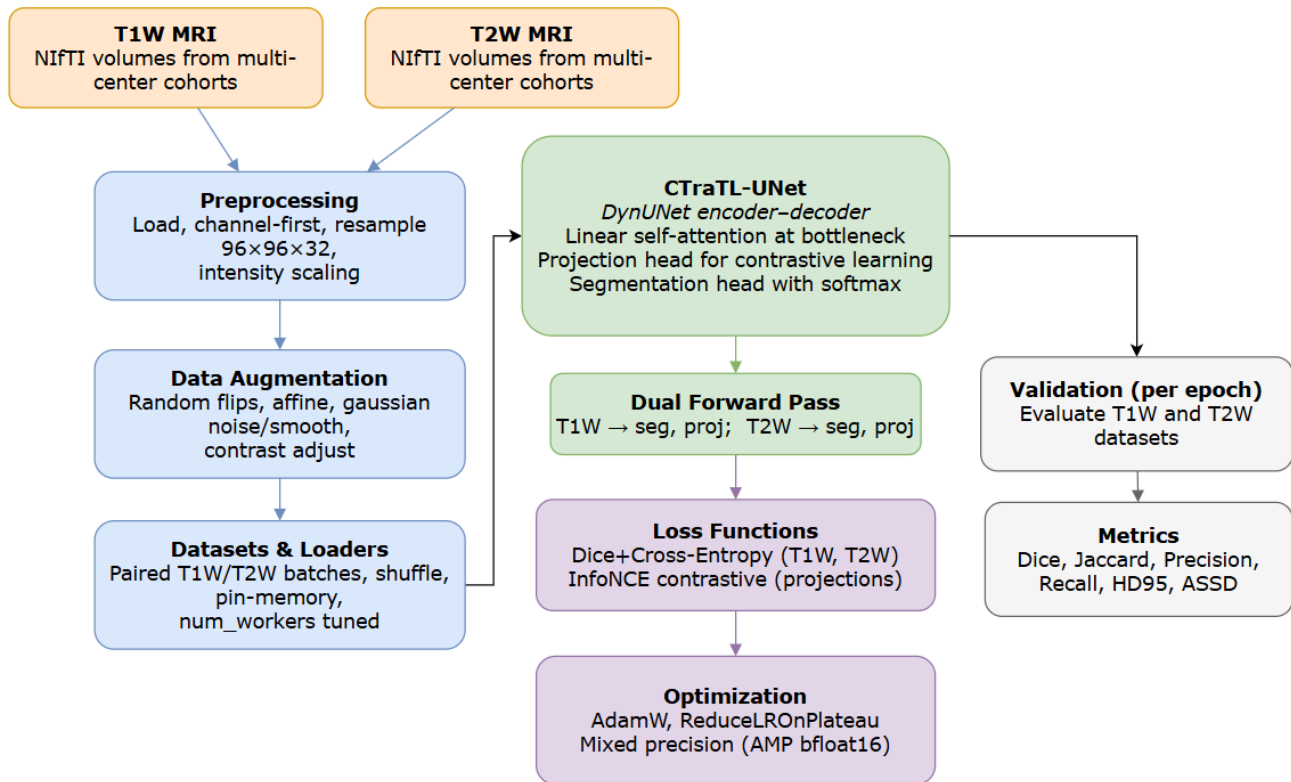


Fig 1. Workflow of Proposed Methodology

2.2 Baseline Architecture: 3D DynUNet

The backbone used is 3D DynUNet, which is an architectural component because it has been utilized previously in medical volumetric segmentation applications. It has an encoder-decoder structure, consisting of several blocks of down-sampling and up-sampling blocks with skip-connections that can capture the rich multi-scale features. It helps retaining the small-scale spatial details. Skip connections mitigate the gradient vanishing, and allow effective combination of low-level and high-level feature representations, which is essential to the correct delineation of pancreatic boundaries. Nevertheless, traditional DynUNet only makes use of local convolutional receptive fields, such that it cannot capture long-range contextual effects that are necessary to segment anatomically variant organs such as the pancreas.

Let $X \in \mathbb{R}^{B \times C \times D \times H \times W}$ denote the input feature map with batch size B , channels C , and spatial dimensions D , H , W . The Linear Self-Attention computes attention as:

$$Z = \sigma \left(\frac{(Q \cdot K^T)}{\sqrt{d_k}} \right) V,$$

with

$$Q = XW_Q,$$

$$K = XW_K,$$

$$V = XW_V$$

where $W_Q, W_K, W_V \in R^{C \times d_k}$ are learnable projection matrices and $\sigma(\cdot)$ denotes the softmax normalization.

