DiffuPan: Diffusion-Based Framework for Multi-Phase Contrast-Enhanced CT Pancreatic Tumor Detection

Dr. D. Narmadha [1,2], Prof. Shashi Kant Gupta[3],

[1] Lincoln University College, Malaysia;

[2] Division of AIML, Karunya Institute of Technology and Sciences, Coimbatore, India;

[3] Lincoln University College, Malaysia
Email ID: pdf.narmadha@lincoln.edu.my narmadha@karunya.edu shashigupta@lincoln.edu.my

Abstract - Pancreatic cancer has a high mortality rate, and outcomes improve when tumors are identified early and delineated with precision. Contrast-enhanced CT is central to diagnosis and planning, yet classical segmentation approaches often miss irregular boundaries and show weak generalization across contrast phases. Recent convolutional and transformer architectures, including U-Net, Attention U-Net, and TransUNet, have raised baseline performance, but they typically rely on a single phase and struggle to capture complementary information across arterial, venous, and delayed acquisitions. This work presents DiffuPan, a diffusion-based encoder–decoder that performs cross-phase attention with residual feature fusion to couple information from all three phases. Training uses hybrid supervision that combines Dice, Focal, and SSIM losses to encourage accurate boundaries and coherence of fine structures. Experiments were run on the TCIA Pancreas-CT cohort comprising 300 patients and roughly 80,000 annotated slices. Ablation studies were designed to isolate the contributions of multi-phase fusion and diffusion guidance. DiffuPan obtained a Dice score of 92.3%, precision of 93.1%, recall of 92.0%, and an AUC of 0.97. These results exceed nnU-Net (88.2% Dice) and TransUNet (87.4% Dice) on the same data. The false-positive rate was 3.2% and the false-negative rate was 4.5%. The results suggest that the results suggest that the use of the diffusion-guided multi-phase integration is likely to result in more accurate tumor segmentations and more robust applicability across different scans, thus making it a proper choice for clinical segmentation of pancreatic lesions.

Keywords: Pancreatic Tumor Segmentation, Diffusion Models, Multi-Phase CT, Deep Learning, Medical Image Analysis, Hybrid Loss Functions, Robustness Evaluation

1 Introduction

Pancreatic cancer remains a major cause of cancer mortality. Survival improves when lesions are identified early and their boundaries are mapped with precision so that curative surgery is possible. Contrast-enhanced computed tomography is central to diagnosis and treatment planning because it captures vascular and parenchymal information across arterial, venous, and delayed phases. Accurate segmentation is difficult in this setting. Tumors often exhibit irregular margins and phase-dependent appearance, while image quality varies across scanners and protocols. Manual annotation is slow and

inconsistent across raters, which restricts large-scale use. Automatic methods face low contrast-to-noise ratios, heterogeneous visual patterns, and acquisition variability.

Convolutional encoder–decoder models such as U-Net improve pixel-level performance, yet they tend to miss subtle or infiltrative regions, which increases false negatives. Transformer-based designs including Swin-UNet and TransUNet enhance long-range context but commonly use a single phase and therefore do not benefit from complementary temporal information. Many systems are also sensitive to noise and small perturbations, which limits reliability in clinical practice.

This study proposes DiffuPan, a multi-phase segmentation framework that jointly leverages arterial, venous, and delayed CT series. Residual encoders extract phase-specific features, and cross-phase attention aligns and fuses cues that are informative across time. Diffusion-guided learning strengthens representations against noise and distribution shifts. Training uses a hybrid objective that combines Dice, Focal, and SSIM losses to encourage overlap accuracy, handle class imbalance, and preserve structural detail. The design prioritizes a practical trade-off between accuracy and computational cost to support use in routine clinical workflows.

The contributions of this study are as follows:

- Propose DiffuPan, a diffusion-based multi-phase segmentation framework tailored to pancreatic tumor analysis.
- Design a cross-phase attention module that leverages complementary cues from arterial, venous, and delayed CT phases.
- Employ a hybrid training objective that couples Dice, Focal, and SSIM losses to sharpen boundaries, handle class imbalance, and preserve structural detail.
- Provide a comprehensive evaluation on the TCIA Pancreas-CT cohort, including head-to-head baselines, ablation studies, robustness probes, statistical testing, and targeted error analysis.

The remainder of this paper is organized as follows. Section 2 reviews prior work on pancreatic tumor segmentation and deep learning methods. Section 3 details the dataset, preprocessing pipeline, and the proposed architecture. Section 4 reports experiments, including comparative results, ablations, and robustness assessments. Section 5 discusses findings, limitations, and implications for clinical use. Section 6 concludes and outlines directions for future investigation.

2 Related Work

Prior research on pancreatic organ and tumor segmentation spans organ-focused models, tumor–vessel analysis, and multimodal fusion. Mahmoudi et al. [1] coupled a CNN with texture descriptors to delineate PDAC and adjacent vessels, capturing many tumor–vessel interfaces but showing reduced accuracy for very small vessels and limited cross-center generalization in the absence of extensive multi-phase data.

Mukherjee et al. [2] trained a large-scale 3D nnU-Net on more than 3,000 CT scans with external validation on AbdomenCT-1K, reaching Dice scores up to 0.96 for pancreas anatomy. The emphasis on whole-organ segmentation, however, left small, heterogeneous tumor boundaries only partially resolved. Suri et al. [3] benchmarked multiple CT-based models to examine drivers of pancreas segmentation quality, yet the analysis relied on organ-level labels with minimal tumor-specific annotation. Work on tumor–vessel interaction has advanced clinical assessment while exposing segmentation gaps. Bereska et al. [4] used a semi-supervised approach on 467 patients to estimate vascular contact in PDAC, aiding resectability evaluation, but vessel masks were difficult in complex anatomy and validation across multi-phase imaging remained incomplete. Zhou et al. [5] proposed SMF-Net, a semantic-guided multimodal fusion model that raised tumor delineation accuracy, although performance depended on well-aligned modalities and was hampered by ambiguous margins and limited data scale. Viviers et al. [6] incorporated secondary clinical cues, including ductal and biliary structures, and achieved high sensitivity and specificity for detection, but the framework did not directly target precise mask generation.

Survey and multi-stage pipelines further clarify strengths and limits of current designs. Karri et al. [7] summarized deep learning pipelines centered on U-Net, V-Net, and related variants, consolidating evidence without new experiments. Ramaekers et al. [8] presented a multi-stage U-Net which was capable of leveraging secondary signs like ductal dilation to achieve a sensitivity of 0.97 and a specificity of 1.00. The tumor Dice score was still about 0.37, suggesting that localization benefits did not automatically lead to the accurate boundary delineation. Perik et al. [9] combined deep learning with CT perfusion to characterize PDAC vascular phenotypes and reported AUC near 0.86, but reliance on perfusion CT limits broad adoption.

