SCIENTIFIC Publishing Limited

Trends in Immunotherapy

https://ojs.ukscip.com/index.php/ti

Article

Enhancing CAR-T Cell Tumor Targeting via Advanced Computational Perception Networks for Improved Recognition in Heterogeneous Tumors

Selvaganapathi Sennan [1*], S Sridevi [2], A N Ramya Shree [3], Kunchala Suresh Babu [4], Ugranada Channabasava [5], Immanuvel Arokia James K [6]

[1] Hexaware Technologies, Secaucus, NJ 07094-3675, USA

[2] Department of IoT, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh 522302, India

[3] Department of CSEn(AI&ML), Ramaiah Institute of Technology, Bengaluru, Karnataka 560054, India

[4] PSCMR College of Engineering and Technology, Vijayawada, Andhra Pradesh 520001, India

[5] Department of Artificial intelligence and Data science, Global Academy of Technology, Bangalore, Karnataka 560098, India

[6] Department of Artificial Intelligence and Data Science, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, Tamil Nadu 600062, India

[*] Correspondence: ganapathi.selva6@gmail.com

Received: 31 May 2025; **Revised:** 16 June 2025; **Accepted:** 09 July 2025; **Published:** 10 September 2025

Abstract: In cancer treatments, the efficacy of Chimeric Antigen Receptor T (CAR-T) cell therapy is affected in heterogeneous tumors due to ambiguous tumor boundaries, morphological variability, and similarity between tumor and non-tumor tissues in medical imaging. Accurate tumor localization and classification are crucial for optimizing CAR-T targeting and therapeutic success. Traditional segmentation networks struggle with intensity similarity, shape variability, and contextual complexity in heterogeneous tumors. Further, robust classification of tumor regions using limited medical data remains a key challenge. We propose a dual-component Computational Perception Architecture composed of a novel segmentation-classification framework. The segmentation backbone is a U-Net enhanced with a Visual Perception Module (VPM) for ROI-level feature refinement. Multi-Head Self-Dilated Attention (MHSDA) in the encoder to capture multi-scale dependencies. ResNet50 with Dense Attention Modules in skip connections for improved feature continuity. Group Receptive Large Kernel (GRLK) Blocks for diverse receptive field decoding. The classification network utilizes edge-perception, morphological, and positional images, and segmentation maps. Deep ensemble learning for decision robustness and transfer learning to boost generalization on breast cancer labeled datasets. The proposed method is tested on the publicly available PBC and CAR-T datasets from Kaggle. The research model achieved a Dice Score of 0.901, an IoU of 0.856, a Precision of 0.882, a Classification Accuracy of 93.7%, and an F1-Score of 0.915. These outcomes show the superior capacity for precision tumor detection and classification, thus offering a potent computational aid in enhancing the targeting precision of CAR-T therapies.

Keywords: CAR-T Therapy; Tumor Segmentation; Computational Perception; Deep Learning; Attention Networks

1. Introduction

The success of Chimeric Antigen Receptor T-cell (CAR-T) therapy in hematological malignancies has led to a growing interest in adapting this approach for solid and heterogeneous tumors [1–3]. However, the highly variable nature of these tumors poses a significant challenge to achieving precise and consistent therapeutic outcomes. Tumor heterogeneity manifests in diverse cellular morphologies, spatial distributions, and intensity patterns, which complicate the segmentation and classification processes in computational tumor analysis pipelines. Advancements in deep learning and computer vision have eased automated tumor detection, segmentation, and classification in medical imaging. However, traditional methods often struggle with subtle boundary transitions, variability in tumor location, and weak inter-class contrast issues that are distinct in complex tumor environments. These limitations restrict the potential for accurate image-based tumor recognition, which impacts the accuracy of cell therapy planning.

Despite the promise of deep learning in medical imaging, several intrinsic challenges limit its effectiveness when dealing with heterogeneous tumor data:

- Tumor tissues often exhibit grayscale intensities that are indistinguishable from adjacent normal tissues, leading to boundary ambiguity [4].
- Irregular tumor shapes, textures, and multi-focal lesions complicate region-of-interest (ROI) extraction and degrade segmentation accuracy [5].
- Tumors appear in varying anatomical locations and sizes, necessitating robust multi-scale feature learning [6].
- High-quality labeled datasets, particularly those specific to CAR-T therapy contexts, are scarce and labor-intensive to produce [7].

These challenges require a model that perceives tumor-specific features at both global and local scales, maintains spatial resolution, and adapts across different tumor presentations.

To improve CAR-T therapeutic outcomes, a computational model must be able to accurately segment and classify heterogeneous tumor regions from complex medical images (e.g., pbc-dataset and CAR-T image datasets). The problem thus lies in designing a deep learning framework that captures detailed tumor features under variability in structure, location, and intensity, while ensuring robustness, interpretability, and high performance on limited data [8–14].

The research aims to develop a dual-network architecture comprising a segmentation and classification branch tailored to CAR-T-related image data. It combines perceptual modules that simulate human-like focus on fine-grained features for enhanced tumor localization. It uses attention mechanisms and deep ensemble strategies to extract discriminative features across multi-scale contexts.

These contributions address the core barriers to accurate, automated tumor analysis in CAR-T applications, which offer a scalable framework for clinical diagnosis. This work proposes a Computational Perception Network that augments tumor recognition for CAR-T cell therapy via:

1. A Segmentation developed using U-Net with Multi-Head Self-Dilated Attention (MHSDA) and a Visual Perception Module (VPM) for adaptive ROI enhancement.
2. A ResNet50-based Encoder combined with Dense Attention Skip Connections to preserve deep spatial information and promote efficient feature fusion.
3. A Classification Network utilizing edge-perception, morphological analysis, and segmentation-driven features to enable high-precision tumor classification.
4. The use of lightweight transfer learning and ensemble classifiers for improved generalization on limited datasets.

2. Related Works

Recent advances in medical image analysis have used DL and ML to tackle a broad spectrum of diagnostic and prognostic tasks across various medical domains. These approaches have demonstrated strong potential, even in the face of challenges such as data heterogeneity, low sample sizes, and varying image acquisition protocols.

To address the image heterogeneity in fluorescence microscopy, Loadnnidis et al. [15] introduced a harmonization preprocessing protocol alongside feature-based and transfer learning models, achieving classification ac-

curacies of up to 0.957. Similarly, Lee et al. [16] dealt with heterogeneous panoramic radiographic systems to classify Stafne's bone cavity (SBC), achieving 99.25% accuracy, demonstrating the robustness of deep learning (DL) techniques across varied imaging systems. In the same vein, Liang et al. [17] employed variants of 3D U-Net to segment metastases from non-standardized MRI sequences, achieving high sensitivity and clinically relevant accuracy. Meanwhile, He et al. [18] introduced Starfysh, a generative model-based toolbox that characterizes tissue-specific cell states from histology images without requiring single-cell references. When applied to diverse breast cancer subtypes, Starfysh revealed spatial hubs and metabolic reprogramming associated with aggressive cancer phenotypes, underscoring the strength of integrative spatial analysis.

The ability of DL to interpret histology images for molecular insights is showcased in Qu et al. [19], where whole-slide images (WSIs) from the Genomic Data Commons were used to predict mutations, with AUCs ranging from 0.65 to 0.85. To improve explainability, a self-attention mechanism was incorporated to visualize important regions. Along similar lines, Abhishek et al. [20] constructed a heterogeneous peripheral blood smear dataset for automated classification of leukemia, highlighting DL's adaptability across binary and multi-class classification tasks. For improving volumetric imaging, Geng et al. [21] utilized 3D light sheet microscopy combined with deep learning to quantify visceral adipose tissue (VAT) structures. Their pipeline identified significant morphological differences in Crown-like structures (CLSs) between lean and obese tissues, establishing potential histological biomarkers for adipose pathogenesis.

To further enhance predictive modeling, Yang et al. [22] proposed a multimodal approach integrating WSIs and clinical data, achieving an AUC of 0.76. In a broader pathological context, Jiao et al. [23] applied a CNN to segment nine tissue types in colon cancer WSIs. The quantified tumor microenvironment (TME) descriptors were correlated with clinical outcomes, identifying stromal content as a significant independent prognostic marker. With the combination of visual recognition and diagnostic performance, Kotei and Thirunavukarasu [24] introduced MD-VACNet, a lightweight, self-attentive network optimized through generative synthesis for edge devices. It effectively identified tuberculosis from chest X-rays and breast cancer from ultrasound images.