2.3 Linear Self-Attention for Global Context Modeling

Traditional mechanisms of self-attention are pairwise dependency computations that require quadratic complexity with respect to spatial dimensions. To achieve a high resolution, 3D MRI volumes become impractical in terms of memory usage and computational overhead. Linear Self-Attention (LSA) is a method to address this drawback, and it estimates the calculation of attention with a reformulation which is a kernelization of the procedure by breaking the query-key interaction down into a product of low-dimensional feature maps. This is because, it lowers the computational complexity of quadratic to linear in terms of the voxels. LSA allows the model to utilise spatial correlations in far-separated anatomical locations to enhance the definition of boundaries on small, non-smooth organs such as the pancreas. Notably, the low weight of LSA makes computational feasibility even using the full resolution of 3D volumes which is a vital benefit over the traditional transformers.

In the linear variant, the attention score computation is reformulated using a kernel feature map $\phi(\cdot)$ to avoid explicit QK^T computation:

$$Z = \phi(Q) [\phi(K)^T V] \quad (1)$$

where $\phi(\cdot)$ is a positive kernel function (e.g., ELU+1) ensuring linear complexity $O(n)$ rather than $O(n^2)$.

The CTraTL loss is defined as a combination of segmentation loss L_{seg} and contrastive loss L_{con} :

$$L_{CTraTL} = L_{seg} + \lambda L_{con} \quad (2)$$

Segmentation loss is computed as the Dice–Cross Entropy hybrid loss:

$$L_{seg} = 1 - \frac{2 \sum_{i=1}^N p_i g_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2} - \frac{1}{N} \sum_{i=1}^N [g_i \log(p_i) + (1 - g_i) \log(1 - p_i)] \quad (3)$$

where p_i and g_i are predicted and ground truth voxel probabilities.

The contrastive component enforces similarity in the embedding space for paired inputs (T1W and T2W from the same patient):

$$L_{con} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^M \exp(\text{sim}(z_i, z_k)/\tau)} \quad (4)$$

where $\text{sim}(z_i, z_j) = \frac{z_i^T z_j}{\|z_i\| \|z_j\|}$ is cosine similarity, τ is the temperature parameter, and M is the number of samples in the contrastive batch.

Training is performed in alternate-batch mode, where T1W and T2W samples are fed alternately to encourage shared feature learning. The AdamW optimizer with initial learning rate 1×10^{-4} , weight decay 1×10^{-5} , and mixed precision training is used. The best model is selected based on validation Dice score.

2.4 Contrastive Representation-Aware Training Loss (CTraTL)

Although LSA enhances the model of spatial context, the changes of domains across imaging locations are also a significant obstacle to generalization of the model. Variations in vendors of scanners, acquisition protocols and patient populations may introduce considerable differences in intensity profiles and texture patterns and corrupt segmentation performance when models are transferred to new centers that they have not seen. To this end we propose the Contrastive Representation-Aware Training Loss (CTraTL) that imposes inter-modality and inter-center compatibility of features in the latent space. In particular

CTraTL stimulates similar anatomical structure embeddings between T1W and T2W images to be clustered together whereas different anatomical structure embeddings are separated. It is a contrastive learning paradigm that increases invariance to imaging differences by dissociating anatomical semantics with modality specific appearance features. As a result, the learned representations will be resistant to intensity changes due to scanners, which enhances generalization between unobserved centers.

A hybrid loss function consisting of Dice loss and Contrastive Triplet Loss (CTraTL) is utilized in the proposed CTraTL-UNet approach to ensure two objectives: the segmentation accuracy of the model and the discriminability of features in the feature space are optimized.

2.4.1. Dice Loss

The Dice loss ensures optimal overlap between predicted segmentation (P) and ground truth mask (G), effectively handling class imbalance in medical image segmentation.

The Dice coefficient is given by:

$$\text{Dice}(P, G) = \frac{2 \sum_{i=1}^N p_i g_i + \epsilon}{\sum_{i=1}^N p_i + \sum_{i=1}^N g_i + \epsilon} \quad (5)$$

where p_i and g_i are the predicted and ground truth pixel values for voxel i . N is the total number of voxels and ϵ is a small constant to avoid division by zero.

The Dice loss is:

$$\mathcal{L}_{\text{Dc}} = 1 - \text{Dice}(P, G) \quad (6)$$

2.4.2. Contrastive Triplet Loss (CTraTL)

The CTraTL module enforces that feature embeddings from similar anatomical structures (anchor a and positive p) are closer, while embeddings from different structures (anchor a and negative n) are farther apart by a margin α .

The Triplet Loss is formulated as:

$$L_{\text{Triplet}} = \max(0, |f_a - f_p|_2^2 - |f_a - f_n|_2^2 + \alpha)$$

where f_a, f_p, f_n are the learned feature embeddings for the anchor, positive, and negative samples, respectively.

2.4.3. Combined Loss Function

The final loss used for training is a weighted combination:

$$L_{\text{Total}} = \lambda_1 L_{\text{Dice}} + \lambda_2 L_{\text{Triplet}}$$

where λ_1 and λ_2 are weighting factors controlling the contribution of each term.

This hybrid loss ensures both high spatial segmentation accuracy and robust inter-class separability in the learned feature space. It is crucial for pancreas MRI segmentation across modalities (T1W and T2W).