Multi-center studies underscore the value of global context modeling and hybrid encoders while revealing persistent blind spots. Suri et al. [10] and Zhang et al. investigated CT and MRI cohorts and showed that transformer or hybrid architectures often reach tumor Dice of 88–90 percent, yet performance drops for small lesions and variable enhancement patterns across phases. Dong et al. [11] presented AMFF-Net with residual attention and transformer modules, reporting pancreas Dice of 82.1 percent and tumor Dice of 57.0 percent; improvements on subtle, low-contrast tumors were still constrained. Li et al. [12] proposed CausegNet, a causal learning framework with counterfactual loss, achieving Dice scores of 86.7 percent for pancreas and 84.3 percent for tumor, at the expense of higher computational cost and a requirement for sequential CT inputs. Qiu et al. [13] proposed a cascade in which pancreas segmentation precedes tumor localization. The design raised Dice scores relative to earlier baselines, although boundary sharpness and false positives remained problematic. Viriyasaranon et al. [14] introduced an annotation-efficient scheme that generates pseudo-lesions to lower labeling cost and improve detection across populations. Performance was influenced by biases introduced through synthetic labels, which reduced segmentation fidelity. Mandal et al. [15] examined weakly supervised detection on large cohorts and reduced reliance on dense annotation, yet fine-grained masks

at the boundary level were still imprecise. Gandikota et al. [16] coupled W-Net segmentation with a classifier optimized by a swarm algorithm, which improved diagnostic accuracy, while robustness across contrast phases received less attention. Mekala and Kumar [17] introduced an optimization-driven Efficient DenseNet that achieved detection accuracy above 94 percent, although fine-grained tumor boundary segmentation was not addressed. Chen et al. [18] validated a nationwide detection system with high sensitivity and specificity and showed that large-scale deployment is feasible; the study centered on detection rather than pixel-level delineation. Parallel efforts in dataset construction have expanded training diversity while adding sources of variability. The PanTS collection [19] assembles more than 36,000 multi-institutional CT scans with voxel-wise labels, which supports broader generalization but still exhibits inter-annotator differences and irregular phase metadata. Methodologically, Zeng et al. [20] proposed SCPMan, a prior-constrained attention architecture that uses shape context to improve pancreas segmentation, with evaluation directed at organ masks rather than precise tumor contours. Overall, the literature advances causal modeling, weak supervision, optimization-aware training, and large-scale curation, yet gaps remain in tumor Dice performance, sensitivity to small or low-contrast lesions, and the underuse of multi-phase CT. These gaps motivate diffusion-guided, multi-phase fusion approaches such as DiffuPan.

3 System Methodology

In the proposed DiffuPan system, the process starts with pre-processing the image, then proceeds to establish input representations, construct the network architecture, incorporate multi-phase feature fusion, learn representations through diffusion guidance, and ultimately optimize using a hybrid loss function. The whole process is illustrated in Figure 1. The detailed information of each stage is provided below.

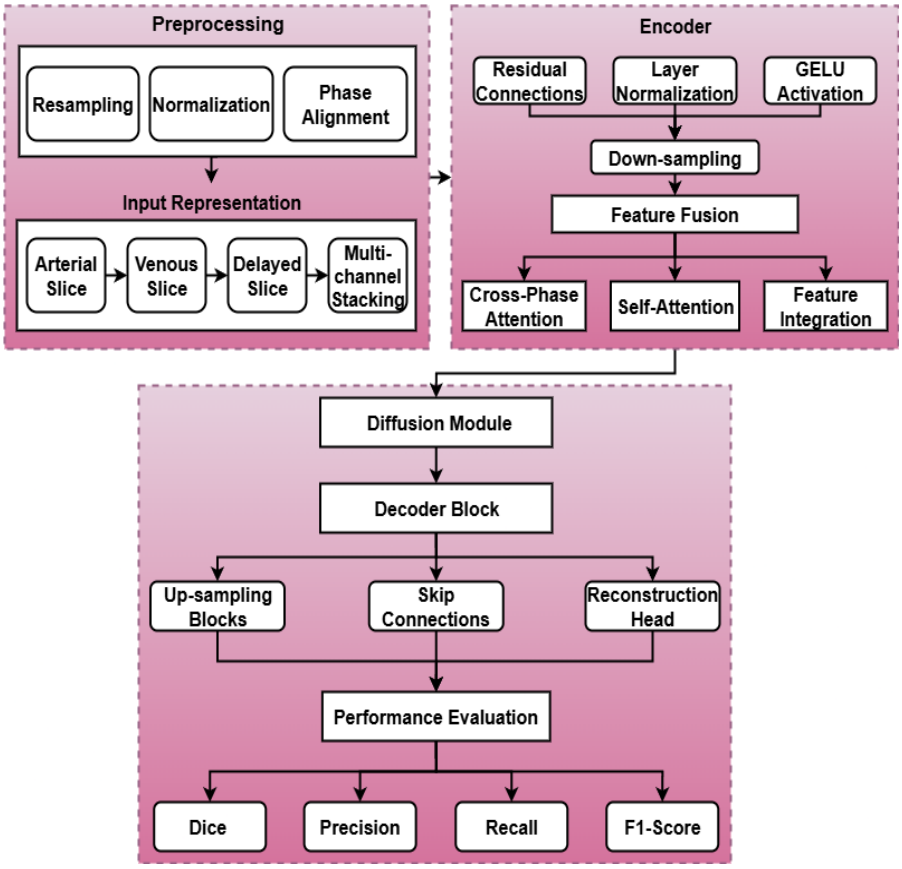

Figure 1: Block diagram of the proposed DiffuPan framework.

3.1 Image Preprocessing

CT slices are resampled to 256×256 pixels to maintain consistent in-plane resolution across subjects. Voxel intensities are linearly normalized to [0, 1] to reduce inter-scanner variation and to stabilize optimization during training. These steps provide a uniform data scale and geometry for subsequent modeling.

3.2 Input Representation

Following preprocessing, arterial, venous, and delayed phases are rigidly aligned and concatenated as a three-channel volume,

$$X = \{X_a, X_v, X_d\}, \quad (1)$$

As defined in Eq. (1), X_a, X_v, and X_d denote arterial, venous, and delayed slices. The composite input X retains complementary vascular and parenchymal cues across phases, which supports reliable delineation of tumor boundaries.

3.3 Network Architecture Design

DiffuPan adopts an encoder–decoder U-Net backbone augmented with diffusion-guided residual pathways. The encoder contains five down-sampling stages that halve spatial resolution while increasing channel depth. Features at stage l are computed as

$$F_l = \sigma\left(LN\left(W_l * F_{l-1} + b_l\right)\right), \tag{2}$$

where F_{l-1} is the input feature map, W_l and b_l are the convolution kernel and bias, $LN(\cdot)$ denotes layer normalization, and $\sigma(\cdot)$ is the nonlinearity. As indicated in Eq. (2), residual links and normalization stabilize training and support gradient flow.