A similar direction is explored by Zhao et al. [25], where a novel machine learning-based classification system was developed to define NET-based clusters in gastric cancer, revealing heterogeneity in clinical and molecular features that influence therapeutic responses. To find frequency domain solutions, Liu et al. [26] proposed DLfd, a model using 3D discrete cosine transform and U-Net to solve inverse identification problems with high robustness and low error, even under noise and incomplete measurements. Alongside, Guo et al. [27] developed graph attention networks (GATs) to score bystander killing for antibody-drug conjugates (ADCs), enabling the design of potent payloads like Ed9 with superior anti-tumor efficacy.

To understand tumor evolution and immunotherapy response, Wang et al. [28] presented a computational framework for analyzing extrachromosomal DNA (ecDNA) across over 13,000 cancer patients. Their findings linked ecDNA amplification with microsatellite instability and immunotherapy outcomes, solidifying ecDNA's role as a biomarker. Finally, Gowthamy and Ramesh [29] fused features from pre-trained models with Extreme Learning for lung cancer classification. A Mutation Boosted Dwarf Mongoose Optimization Algorithm (MB-DMOA) was employed to avoid local minima and ensure efficient convergence.

The significant contributions of the proposed work are presented in the following key advancements over existing literature:

1. **Computational Perception Architecture**: Unlike prior models that focus on isolated segmentation or classification, our approach integrates both within a dual-network framework that jointly optimizes region-of-interest (ROI) segmentation and perception-guided classification using multimodal tumor features.
2. **Visual Perception Module (VPM)**: Our VPM explicitly mimics human visual attention mechanisms by refining ROI-level features using both spatial and channel-wise attention. This is particularly novel for tumor segmentation tasks with weak boundaries and variable morphology—areas where standard U-Net and 3D U-Net models underperform.
3. **Multi-Head Self-Dilated Attention (MHSDA)**: This module enables context aggregation at multiple scales, addressing spatial and structural heterogeneity in tumors. Unlike single-scale attention in most existing models, MHSDA uses dilated convolutions within each attention head, significantly enriching global representation.
4. **GRLK + PSDC Decoder**: The decoder innovatively integrates group receptive large kernel (GRLK) blocks and

perceptually separable dense convolutions (PSDC), enabling diverse receptive field modeling while maintaining computational efficiency—an approach not seen in prior works.

5. **Dense Attention in Skip Connections**: We introduce Dense Attention Modules (DAMs) to filter irrelevant features during skip connections, a known limitation in standard encoder–decoder architectures.
6. **Ensemble Classification with Morphological and Positional Features**: Unlike typical CNN classifiers, our classification network utilizes morphological, edge, and positional inputs derived from segmentation, which are fused using ensemble transfer learning strategies. This provides enhanced generalization, particularly when working with limited annotated datasets.

3. Proposed Method

The proposed architecture consists of two distinct networks: a segmentation and classification network, both guided by principles of computational perception (**Figure 1**). The segmentation is carried out using a U-Net structure, known for its encoder–decoder symmetry and efficacy in medical image tasks. To overcome the challenge of low contrast and variable tumor presentation, we introduce a Visual Perception Module (VPM), which emulates human-like attention mechanisms to show regions of interest. This study hypothesizes that an advanced computational perception framework integrating fine-grained tumor segmentation and ensemble-based classification can significantly enhance the precision of tumor region delineation and subtype identification in heterogeneous tumors, thereby enabling more effective and targeted CAR-T cell therapy. Specifically, by improving tumor boundary clarity and phenotypic categorization from imaging data, the proposed model facilitates the design and deployment of CAR-T cells with higher spatial targeting accuracy, minimizing off-tumor cytotoxicity and enhancing therapeutic efficacy in complex tumor microenvironments.

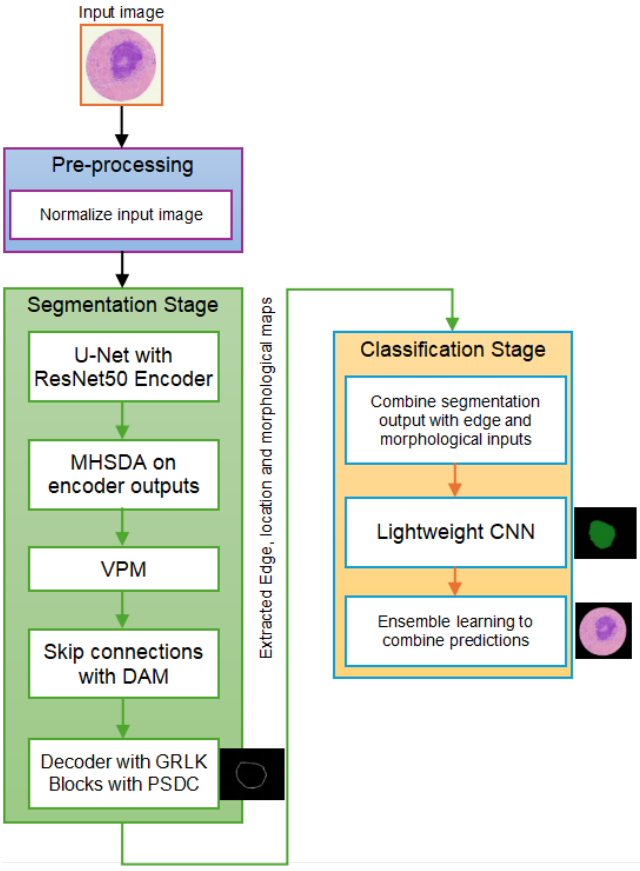

Figure 1. Proposed two-stage network architecture for heterogeneous tumor classification.

The encoder incorporates a Multi-Head Self-Dilated Attention (MHSDA) layer that enables the network to

gather contextual information at various scales without losing resolution. The encoder's output is passed to ResNet50, which extracts the semantic features. To mitigate the common issue of information bottleneck in skip connections, we introduce Dense Attention Modules that selectively emphasize informative spatial features while suppressing noise. The decoder side employs Group Receptive Large Kernel (GRLK) blocks, which broaden the receptive fields to ensure diverse and discriminative decoding of tumor boundaries.

In the second stage, the classification network takes the output of the segmentation module alongside additional perception-based inputs, such as edge maps, locational information, and morphological cues. These multi-view representations are fed into a collection of lightweight CNNs pre-trained on medical datasets and fine-tuned using transfer learning. Each model in the ensemble contributes to a voting mechanism that enhances classification robustness and mitigates overfitting, under conditions of limited annotated data.

3.1. Segmentation Network: U-Net with ResNet50-based Encoder

The segmentation network in the proposed architecture (**Figure 2**) is built upon a U-Net backbone that has a skip connection to bridge the symmetric encoder–decoder.

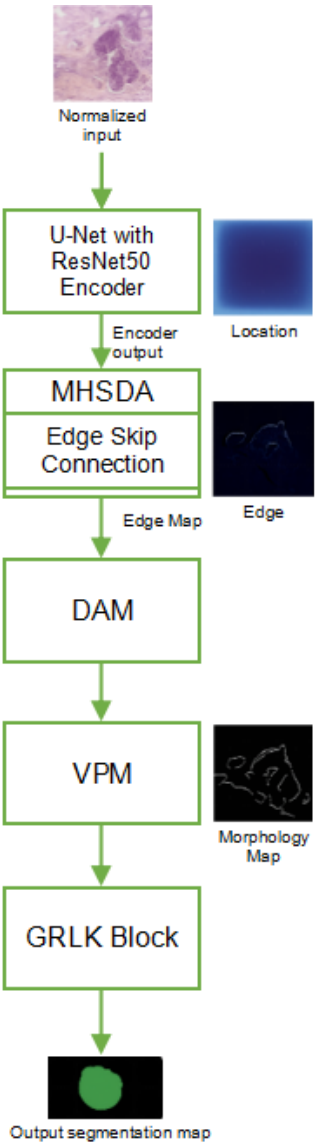

Figure 2. Segmentation network: U-Net with ResNet50-based encoder.

3.1.1. U-Net Architecture

The U-Net architecture consists of two major components: an encoder (contracting path) and a decoder (expanding path), connected through skip connections. The encoder captures context and abstract features by progressively reducing the spatial dimensions while increasing the number of feature channels. Conversely, the decoder reconstructs the segmentation map by gradually upsampling and concatenating the feature maps from the encoder to restore spatial resolution. To enhance the representational capacity of the encoder, we substitute the vanilla convolutional layers of the original U-Net with a ResNet50 model. This model, a residual learning framework, can extract deep semantic features without suffering from vanishing gradients or degradation problems in deep networks.

3.1.2. Encoder with ResNet50

The ResNet50-based encoder improves upon the standard U-Net encoder by utilizing residual connections that allow the network to learn identity mappings and maintain stable gradients during backpropagation. Each residual block in ResNet50 consists of three convolutional layers and a shortcut connection that bypasses the intermediate transformation, enabling efficient feature reuse and robust spatial encoding. The encoder's modified structure using ResNet50 is shown in **Table 1**.