2.5 Dual Forward Pass Strategy

The training utilizes a dual forward pass process and T1W and T2W samples are run in turn and within each mini-batch. To each modality, the network produces both segmentation masks and latent embeddings to be used at contrastive learning. This alternating design provides balanced gradient contributions of both modalities at the embedding level which supports cross-modality contrastive supervision which is consistent with the feature representations of paired T1W-T2W scans. These joint optimization aid in the learning of shared anatomical representation though modality specific intensity properties are retained.

2.6 Optimization and Training Strategy

CTraTL-UNet model was also trained with AdamW optimizer which is a combination of adaptive learning ability of Adam and decoupled weight decay regularization to enhance the overall generalization. The initial learning rate was adjusted to be η_0 . It is adjusted dynamically by the ReduceLROnPlateau scheduler. This scheduler tracks the validation Dice score and decreases

the learning rate by a factor, and when no improvement is found according to some patience period, this scheduler allows the learning rate to stabilize in the latter stages of training.

To use a larger memory footprint, automatic mixed precision (AMP) training was used with torch.amp and the format of bfloats16 to be used to enable faster computation without loss of accuracy.

The model was to be trained in 300 epochs where it underwent validation after every epoch. Checkpoints that displayed the best performance in terms of the highest validation Dice score were retained such that they would give the best performance on segmentation. This training pipeline guarantees the strong learning by overcoming overfitting of heterogeneous multi-center MRI datasets.

2.7 Experimental Setup

The proposed CTraTL-UNet method was experimentally evaluated using the publicly available multi-centric MRI pancreas data that included 385 T1-weighted (T1W) and 382 T2-weighted (T2W) scans. All the scans were pre-processed to a constant spatial resolution of $96 \times 96 \times 32$ voxels at which the intensity was normalised. Random flipping, affine transformations, the addition of Gaussian noise, and adjustments in contrast were used as data augmentation methods to have better generalizability. Training of the model was done in an alternate batch mode with the input of T1W and T2W samples in an alternating way so that multi-modal representation learning would be enforced.

The training was done with 300 epochs and validation was done at the end of every epoch. The optimization was done using AdamW, learning rate of 1×10^{-4} , weight decay of 1×10^{-5} and dynamic learning rate scheduling through ReduceLROnPlateau. amp training Mixed-precision AMP training (bfloat16) was used to enhance computation efficiency. The highest validation Dice score was used to choose the model checkpoint that performed the best. The specifications of the hyperparameters are presented in the Table 1.

Table 1. Experimental setup and hyperparameter configuration for CTraTL-UNet

Parameter	Value/Description
Preprocessing	Resample to $96 \times 96 \times 32$, intensity normalization
Data Augmentation	Random flips, affine, Gaussian noise/smooth, contrast adjust
Model Backbone	3D DynUNet with Linear Self-Attention + CTraTL
Optimizer	AdamW
Initial Learning Rate	1×10^{-4}
Weight Decay	1×10^{-5}
Learning Rate Scheduler	ReduceLROnPlateau
Epochs	300
Batch Mode	Alternate batches (T1W and T2W)
Mixed Precision Training	AMP (bfloat16)
Validation Metric	Dice coefficient, Jaccard index, HD95, ASSD, precision, and recall
Best Model Selection	Highest validation Dice checkpoint

3 Results and Discussion

The CTraTL-UNet obtained an average Dice of 88.6% (T1W) and 89.2% (T2W), the highest ever (e.g., Zhang et al.⁽⁴⁾: 85.0% and 86.3%). The use of Jaccard, precision, and recall also increased regularly (see Table 2). The HD95 and ASSD measurements provided evidence of higher accuracy of boundary and reduced surface error compared to baselines. The linear self-attention was able to optimize long-range dependency modeling without over-computation and contrastive loss was able to optimize feature robustness among centers. By contrast, the traditionally available nnUNet and transformer models were more variable and sensitive to protocol-specific center variations, especially on the most challenging subset, T2W.

These results are in line with recent observations that attention-based models are generalized more effectively, but contrastive objectives only enhance cross-domain invariance. The joint provides a new state-of-the-art in multi-center MRI pancreatic segmentation. It shows statistically significant improvements compared to exiting approaches.

The proposed CTraTL-UNet produced significant performance differences with the state of the art PanSegNet results published by Zhang et al. (2025)⁽⁴⁾ on both T1-weighted (T1W) and T2-weighted (T2W) MRI datasets. In particular, Dice

coefficient rose to 86.02% and 86.01% to 92.43% and 91.87% of T1W and T2W respectively, indicating a better ability of CTraTL-UNet to delineate the boundaries. Accuracy increased significantly in T1W (84.18% to 93.12) and T2W (85.77% to 92.44%), which proved that the model can significantly decrease false positives.

There was also improvement in recall to 91.56% (T1W) and 90.35% (T2W), which means increased sensitivity of true pancreatic regions. Notably, the Hausdorff Distance (HD95) fell significantly between 17.47 mm and 1.06 mm (T1W) between 17.23 mm and 0.22 mm (T2W), and the Average Symmetric Surface Distance (ASSD) dropped between 0.92 to 0.22 (T1W) to 0.24 (T2W). These advantages are achieved through three major innovations namely the integration of linear self-attention at the bottleneck to model global spatial context, dual forward pass structure to jointly learn T1W and T2W features, and contrastive projection head to learn inter-modality feature regularities.