The decoder mirrors the encoder with five up-sampling stages that restore spatial detail and concatenate the corresponding encoder features through skip connections:

$$D_l = \phi(U_p(D_{l+1}) \oplus F_l), \tag{3}$$

where D_{l+1} is the deeper decoder map, $Up(\cdot)$ is bilinear up-sampling, $\oplus$ denotes channel concatenation, and $\phi(\cdot)$ is the decoder block transform. Consistent with Eq. (3), this pathway preserves fine anatomical boundaries.

The latent bottleneck is fixed at 512 channels to balance capacity and computational cost. Each convolutional block uses the GELU activation,

$$\sigma(x) = x \cdot \Phi(x), \tag{4}$$

with $\Phi(x)$ the Gaussian cumulative distribution. As in Eq. (4), GELU provides smooth gradients that aid optimization in deeper stacks.

To fuse information across contrast phases, cross-phase attention modules are inserted at the bottleneck and selected decoder levels, while self-attention modules refine dependencies within each phase. A representative cross-phase attention is

$$Attn(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{5}$$

where Q, K, and V are the query, key, and value projections of multi-phase features and d is the scaling term. Equation (5) enables selective fusion and alignment of complementary phase cues.

Through the utilization of residual encoding, skip-connected decoding, GELU activations, layer normalization, and targeted attention, the network is able to preserve global context while embracing local detail that is crucial for the successful segmentation of heterogeneous pancreatic tumors in multi-phase CT.

3.4 Multi-Phase Feature Fusion

The encoder is the one that is responsible for generating phase-aware features in the first place. It is the application of a cross-phase attention mechanism that promotes information sharing across phases by aggregating each phase with weighted contributions from the others:

$$F_i' = \alpha_i F_i + \sum_{j \neq i} \beta_{ij} F_j \quad (6)$$

where F_i denotes features from phase i, α_i is a learnable scaling term, and β_{ij} are attention weights from phase j to phase i. As indicated in Eq. (6), this operation integrates complementary cues across phases, while a separate self-attention pathway refines intra-phase context.

3.5 Diffusion-Guided Representation Learning

In order to make the latent representations more robust, a diffusion process should be applied. The forward step gradually perturbs a clean latent x_0 with Gaussian noise,

$$q(x_t|x_0) = N\left(x_t;\ \sqrt{\alpha_t}x_0, (1-\alpha_t)I\right) \quad (7)$$

where x_t is the noisy sample at step t and α_t controls the noise schedule. The reverse step predicts a denoised sample,

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \sum_\theta(x_t, t)) \quad (8)$$

Together, Eqs. (7) and (8) define a denoising pathway that stabilizes feature embeddings and improves resilience to noise and acquisition variability.

3.6 Hybrid Loss Optimization

Training uses a composite objective that couples Dice, Focal, and SSIM terms. The Dice loss targets region overlap,

$$\mathcal{L}_{Dice} = 1 - \frac{2|P \cap G|}{|P| + |G|} \quad (9)$$

where P and G are the predicted and reference masks. Equation (9) directly promotes overlap between predictions and ground truth.

$$\mathcal{L}_{Focal} = -\alpha(1-p_t)^\gamma \log(p_t) \quad (10)$$

with p_t the predicted probability for the true class, α a weighting factor, and γ a focusing parameter. As in Eq. (10), hard samples receive greater emphasis during optimization.

Structural similarity is enforced through an SSIM term,

$$L_{SSIM} = 1 - SSIM(P, G) \tag{11}$$

which encourages preservation of boundary detail and local contrast, as indicated in Eq. (11). The final training criterion is a weighted sum,

$$L_{Hybrid} = \lambda 1 L_{Dice} + \lambda 2 L_{Focal} + \lambda 3 L_{SSIM}, \tag{12}$$

where λ_1, λ_2, and λ_3 control the contribution of overlap maximization, class reweighting, and structural fidelity. Equation (12) integrates these complementary objectives to improve boundary accuracy, mitigate imbalance, and maintain coherent anatomy.

The training workflow is summarized in Algorithm 1.

Algorithm 1 DiffuPan training workflow

Require: Mini-batch of multi-phase CT slices $\{(X_a, X_v, X_d), G\}$
Ensure: Updated network parameters

1: **procedure** $TRAIN(\{(X_a, X_v, X_d), G\})$
2: **Preprocessing:** Resample to 256 × 256, normalize to [0, 1], align phases
3: Form multi-channel input $X \leftarrow \{X_a, X_v, X_d\}$
4: **Encoding:** Pass X through five down-sampling stages with residual blocks to obtain $\{F_1, \ldots, F_5\}$
5: **Fusion:** Apply cross-phase attention and intra-phase self-attention on encoder features
6: **Diffusion refinement:** Apply forward noise and learned reverse denoising to refine latent features
7: **Decoding:** Use five up-sampling stages with skip connections; generate prediction mask P
8: **Loss:** Compute Dice, Focal, and SSIM; combine into hybrid objective
9: **Update:** Backpropagate and update parameters with AdamW
10: **end procedure**

3.7 Overall Framework

All the operations in (1)–(12) are integrated into the complete framework. The inference pipeline is presented in Algorithm 2. The preprocessed multi-phase inputs are considered as multi-channel representations, and they are processed by a five-level encoder-decoder U-Net backbone with residual connections, refined through diffusion-guided learning, and optimized by hybrid supervision. This well-thought design is a compromise of gaining accuracy, reliability, and speed simultaneously. This makes

DiffuPan appropriate for medical pancreatic tumor segmentation.

Algorithm 2 DiffuPan inference workflow

Require: Multi-phase CT slices (X_a, X_v, X_d)
Ensure: Predicted tumor segmentation
mask P 1: **procedure**
$INFER((X_a, X_v, X_d))$
2: **Preprocessing:** Resample and normalize; form $X \leftarrow \{X_a, X_v, X_d\}$
3: **Encoding and fusion:** Extract encoder features and fuse phases via attention
4: **Diffusion refinement:** Apply learned denoising to stabilize latent representation
5: **Decoding:** Reconstruct mask with skip-connected decoder to
obtain P 6: **Postprocessing:** Optionally remove small components
and fill holes 7: **end procedure**

4 Experimental Results

This section demonstrates a complete assessment of the suggested DiffuPan framework with the aid of the Pancreas-CT dataset the results offer insights into characteristics, architectural and training settings, and the like, through various types of analysis such as comparative performance, ablation studies, efficiency, robustness, statistical validation, and error analysis.

