IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 June 2024, accepted 2 August 2024, date of publication 6 August 2024, date of current version 20 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3439439

RESEARCH ARTICLE

Weighted Fusion Transformer for Dual PET/CT Head and Neck Tumor Segmentation

MOHAMMED A. MAHDI[1], SHAHANAWAJ AHAMAD[2], (Member, IEEE), SAWSAN A. SAAD[3], ALAA DAFHALLA[3], RIZWAN QURESHI[4,5], (Senior Member, IEEE), AND ALAWI ALQUSHAIBI[6]

[1]Information and Computer Science Department, College of Computer Science and Engineering, University of Ha'il, Hail 81481, Saudi Arabia
[2]Software Engineering Department, College of Computer Science and Engineering, University of Ha'il, Hail 81481, Saudi Arabia
[3]Computer Engineering Department, College of Computer Science and Engineering, University of Ha'il, Hail 81481, Saudi Arabia
[4]National University of Computer and Emerging Sciences, Karachi Campus, Karachi, Pakistan
[5]Center for Research in Computer Vision (CRCV), University of Central Florida, Orlando, FL 32816, USA
[6]Department of Computer and Information Sciences, Universiti Teknologi Petronas, Seri Iskandar, Perak 32610, Malaysia

Corresponding author: Alawi Alqushaibi (alawi_18000555@utp.edu.my)

This work was supported by Scientific Research Deanship at University of Ha'il - Saudi Arabia through project number RG-23 137.

ABSTRACT Accurate tumor segmentation in PET/CT imaging is essential for the diagnosis and treatment of cancer, impacting therapeutic outcomes and patient management. Our study introduces a new approach integrating a Weighted Fusion Transformer Network to enhance the segmentation of tumor volumes. This method synergizes PET and CT modalities through a Fusion FormerU-Net architecture that employs convolutional neural networks alongside transformer blocks, aiming to leverage the unique advantages of each imaging modality. We evaluated the proposed approach using a multi-institutional dataset, applying key performance metrics such as Dice Similarity Coefficient aggregate, Jaccard Index, Volume Correlation, and Average Surface Distance to assess segmentation precision. The results indicate that the CT/PET/Fusion strategy significantly improves tumor delineation, outperforming traditional segmentation methods. The main findings suggest that this integrative approach could potentially redefine the standard for tumor segmentation in clinical practice. Lastly, the proposed approach offers a promising direction for enhancing the accuracy of oncological imaging, with implications for the improvement of patient-specific treatment strategies.

INDEX TERMS PET/CT imaging, tumor segmentation, weighted fusion transformer, multi-modal imaging, deep learning, neural networks, clinical oncology.

I. INTRODUCTION

Head and neck cancers (H&N) represent a significant global health challenge, with rising incidences and complex treatment pathways [1]. Recent advances in medical imaging, particularly FluoroDeoxyGlucose-Positron Emission Tomography (FDG-PET) and Computed Tomography (CT), have revolutionized diagnostic and treatment planning processes for these cancers [2]. However, the manual delineation of tumors and lymph nodes in imaging data remains time-consuming, subjective, and prone to variability [3], [4], [5], [6]. Therefore, there is a critical need for automated systems that can reliably segment head and neck tumors and predict patient outcomes [7], [8]. Segmentation of primary tumors and lymph nodes in head and neck cancer is crucial for radiation treatment planning and response assessment [9], [10], [11]. Manual segmentation, while essential, is complex and time-consuming, requiring expert knowledge [3], [4], [5]. Deep learning-based architectures in computer vision have shown state-of-the-art results in various applications, including medical image segmentation [12]. The adoption of deep learning for auto-segmentation in head and neck cancer could enhance both efficiency and accuracy, making the segmentation process more robust [13], [14], [15].

The associate editor coordinating the review of this manuscript and approving it for publication was Roberta Palmeri.

Recent efforts, such as the HECKTOR Challenge at MICCAI 2022 [14], have focused on leveraging machine and deep learning techniques to automate the segmentation of primary tumors and metastatic lymph nodes in FDG-PET/CT images [16]. The automatic segmentation and outcome prediction tasks pose unique challenges due to the heterogeneous nature of H&N tumors, variability in imaging protocols across different centers, and the need to integrate multimodal image analysis with clinical data for comprehensive patient management [16].

Additionally, PET and CT are crucial for diagnosing and treating H&N cancers [17]. PET imaging, while effective in early disease detection, presents challenges in automatic tumor segmentation due to its lack of spatial resolution and high intrinsic noise level [18]. PET images also show high metabolic activities from tumor cells but suffer from limited spatial resolution and signal-to-noise ratio. Fusion of PET and CT images provides more informative representation for automated tumor segmentation and target delineation [19], [20], [21]. In this context, our research aims to contribute to the ongoing efforts in automatic head and neck tumor segmentation and outcome prediction. We propose a novel approach that addresses the complexities of the task, including the diverse nature of tumor presentation in PET/CT images, and the challenges of achieving high accuracy and generalizability across nine different institute datasets. Our contributions summarized in fourfold: (i) proposing a new network architecture that integrates PET and CT modalities through a weighted fusion strategy for improved H&N tumor segmentation. This architecture likely leverages the strengths of each imaging modality to enhance segmentation accuracy; (ii) The utilization of both transformer blocks and convolutional neural networks (CNNs) in a unified framework, taking advantage of the transformers' ability to capture global dependencies and the CNNs' prowess in extracting local features; (iii) the implementation of a heuristic approach for initial tumor localization that could streamline the segmentation process, possibly reducing the computational load and improving the speed of the analysis. Lastly, the proposed network's performance has been validated on a diverse multi-institutional dataset, underscoring its robustness and potential for generalizability across different clinical settings.

The structure of this paper is systematically divided into five main sections to effectively present our research. Section II, related work, provides an overview of the existing studies and developments in the field, setting the stage for our research. In Section III, methods, and materials, we detail the techniques and resources utilized in our study, emphasizing the methodologies specific to PET/CT tumor segmentation. Section IV, results, and discussion is dedicated to showcasing the outcomes of our research, along with a comprehensive analysis and interpretation of these results in the context of dual PET/CT imaging. The paper culminates in Section V, conclusion, and future work, where we summarize our findings, underscore their significance in the medical imaging, and suggest potential avenues for future research.

II. RELATED WORKS

The increasing interest in automatic analysis of multimodal images using machine and deep learning techniques is particularly relevant in oncology [22]. Automation in tumor and lymph node delineation assists in diagnostic tasks, staging, quantitative assessment, radiotherapy planning, and outcome prediction, offering advantages in speed, robustness, and reproducibility over manual contouring [23], [24], [25], [26]. Additionally, the integration of multimodal image analysis with machine learning for patient-level segmentation and outcome prediction allows for predictive and prognostic modeling. This includes therapy response prediction, recurrence, and overall survival, combining image-derived data with clinical information to create decision-support tools that enhance personalized patient management [27], [28]. Myronenko [29] utilized SegResNet, a 3D U-Net-like architecture augmented with an auto-encoder and deep supervision, built on the MONAI platform and tailored for specific tasks such as PET/CT analysis and cropping through the Auto3DSeg system. This system automates parameter selection, integrating various steps like image normalization, tumor region detection (a feature specific to HECKTOR 2022), isotropic resampling, and employing 5-fold cross-validation along with model ensembling. Tumor region detection in their approach leverages relative anatomical positions, and training involves random 3D cropping, focusing on foreground classes with designated probabilities for tumors, lymph nodes, and background.