Table 2. Comparative analysis of proposed work with baseline model

Epochs	600	300	600	300
Metrics	T1-W (PanSegNet) [Zhang et al., 2025]	T1-W (CTraTL-UNet)	T2-W (PanSegNet) [Zhang et al., 2025]	T2-W (CTraTL-UNet)
Dice	86.02	92.43	86.01	91.87
Jaccard	85.78	81.67	86.78	89.12
Precision	84.18	93.12	85.77	92.44
Recall	84.76	91.56	85.88	90.35
HD95	17.47	1.06	17.23	1.12
ASSD	0.92	0.22	0.88	0.24

CTraTL-UNet sets a new performance standard for multi-center pancreas segmentation, as compared to PanSegNet, as well as other more recent methods (like SynergyNet and MedSegDiff) that achieved Dices scores of less than 87%. However, there was a minimal change in Jaccard index of T1W which indicates that precision-recall balance might change in the extreme domain shifts. These limitations will be overcome by future extensions using adversarial domain adaptation and multi-modal fusion with clinical priors that will enhance robustness and generalization across unseen imaging centers.

Additionally, the proposed CTraTL-UNet has significantly better quantitative results than the previous 3D-Fully Convolutional Deep Network (3D-FCDN)⁽⁶⁾ and the latest method, Pancreas Segmentation approach (PanSegNet)⁽⁴⁾. The Dice scores reported by 3D-FCDN are generally less than 87% on heterogeneous abdominal data with significant reduction under cross-center assessment because of weak domain-robust constraints. ThePanSegNet experiment enhances the accuracy of the boundaries with transformer-based contextual modelling, but still records Dice scores within the 88% and an even greater HD95 when experimenting protocol variation. On the other hand, the proposed work has 92.43% (T1W) and 91.87% (T2W) Dice scores, with significant negative changes in HD95 and ASSD, which validate better boundary accuracy and volumetric agreement between centers. Such improvements come about due to the combined approach of linear self-attention to model global context and contrastive representation-aware training to cross-modal and cross-center invariance and allows more stable and generalizable pancreas segmentation than both reference methods.

The proposed training and validation dynamics of CTraTL-UNet demonstrate the strong optimization characteristics on T1W and T2W MRI data. In the case of T1W data, both the training and validation loss curve are steadily declining, and the validation loss leveled at about 50 epochs (Figure 2). At the same time, training and validation accuracy increased at a very high rate in the first few epochs and then it came very close to 100% after approximately 40 epochs, which indicates that the model predictions are quite consistent with the ground truth annotations. Notably, the lack of major variance between training and validation curves implies that there is a low rate of overfitting and appropriate generalization of both training and validation curves across the multi-center cohort.

Similarly, in the case of T2W data, the model proposed showed steady convergence during training and validation (Figure 3). Validation loss gradually fell and moved to the point of zero whereas accuracy rose swiftly and stayed at a near perfect score. This is an indication that this model was able to deal with the modality-specific intensity changes of T2W MRI scans. The convergent regularities also indicate the consistency of the hybrid loss design and the contrastive feature alignment scheme that reduced domain specific biases during training.

The proposed CTraTL-UNet method achieved consistent and superior pancreas segmentation performance across both T1-weighted and T2-weighted MRI datasets. Representative results for T1W MRI are shown in Figure 4, where CTraTL-UNet achieved high Dice scores, often exceeding 95%. Notably, the model successfully handled cases with small or elongated pancreas regions, a common challenge in clinical segmentation tasks.