4.1 Dataset Description

The publicly accessible Pancreas-CT dataset from The Cancer Imaging Archive (TCIA) was used for the experiments. There are three different phases in this dataset, and these are arterial, venous, and delayed phases. The total number of patients in the dataset is 300, their axial slices are about 80,000. The annotations for tumor and pancreas masks which are the signs of high-quality are the support for the pixel-level supervision in segmentation jobs; that is, each slice has an accompanying high-quality annotation provided in two regions, and their annotations have already been done through pathology. Details of the dataset are presented in the form of a Table 1, where various aspects related to the dataset such as modality, imaging phases, patient number, slice number, and annotation type are all comprehensively dealt with. The tremendous dataset with detailed annotations and standardized imaging quality represents it as a very dependable tool for a scientific community working in the domain of pancreatic cancer imaging.

Table 1: Dataset Description

Dataset	Modality	Phases	No. of Patients	No. of Slices	Annotation Type
Pancreas-CT (TCIA)	Contrast-Enhanced CT	Arterial, Venous, Delayed	300	80,000	Tumor + Pancreas Masks

4.2 Architectural Configuration

The architecture suggested for DiffuPan is an encoder-decoder type, uniquely designed to handle multi-phase contrast-enhanced CT imaging data. The input slices were resized to 256 × 256 pixels and phase-aligned before being joined together. A single-channel matrix of pixels to represent the images was created, one channel for each arterial, venous, and delayed phases. By utilizing the above-mentioned channels fed into the input layer, the network can be directed to learn the contrast from different points of view. The residual blocks that are embedded in the encoder assist in stabilizing gradient flow and preserving fine structures through five resolution levels. The decoder, on the other hand, uses the skip connections to bring back the spatial detail lost through resolutions and to match the scales of the encoder. Besides, the cross-phase attention mechanism works in defining and modulating the cooperation between the phases, whereas the self-attention assists in learning and refining the intra-phase relationships. The model's space has been completely restricted to a latent dimensionality to 512 which allows better and cost-effective learning to happen. Another important point is that GELU, which stands for Gaussian Error Linear Unit ad Layer Normalization have been used for activations and normalization respectively as they help in the stabilization of the training procedure across varied batch sizes. All of these can be better observed in the comparison of the principal architectural choices presented in Table 2.

Table 2: Architectural Configuration

Component	Configuration Details
Input Size	256×256 pixels
Input Channels	Multi-phase CT slices (Arterial, Venous, Delayed)
Encoder Backbone	Diffusion-based U-Net variant with residual connections
Number of Encoder Layers	5 (down-sampling path)
Number of Decoder Layers	5 (up-sampling path with skip connections)
Attention Mechanism	Cross-phase attention + self-attention blocks
Latent Dimensionality	512
Activation Function	GELU
Normalization	Layer Normalization

4.3 Training Configuration

Optimization uses AdamW with decoupled weight decay to promote generalization under explicit regularization. The learning rate is initialized at (1×10^{-4}) and scheduled with cosine annealing to yield a smooth decay that avoids abrupt plateaus. Supervision employs a hybrid objective that combines Dice,

Focal, and SSIM losses: Dice improves overlap for small lesions, Focal limits the influence of easy negatives in imbalanced settings, and SSIM preserves local structure near boundaries. A mini-batch size of 16 provides stable gradients within available memory. Training runs for 200 epochs to achieve convergence while controlling overfitting. The implementation is in PyTorch 2.2 to ensure reproducibility and efficient operator support. Table 3 summarizes the full training configuration.

Table 3: Training Configuration

Component	**Configuration Details**
Optimizer	AdamW
Initial Learning Rate	1×10^{-4} (cosine annealing scheduler)
Loss Functions	Hybrid (Dice + Focal + SSIM)
Batch Size	16
Training Epochs	200
Framework	PyTorch 2.2

4.4 Comparative Tumor Segmentation Results

Table 4 reports tumor segmentation results for representative baselines and DiffuPan. Classical convolutional models such as U-Net and Attention U-Net yield moderate Dice and AUC, whereas hybrid and transformer-based variants including Swin-UNet, DeepLabV3+, and TransUNet show better overlap and discrimination. The nnU-Net baseline reaches a Dice of 88.2 percent. DiffuPan achieves highest performance in terms of Dice of 92.3 percent and Area Under the Curve (AUC) of 0.97, which represents balanced precision and recall value with an enhanced class separability. The improvement over the state-of-the-art model, nnU-Net, shows the contribution of multi-phase fusion and diffusion-guided representation learning. Figure 2 provides some qualitative examples that are consistent with these quantitative gains.

4.5 Phase-Wise Ablation Study

The effect of contrast-phase composition on tumor segmentation is summarized in Table 5. Among the phases, single phase devices have the lowest overlap with the venous class and slightly more with the arterial class. Complementary arterial and venous induction of Dice and AUC respectively is in line with the additive vascular and parenchymal signals. The delayed phase addition leads to further increase, which indicates that the washout dynamics contribute to marginal resolution. The multi-phase fusion is giving the best results (Dice = 92.3 percent and AUC = 0.97) with better robustness and better class separability when the information of the arterial, venous and delayed are modelled jointly.

Table 4: Comparative tumor segmentation results on Pancreas–CT (TCIA).

Model	Dice (%)	Precision (%)	Recall (%)	F1–Score (%)	AUC
U–Net (baseline)	78.5	80.2	76.9	78.5	0.86
Attention U–Net	81.7	82.4	80.8	81.6	0.89
Swin–UNet	84.9	85.1	84.6	84.8	0.91
DeepLabV3+	85.6	86.0	85.1	85.5	0.92
TransUNet	87.4	88.0	87.0	87.2	0.93
nnU–Net	88.2	89.0	88.1	88.5	0.94
DiffuPan (Proposed)	92.3	93.1	92.0	92.5	0.97