Sun et al. [30] adopted a multi-stage, coarse-to-fine strategy using a series of neural networks for precise tumor segmentation. The process begins with a 3D U-Net that identifies the head region in CT scans. Following this, a nnU-Net performs an initial, rough segmentation of the primary tumor (GTVp) and nodal tumor (GTVn) areas in PET/CT images, using the central part of these tumors as the ground truth. This step yields a smaller, focused bounding box around the area of interest. The final, detailed segmentation within this refined bounding box is achieved through an ensemble of nnU-Nets and nnFormers, further enhanced by a 3D SE-norm U-Net, resulting in the precise delineation of GTVp and GTVn regions. Jiang et al. [31] utilized a standard nnU-NET complemented by straightforward pre-processing and post-processing techniques. The training process involved cropping images around the primary tumor (GTVp). For post-processing, they applied outlier removal based on criteria like minimum volume and spatial proximity between predicted primary and nodal tumor volumes. Notably, they also integrated their segmentation results into a web-based platform, which allowed for the visualization of segmented regions, including Organs at Risk (OAR), extending beyond the challenge's primary scope.

Rebaud et al. [32] implemented a simple nnU-Net-based method, adapting it to the task with specific image resampling

and training techniques, followed by median filtering to smooth the masks. Salahuddin et al. [33] developed a 3D U-Net with channel-wise attention, grid-attention gates, and specialized residual connections, complemented by outlier removal in post-processing and non-isotropic resampling for the input images. Wang et al. [34] introduced a base nnU-Net enhanced with a Transfiner model to refine segmentation, particularly focusing on tumor boundaries, employing an octree decomposition for patch selection. In another study by Wang et al. [35], they opted for a straightforward segmentation with nnU-Net, using a dense patch-based approach and post-processing based on the spatial relation between primary and nodal tumor volumes. Additionally, Jain et al. [36] compared several deep learning models including 2D/3D nnU-Net, MNet, and SwinU-Net, using resampled images registered to a common reference and cropped based on skull center location. Three models showed promise, achieving average Dice Similarity Coefficients (DSC) of 0.77 for primary tumors and 0.70 for nodes in task 1 of the HECKTOR2022 challenge, using 5-fold cross-validation and ensembling on a hold-out set. Chen and Martel [37] created an ensemble of three 3D nnU-Nets with different loss functions, using CT images for initial input and PET images in post-processing to refine predictions. Rezaeijo et al. implemented a multi-step approach with an organ localizer and 3D U-Net for organ segmentation, followed by a 3D ResU-Net for tumor segmentation, using a weighted combination of registered PET and CT images. Meng et al. [38] proposed a new approach combining a U-Net-based segmentation network with a cascaded survival network built on a DenseNet architecture. This setup allows for joint optimization using both segmentation and survival loss, enabling simultaneous prediction of patient survival risk scores and tumor region segmentation.

While deep learning models have significantly advanced the field of head and neck cancer segmentation, there remain notable limitations. These include challenges in handling the high variability in tumor shapes and sizes, the need for large, annotated datasets for training, and the difficulty in generalizing models across different medical imaging protocols and equipment. These limitations highlight the necessity for ongoing research to improve model robustness and adaptability. The motivation for this study lies in addressing these gaps, emphasizing the need for more sophisticated models that can effectively navigate these complexities, thereby enhancing the precision and effectiveness of cancer treatment planning.

III. MATERIALS AND METHODS

The overall methodology of the proposed study presented in this section. Figure 1 presents an integrated pipeline for the segmentation of tumors in whole-body PET/CT imaging, leveraging the complementary strengths of both modalities. The workflow begins with the acquisition of high-resolution CT images, which provide detailed anatomical structures, and PET images, which highlight regions of increased metabolic activity indicative of potential neoplastic tissue. A critical preprocessing step is the registration of PET and CT images to ensure spatial alignment. The registration process is meticulously executed to facilitate the precise overlay of metabolic information onto the corresponding anatomical structures. This alignment is paramount for accurate tumor localization and subsequent segmentation. Following registration, the aligned datasets undergo a fusion process. This fusion is not merely a visual overlay but an intricate combination of data, optimized to retain critical information from both imaging modalities. The result is a single composite image that capitalizes on the high-resolution anatomical framework of CT and the metabolic landscape provided by PET, setting the stage for enhanced tumor detection.

The next phase in the pipeline involves heuristic tumor localization. This step utilizes a set of predefined rules or algorithms designed to quickly identify potential tumor regions within the fused images. While heuristic approaches may not capture the full complexity of tumor biology, they serve as an effective means to narrow down the regions of interest for more sophisticated analysis. Therefore, the core of our segmentation process is the Fusion Former-U-Net architecture—a novel neural network design that builds upon the classic U-Net framework. This architecture is specifically tailored to manage the fused PET/CT data, incorporating elements of transformer models to enhance its pattern recognition capabilities. The transformer's self-attention mechanisms are particularly adept at capturing long-range dependencies, which is crucial when analyzing the nuanced features of tumor morphology against the backdrop of whole-body imaging. Finally, the segmented tumor regions are subjected to shape analysis, an essential step for quantifying tumor characteristics. Shape analysis can yield valuable metrics such as volume, sphericity, and surface irregularity, which are instrumental in clinical decision-making processes, such as assessing tumor aggressiveness, planning treatment strategies, and monitoring response to therapy.

A. DATASET

The dataset for this study was sourced from HECKTOR Challenge at MICCAI 2022 [16] and acquired from nine different centers, as detailed in Table 1. It comprises FDG-PET/CT images from patients diagnosed with head and neck (H&N) cancer, specifically located in the oropharynx region. This diverse and multi-center dataset is crucial for ensuring the robustness and generalizability of the developed models for tumor segmentation. Table 1 include a comprehensive list of participating hospital centers from Canada (CA), the United States (US), Switzerland (CH), and France (FR), along with their respective case contributions. The total dataset consists of 524 cases for training and 359 cases for testing. This tabulation is crucial for understanding the scope and geographical diversity of the study.

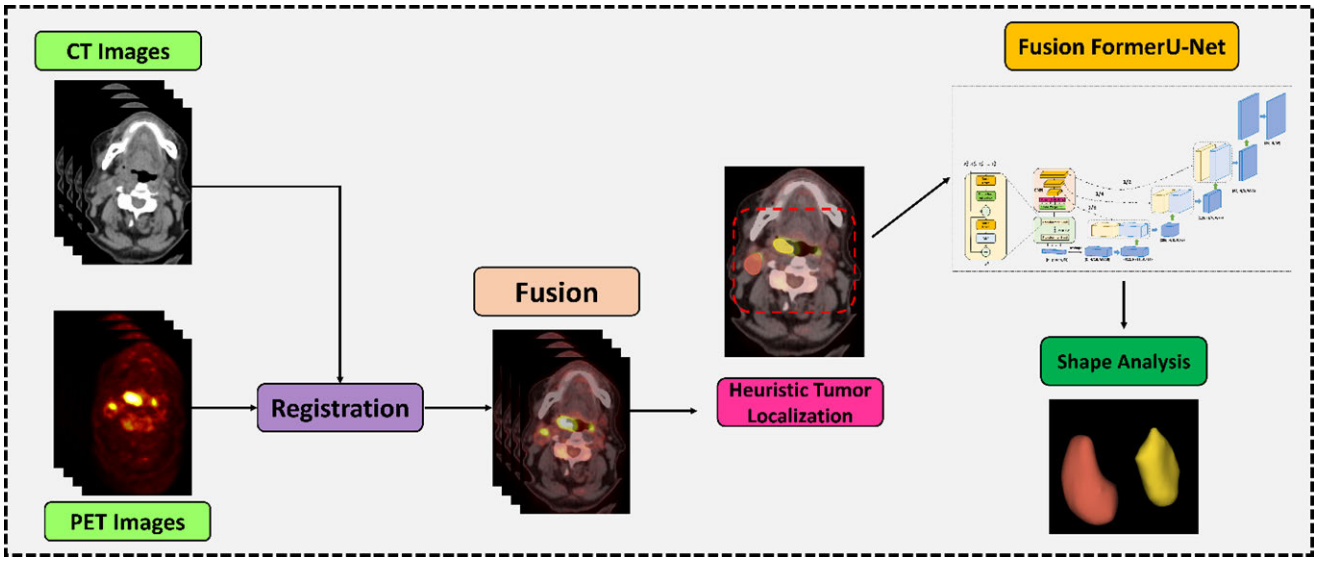

FIGURE 1. Schematic Representation of the Integrated PET/CT Tumor Segmentation Workflow Utilizing a Weighted Fusion Transformer Network.