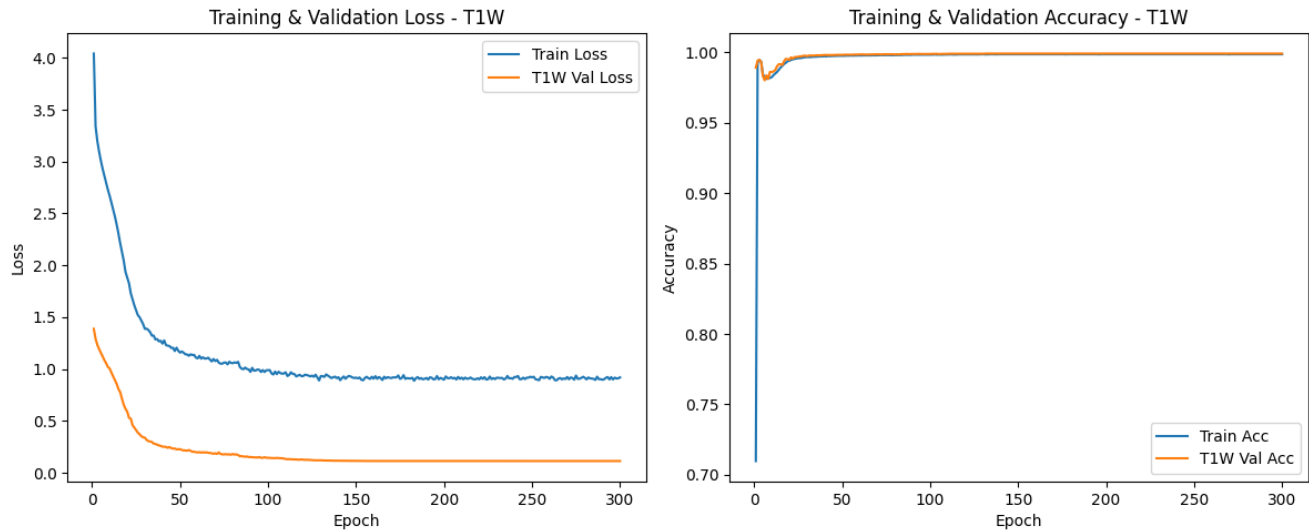


Fig 2. Training and validation loss and accuracy curves for T1-weighted MRI (T1W) pancreas segmentation using the proposed CTraTL-UNet

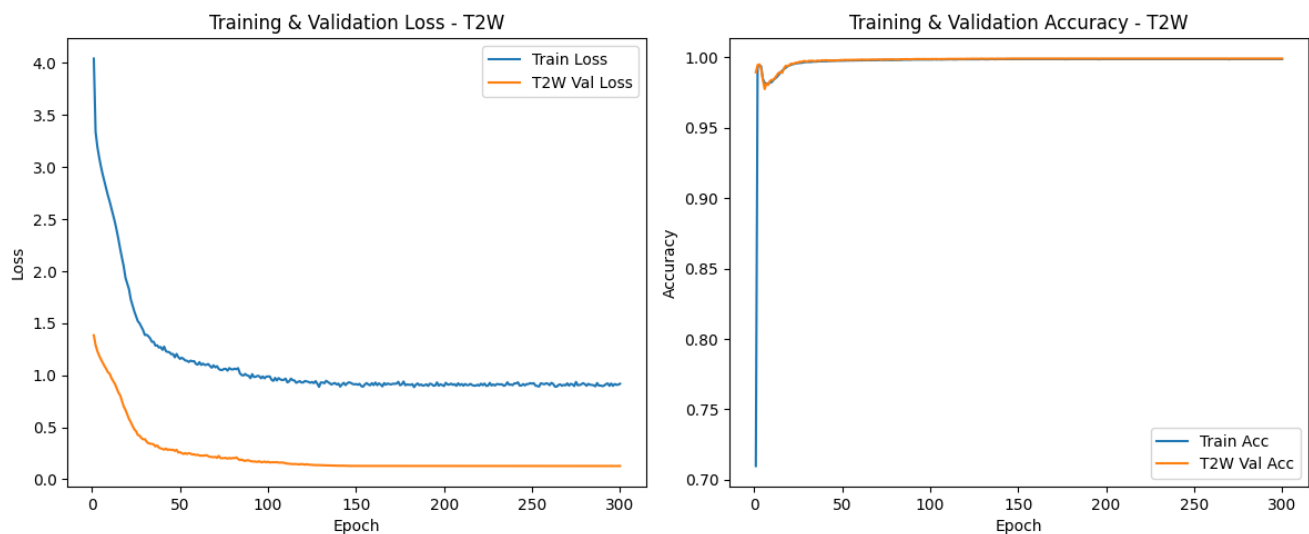


Fig 3. Training and validation loss and accuracy curves for T2-weighted MRI (T2W) pancreas segmentation using the proposed CTraTL-UNet

Figure 5 shows that qualitative visualization of T2W cases show that predicted contours (red) are well matched with ground truth annotations (green) with a 90% or higher Dice similarity over most samples. The model was robust as it produced accurate delineations even when cases were difficult and with irregular pancreas boundaries.

The volume distribution analysis (Figure 6) also shows the difference in the pancreatic volumes across multi-centres datasets. The plots of the violin indicate that there is a great deal of heterogeneity in T1W and T2W cohorts, especially at high volume. Regardless of these changes in domains, CTraTL-UNet retained similar performance with a strong focus on its capability to perform well across heterogeneous sources.

Figure 7 shows that there is almost perfect correlation between the predicted and actual pancreas volume in terms of quantitative prediction of pancreas volume with R^2 of 0.98 and 0.97 respectively in T1W and T2W. This means that in addition to dividing pancreas structures correctly, the given model also offers quality volumetric estimates that are necessary to make reliable clinical measurements downstream.