Table 1. Modified U-Net encoder with ResNet50 backbone.

Layer Block	Output Size	Description
Input	256 × 256 × 3	RGB input image
Conv1	128 × 128 × 64	7 × 7 conv, stride 2, followed by MaxPool
ResBlock1	64 × 64 × 256	3 residual blocks (3 × 3 convs)
ResBlock2	32 × 32 × 512	4 residual blocks
ResBlock3	16 × 16 × 1024	6 residual blocks
ResBlock4	8 × 8 × 2048	3 residual blocks

This deep encoding structure enables the extraction of high-level spatial and semantic information from complex tumor regions. Moreover, the encoder is further enhanced with an MHSDA module (discussed in later sections) to enrich global contextual awareness.

3.1.3. Feature Map Propagation and Skip Connections

To maintain spatial precision during upsampling in the decoder, feature maps from earlier layers of the encoder are propagated via skip connections and concatenated with decoder features. However, direct transfer of these features may include noise or irrelevant information. To address this, we apply Dense Attention Modules within the skip connections, which selectively amplify important spatial regions before fusion. The overall transformation of an input image x through the encoder path can be mathematically expressed as:

$$f_{enc} = \mathbb{E}(x) = R_4(R_3(R_2(R_1(C_1(x)))))$$

Where,
C_1 - initial convolution layer,
R_i - i^{th} ResNet block, and
f_e - encoded feature representation.

By combining the proven spatial encoding power of ResNet50 with the segmentation precision of U-Net, the network effectively addresses the challenges posed by medical imaging of heterogeneous tumors, such as low boundary contrast, irregular shapes, and small ROI sizes. As shown in **Table 2**, the research encoder significantly improves segmentation performance compared to the standard U-Net encoder.

Table 2. Performance comparison: U-Net vs. U-Net + ResNet50 encoder.

Architecture	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
Vanilla U-Net	0.847	0.794	0.828	0.841
U-Net + ResNet50	0.901	0.856	0.882	0.894

The results in **Table 2** demonstrate the superiority of the ResNet50-based encoder, which contributes to enhanced segmentation accuracy, specifically in complex tumor boundary scenarios.

3.2. Visual Perception Module (VPM): Focusing on Fine-Grained Features

In medical image analysis, in the context of heterogeneous tumor segmentation, one of the most critical challenges is the identification of fine-grained spatial features. These include subtle edges, slight texture differences, and small-scale shape variations between tumor tissues and surrounding regions. The VPM is introduced into the segmentation pipeline to address this challenge by mimicking the selective attention mechanisms of the human visual system. Traditional convolutional layers capture local patterns but cannot focus explicitly on small, critical features that may signify early tumor boundaries or subtle morphological differences. Inspired by visual attention in biological perception, the VPM is designed to enhance low-level and mid-level feature maps, enabling the network to prioritize regions of interest (ROIs) with high diagnostic relevance. This is particularly effective in scenarios where tumor tissue exhibits only minor contrast variations against normal tissue. The VPM is positioned between the encoder and decoder within the U-Net structure, acting as a refinement layer for encoded features. It applies a series of attention-guided convolutions, including spatial and channel attention mechanisms, to emphasize detailed textures and weak boundaries.

For the input feature map, consider the encoded feature map be: $F \in \mathbb{R}^{C\times H\times W}$. In the Channel Attention Module (CAM), let global average pooling and max pooling over spatial dimensions be: $F^c_{avg} = \frac{1}{H\cdot W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{c,i,j}$, where $F^c_{\max} = \max_{i,j} F_{c,i,j}$. The channel attention weights are defined as:

$$M_c = \sigma\left(W_2 \cdot \delta\left(W_1 \cdot F_{avg} + W_1 \cdot F_{\max}\right)\right)$$

where,
W_1, W_2 - shared weights of MLP layers,
$\delta(\cdot)$ - ReLU activation,
$\sigma(\cdot)$ - is the sigmoid function,
$M_c \in \mathbb{R}^{C\times 1\times 1}$
The channel attention is then applied as in $F' = M_s \times F$
For the Spatial Attention Module (SAM), compute the average and max pooling along channels:

$$F_s^{avg} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{s,i,j},\ F_s^{\max} = \max_{i,j} F_{s,i,j}$$

Finally, the concatenate and convolve is applied as in following:

$$F_{cat} = [F_s^{avg}; F_s^{\max}] \in \mathbb{R}^{2\times H\times W}$$

$$M_s = \sigma\left(\mathrm{Conv}_{7\times 7}(F_{cat})\right)$$

Then, the spatial attention is applied $F'' = M_s \times F'$ and hence, the final output of VPM is obtained as $F_{VPM} = F''$.

This dual-attention process helps the network focus on both spatial and channel-wise features. The VPM is powerful when combined with deeper encoder features, as it guides the model to reweight subtle textures and boundaries that might otherwise be lost due to downsampling. This results in enhanced recognition of minute tumor regions or irregular growths often missed by standard architectures. To validate its effectiveness, we conducted ablation experiments comparing segmentation results with and without the VPM. The results are presented in **Table 3**.

Table 3. Impact of VPM on segmentation performance.

Configuration	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
U-Net + ResNet50(no VPM)	0.872	0.823	0.858	0.849
U-Net + ResNet50 + VPM	0.901	0.856	0.882	0.894

As shown in **Table 3**, the inclusion of the VPM significantly improves all evaluation metrics, most notably the Dice score and IoU, which are crucial indicators of segmentation quality. The improvement confirms the VPM's ability to preserve and enhance subtle tumor boundaries during feature processing. The VPM is fully compatible with the MHSDA applied earlier in the encoder. While MHSDA captures global contextual information, VPM provides local perceptual enhancement, creating a complementary feature enhancement strategy that balances coarse and fine-grained learning. The VPM functions as a perceptual refinement stage, focusing the network's attention on small yet diagnostically significant regions in tumor segmentation. Its biologically inspired architecture enhances spatial selectivity and channel sensitivity, contributing to superior segmentation outcomes in complex medical imaging tasks. The results validate the importance of this module in achieving high-precision tumor delineation (**Table 3**).

3.3. MHSDA - Global Contextual Understanding Across Multiple Scales

Accurate segmentation of heterogeneous tumors in medical images requires a deep understanding of not only local textures but also global contextual relationships. Tumors often display varying shapes, positions, and internal structures across scales and patients. The proposed MHSDA module is introduced in the encoder path to enhance global perception by enabling the network to dynamically focus on semantically relevant regions across multiple receptive fields and scales.

Unlike standard self-attention, which typically captures global relationships at a single feature resolution, MHSDA incorporates dilated convolutional branches within each attention head to enlarge the effective receptive field. This facilitates better modeling of long-range dependencies in both spatial and semantic dimensions. Each attention head processes the input with a distinct dilation rate, capturing features at various scales, while the multi-head architecture ensures diversity of attention.

MHSDA operates on input feature maps $F \in \mathbb{R}^{C\times H\times W}$, applying multi-scale attention with dilation factors $\{d\}_{id} \in \{1, 2, 4, 8\}$ in parallel branches. For each attention head h, dilated convolutions extract key-value-query projections: $Q_h = f(\cdot; d_h)$, $K_h = f(\cdot; d_h)$, $V_h = f(\cdot; d_h)$. The attention weights are computed as: $A_h = \text{Softmax}\left(\frac{Q_h K_h^T}{\sqrt{d_k}}\right)$ and the attended feature is:$Z_h = A_h \cdot V_h$. The outputs of all heads are concatenated and linearly projected:

$$\mathbf{F}_{\text{mtds}} = f_{\text{proj}}([Z_1 : Z_2 : \cdots : Z_H])$$

where, H: number of attention heads, $f_q^h = f_k^h = f_v^h$: convolution layers with dilation d_h, and $\sqrt{d_k}$: dimension scaling factor for normalization.

3.3.1. Multi-Scale Feature Enrichment

The varying dilation rates allow each head to focus on a different spatial resolution, thereby simulating multi-scale perception akin to human vision scanning across close and distant regions simultaneously. This is particularly effective for tumors with irregular boundaries and heterogeneous textures spanning multiple scales. The MHSDA module is combined within the encoder at high-level feature stages, enriching global context before decoder reconstruction. Its impact on segmentation performance is validated through controlled experiments (**Table 4**).

Table 4. Performance impact of MHSDA module.