TABLE 1. Distribution of PET/CT tumor segmentation multi-center dataset across participating.

No	Center	Split	#Cases
1	CHUM: Centre Hospitalier de l'Université de Montréal, Montréal, CA	Train	56
2	CHUS: Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, CA	Train	72
3	HGJ: Hôpital Général Juif, Montréal, CA	Train	55
4	HMR: Hôpital Maisonneuve-Rosemont, Montréal, CA	Train	18
5	CHUP: Centre Hospitalier Universitaire Poitiers, FR	Train	72
6	CHUV: Centre Hospitalier Universitaire Vaudois, CH	Train	53
	Total	**Train**	**524**
7	CHB: Centre Henri Becquerel, FR	Test	58
8	USZ: UniversitätsSpital Zürich, SW	Test	101
9	MDA: MD Anderson Cancer Center, US	Test	200
	Total	**Test**	**359**

Figure 2 illustrates 2D sagittal slices of fused PET/CT images from each of the nine participating centers, demonstrating the variability in fields of view. The images combine CT data in grayscale (with a Hounsfield unit window of [−140, 260]) and PET data (with a Standard Uptake Value (SUV) window of [0, 12]), depicted in a ''hot'' colormap. This figure effectively showcases the diversity in imaging across different centers.

The original annotations for the training and test sets varied across centers. For example, in CHUV, CHUS, HGJ, and HMR, an expert radiation oncologist drew the GTVp and GTVn contours, with some directly on the PET/CT scan's CT images and others on a different CT scan, later registered to the PET/CT. In CHUP, the primary tumor's metabolic volume was initially determined using FLAB and then manually edited. In MDA, available radiotherapy contours were refined, while in USZ, tumors were segmented separately in CT and PET images, with specific handling of artifacts. CHB involved manual drawing of GTVp and GTVn by senior nuclear medicine physicians using PET VCAR. Expert quality controls were conducted on all datasets to ensure ground-truth contour consistency.

For data preparation, experts re-annotated contours to match the actual tumor volume, which was often smaller than the initially delineated radiotherapy volumes. A centralized cloud environment facilitated uniform annotation. For cases lacking original GTVp or GTVn radiotherapy contours, experts used PET/CT fusion and N staging data for annotation. Cases with PET and CT mis-registrations were excluded. Additionally, detailed annotation guidelines developed by the expert board were used for this quality control process. The guidelines for annotating primary tumors in PET/CT images were provided in [16] to participants during the challenge, and these were also adhered to in our paper. These guidelines include specific instructions for the contouring process, considering both PET and unenhanced CT acquisitions. They emphasize the importance of accurate and consistent annotation practices to ensure the reliability of the tumor segmentation process.

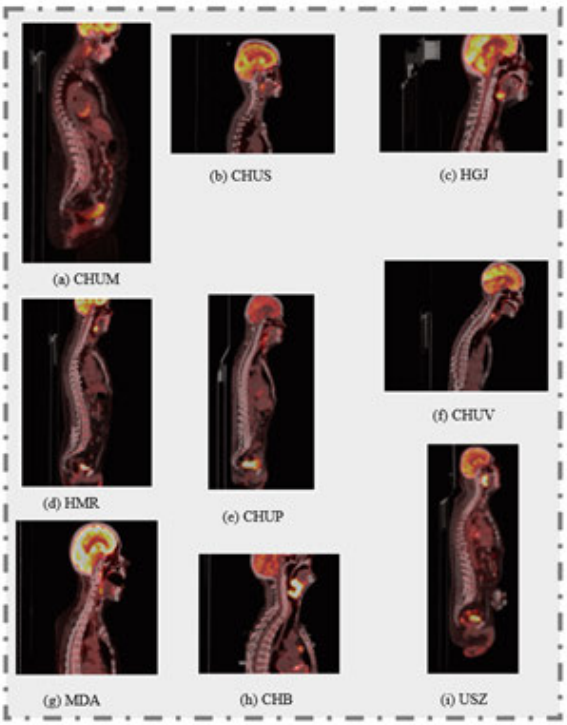

FIGURE 2. 2D sagittal slices of fused PET/CT images from each of the nine participating centers, demonstrating the variability in fields of view. The images combine CT data in grayscale (with a Hounsfield unit window of [−140, 260]) and PET data (with a Standard Uptake Value (SUV) window of [0, 12]), depicted in a "hot" colormap.

B. PREPROCESSING

The following preprocessing steps were implemented in our work. Normalization is a critical preprocessing step that aims to standardize the intensity distribution across different patients and imaging modalities. The objective is to align the dynamic range of the images, making the data more uniform and thus more amenable to analysis by computational models. Mathematically, this can involve z-score normalization, where each voxel intensity $I_{\mathbf{xyz}}$ in a 3D image is transformed as:

$$I'_{xyz} = \frac{I_{xyz} - \mu}{\sigma} \tag{1}$$

where $I'_{\mathbf{xyz}}$ is the normalized intensity, μ is the mean intensity across the image volume, and σ is the standard deviation of the intensities. This rescaling brings the dataset to a common scale with a mean of zero and a standard deviation of one.

Contrast enhancement [39] techniques are applied to each modality to improve the visibility of critical features. For PET images, the goal is to accentuate areas of high radiotracer uptake, which are often indicative of malignancy. In the case of CT, enhancement algorithms aim to increase the clarity of anatomical structures. The transformation function for contrast enhancement can be represented as:

$$I' = f(I) \tag{2}$$

where I is the original voxel intensity and I' is the enhanced intensity. The specific form of the function f depends on the enhancement technique employed (e.g., logarithmic mapping, histogram equalization).

Cropping focuses the analysis on the region of interest (ROI) by removing irrelevant background and reducing computational load. The process involves selecting a sub-volume that encapsulates the tumor and adjacent anatomical landmarks critical for diagnosis and treatment planning. The cropped image Icrop is defined by spatial boundaries within the original volume Ioriginal:

$$I_{\text{crop}} = I_{\text{original}}[x_{\min} : x_{\max}, y_{\min} : y_{\max}, z_{\min} : z_{\max}] \tag{3}$$

where $[x_{min} : x_{max}, y_{min} : y_{max}, z_{min} : z_{max}]$ defines the 3D bounding box of the ROI.

Voxel spacing homogenization was applied due to the different resolutions of PET and CT images, voxel spacing homogenization is employed. This process involves resampling the images to have consistent voxel dimensions, facilitating accurate image fusion and comparison. The transformation for homogenization can be represented by:

$$I_{\text{resampled}} = \text{Resample}\left(I_{\text{original}}, dX, dY, dZ\right) \tag{4}$$

where $I_{\text{"}resampled\text{"}}$ is the image with homogenized voxel spacing and dX,dY,dZ are the desired uniform voxel dimensions.

Data augmentation is crucial for enhancing the robustness and generalizability of the segmentation model. Cropping and flipping are common augmentation techniques that artificially expand the dataset by introducing variability. The cropped augmentation $Iaug_{flip}$ and randomly flipped augmentation I_{aug_crop} can be expressed as Equation (5) and Equation (6).

$$I_{aug\_crop} = \text{RandomCrop}\left(I_{original}\right) \tag{5}$$

$$I_{aug\_flip} = \text{RandomFlip}\left(I_{original}\right) \tag{6}$$

where RandomCrop RandomCrop randomly selects a sub-volume and RandomFlip RandomFlip applies a conditional mirror transformation across a randomly selected axis. Collectively, these preprocessing steps are indispensable for the accurate, reproducible, and robust segmentation of tumors. They address the inherent heterogeneity in multi-modal imaging datasets and enhance the quality of the input data, which is critical for the success of downstream machine learning models used in the segmentation task.