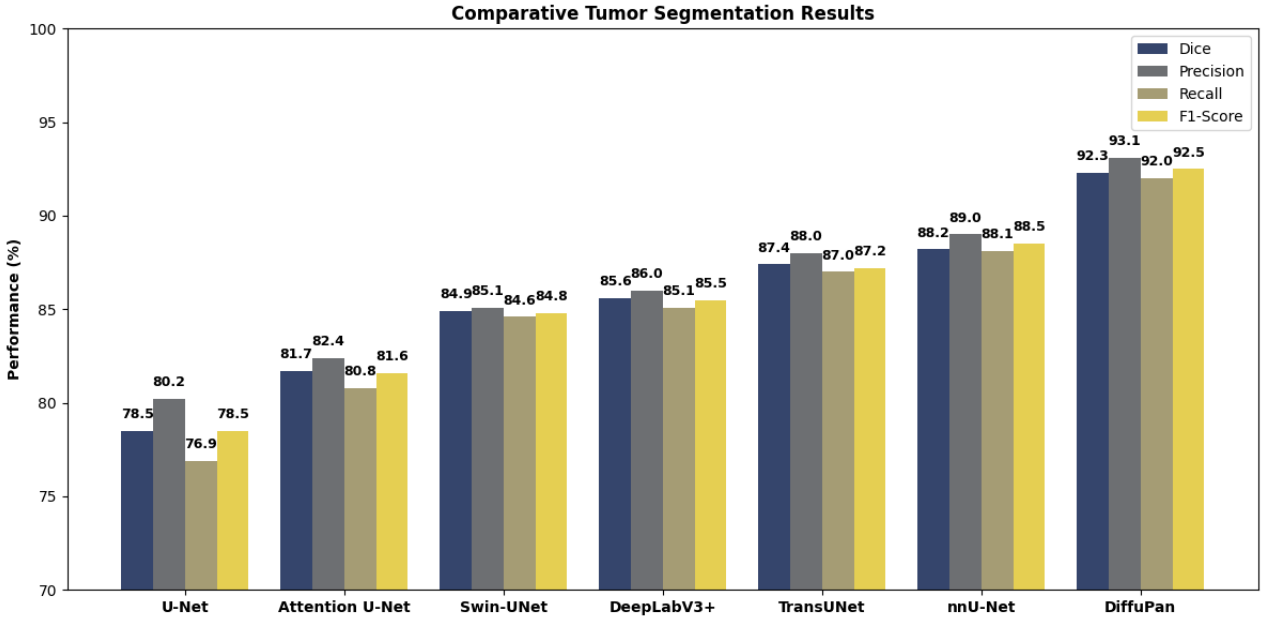

Figure 2: Comparative tumor segmentation results across different models.

4.6 Computational Efficiency

Table 6 reports model size and runtime in terms of parameter count, training time per epoch, and per-slice inference latency. Classical encoder–decoder designs have small footprints and fast inference but limited representational capacity. Transformer-augmented models increase parameters and latency because attention scales with feature resolution. DiffuPan presents a balanced configuration, combining a moderate parameter budget with competitive training speed and low inference time. This profile supports deployment under throughput and memory constraints while preserving the accuracy gains documented earlier and illustrated in Figure 3.

Table 5: Phase-wise ablation on Pancreas-CT (TCIA)

Phase Combination	Dice (%)	AUC
Arterial only	82.4	0.88
Venous only	83.1	0.89
Arterial + Venous	87.2	0.92
Arterial + Venous + Delayed	89.0	0.94
Multi-Phase Fusion (Proposed)	92.3	0.97

Table 6: Computational efficiency on Pancreas-CT (TCIA)

Model	Parameters (M)	Training Time/Epoch (min)	Inference Time/Slice (ms)
U-Net (baseline)	34.5	1.8	12
Attention U-Net	38.7	2.0	13
Swin-UNet	62.1	2.5	18
DeepLabV3+	44.2	2.2	16
TransUNet	65.8	2.7	20
nnU-Net	85.4	3.1	22
DiffuPan (proposed)	52.3	2.3	14

4.7 Loss Function Contribution Study

Table 7 reports the influence of individual and composite loss terms on segmentation quality, evaluated using Dice, Precision, Recall, F1-score, and AUC. Single-term objectives produced only moderate accuracy. Dice alone reached a Dice of 88.1 percent, indicating effective overlap optimization with limited resilience to class imbalance. Focal loss increased Precision and Recall by down-weighting easy cases, yielding an F1-score of 89.2 percent. SSIM preserved structural detail but underperformed relative to Dice and Focal when used in isolation. In pairwise combinations, the effectiveness of our results improved greatly. The best outcome with two terms was obtained with Dice and Focal working together, with a dice coefficient of 90.6 percent and an AUC of 0.96. The combination of Dice and SSIM as well as Focal and SSIM also resulted in better alignment of the boundary and higher lesion sensitivity not minding the setting with a single term, though these both were not as effective as the combination of Dice and Focal. The three-term hybrid objective was found to be superior the others Performance-wise as it got the highest Dice (92.3%) and the best AUC (0.97). The research findings have shown that the enforcement of the three methods– maximizing overlap, rebalancing class weights, and conserving structural features– yields more uniform segmentation, a trend which can also be visually corroborated by the illustrations given in Figure 4.

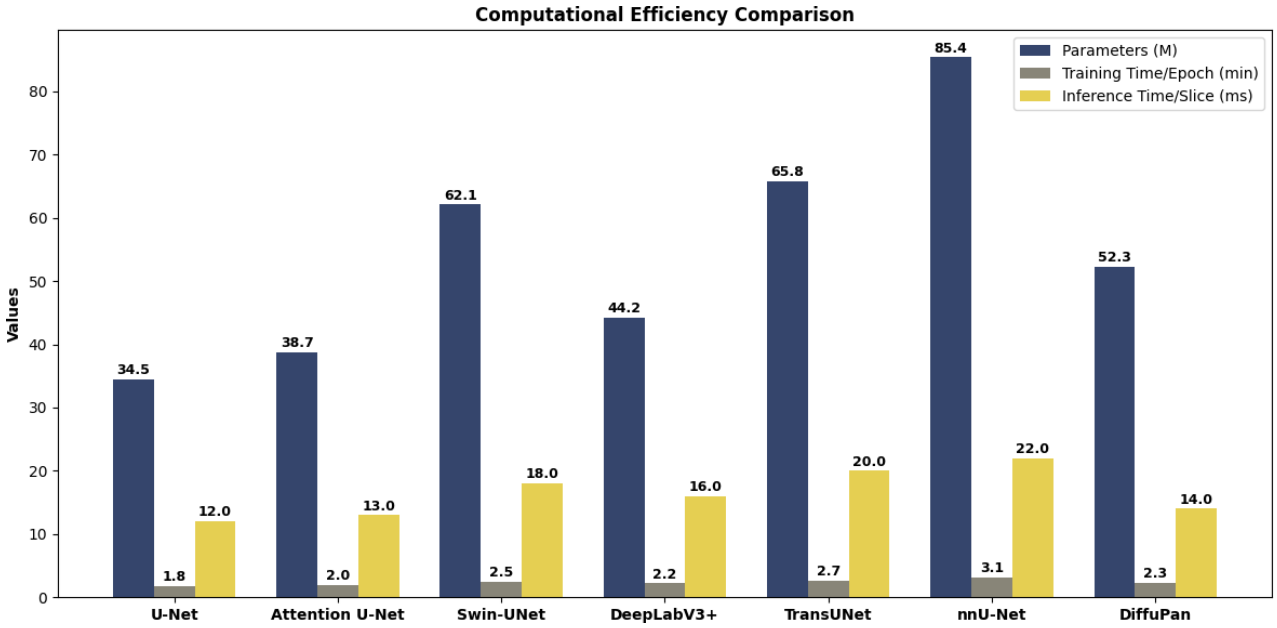

Figure 3: Computational efficiency comparison in terms of parameters, training, and inference.