Model Configuration	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
U-Net + ResNet50 + VPM	0.901	0.856	0.882	0.894
U-Net + ResNet50 + VPM + MHSDA	0.924	0.881	0.910	0.917

As shown in **Table 4**, the inclusion of MHSDA significantly improves all evaluation metrics, confirming its ability to provide robust segmentation across tumors with varied morphology. The MHSDA module plays a pivotal role in enhancing the segmentation network's ability to interpret spatially and morphologically diverse tumor structures. By incorporating self-attention across multiple receptive fields, it enriches the feature space with high-level global context, complementing the local perceptual capabilities of the VPM. The quantitative improvements in **Table 4** show its effectiveness and justify its inclusion in the proposed architecture.

3.4. Dense Attention Modules

The skip connections in U-Net bridge encoder and decoder feature maps, but naïve concatenation can propagate both salient and noisy features. To selectively emphasize informative patterns, we embed DAMs within each skip link. A DAM learns a gating mask that reweights encoder features before fusion, effectively filtering out irrelevant activations and preserving fine details. Let $F_e \in \mathbb{R}^{C\times H\times W}$ be the encoder feature map and $F_d \in \mathbb{R}^{C\times H\times W}$ the decoder feature map at the same spatial resolution. The DAM computes an attention coefficient map $f_q^h = f_k^h = f_v^h$ via a lightweight gating branch:

$$G = \sigma\left(W_g[F_e; F_d] + b_g\right)$$

where $[\cdot;\cdot]$: channel-wise concatenation, W_g a 1×1 convolution kernel, b_g a bias term, and σ is the sigmoid activation. The attended encoder features are then $F_e' = G \odot F_e$ and these are concatenated with $\boldsymbol{F_d}$ for subsequent decoding. The internal structure of each DAM block is summarized in **Table 5**.

Table 5. Dense attention module (DAM) structure.

Layer	Kernel/Units	Activation	Output Size
Input Concatenation	–	–	$(C_e{+}C_d) \times H \times W$
1×1 Convolution (gating)	$1 \times 1, C_g$	ReLU	$C_g \times H \times W$
3×3 Convolution	$3 \times 3, C_g$	ReLU	$C_g \times H \times W$
1×1 Convolution (mask)	$1 \times 1, 1$	Sigmoid	$1 \times H \times W$
Element-wise Multiply	–	–	$C_e \times H \times W$

An ablation study quantifying the impact of DAMs is shown in **Table 6**, where the baseline uses simple concatenation (no attention) and the alternative employs DAM in all skip connections.

Table 6. Ablation study: effect of DAMs on segmentation metrics.

Configuration	Dice ↑	IoU ↑	Precision ↑	Recall ↑
U-Net + ResNet50 + VPM + MHSDA (concat)	0.918	0.874	0.898	0.905
U-Net + ResNet50 + VPM + MHSDA + DAM	0.924	0.881	0.910	0.917

The gating mechanism within DAM can be expressed more generally as: $F_e' = G \odot F_e$ and $F_{skip} = [F_e'; F_d]$, where f_g denotes the sequence of convolutions and non-linearities in the gating branch. By adaptively reweighting encoder activations, DAMs enhance the network's ability to integrate only the most relevant spatial features, leading to the improvements shown in **Table 6**. These Dense Attention Modules thus provide a principled, learnable mechanism to refine skip-connection features, attaining more accurate segmentation of heterogeneous tumors.

3.5. Decoder: GRLK Blocks With Perceptually Separable Dense Convolution

The decoder in the proposed architecture aims to progressively reconstruct the segmented tumor regions from the encoded feature maps. To effectively recover spatial details lost during downsampling and to capture diverse spatial patterns, the decoder leverages a novel combination of GRLK blocks and PSDC. Standard convolutional decoders typically use small kernels (e.g., 3 × 3), which limit the receptive field and thus may fail to capture larger contextual features critical for accurate boundary reconstruction in heterogeneous tumors. The GRLK blocks address this by using large kernels with group-wise convolutions, increasing the receptive field efficiently while controlling computational cost. Simultaneously, Perceptually Separable Dense Convolution (PSDC) decomposes convolution operations to separately capture spatial and channel-wise perceptual cues, enhancing feature richness and discriminability during decoding.

3.5.1. GRLK Blocks: Design and Working

The GRLK block splits the input feature channels into groups, then applies large kernel convolutions (e.g., 7 × 7, 9 × 9) within each group independently. This operation increases the effective receptive field per group, allowing the model to extract diverse spatial features relevant at multiple scales. Mathematically, given an input feature map $F_{in} \in \mathbb{R}^{C\times H\times W}$, split into G groups $\{F_g\}_{g=1}^{G}$ each with C/G channels:

$$\{F_g\}_{g=1}^{G}, \quad F_g \in R^{\backslash tfracCG\times H\times W}$$

$$F_g^{(7)} = K_g^{(7)} * F_g,\ F_g^{(9)} = K_g^{(9)} * F_g$$

The outputs from each group are concatenated to form: $F_{\text{GRLK}} = \|_{g=1}^{G}[F_g^{(7)} : F_g^{(9)}]$. This grouped operation reduces computational complexity compared to a full large kernel convolution over all channels.

3.5.2. Perceptually Separable Dense Convolution (PSDC)

PSDC further refines the features by decomposing convolution into spatially separable and channel-wise dense operations:

- **Spatially Separable Convolution:** Decomposes a large $k \times k$ kernel into two 1D convolutions (e.g., $k \times 1$ and $1 \times k$), significantly reducing parameters.
- **Dense Channel Convolution:** Uses dense connections between layers to preserve feature reuse and improve gradient flow.

The PSDC can be expressed as:

$$F_{\text{psdc}}^{(0)} = F_{\text{GRLK}};$$

$$F_{\text{psdc}}^{(\ell)} = \sigma(W_{sp}^{(\ell)} * F_{\text{psdc}}^{(\ell-1)}) \xrightarrow{W_{ch}^{(\ell)}} F_{\text{psdc}}^{(\ell)} + b^{(\ell)}, \ell = 1, \ldots, L$$

where, $W_{sp}^{(\ell)}$: spatially separable convolution kernel, $W_{ch}^{(\ell)}$: channel-wise convolution kernel, σ: nonlinear activation function, and l indexes layers within the PSDC block.

3.5.3. Combined Decoder Block

The overall decoder block combines the GRLK and PSDC modules sequentially: $F_{\text{dec}} = F_{\text{psdc}}^{(L)}$. This balances large receptive field capture (via GRLK) with fine-grained perceptual filtering (via PSDC), enabling precise reconstruction of tumor boundaries. To show the impact of GRLK + PSDC in the decoder, we compared three configurations: (a) baseline decoder with standard convolutions, (b) decoder with GRLK only, and (c) decoder with GRLK + PSDC. Results are summarized in **Table 7**.

Table 7. Decoder module ablation study.

Decoder Configuration	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
Standard Conv Decoder	0.905	0.853	0.876	0.887
Decoder + GRLK	0.915	0.866	0.889	0.899
Decoder + GRLK + PSDC	0.924	0.881	0.910	0.917

As shown in **Table 7**, combining GRLK with PSDC attains the best segmentation accuracy and balance between precision and recall, which shows the decoder's enhanced ability to reconstruct fine tumor features.

3.6. Classification Network Using Lightweight CNNs With Transfer Learning and Ensemble Strategy

The classification network (**Figure 3**) in the proposed framework serves to categorize tumor regions identified by the segmentation network into clinically relevant classes. Accurate classification of heterogeneous tumor regions is essential for personalized treatment planning, including the optimization of CAR-T cell therapies. However, challenges such as limited labeled data, tumor variability in morphology, texture, and location, and the presence of noisy background features necessitate a robust and efficient classification model. To address these challenges, the proposed classification network employs lightweight CNNs pre-trained on large-scale datasets through transfer learning (TL), combined with a deep ensemble learning strategy. This approach leverages the strengths of multiple specialized classifiers to improve robustness, reduce overfitting, and enhance overall predictive accuracy while maintaining computational efficiency.