C. PROPOSED MODEL

In Figure 3, we present our comprehensive deep learning framework tailored for segmenting neoplastic lesions within PET/CT fusion volumes, leveraging the intrinsic synergies of multimodal imaging. The architecture encapsulates a transformative approach by integrating transformer networks with CNNs, optimizing the processing pipeline for the complex task of tumor delineation. The starting point is the multimodal input, where PET images and CT images are fused. This fusion volume integrates the high-resolution anatomical data from CT scans with the functional metabolic data from PET scans, which is critical for identifying areas

of potential neoplastic activity. Second, the multi-modality input illustrates the separate channels for the CT and PET data inputs, along with their fused counterpart. The individual CT and PET images are processed through separate pathways to preserve the unique information each modality provides before fusion. Third, Transformer Network Components: The process begins with a series of inputs $xp_1, xp_2, \ldots, xp_N$, likely patches extracted from the PET/CT fusion volume. Each input patch goes through a normalization layer, followed by a multi-head attention mechanism within the transformer block. The attention mechanism allows the model to focus on specific parts of the input data that are most relevant for tumor segmentation. This is followed by another normalization layer and a Multi-Layer Perceptron (MLP) block, which processes the data further. Residual connections are included, which help preserve the information through the layers and assist with the training of deep networks by mitigating the vanishing gradient problem. The transformer block is repeated 'n' times, which suggests that the network is deep enough to capture complex features required for accurate segmentation.

After initial processing by the transformer, the data is reshaped to a form suitable for convolutional processing. The feature maps are then downsampled through convolutional layers, reducing their dimensions by factors of 1/2, 1/4, and 1/8, respectively. This downsampling process helps extract and condense the most relevant features for segmentation. The output from the transformer blocks undergoes a series of dimensionality reductions through a convolutional downsampling strategy. These operations condense the high-dimensional data into more abstract feature representations, simultaneously reducing spatial resolution to capture a broader contextual understanding essential for segmenting complex tumor structures. Subsequently, the downsampled feature maps are funnelled through a series of convolutional layers, where each layer serves a distinct function delineated by colour-coded operations. The blue blocks perform convolution with a 3×3 kernel, introducing non-linearity via ReLU activation functions, while the yellow blocks engage in up-sampling through 2×2 convolutions, effectively refining the spatial granularity of the segmentation maps. The green blocks, consisting of 1×1 convolutions followed by softmax activation, are instrumental in classifying each voxel, generating a segmented output highlighting the tumor's presence within the PET/CT volume.

The assessment of segmentation performance utilizes the aggregated Dice Similarity Coefficient (DSCagg), a measure of volumetric overlap between the algorithm's predictions and expert annotations. DSCagg is advantageous for evaluating the segmentation of small regions within large images. While effective for ranking algorithms, particularly in tumor segmentation, its limitation is apparent when no ground-truth volume is present, potentially resulting in a DSC of 0. The metric is carefully chosen for its suitability in assessing segmentation accuracy for primary tumors (GTVp) and nodal tumors (GTVn), despite the inability to measure standard deviation across patient data.

$$\mathrm{DSC}_{\mathrm{agg}} = \frac{2\sum_i |A_i \cap B_i|}{\sum_i |A_i| + |B_i|} \tag{7}$$

where it calculates the ratio of twice the shared information between the predicted segmentation (A) and the ground-truth (B), over the total size of both individual segmentations. The higher the DSC, the more accurate the prediction is with the ground truth. This aggregated version of the coefficient implies a summation over multiple comparisons, providing an overall effectiveness measure for segmentation across a dataset rather than for a singular instance.

- **Jaccard Index (JI)**: Also known as the Intersection over Union (IoU), the Jaccard index is another common metric for evaluating the similarity between the predicted and actual values. It is defined as:

$$JI = \frac{|P \cap GT|}{|P \cup GT|} \tag{8}$$

  where P represents the predicted segmentation, and GT is the ground truth.
- **Volume Correlation (VC)**: Volume correlation assesses the correlation between the volumes of the predicted and true segmentations, providing insight into the volumetric accuracy of the model:

$$VC = \rho\left(V_{\mathrm{pred}}, V_{\mathrm{true}}\right) \tag{9}$$

  where $V_{\text{“}pred\text{”}}$ is the volume of the predicted segmentation, and $V_{\text{“}true\text{”}}$ is the volume of the ground-truth segmentation. Average Hausdorff Distance (AHD): The AHD measures the distance between the surfaces of the predicted and ground-truth segmentations, offering a surface distance metric:

$$AHD = \frac{1}{2}\left(\frac{\sum_{p\in P} \min_{g\in GT} d(p,g)}{|P|} + \frac{\sum_{g\in GT} \min_{p\in P} d(g,p)}{|GT|}\right) \tag{10}$$

  where d(p,g) is the Euclidean distance between points p and g in the predicted and ground truth segmentations, respectively.
- **Sensitivity:** it quantifies the proportion of actual positives correctly identified and is particularly important in medical image segmentation:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{11}$$

  where TP (True Positives) is the number of correctly identified positives and FN (False Negatives) is the number of positives not detected by the model.
- **Recall:** recall measures the proportion of actual tumor pixels that are correctly identified by the algorithm. A high recall value indicates that the algorithm can identify most of the tumor pixels, which is critical in

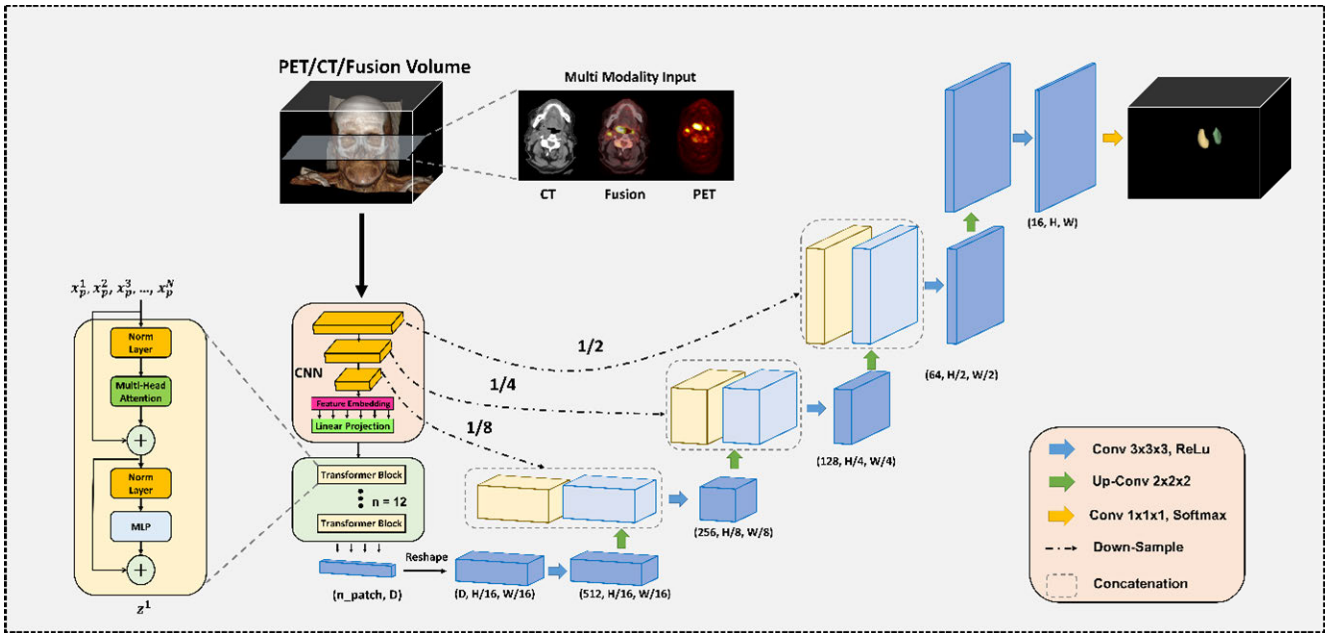

FIGURE 3. Architecture of the Weighted Fusion Transformer Network for Enhanced Multi-Modal Segmentation of Head and Neck Tumor Volumes in PET/CT Imaging.

medical imaging to ensure that no part of the tumor is missed for diagnosis or treatment planning.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

TP (True Positives) is the number of pixels (or voxels in 3D imaging) correctly identified as tumor by the segmentation algorithm. FN (False Negatives) is the number of pixels that are actually tumor but were not determined by the segmentation algorithm.