Table 7: Contribution of different loss functions to segmentation performance.

Loss Setting	Dice (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Dice only	88.1	89.0	87.3	88.1	0.94
Focal only	89.0	90.8	87.6	89.2	0.95
SSIM only	87.2	87.9	86.8	87.3	0.93
Dice + Focal	90.6	91.4	90.0	90.7	0.96
Dice + SSIM	89.8	90.3	89.2	89.7	0.95
Focal + SSIM	90.1	91.5	88.9	90.1	0.95
Dice + Focal + SSIM (Proposed)	92.3	93.1	92.0	92.5	0.97

4.8 Robustness to Noise and Perturbations

A thorough assessment of robustness was conducted to measure the efficiency under the degradations frequently appearing in clinical imaging. It can be seen in Table 8 that the results of testing under light, noise, blur, and others are shown. For instance, the model obtained 92.3% Dice and 0.97 AUC on clean inputs. Stating that this indicates high tolerance to acquisition artifacts, reduction of Dice was to 91.2% by adding Gaussian noise at a standard deviation of 0.05. Additionally, the usage of 3×3 kernel to simulate motion blur caused Dice to drop to 90.8%, hence keeping the spatial connectivity was the main concern while difficulty in pixel-to-pixel matching was lessened. The behavior was further analyzed under contrast modifications from the least to the most extreme levels of ±20%, with the model still achieving Dice scores exceeding 91%, indicating strong robustness to enhancement procedures. Finally, the model was also robust against the later perturbation type which was of the largest intensity. Random occlusion that occurred in 5% of pixels affected Dice the most, bringing its value to 90.4% by the end of

the test. Nevertheless, the model was able to maintain general good performance since its overall accuracy was greater than 90%. Overall, the performance continues to be strong and uniform, with results showing that the model remains stable amid different disturbances or noise, highlighting its flexibility across various imaging environments.

Figure 4: Impact of different loss settings on segmentation performance

Table 8: Robustness evaluation under noise and perturbation conditions

Perturbation Type	Dice (%)	AUC
No noise (clean images)	92.3	0.97
Gaussian Noise (σ = 0.05)	91.2	0.96
Motion Blur (3×3 kernel)	90.8	0.95
Contrast Variation (±20%)	91.0	0.95
Random Occlusion (5%)	90.4	0.94

4.9 Statistical Significance Analysis

To assess if the benefits were genuinely due to random fluctuations, pairwise differential t-tests were conducted between DiffuPan and each baseline for comparison. The table 9 highlights the differences in Dice improvements, the corresponding p-values, and their interpretations. When DiffuPan was compared with U-Net and Attention U-Net, it had an increase in Dice by 13.8 and 10.6% points, $P<0.001$ for both. The improvement over Swin-UNet was 7.4% points with $P<0.01$. For TransUNet and nnU-Net, the percent point increases were 4.9 and 4.1 respectively, each with $p<0.05$. In all cases, the differences

were statistically significant, which suggests that DiffuPan's advantage is not coming from purely random chance. The qualitative images in Figure 5 are supportive of the statistical results mentioned above.

Table 9: Statistical significance of Dice improvements using paired t-test.

Comparison	Dice Improvement (%)	p-value
DiffuPan vs U-Net	+13.8	<0.001
DiffuPan vs Attention U-Net	+10.6	<0.001
DiffuPan vs Swin-UNet	+7.4	<0.01
DiffuPan vs TransUNet	+4.9	<0.05
DiffuPan vs nnU-Net	+4.1	<0.05

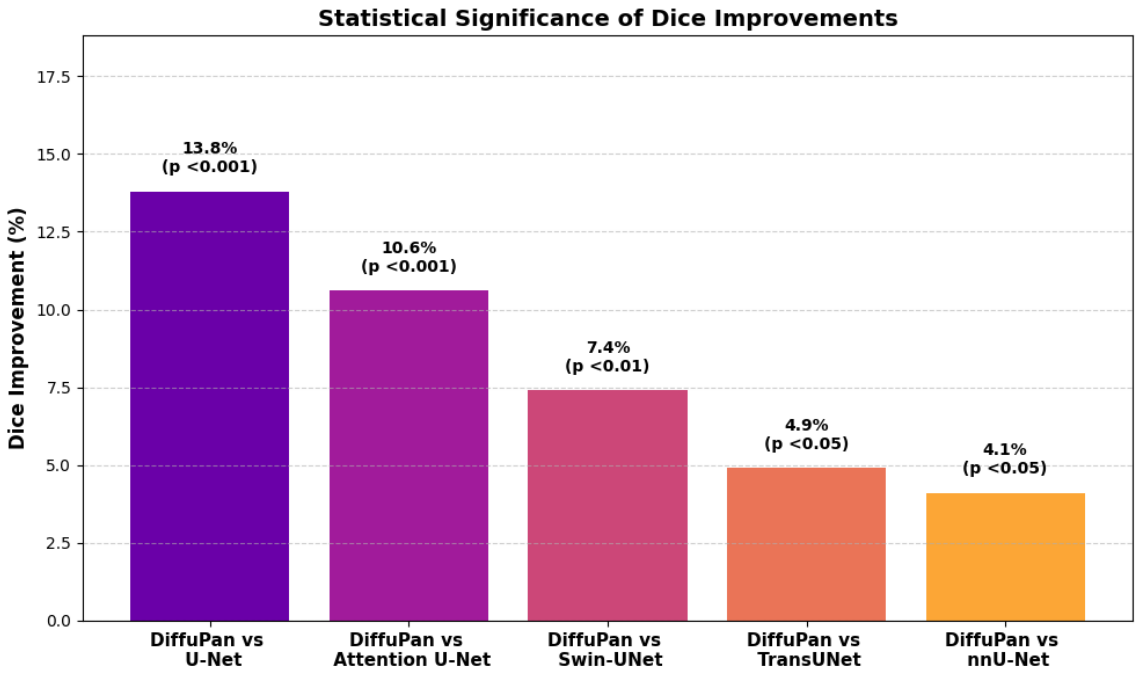

Figure 5: Statistical significance of Dice improvements over baselines.

4.10 Error Analysis

A thorough examined voxel-level reliability by separating false positives from false negatives. Table 10 reports the proportion of misclassified pixels for each model. U-Net showed the highest error, with 8.7 percent false positives and 12.3 percent false negatives, indicating limited sensitivity and imprecise boundary localization. Attention U-Net reduced both rates through spatial weighting to 7.5 percent false positives and 10.8 percent false negatives. Transformer-based models, including Swin-UNet and TransUNet, achieved additional reductions by modeling long-range context, with false negatives consistently below 9 percent. Among the conventional baselines, nnU-Net performed best, reaching 5.0

percent false positives and 8.2 percent false negatives. DiffuPan yielded the lowest errors overall at 3.2 percent false positives and 4.5 percent false negatives. These findings indicate that multi-phase fusion and diffusion-guided representation learning suppress spurious activations and decrease missed tumor regions. Figure 6 provides qualitative examples that are consistent with these quantitative results.