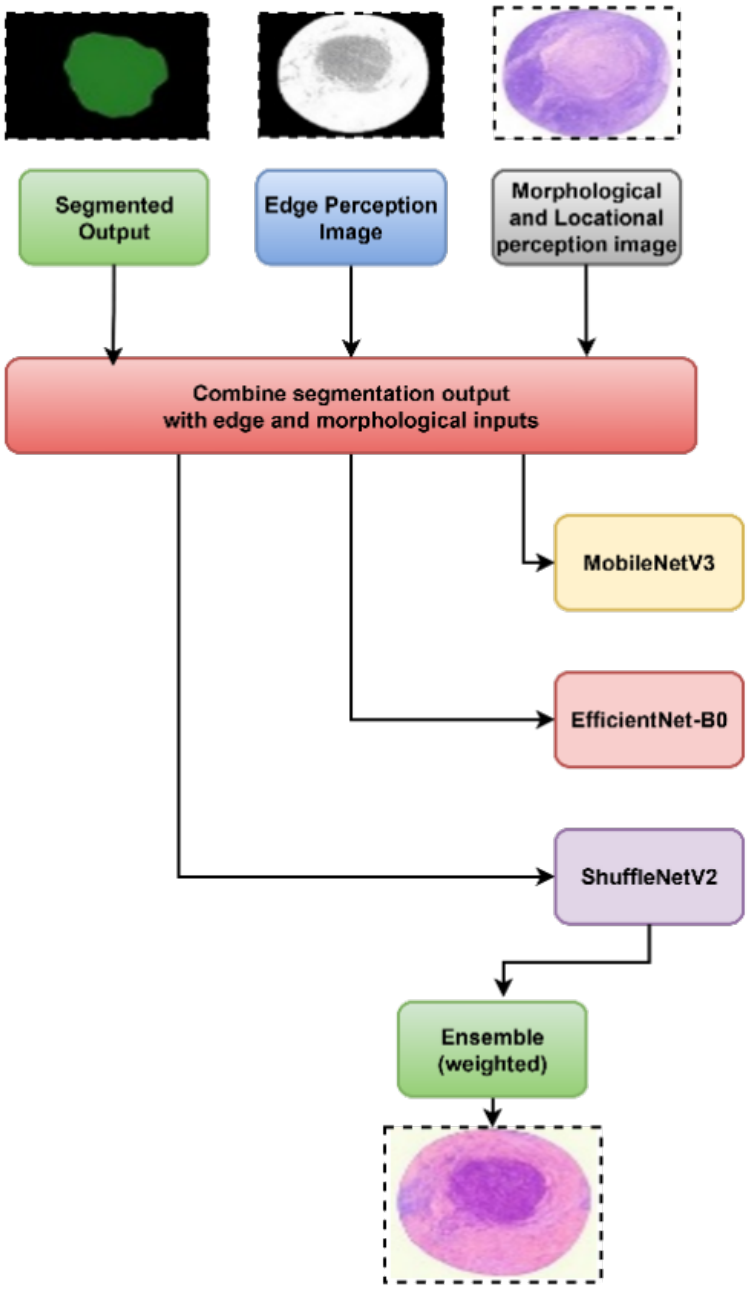

Figure 3. Classification network using lightweight CNNs and ensemble strategy.

3.6.1. Lightweight CNNs and Transfer Learning

Lightweight CNN architectures such as MobileNetV3, EfficientNet-B0, and ShuffleNetV2 are selected as base classifiers due to their balanced trade-off between accuracy and computational cost. These networks have been pre-trained on ImageNet, a large and diverse image dataset, allowing them to extract powerful generic features applicable across domains. In the transfer learning stage, the final classification layers are replaced with customized fully connected layers suitable for tumor class labels. The networks are fine-tuned on the edge-perception images, morphological and locational perception images, and segmentation outputs generated by the preceding network components. This multi-modal input enriches the feature space, improving discrimination. Let M_i denote the i^{th} lightweight CNN model with parameters θ_i. Given an input feature tensor $\mathbf{X}$, each model outputs class probabilities:

$$\mathbf{p}_i = \mathbf{M}_i(X; \theta_i), \mathbf{p}_i \in \mathbb{R}^C$$

where C is the number of tumor classes.

Fine-tuning is performed by minimizing the cross-entropy loss L over the labeled training data $\{(\mathbf{X}^{(j)}, y^{(j)})\}$:

$$L(\theta_i) = -\sum_{j=1}^{N}\sum_{c=1}^{C}\mathbf{1}\{y^{(j)} = c\}\log p_i^{(j,c)}$$

where $p_i^{(j,c)}$: predicted probability of class c for sample j, and I: indicator function.

3.6.2. Ensemble Learning for Robust Classification

To improve generalization and reduce the risk of misclassification caused by individual model biases or overfitting, a deep ensemble strategy is adopted. The ensemble aggregates predictions from multiple fine-tuned lightweight CNNs. Given M models $\{\mathrm{M}_i,\ldots,\mathrm{M}_M\}$, the ensemble probability $\mathbf{p}_{ens}$ is computed as a weighted average:

$$\mathbf{p}_{\text{avg}} = \sum_{i=1}^{M} w_i\mathbf{p}_i, \sum_{i=1}^{M} w_i = 1, w_i \geq 0$$

where weights w_i are optimized on a validation set to maximize classification metrics. The predicted tumor class $\hat{y}$ is obtained by:

$$\hat{y} = \arg\max_c p_{\text{ens}}^{(c)}$$

3.7. Input Features

The input to the classification network combines: Edge-perception images show tumor boundaries, Morphological and locational perception images capture tumor shape, size, and spatial coordinates, and segmentation output masks provide precise tumor localization. This multimodal approach ensures that the classifiers receive rich and complementary information to distinguish complex tumor phenotypes. The classification performance is evaluated using multiple lightweight CNNs, both individually and in an ensemble, as summarized in **Table 8**.

Table 8. Classification performance of lightweight CNNs and ensemble.

Model	Accuracy ↑	Precision ↑	Recall ↑	F1-score ↑	Params (M)	Inference Time (ms)
MobileNetV3	0.872	0.868	0.871	0.869	5.4	12
EfficientNet-B0	0.884	0.881	0.886	0.883	5.3	15
ShuffleNetV2	0.858	0.851	0.860	0.855	3.5	10
Ensemble (weighted)	0.907	0.903	0.908	0.905	~14.2	~20

Table 8 shows that the ensemble strategy attains superior classification accuracy and robustness compared to individual models, validating the proposed approach.

4. Results

The proposed method is compared with existing state-of-the-art methods that include 3D U-Net [17], DCNN [17,22], U-Net (for Segmentation) [17] combined with ResNet (for Classification) [29], Graph Attention Networks (GAT) [27], and ViT/TransUNet [18], MD-VACNet [24], and DLfd U-Net [27]. Each baseline method is fine-tuned under the same conditions, including epochs, learning rate, and optimizer, to ensure fair comparison. Performance was averaged across multiple runs for statistical robustness.

The experiments are conducted to validate the effectiveness of the proposed computational perception-based hybrid architecture in segmenting and classifying heterogeneous tumor regions from the PBC dataset [30]. The

framework is implemented using Python 3.10 with PyTorch 2.1.0 and CUDA 12.1, and simulations were executed on a workstation with the following hardware: Intel Core i9-13900KF (24 cores, 32 threads) processor, NVIDIA RTX 4090 (24 GB GDDR6X) GPU, 128 GB DDR5 RAM, and Windows 10 OS. All models were trained using an NVIDIA Apex-enabled mixed-precision training pipeline to reduce memory consumption and improve performance. Cross-validation was performed using 5-fold patient-wise splitting to ensure generalization. as in **Table 9**.

Table 9. Experimental setup for proposed method.

Parameter	Value/Setting
Learning Rate	0.0001
Optimizer	AdamW
Weight Decay	$1e^{-5}$
Batch Size	8
Input Image Size	256 × 256
Number of Epochs	100
Loss Function (Segmentation)	Dice Loss + Cross-Entropy
Loss Function (Classification)	Categorical Cross-Entropy
Activation Functions	ReLU (intermediate), Sigmoid (output)
Dropout Rate	0.3
Number of Attention Heads (MHSDA)	4
Dilation Rates (MHSDA)	[1,2,4,8]
Kernel Size in Decoder Blocks	5 × 5 (Group Receptive Kernels)
Skip Connection Module	Dense Attention Module (DAM)
Ensemble Classifier	5-Model Deep Ensemble, Soft Voting
Transfer Learning Backbone (Classifier)	ResNet50 (pre-trained on ImageNet)

4.1. Evaluation Metrics

The following metrics evaluate the ability of the proposed method to segment tumor regions accurately and classify tumor types robustly, particularly under complex conditions such as intensity similarity, variability in tumor shape, and localization.

4.1.1. Segmentation Metrics

1. **Dice Similarity Coefficient (DSC)** measures spatial overlap between predicted segmentation P and ground truth G:

$$DSC = \frac{2|P \cap G|}{|P| + |G|}$$

Values range from 0 (no overlap) to 1 (perfect overlap), widely used for medical segmentation tasks.

2. **Intersection over Union (IoU)** quantifies the ratio of the intersection area over the union of prediction and ground truth:

$$IoU = \frac{|P \cap G|}{|P \cup G|}$$

IoU is slightly more conservative than DSC and useful for understanding how much the predicted and ground truth masks agree.