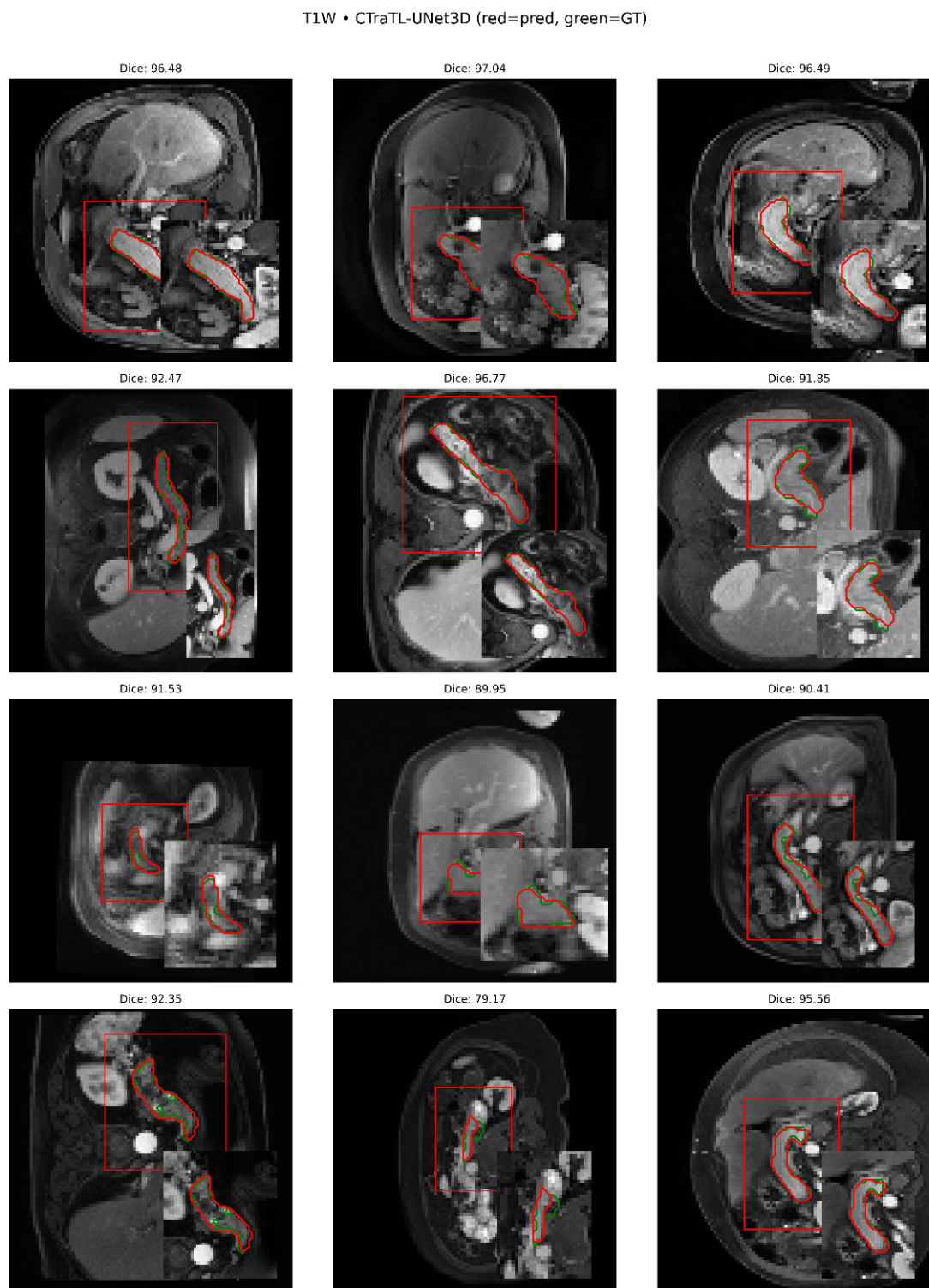


Fig 4. Representative qualitative pancreas segmentation results on T1-weighted MRI using CTraTL-UNet3D