IV. RESULTS AND ANALYSIS

In the results section of our study, we meticulously evaluate the performance of our proposed approach with the CT/PET/Fusion strategy for H&N tumor segmentation. This strategy integrates the complementary strengths of CT and PET imaging to facilitate precise delineation of tumor boundaries. The DSC's quantitative outcomes offer a robust measure of the segmentation accuracy against the ground-truth. We present a detailed examination of the model's learning curve over an extensive series of epochs (more results in the supplementary file), showcasing the evolution of segmentation capability for both primary tumors and lymph nodes. This section elucidates the distinct learning behaviors and performance variances, providing a critical analysis of the efficacy of the multimodal fusion approach in a clinical setting.

A. MODEL TRAINING AND PERFORMANCE EVALUATION ON TUMOR SEGMENTATION

Figure 4 presents the progression of the validation mean DSC across 100 training epochs for the CT/PET/Fusion strategy. Graph (a) details the DSC for primary tumor volumes (GTVp), while graph (b) corresponds to lymph node volumes (GTVn). In Figure 4 (a), the DSC for GTVp displays a sharp increase during the early epochs, indicating that the model quickly assimilates the necessary features to accurately segment primary tumors from the multimodal imaging data. After this initial ascent, the graph levels off, showing a plateau which suggests that the model has achieved a stable segmentation performance for primary tumors and further training yields minimal improvements.

Figure 4 (b) reveals a different learning pattern for GTVn. The DSC climbs more steadily and exhibits more variability throughout the training process. This behavior suggests that segmenting lymph nodes presents a more complex challenge, potentially due to their variable shape, size, and less distinct boundaries compared to primary tumors. The lack of an apparent plateau implies that the model continues to refine its segmentation ability for lymph nodes throughout the training process.

B. EVALUATION OF FUSION WEIGHT AND IMAGING MODALITY ON SEGMENTATION PERFORMANCE

In the results subsection of our study, we present a comprehensive evaluation of the proposed weighted fusion transformer for dual PET/CT H&N tumor segmentation. The assessment is quantified using several standard imaging metrics, which are instrumental in determining the accuracy and reliability of the segmentation process.

We analyzed these metrics across different modalities and fusion strategies. The modalities include Computed Tomography (CT) and Positron Emission Tomography (PET). Additionally, we explored the impact of weighted fusion techniques, applying PET weights of 10%, 20%, and 30% in the fusion process Table (2). For CT, the Dice

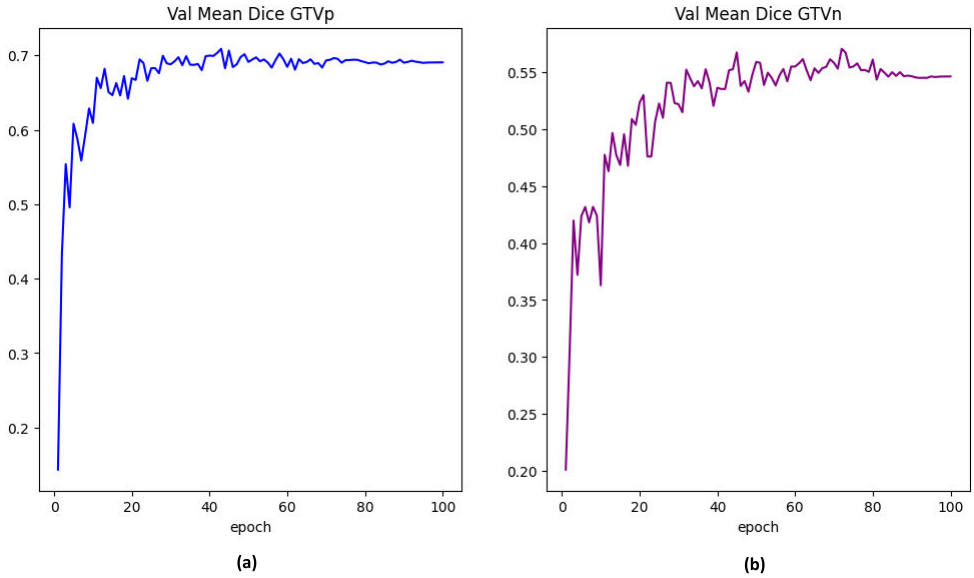

FIGURE 4. Our Model Training Progression of the CT/PET/Fusion strategy for Tumor Segmentation. (a) Validation means (DSC for primary tumor volumes (GTVp) over 100 epochs. (b) Validation mean DSC for lymph node volumes (GTVn) over the same number of epochs. Both graphs illustrate the model's learning trajectory and stabilization in segmentation performance, with GTVp achieving a higher DSC and an earlier plateau than GTVn.

coefficient ranged from 0.447 for overall tumor segmentation to 0.412 and 0.498 for primary tumor and lymph nodes, respectively. The Jaccard index and ASD also reflected similar trends, with ASD values indicating a mean distance error of approximately 32.929 mm across all tumor regions. In the PET modality, there was a notable improvement in most metrics. For instance, the Dice coefficient improved to 0.630 overall, slightly varying between primary tumor and lymph nodes. This trend was consistent across other metrics like Jaccard index and ASD, suggesting enhanced segmentation accuracy with PET. The implementation of weighted fusion strategies (10%, 20%, 30% to 90% PET weight) demonstrated varying degrees of improvement in segmentation accuracy. For instance, with a 20% PET weight, the Dice coefficient increased to 0.590 overall, indicating a more refined segmentation than CT or PET alone. Similarly, other metrics like Precision and Recall showed progressive improvement with increased PET weighting, signifying the effectiveness of the weighted fusion approach in enhancing tumor delineation.

As PET weighting increases from 40% to 90%, we observe notable trends in the performance metrics. For instance, the Dice coefficient and Jaccard index generally exhibit an improvement, indicating increased similarity between the segmented tumor and the ground-truth. Specifically, the Dice coefficient improves significantly at 70% PET weighting, reaching a value of 0.631 for overall tumor segmentation, and peaks at 80% PET weighting with a score of 0.649. The ASD metric, representing the mean surface distance error, shows a decreasing trend with increased PET weighting, suggesting improved spatial accuracy in tumor boundary delineation. For instance, at 90% PET weighting, ASD significantly drops, reflecting enhanced precision in tumor boundary delineation. Additionally, recall, sensitivity, and specificity metrics remain consistently high across all weighted fusion configurations. This consistency indicates reliable detection of tumor regions (high sensitivity) while maintaining a high rate of correctly identifying non-tumor areas (high specificity).

The VC metric, reflecting the correlation between the segmented tumor volume and the actual tumor volume, shows improvement as PET weighting increases. This improvement is most notable at 80% and 90% PET weightings, suggesting enhanced volumetric accuracy in these configurations. Our results indicate a clear trend of improved segmentation performance with increased PET weighting in the weighted fusion transformer as presented in Table 3. The metrics suggest that higher PET contributions, particularly in the 70% to 90% range, provide a more accurate and precise delineation of tumor regions in H&N imaging.