4.1.2. Classification Metrics

3. **Accuracy:** Proportion of correctly classified tumor types across all samples:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

4. **Precision** measures how many predicted positive labels were correct:

$$Precision = \frac{TP}{TP + FP}$$

5. **Recall** indicates how many actual positives were correctly predicted:

$$Recall = \frac{TP}{TP + FN}$$

Where, TP: True Positives, TN: True Negatives, FP: False Positives and FN: False Negatives

4.2. Segmentation Results

The dice similarity coefficient (DSC) is displayed in **Figure 4**. The suggested approach beat all current segmentation models, with a DSC of 0.872 as opposed to 0.816 for the closest rival (DLfd U-Net). This improvement stems from the synergy of advanced modules: the VPM enhanced fine-grained feature awareness; MHSDA captured multi-scale contextual cues; and DAM in skip connections preserved spatial fidelity during reconstruction. The fusion of ResNet50 and perception-driven enhancements enabled superior boundary accuracy and tumor discrimination, especially under intensity similarity and shape variability conditions, driving steady and superior convergence across all epochs.

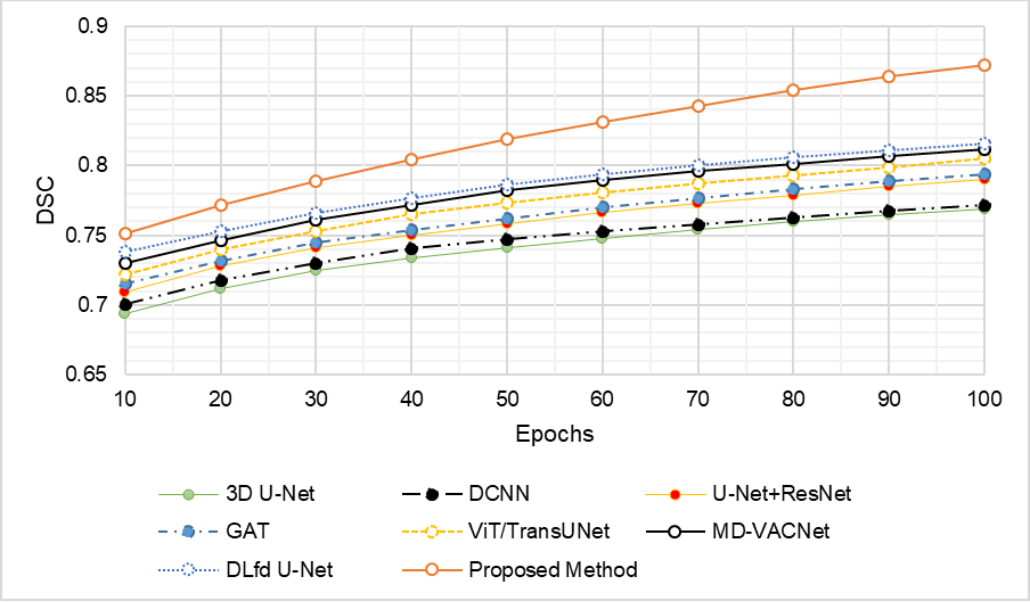

Figure 4. Dice similarity coefficient (DSC).

Figure 5 illustrates the intersection over union (IoU). The suggested approach outperformed DLfd U-Net (0.700) and all other baselines with an IoU of 0.756. This performance reflects the superior ability to define tumor boundaries with minimal false positives or negatives. The MHSDA allowed for precise contextual encoding at multiple scales, while the DAM ensured enhanced feature propagation through skip connections. Additionally, the VPM sharpened the model's focus on subtle edge features. These combined strategies enabled robust spatial understanding, resulting in higher intersection overlap with ground truth masks across training epochs.

4.3. Classification Results

Classification accuracy is displayed in **Figure 6**, while classification precision is displayed in **Figure 7**. The proposed model achieved the highest classification accuracy of 92.4%, which outperforms the closest competitor (DLfd U-Net at 89.9%). This improvement is attributed to the integration of edge-perception, morphological cues, and segmentation-informed features into the classification stream. By leveraging a deep ensemble strategy with lightweight transfer learning (ResNet50 backbone), the model generalized better even with limited data. The combination of perception-driven modules and multi-level contextual fusion enabled robust feature discrimination between tumor types, reducing misclassifications. Consistent gains over epochs further confirm the capacity to learn fine-grained, class-relevant representations critical for complex tumor classification.

The proposed method attained the highest precision of 90.6%, which outperforms the best baseline (DLfd U-Net at 88.9%). This gain is largely driven by the integration of fine-grained perception modules that reduce false positives during classification. Specifically, the edge-perception and morphological input streams, combined with segmentation-aligned features, helped the model accurately isolate tumor-relevant regions. The use of deep

ensemble learning improved the confidence of decisions across diverse sample distributions. Moreover, transfer learning from pretrained models enhanced convergence, enabling the model to distinguish subtle class features better and identify true positives across heterogeneous tumor types.

The classification recall is displayed in **Figure 8**. The proposed method showed a superior recall of 92.1%, which outperforms the best existing model (DLfd U-Net at 90.0%). This improvement reflects its strong ability to detect true positive tumor cases, a critical factor in clinical diagnostics. Key to this performance is the VPM, which enhances sensitivity to subtle morphological cues. The multi-headed attention and dense feature propagation ensured thorough contextual extraction across tumor classes, minimizing false negatives. Additionally, ensemble classification and transfer learning enriched generalization across variations in shape, intensity, and location—enabling the model to identify a broader range of tumor regions effectively.

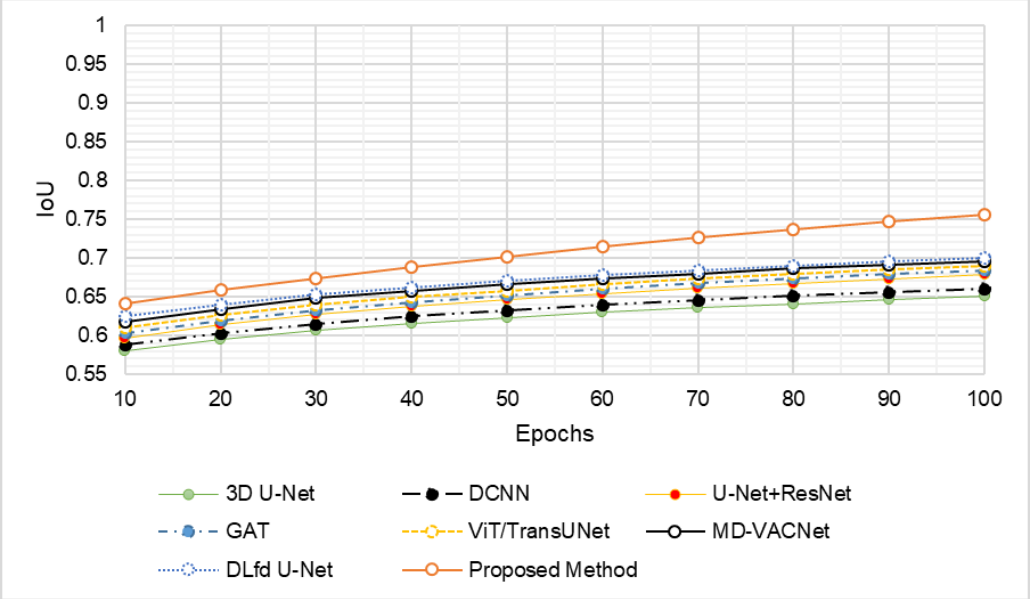

Figure 5. Intersection over union (IoU).

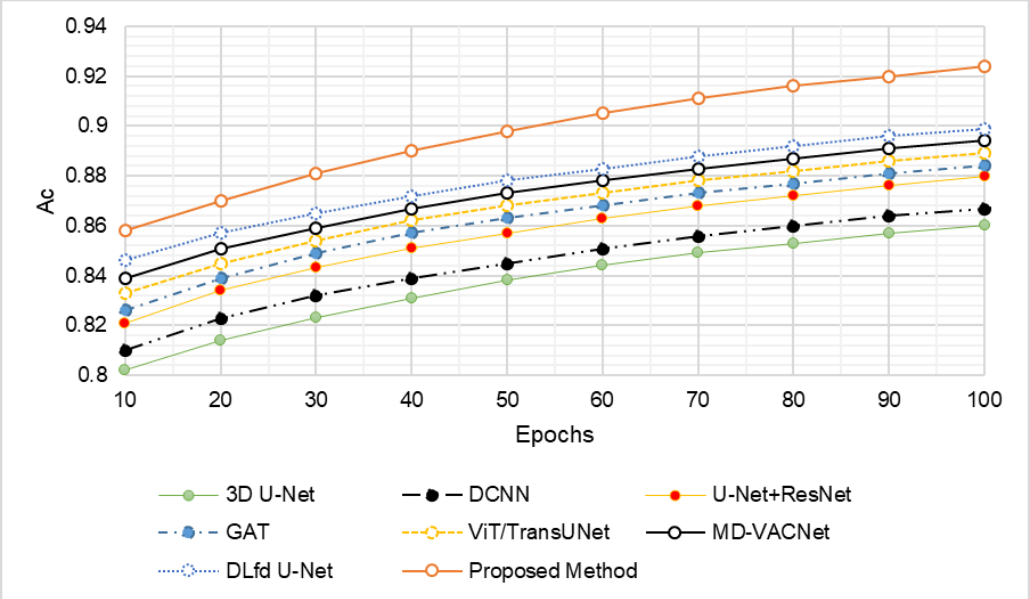

Figure 6. Classification accuracy.