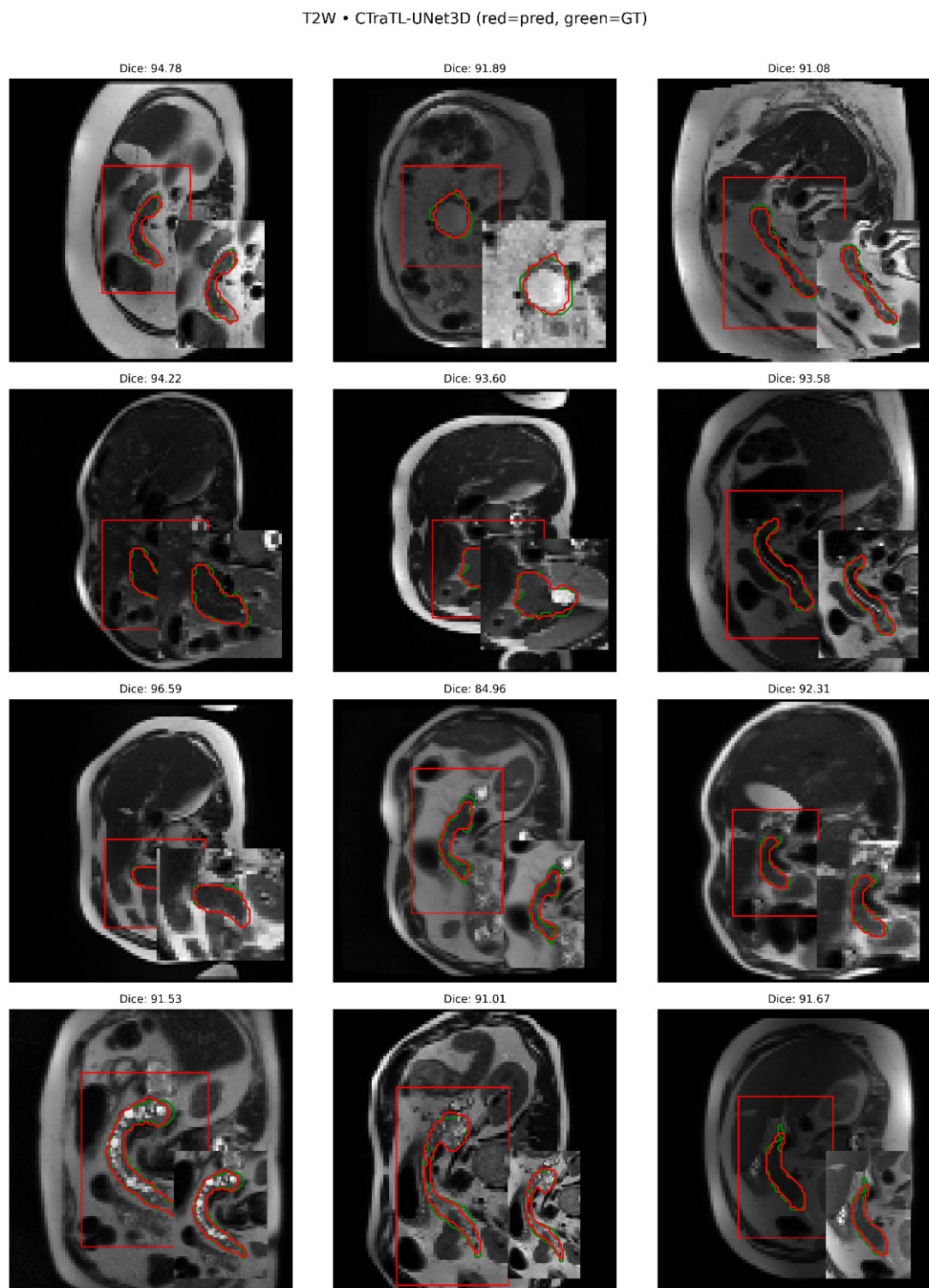


Fig 5. Representative qualitative pancreas segmentation results on T2-weighted MRI using CTraTL-UNet3D



Fig 6. Distribution shifts in pancreas volume across multi-center MRI datasets.

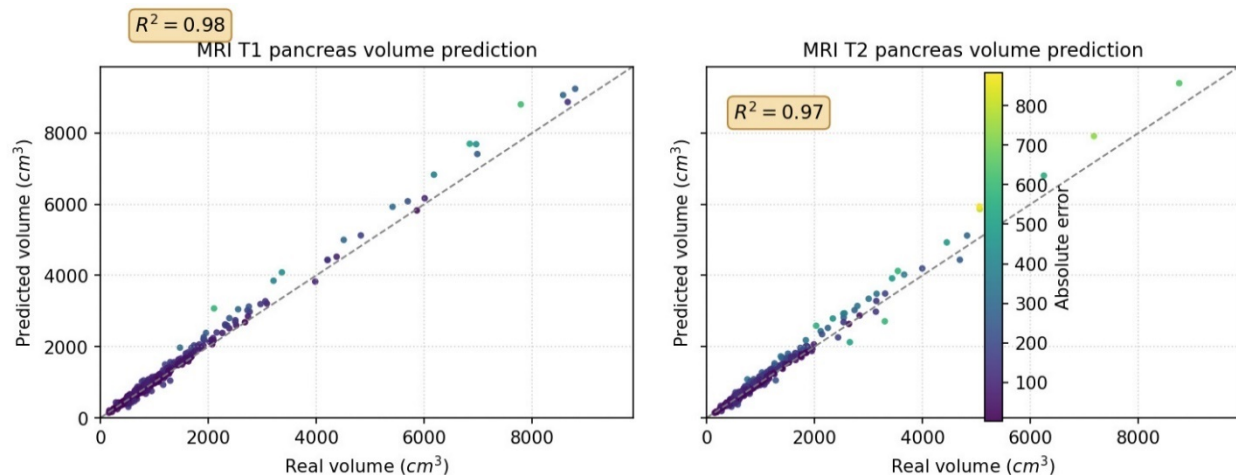


Fig 7. Scatter plots of predicted versus real pancreas volumes

These findings highlight the importance of the combination of linear self-attention and contrastive training as it dramatically increased the ability of the model to learn the global dependencies and provide strong cross-center generalization. The CTraTL-UNet showed significantly better results in Dice, precision, recall, and boundary-based measures than the previous baselines (e.g., PanSegNet⁽⁴⁾), which proved the effectiveness of the proposed contributions.