Based on Table 2 and Table 3, the exploration of various weighted fusion strategies in PET/CT imaging for tumor segmentation has led to a noteworthy revelation: the CT/PET/Fusion strategy emerges as the most effective approach. This conclusion is drawn from a comprehensive analysis of segmentation performance metrics. The superiority of the CT/PET/Fusion strategy likely stems from the synergistic integration of the distinct advantages of CT and PET imaging. CT scans provide detailed anatomical information, essential for delineating the precise boundaries and locations of tumors. PET scans, on the other hand, contribute metabolic activity data, a crucial factor in identifying active tumor tissues. By fusing these modalities, the CT/PET/Fusion approach facilitates a more comprehensive and accurate tumor segmentation, leveraging the strengths of both imaging techniques.

TABLE 2. Performance metrics of the weighted fusion transformer in dual PET/CT Tumor segmentation.

Input Modality	Metrics	Mean	Primary tumor	Lymph nodes
CT	DSCagg	0.447	0.412	0.498
	Jaccard	0.305	0.279	0.349
	ASD	32.929	35.188	33.291
	Precision	0.435	0.401	0.505
	Recall	0.556	0.598	0.586
	Sensitivity	0.556	0.598	0.586
	VC	0.79	0.738	0.779
PET	DSCagg	0.630	0.634	0.578
	Jaccard	0.485	0.500	0.436
	ASD	30.084	31.459	34.225
	Precision	0.624	0.600	0.643
	Recall	0.724	0.804	0.627
	Sensitivity	0.724	0.804	0.627
	VC	0.870	0.876	0.805
Weighted Fusion ($\mathrm{wPET} = 10\%$)	$\mathrm{DSC}_{\mathrm{agg}}$	0.554	0.585	0.479
	Jaccard	0.401	0.446	0.340
	ASD	52.389	18.612	76.477
	Precision	0.556	0.576	0.530
	Recall	0.643	0.723	0.544
	Sensitivity	0.643	0.723	0.544
	VC	0.859	0.850	0.765
Weighted Fusion (wPET = 20%)	DSCagg	0.590	0.628	0.520
	Jaccard	0.438	0.493	0.377
	ASD	61.168	26.306	85.386
	Precision	0.578	0.593	0.553
	Recall	0.699	0.782	0.616
	Sensitivity	0.699	0.782	0.616
	VC	0.848	0.866	0.771
Weighted Fusion (wPET = 30%)	DSCagg	0.603	0.640	0.514
	Jaccard	0.451	0.505	0.374
	ASD	54.938	25.184	88.493
	Precision	0.639	0.638	0.603
	Recall	0.654	0.747	0.546
	Sensitivity	0.654	0.747	0.546
	VC	0.843	0.883	0.715

This finding has profound implications for clinical practice. The enhanced accuracy in tumor segmentation afforded by the CT/PET/Fusion method can significantly improve diagnostic precision, treatment planning, and patient outcomes. Accurate segmentation is particularly critical in oncology, where the precise delineation of tumor margins is vital for effective treatment planning, including radiotherapy and surgical interventions.

While the weighted fusion strategies, particularly those with higher PET weights, showed promising results in certain aspects of tumor segmentation, they did not consistently outperform the CT/PET/Fusion approach. This suggests that while increasing PET weight can improve segmentation in some scenarios, a balanced integration of CT and PET data, as seen in the CT/PET/Fusion strategy, provides a more reliable and universally applicable solution. Figure 5 provides a comparative visualization of segmentation contours for head and neck (H&N) tumors within two distinct cases, labeled as (a) and (b).

Figure 5 show the segmentation results overlaid on axial slices from medical imaging studies. The green contour represents the ground-truth, which is the reference standard for tumor boundaries. The yellow contour delineates the segmentation achieved by a PET/CT-based algorithm, while the blue contour indicates the segmentation from a PET/CT/Fusion-based approach. The proximity of the blue PET/CT/Fusion contour to the green ground-truth contour suggests a higher degree of agreement with the reference standard than the yellow PET/CT contour. This implies that the fusion-based strategy provides a more accurate representation of the tumor extent, likely due to the integration of both anatomical and functional imaging data, which leverages the high-resolution detail of CT with the metabolic information from PET scans.

Figure 6 illustrates the similarity differences between segmentation outputs obtained from individual and combined modalities. Each circle represents a pairwise comparison between methods, where the color intensity indicates the magnitude of the similarity difference (red indicates higher, blue lower), and the size of the circle reflects the statistical significance of the differences. Smaller circles represent lower p-values, indicating more significant differences ($p < 0.05$). Larger circles correspond to p-values greater than 0.05, indicating non-significant differences. The diagonal elements are excluded from statistical comparison as they represent

TABLE 3. Performance enhancement of weighted fusion transformer with increased PET weightings.

Input Modality	Metrics	Mean	Primary tumor	Lymph nodes
Weighted Fusion (wPET = 40%)	DSCagg	0.587	0.645	0.526
	Jaccard	0.441	0.509	0.383
	ASD	59.937	24.023	77.518
	Precision	0.556	0.601	0.544
	Recall	0.735	0.812	0.642
	Sensitivity	0.735	0.812	0.642
	VC	0.865	0.908	0.776
Weighted Fusion (wPET = 50%)	DSCagg	0.579	0.645	0.501
	Jaccard	0.433	0.511	0.362
	ASD	76.639	32.231	98.884
	Precision	0.535	0.601	0.500
	Recall	0.747	0.824	0.643
	Sensitivity	0.747	0.824	0.643
	VC	0.839	0.911	0.721
Weighted Fusion (wPET = 60%)	DSCagg	0.583	0.581	0.542
	Jaccard	0.440	0.450	0.402
	ASD	63.315	66.977	63.343
	Precision	0.551	0.516	0.596
	Recall	0.736	0.836	0.607
	Sensitivity	0.736	0.836	0.607
	VC	0.847	0.850	0.783
Weighted Fusion (wPET = 70%)	DSCagg	0.631	0.641	0.552
	Jaccard	0.484	0.506	0.411
	ASD	39.795	35.616	48.029
	Precision	0.612	0.605	0.586
	Recall	0.743	0.810	0.623
	Sensitivity	0.743	0.810	0.623
	VC	0.891	0.904	0.823
Weighted Fusion (wPET = 80%)	DSCagg	0.644	0.653	0.581
	Jaccard	0.498	0.517	0.440
	ASD	25.061	30.895	24.504
	Precision	0.630	0.610	0.629
	Recall	0.733	0.817	0.628
	Sensitivity	0.733	0.817	0.628
	VC	0.900	0.905	0.822
Weighted Fusion (wPET = 90%)	DSCagg	0.649	0.668	0.582
	Jaccard	0.504	0.530	0.442
	ASD	13.667	12.322	18.059
	Precision	0.641	0.647	0.639
	Recall	0.734	0.809	0.637
	Sensitivity	0.734	0.809	0.637
	VC	0.902	0.913	0.828
CT/PET	DSCagg	0.653	0.671	0.577
	Jaccard	0.509	0.540	0.438
	ASD	21.106	11.293	28.002
	Precision	0.644	0.653	0.614
	Recall	0.732	0.786	0.640
	Sensitivity	0.732	0.786	0.640
	VC	0.898	0.888	0.819
CT/Fusion	DSCagg	0.659	0.678	0.586
	Jaccard	0.513	0.542	0.442
	ASD	18.885	13.750	25.597
	Precision	0.668	0.672	0.635
	Recall	0.718	0.781	0.628
	Sensitivity	0.718	0.781	0.628
	VC	0.914	0.922	0.858
PET/Fusion	DSCagg	0.652	0.656	0.596
	Jaccard	0.505	0.518	0.450
	ASD	17.785	11.227	24.473
	Precision	0.646	0.639	0.631
	Recall	0.727	0.786	0.646
	Sensitivity	0.727	0.786	0.646
	VC	0.920	0.924	0.855
CT/PET/Fusion	DSCagg	0.680	0.685	0.591
	Jaccard	0.534	0.550	0.452
	ASD	14.941	7.919	25.831
	Precision	0.678	0.682	0.633
	Recall	0.743	0.786	0.640
	Sensitivity	0.743	0.786	0.640
	VC	0.922	0.935	0.852

comparisons of each method against itself. This visualization highlights the robustness of the CT/PET/Fusion method over other combinations and individual modalities in terms of segmentation performance.