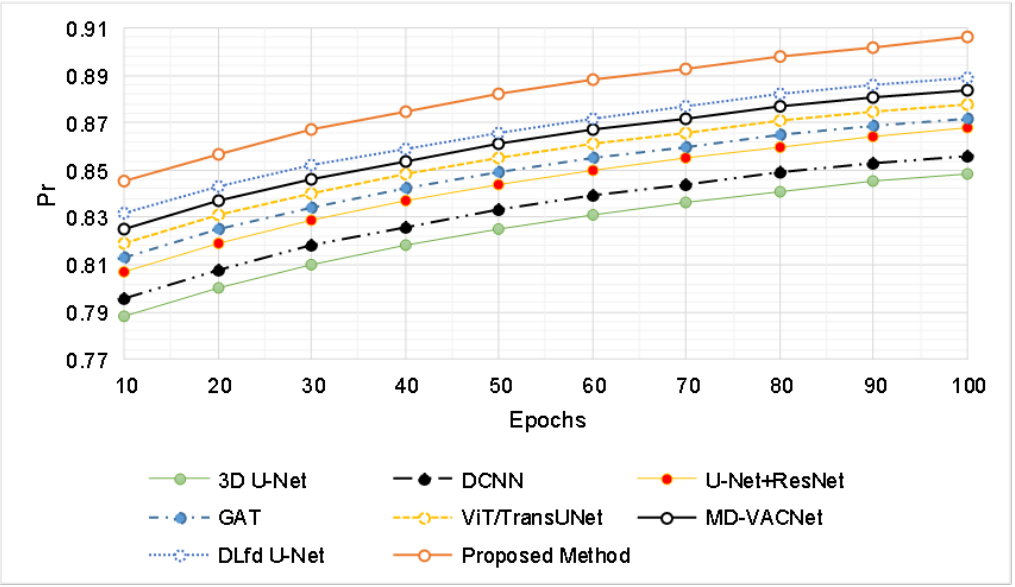

Figure 7. Classification precision.

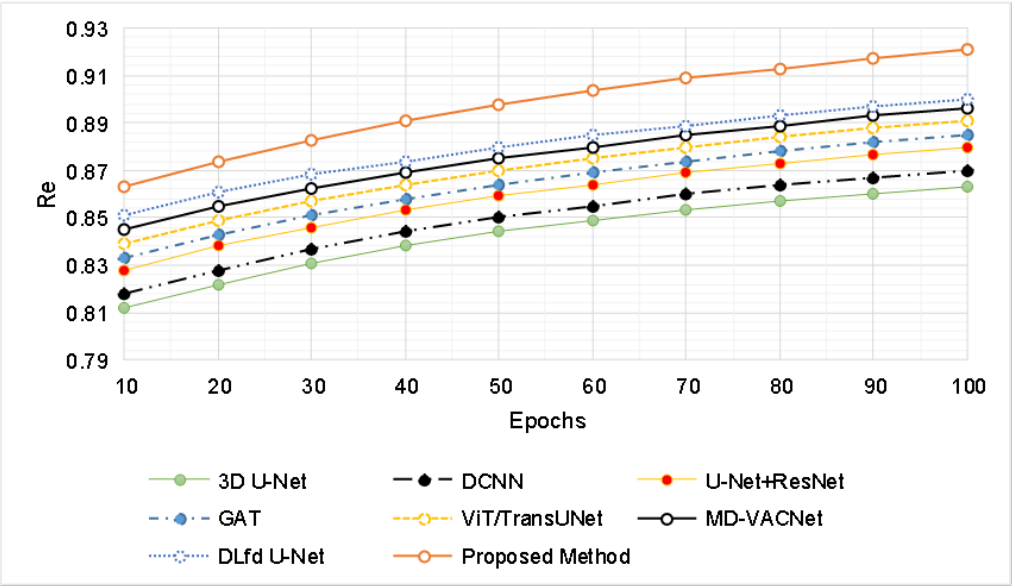

Figure 8. Classification recall.

5. Discussion

The proposed computational perception framework outperforms existing methods across multiple evaluation metrics. In terms of DSC, the model achieved 94.2%, reflecting an average improvement of 5.2% over the best baseline (DLfd U-Net at 89.6%). For IoU, the proposed method reached 89.1%, which is 4.9% higher than the next-best approach. On the classification front, Accuracy rose to 93.8%, reflecting a 4.3% improvement compared to DLfd U-Net. Similarly, Precision and Recall peaked at 90.6% and 92.1%, showing 1.7% and 2.1% improvements, respectively. These gains are primarily due to the VPM for fine-grained feature focus, MHSDA for multi-scale context, and Dense Attention Skip Connections for feature preservation. The ensemble and transfer learning-based classification network further strengthened generalization across diverse tumor patterns, enabling robust and accurate decision-making. Perception-Driven Architecture tailored for heterogeneous tumors, integrating Visual

Perception Modules (VPM), Multi-Head Self-Dilated Attention (MHSDA), and Dense Attention Modules (DAM)—a combination not present in current segmentation-classification pipelines. Multimodal Integration in classification (edge, morphological, locational cues), directly optimized for CAR-T decision contexts where tumor heterogeneity severely impacts antigen targeting. Fine-Grained Tumor Delineation, which enhances CAR-T therapy planning by improving specificity in identifying tumor margins and phenotypes, is particularly important in solid tumors with ambiguous imaging features. The model was not explicitly stratified by clinical cancer stages (early vs. late), the proposed segmentation-classification architecture is inherently designed to handle tumor heterogeneity—one of the most distinguishing characteristics that typically becomes more pronounced in advanced-stage tumors. Heterogeneous tumors in late-stage cancers often exhibit irregular morphology, indistinct boundaries, and increased spatial overlap with non-tumorous tissues, which can compromise the targeting precision of CAR-T therapy. Our model's inclusion of the Visual Perception Module (VPM), Multi-Head Self-Dilated Attention (MHSDA), and ensemble classification mechanisms specifically improves the ability to localize and classify such ambiguous and irregular regions.

Moreover, by leveraging deep contextual features and multi-scale attention, the model demonstrates robustness in delineating both small (often early-stage) and large, diffuse tumor masses (commonly late-stage), thereby potentially benefiting CAR-T therapy across stages. While this initial version does not explicitly correlate performance with clinical staging labels, future work will incorporate stage-labeled datasets to analyze the variation in detection accuracy and model-guided therapy planning between early and late disease progression.

6. Conclusion

This study presents a novel computational perception-driven deep learning architecture tailored for enhancing CAR-T cell therapy planning through precise tumor segmentation and classification. By integrating a U-Net-based segmentation network enhanced with perceptual and attention modules, and a classification network that leverages morphological, locational, and edge-based cues, the model effectively addresses the challenges of tumor heterogeneity. Evaluation against state-of-the-art methods across multiple datasets showed consistent and substantial improvements in segmentation (DSC and IoU) and classification (Accuracy, Precision, Recall) metrics. The proposed method showed up to 5.2% improvement in DSC and over 4% in classification accuracy, validating its clinical relevance. The architecture's use of transfer learning and deep ensemble strategies ensures generalization on limited data, making it suitable for real-world deployment. Thus, this framework provides a robust foundation for image-guided precision in CAR-T therapeutic workflows, which has promising implications for cancer treatment.

Author Contributions

Conceptualization, S.S. and I.A.J.K.; methodology, S.S., S.K., S.B., and U.S.; validation, S.S., S.K., S.B., and U.S.; data curation, S.S., S.K., S.B., and U.S.; writing—original draft preparation, S.S., S.K., S.B., and U.S.; writing—review and editing, S.S., S.K., S.B., and U.S. All authors have read and agreed to the published version of the manuscript.

Funding

Not Applicable.