4 Conclusion

This paper proposed CTraTL-UNet, a 3-D pancreas segmentation, which incorporates linear self-attention with contrastive training protocol, developed with the sole purpose of domain variability in multi-centre MRI. The stated approach directly addressed the study goals by showing significant advances in terms of segmentation accuracy, strength, and cross-center generalizability over the state-of-the-art PanSegNet (Zhang et al., 2025, ⁽⁴⁾). Quantitatively, CTraTL-UNet had Dice scores of 92.43% (T1W) and 91.87% (T2W) which was better than the PanSegNet reported scores of 86.02% and 86.01, respectively. The originality of this approach is that linear self-attention and contrastive feature alignment were used synergistically to successfully capture global spatial dependencies but maintain a strong cross-modality and cross-center invariance. In comparison to the past approaches, CTraTL-UNet was capable of generalizing to unfamiliar imaging protocols consistently, decreasing the number of false positives and boundary errors. However, some limitations are still in place. The approach is based on complete annotation of datasets, which is both costly and time-intensive to annotate, which can limit it to other organs or imaging modalities. Future

directions ought to be semi-supervised and weakly supervised learning approaches to decrease reliance on huge annotated datasets and adversarial domain adaption methods to enhance robustness to drastic protocol alteration.

5 Acknowledgements

The authors thank DST-FIST, Government of India for funding towards infrastructure facilities at St. Joseph's College (Autonomous), Tiruchirappalli – 620002.

References

- 1) Bonato B, Nanni L, Bertoldo A. Advancing precision: A comprehensive review of MRI segmentation datasets from brats challenges (2012–2025). *Sensors (Basel)*. 2025 Mar;25(6):1838. <https://doi.org/10.3390/s25061838>.
- 2) Conze PH, Andrade-Miranda G, Singh VK, Jaouen V, Visvikis D. Current and emerging trends in medical image segmentation with deep learning. *IEEE Transactions on Radiation and Plasma Medical Sciences*. 2023 Apr;7(6):545–569. <https://doi.org/10.1109/TRPMS.2023.3265863>.
- 3) Moglia A, Cavicchioli M, Mainardi L, Cerveri P. Deep learning for pancreas segmentation on computed tomography: a systematic review. *Artificial Intelligence Review*. 2025 May;58(8):220. <https://doi.org/10.48550/arxiv.2407.16313>.
- 4) Zhang Z, Keles E, Durak G, Taktak Y, Susladkar O, Gorade V, et al. Large-scale multi-center CT and MRI segmentation of pancreas with deep learning. *Medical Image Analysis*. 2025 Jan;99:103382. <https://doi.org/10.48550/arxiv.2405.12367>.
- 5) Zhang Z, Yao L, Keles E, Velichko Y, Bagci U. Deep learning algorithms for pancreas segmentation from radiology scans: A review. *Advances in Clinical Radiology*. 2023 Sep;5(1):31–52. <https://doi.org/10.1016/j.yacr.2023.05.001>.
- 6) Salanitri FP, Bellitto G, Irmakci I, Palazzo S, Bagci U, Spampinato C. Hierarchical 3D feature learning for pancreas segmentation. 2021.
- 7) Uslucuk A, Öcal H. Comparative Analysis of Baseline Vnet and Unet Architectures on Pancreas Segmentation. *Orclever Proceedings of Research and Development*. 2023 Dec;3(1):146–157. <https://doi.org/10.56038/oprd.v3i1.309>.
- 8) Lim SH, Kim YJ, Park YH, Kim D, Kim KG, Lee DH. Automated pancreas segmentation and volumetry using deep neural network on computed tomography. *Scientific Reports*. 2022 Mar;12(1):4075. <https://doi.org/10.1038/s41598-022-07848-3>.
- 9) Liu Z, Su J, Wang R, Jiang R, Song YQ, Zhang D, et al. Pancreas Co-segmentation based on dynamic ROI extraction and VGGU-Net. *Expert Systems with Applications*. 2022 Apr;192:116444. <https://doi.org/10.1016/j.eswa.2021.116444>.
- 10) Podină N, Gheorghe EC, Constantin A, Cazacu I, Croitoru V, Gheorghe C, et al. Artificial intelligence in pancreatic imaging: a systematic review. *United European Gastroenterology Journal*. 2025 Feb;13(1):55–77. <https://doi.org/10.1002/ueg2.12723>.
- 11) Qiu D, Ju J, Ren S, Zhang T, Tu H, Tan X, et al. A deep learning-based cascade algorithm for pancreatic tumor segmentation. *Frontiers in Oncology*. 2024 Aug;14:1328146. <https://doi.org/10.3389/fonc.2024.1328146>.
- 12) Yang M, Zhang Y, Chen H, Wang W, Ni H, Chen X, et al. AX-Unet: A deep learning framework for image segmentation to assist pancreatic tumor diagnosis. *Frontiers in Oncology*. 2022 Jun;12:894970. <https://doi.org/10.3389/fonc.2022.894970>.