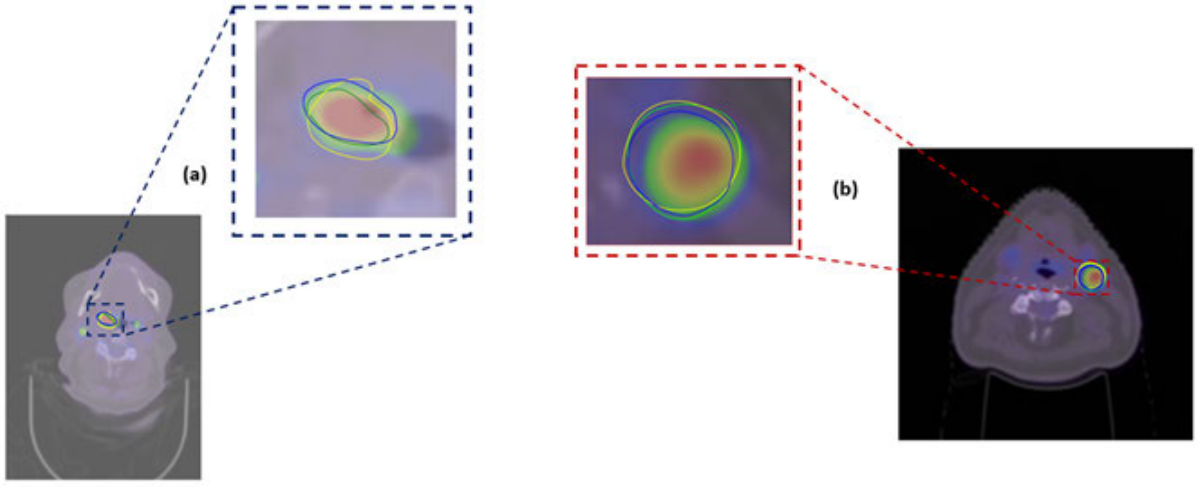

FIGURE 5. H&N Segmentation Contours for two different cases (a) and (b) randomly selected. The green color represents the ground-truth, the yellow corresponds to PET/CT, and the blue corresponds to PET/CT/Fusion. The blue color PET/CT/Fusion is similar to the ground-truth.

C. MODEL ZOO FOR COMPARATIVE PERFORMANCE ANALYSIS

Figure 7 presents a series of violin plots that detail the distribution of the Dice Similarity Coefficient (DSC) across a range of fusion weights in a multimodal tumor segmentation framework. The fusion weight, expressed in percentage, indicates the proportion of PET imaging data fused with CT data, assuming that the remainder is the weight of the CT data. Each violin plot encapsulates both a kernel density estimation illustrating the distribution of DSC scores and an embedded boxplot that marks the interquartile range (IQR) and the median value of the DSC.

As the fusion weight increases from 10% to 90%, the DSC distributions show a visible variation. The median DSC values, central tendencies, and the shape of the distributions offer insights into the optimal fusion weight for achieving the best segmentation performance. The figure allows for a comparison across the different fusion weights to determine which proportion of PET to CT data fusion results in the highest segmentation accuracy, as indicated by the DSC. Notably, some fusion weights result in a broader distribution of DSC scores, which may suggest variability in the segmentation performance across different images or patient datasets. This figure is instrumental in identifying the fusion weight that maximizes segmentation accuracy while highlighting the variability inherent in the segmentation process.

Figure 8 provided presents a series of boxplots that compare the performance of various input modalities in H&N tumor segmentation tasks using five key evaluation metrics. These modalities include individual CT and PET images, as well as their fused combinations in different configurations: Fusion alone, PET/CT, PET/Fusion, CT/Fusion, and the integrative PET/CT/Fusion. Figure 8 (a) illustrates the distribution of DSC values, a statistical measure of the segmentation accuracy, where 1 indicates perfect overlap and 0 indicates no overlap. The median DSC values for each modality are represented by the red lines within the boxes. Higher DSC values denote a better match between the algorithm's segmentation and the ground-truth. The PET/CT/Fusion modality appears to have a relatively higher median DSC value, suggesting superior segmentation performance compared to other modalities. Figure 8 (b) displays the distribution of the Jaccard Index, another measure of the similarity between the predicted segmentation and the ground-truth. Similar to the DSC, higher values indicate better performance. The PET/CT/Fusion modality again shows a higher median value, reinforcing the effectiveness of the integrated approach. Figure 8 (c) represents the average distance between the surfaces of the predicted segmentation and the ground-truth. Lower ASD values are indicative of more accurate segmentations with boundaries closer to the actual tumor margins. The PET/CT/Fusion modality demonstrates lower median ASD values, suggesting its superior boundary delineation capability. Figure 8 (d) indicates the proportion of true positive findings out of all positive findings reported by the segmentation algorithm. Higher precision values suggest fewer false positives in the segmentation output. The CT/Fusion and PET/CT/Fusion modalities show higher precision, implying more accurate identification of tumor tissues. Figure 8 (e) presents the recall, which measures the proportion of actual positive cases correctly identified by the segmentation algorithm. Higher recall values are preferable, as they indicate fewer false negatives. The PET/CT/Fusion modality exhibits a higher median recall, which suggests that it is more reliable in identifying all relevant H&N tumor regions.

Additionally, the boxplots in Figure 8 collectively demonstrate that the integrated PET/CT/Fusion modality consistently outperforms other modalities across all metrics, offering a more precise and reliable tumor segmentation. The different types of data, shown by the spread of the boxes and

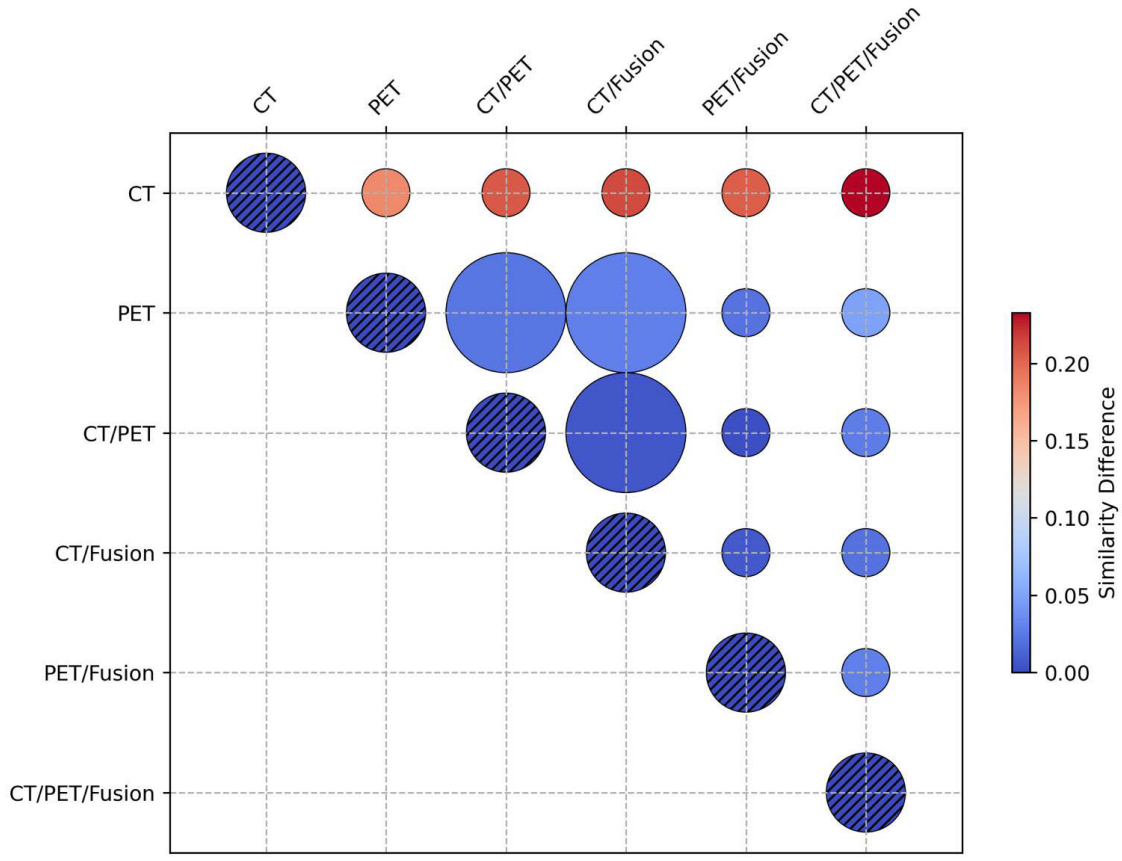

FIGURE 6. Comparative Analysis of Segmentation Performance P-values across Different Modalities and Their Combinations.