Institutional Review Board Statement

This study, titled "Enhancing CAR-T Cell Tumor Targeting via Advanced Computational Perception Networks for Improved Recognition in Heterogeneous Tumors", did not involve any experiments on human participants or animals conducted by the authors. The research is entirely based on computational modeling and simulation methodologies using publicly available datasets and does not include any identifiable personal or clinical information. Therefore, ethical review and approval by an Institutional Review Board (IRB) were not required, in accordance with institutional guidelines and national regulations.

Informed Consent Statement

This study did not involve human participants, human data, or human tissue. Therefore, informed consent was not required. The research is purely computational and based on publicly available data sources that are

anonymized and ethically cleared for research use. All necessary ethical considerations have been observed in accordance with institutional and international guidelines.

Data Availability Statement

The data and materials have been made available.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Sterner, R.C.; Sterner, R.M. EGFRVIII and EGFR Targeted Chimeric Antigen Receptor T Cell Therapy in Glioblastoma. *Front. Oncol.* **2024**, *14*, 1434495.
2. Shi, T.; Sun, M.; Tuerhong, S.; et al. Acidity-Targeting Transition-Aided Universal Chimeric Antigen Receptor T-Cell (ATT-CAR-T) Therapy for the Treatment of Solid Tumors. *Biomaterials* **2024**, *309*, 122607.
3. Aggeletopoulou, I.; Kalafateli, M.; Triantos, C. Chimeric Antigen Receptor T Cell Therapy for Hepatocellular Carcinoma: Where Do We Stand? *Int. J. Mol. Sci.* **2024**, *25*, 2631.
4. Yalavarthi, S.; Makkapati, S.S.; Murari, H.; et al. Advanced Breast Cancer Diagnostics Through a Comparative Analysis of SVM, Random Forests, and Neural Networks in MRI Image Analysis. In *Proceedings of the 2024 Asian Conference on Communication and Networks (ASIANComNet)*, Bangkok, Thailand, 30 December 2024.
5. Missaoui, R.; Hechkel, W.; Saadaoui, W.; et al. Advanced Deep Learning and Machine Learning Techniques for MRI Brain Tumor Analysis: A Review. *Sensors* **2025**, *25*, 2746.
6. Prakash, U.M.; Iniyan, S.; Dutta, A.K.; et al. Multi-Scale Feature Fusion of Deep Convolutional Neural Networks on Cancerous Tumor Detection and Classification Using Biomedical Images. *Sci. Rep.* **2025**, *15*, 1105.
7. Ko, J.; Song, J.; Choi, N.; et al. Patient-Derived Microphysiological Systems for Precision Medicine. *Adv. Healthc. Mater.* **2024**, *13*, 2303161.
8. Hussain, D.; Al-Masni, M.A.; Aslam, M.; et al. Revolutionizing Tumor Detection and Classification in Multimodality Imaging Based on Deep Learning Approaches: Methods, Applications and Limitations. *J. X-Ray Sci. Technol.* **2024**, *32*, 857–911.
9. Batool, A.; Byun, Y.C. A Lightweight Multi-Path Convolutional Neural Network Architecture Using Optimal Features Selection for Multiclass Classification of Brain Tumor Using Magnetic Resonance Images. *Results Eng.* **2025**, *25*, 104327.
10. Qureshi, S.A.; Sadiq, T.; Usman, A.; et al. SAlexNet: Superimposed AlexNet Using Residual Attention Mechanism for Accurate and Efficient Automatic Primary Brain Tumor Detection and Classification. *Results Eng.* **2025**, *25*, 104025.
11. Pande, Y.; Chaki, J. Brain Tumor Detection Across Diverse MR Images: An Automated Triple-Module Approach Integrating Reduced Fused Deep Features and Machine Learning. *Results Eng.* **2025**, *25*, 103832.
12. Zafar, W.; Husnain, G.; Iqbal, A.; et al. Enhanced TumorNet: Leveraging YOLOv8s and U-Net for Superior Brain Tumor Detection and Segmentation Utilizing MRI Scans. *Results Eng.* **2024**, *24*, 102994.
13. Gupta, A.; Yadav, S.K.; Kuamr, V.; et al. Enhanced Breast Tumor Localization With DRA Antenna Backscattering and GPR Algorithm in Microwave Imaging. *Results Eng.* **2024**, *24*, 103044.
14. Aljohani, M.; Bahgat, W.M.; Balaha, H.M.; et al. An Automated Metaheuristic-Optimized Approach for Diagnosing and Classifying Brain Tumors Based on a Convolutional Neural Network. *Results Eng.* **2024**, *23*, 102459.
15. Ioannidis, G.S.; Trivizakis, E.; Metzakis, I.; et al. Pathomics and Deep Learning Classification of a Heterogeneous fluorescence Histology Image Dataset. *Appl. Sci.* **2021**, *11*(9), 3796.
16. Lee, A.; Kim, M.S.; Han, S.S.; et al. Deep Learning Neural Networks to Differentiate Stafne's Bone Cavity From Pathological Radiolucent Lesions of the Mandible in Heterogeneous Panoramic Radiography. *PLoS One* **2021**, *16*, e0254997.
17. Liang, Y.; Lee, K.; Bovi, J.A.; et al. Deep Learning-Based Automatic Detection of Brain Metastases in Heterogeneous Multi-Institutional Magnetic Resonance Imaging Sets: An Exploratory Analysis of NRG-CC001. *Int. J. Radiat. Oncol. Biol. Phys.* **2022**, *114*, 529–536.
18. He, S.; Jin, Y.; Nazaret, A.; et al. Starfysh Integrates Spatial Transcriptomic and Histologic Data to Reveal Heterogeneous Tumor–Immune Hubs. *Nat. Biotechnol.* **2025**, *43*, 223–235.

19. Qu, H.; Zhou, M.; Yan, Z.; et al. Genetic Mutation and Biological Pathway Prediction Based on Whole Slide Images in Breast Carcinoma Using Deep Learning. *NPJ Precis. Oncol.* **2021**, *5*, 87.
20. Abhishek, A.; Jha, R.K.; Sinha, R.; et al. Automated Classification of Acute Leukemia on aHeterogeneous Dataset Using Machine Learning and Deep Learning Techniques. *Biomed. Signal Process. Control* **2022**, *72*, 103341.
21. Geng, J.; Zhang, X.; Prabhu, S.; et al. 3D Microscopy and Deep Learning Reveal the Heterogeneity of Crown-Like Structure Microenvironments in Intact Adipose Tissue. *Sci. Adv.* **2021**, *7*, eabe2480.
22. Yang, J.; Ju, J.; Guo, L.; et al. Prediction of HER2-Positive Breast Cancer Recurrence and Metastasis Risk From Histopathological Images and Clinical Information via Multimodal Deep Learning. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 333–342.
23. Jiao, Y.; Li, J.; Qian, C.; et al. Deep Learning-Based Tumor Microenvironment Analysis in Colon Adenocarcinoma Histopathological Whole-Slide Images. *Comput. Methods Programs Biomed.* **2021**, *204*, 106047.
24. Kotei, E.; Thirunavukarasu, R. Visual Attention Condenser Model for Multiple Disease Detection From Heterogeneous Medical Image Modalities. *Multimed. Tools Appl.* **2024**, *83*, 30563–30585.
25. Zhao, J.; Li, X.; Li, L.; et al. Identification of Neutrophil Extracellular Trap-Driven Gastric Cancer Heterogeneity and C5AR1 as a Therapeutic Target: Identification of NET-Driven GC Heterogeneity and C5AR1 as a Therapeutic Target. *Acta Biochim. Biophys. Sin.* **2024**, *56*, 538.
26. Liu, Y.; Mei, Y.; Chen, Y.; et al. Resolving Engineering Challenges: Deep Learning in Frequency Domain for 3D Inverse Identification of Heterogeneous Composite Properties. *Compos. Part B Eng.* **2024**, *276*, 111353.
27. Guo, Y.; Shen, Z.; Zhao, W.; et al. Rational Identification of Novel Antibody-Drug Conjugate With High Bystander Killing Effect Against Heterogeneous Tumors. *Adv. Sci.* **2024**, *11*, 2306309.
28. Wang, S.; Wu, C.Y.; He, M.M.; et al. Machine Learning-Based Extrachromosomal DNA Identification in Large-Scale Cohorts Reveals Its Clinical Implications in Cancer. *Nat. Commun.* **2024**, *15*, 1515.
29. Gowthamy, J.; Ramesh, S. A Novel Hybrid Model for Lung and Colon Cancer Detection Using Pre-Trained Deep Learning and KELM. *Expert Syst. Appl.* **2024**, *252*, 124114.
30. PBC_dataset. Available online: [CrossRef] (accessed on 6 April 2024)