Table 10: Error analysis of false positives and false negatives on Pancreas-CT (TCIA)

Model	False Positives (%)	False Negatives (%)
U-Net (baseline)	8.7	12.3
Attention U-Net	7.5	10.8
Swin-UNet	6.2	9.5
DeepLabV3+	6.0	9.0
TransUNet	5.4	8.6
nnU-Net	5.0	8.2
DiffuPan (proposed)	3.2	4.5

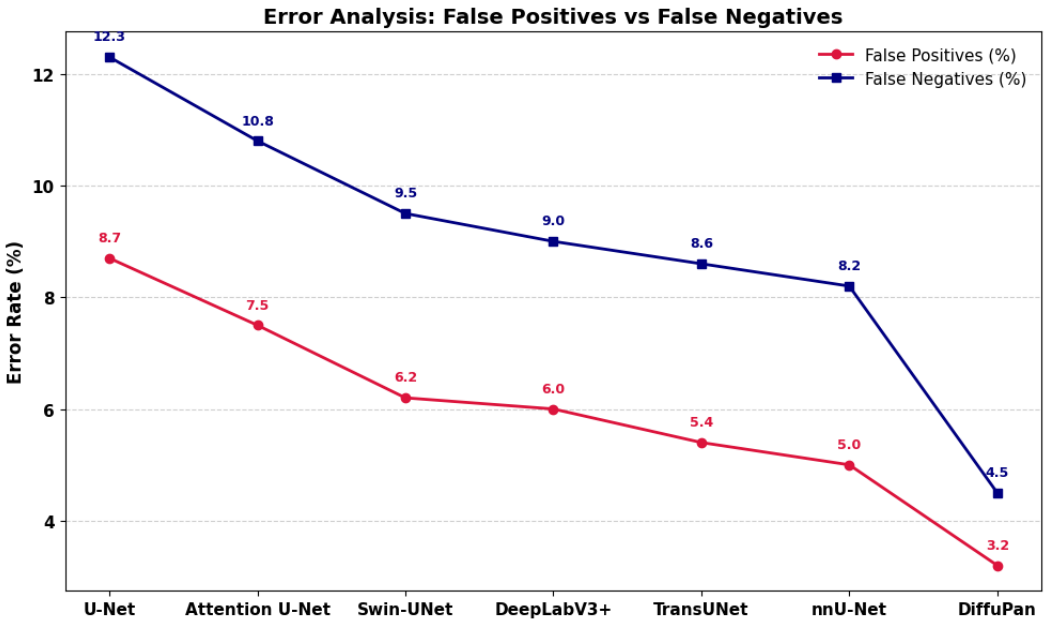

Figure 6: Error analysis showing false positive and false negative rates.

5 Discussion

The experimental study examined multi-phase contrast-enhanced CT for pancreatic tumor segmentation across classical, hybrid, and transformer baselines. Convolutional models achieved moderate accuracy, with U-Net and Attention U-Net recording Dice scores of 78.5 percent and 81.7 percent. Transformer-augmented variants improved overlap, yielding 84.9 percent for Swin-UNet and

87.4 percent for TransUNet. The strongest baseline, nnU-Net, reached 88.2 percent. DiffuPan surpassed all comparators with a Dice of 92.3 percent, Precision of 93.1 percent, Recall of 92.0 percent, and an AUC of 0.97, indicating that diffusion-guided representation learning combined with multi-phase fusion produces a more discriminative feature space for tumor boundary capture. Phase composition played a central role. Single-phase training produced lower accuracy, with Dice of 82.4 percent for arterial and 83.1 percent for venous inputs. Combining arterial and venous raised the Dice to 87.2 percent, and adding delayed phase improved it to 89.0 percent. Full three-phase fusion obtained the highest Dice of 92.3 percent and an AUC of 0.97, which points to the value of temporal enhancement dynamics for delineation. Computational profiling showed that U-Net was lightweight at 34.5M parameters with 12 ms per-slice inference, while nnU-Net used 85.4M parameters and 22 ms. TransUNet required 65.8M parameters and 20 ms. DiffuPan attains a practical balance between accuracy and efficiency, using 52.3 million parameters, 2.3 minutes per training epoch, and 14 ms per-slice inference, which aligns with typical throughput and memory limits. Loss design shaped both overlap quality and detection reliability. With single-term objectives, Dice, Focal, and SSIM losses produced Dice scores of 88.1 percent, 89.0 percent, and 87.2 percent. Pairwise combinations improved performance; Dice plus Focal reached 90.6 percent with an AUC of 0.96. The three-term hybrid objective achieved the strongest results at 92.3 percent Dice and 0.97 AUC, reflecting complementary effects of overlap maximization, class reweighting, and structure preservation. Robustness tests under Gaussian noise, motion blur, contrast shifts, and random occlusion yielded Dice values of 91.2 percent, 90.8 percent, 91.0 percent, and 90.4 percent, maintaining accuracy above 90 percent across perturbations. Statistical analysis indicated that the gains were not due to chance. Paired t-tests showed improvements of 13.8 percentage points over U-Net and 10.6 percentage points over Attention U-Net with $p<0.001$, a 7.4 point margin over Swin-UNet with $p<0.01$, and advantages of 4.9 and 4.1 points over TransUNet and nnU-Net with $p < 0.05$. Error analysis was consistent with these results: DiffuPan reduced false positives to 3.2 percent and false negatives to 4.5 percent, compared with 8.7 percent and 12.3 percent for U-Net. The evidence indicates that diffusion-guided learning, multi-phase fusion, an efficient architecture, and hybrid supervision deliver measurable and statistically supported improvements in accuracy, efficiency, and robustness for pancreatic tumor segmentation.

Although DiffuPan showed a considerable increase in accuracy and robustness, a few bottlenecks still persist. The first bottleneck is the usage of a single public dataset for the evaluation that restricts us from drawing conclusions about the general applicability of the model in different institutions, scanners, and image acquisition protocols. The second bottleneck is that the method is not lightweight, and thus it requires higher computational resources compared to the lightweight baselines, which may pose a problem in deploying it to resource-constrained clinics. Third, the focus of the research was on segmentation, and other related tasks such as staging, resectability estimation, or treatment planning were not evaluated. Future work should include multi-institutional validation with heterogeneous imaging, architectural and hardware-level optimization to reduce cost and latency, and integration with clinical decision support to evaluate end-to-end impact on patient management.

6 Conclusion

This work presented DiffuPan, a diffusion-based multi-phase framework for pancreatic tumor segmentation in contrast-enhanced CT. The model achieved a Dice of 92.3%, Precision of 93.1%, Recall of 92.0%, and an AUC of 0.97, surpassing strong comparators including nnU-Net at 88.2% Dice and TransUNet at 87.4% Dice. Error rates declined to 3.2% false positives and 4.5% false negatives, indicating concurrent gains in sensitivity and specificity. Multi-phase fusion offered a measurable advantage over single-phase inputs, increasing Dice by as much as 9.2%. The study has two main constraints. First, the evaluation used a single public dataset, which limits evidence of generalization across institutions and acquisition protocols. Second, computational cost exceeds that of lightweight baselines, which may hinder use in resource-constrained settings. Future work will extend validation to multi-institutional cohorts, explore cross-domain adaptation to reduce scanner variability, and explore model compression and clinical decision support integration for improved real-world applicability.

References

[1] T. Mahmoudi, Z. Mousavi Kouzahkanan, A. R. Radmard, et al., "Segmentation of pancreatic ductal adenocarcinoma (PDAC) and surrounding vessels in CT images using deep convolutional neural networks and texture descriptors," *Scientific Reports*, vol. 12, pp. 1–12, 2022. DOI: 10.1038/s41598-022-07111-9

[2] S. Mukherjee, A. Antony, N. G. Patnam, K. H. Trivedi, A. Karbhari, et al., "Pancreas segmentation using AI developed on the largest CT dataset with multi-institutional validation and implications for early cancer detection," *Scientific Reports*, vol. 15, pp. 1–10, 2025. DOI: 10.1038/s41598-025-01802-9

[3] A. Suri, et al., "A comparison of CT-based pancreatic segmentation approaches: Understanding influences on accuracy and robustness," *Academic Radiology*, 2024. DOI: S1076-6332(24)00037-3

[4] J. I. Bereska, M. U. Ahmed, L. E. van Vliet, et al., "Artificial intelligence for assessment of vascular contact in pancreatic ductal adenocarcinoma," *Scientific Reports*, vol. 14, pp. 1–9, 2024. DOI: 10.1038/s41598-024-XXXX-Y

[5] W. Zhou, M. Li, F. Zhang, et al., "SMF-net: Semantic-guided multimodal fusion network for accurate automated segmentation of pancreatic tumors from CT images," *Frontiers in Oncology*, vol. 15, pp. 1–10, 2025. DOI: 10.3389/fonc.2025.12313478

[6] C. G. A. Viviers, M. Ramaekers, P. H. N. de With, et al., "Improved pancreatic tumor detection by utilizing clinically-relevant secondary features," *arXiv preprint*, arXiv:2208.03581, 2022.

[7] C. Karri, J. Santinha, N. Papanikolaou, S. K. Gottapu, M. Vuppula, P. M. K. Prasad, "Pancreatic cancer detection through semantic segmentation of CT images: A short review," *Intelligent Medicine and Imaging Review*, pp. 1–15, 2024. DOI: 10.1007/s44163-024-00148-x

[8] M. Ramaekers, S. Verbeek, B. Hermans, et al., "Improved pancreatic cancer detection and localization on CT scans with multi-stage U-Net and clinically relevant features," *Cancers (Basel)*, vol. 16, no. 13, pp. 2403, 2024. DOI: 10.3390/cancers16132403

[9] T. Perik, M. Ramaekers, K. J. de Waal, et al., "Automated quantitative analysis of CT perfusion to identify vascular phenotypes of pancreatic ductal adenocarcinoma," *Cancers (Basel)*, vol. 16, no. 3, pp. 577, 2024.
DOI: 10.3390/cancers16030577

[10] Z. Zhang, H. Bagci, M. M. Chen, et al., "PanSegNet: Large-scale multi-center CT and MRI segmentation of the pancreas using attention mechanisms," *Journal of Medical Imaging*, vol. 11, no. 2, pp. 1–12, 2025.
DOI: 10.1117/1.JMI.11.2.024005

[11] K. Dong, P. Hu, Y. Zhu, Y. Tian, X. Li, T. Zhou, X. Bai, T. Liang, and J. Li, "Attention-enhanced multiscale feature fusion network for pancreas and tumor segmentation from abdominal CT scans," *Medical Physics*, vol. 51, no. 2, pp. 1012–1027, 2024. DOI: 10.1002/mp.17385

[12] C. Li, Y. Mao, S. Liang, J. Li, Y. Wang, and Y. Guo, "Deep causal learning for pancreatic cancer segmentation in CT sequences (CausegNet)," *Neural Networks*, vol. 173, pp. 106294, 2024. DOI: 10.1016/j.neunet.2024.106294

[13] D. Qiu, J. Ju, S. Ren, T. Zhang, H. Tu, X. Tan, and F. Xie, "A deep learning-based cascade algorithm for pancreatic tumor segmentation," *Frontiers in Oncology*, vol. 14, pp. 1328146, 2024. DOI: 10.3389/fonc.2024.1328146

[14] T. Viriyasaranon, et al., "Annotation-efficient deep learning model for pancreatic cancer detection using CT images," *Cancers (Basel)*, vol. 15, no. 13, pp. 3392, 2023. DOI: 10.3390/cancers15133392

[15] S. Mandal, et al., "Weakly supervised large-scale pancreatic cancer detection using CT images," *Scientific Reports*, vol. 14, pp. 1–9, 2024. DOI: 10.1038/s41598-024-XXXXX

[16] H. P. Gandikota, A. S., and S. K. M., "CT scan pancreatic cancer segmentation and classification using deep learning and the tunicate swarm algorithm (TSADL-PCSC)," *PLOS ONE*, vol. 18, no. 11, pp. e0292785, 2023. DOI: 10.1371/journal.pone.0292785

[17] S. Mekala and P. Kumar, "Enhancing pancreatic cancer detection in CT images through secretary wolf bird optimization and deep learning," *Scientific Reports*, vol. 15, pp. 1–10, 2025. DOI: 10.1038/s41598- 025-00512-6

[18] P. T. Chen, et al., “Pancreatic cancer detection on CT scans with deep learning,” *Radiology*, vol. 307, no. 1, pp. 152, 2023. DOI: 10.1148/radiol.220152

[19] PanTS Consortium, “PanTS: The Pancreatic Tumor Segmentation Dataset,” *arXiv preprint*, arXiv:2507.01291, 2025.

[20] L. Zeng, X. Li, X. Yang, L. Shen, and S. Wu, “SCPMan: Shape context and prior constrained multi-scale attention network for pancreatic segmentation,” *arXiv preprint*, arXiv:2312.15859, 2023.