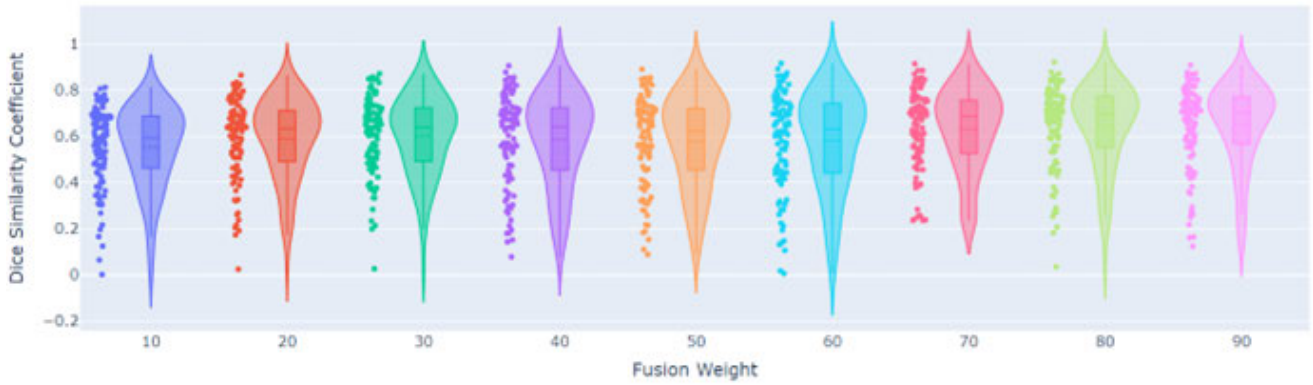

FIGURE 7. Impact of Fusion Weight on Dice Similarity Coefficient for H&N Tumor Segmentation. Violin plots illustrate the distribution of Dice Similarity Coefficient (DSC) across different fusion weights (ranging from 10% to 90%) applied within a multimodal segmentation framework. Each plot combines a boxplot, detailing the interquartile range (IQR) and median DSC (central mark), with a kernel density plot that depicts the distribution shape of the DSC scores. The width of the plots corresponds to the frequency of data points, providing insight into the density of the coefficient distribution at each fusion weight level.

whiskers, show that the PET/CT/Fusion method works well in a lot of different situations.

D. APPLICABILITY OF THE PROPOSED METHOD

Precise segmentation of tumor volumes is indeed critical in specific clinical applications that our study aims to support. Two primary areas where precise segmentation from PET/CT is invaluable include radiotherapy treatment planning and volumetric assessment for tumor response evaluation.

1) Radiotherapy Treatment Planning: Accurate tumor delineation is essential for targeting and dosimetry in radiotherapy. Our method improves the precision of segmenting the tumor boundaries, thus enabling more effective and tailored radiation treatment, which

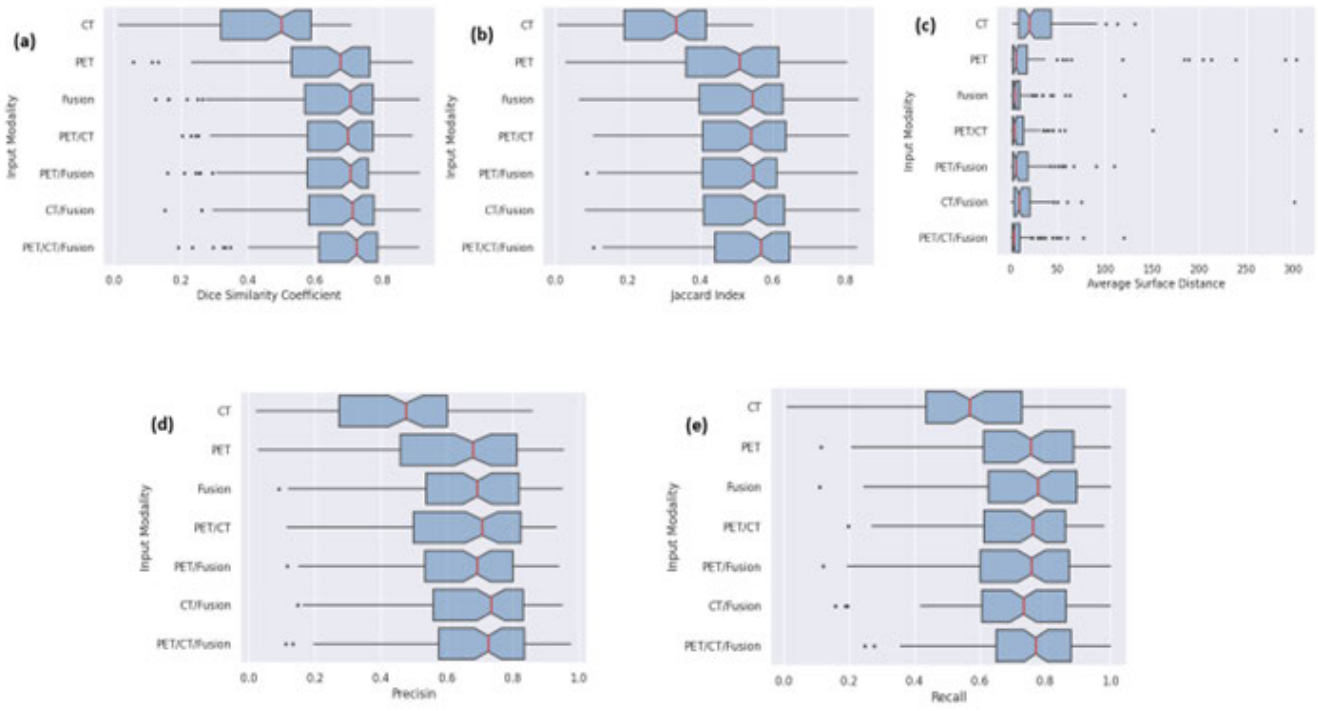

FIGURE 8. Comparative Performance Metrics for Tumor Segmentation Using Various Input Modalities. Boxplots represent the distribution of (a) Dice Similarity Coefficient, (b) Jaccard Index, (c) Average Surface Distance, (d) Precision, and (e) Recall across different imaging modalities, including CT, PET, and their fusion combinations. The central red line in each boxplot indicates the median value, while the box boundaries denote the interquartile range (IQR). Outliers are depicted as individual points outside the whiskers.

is critical for sparing healthy tissue while adequately targeting tumor masses.

2) Volumetric Assessment for Tumor Response: Monitoring how tumors respond to treatment over time requires precise measurement of tumor volume. Our segmentation approach provides a reliable method to quantify changes in tumor size and density, which are pivotal in assessing the efficacy of the treatment regimen.

V. CONCLUSION

To conclude, this study has systematically explored the influence of varying fusion weights on the accuracy of tumor segmentation within PET/CT imaging, leveraging a sophisticated Weighted Fusion Transformer Network. Our findings reveal that the integrative approach of PET/CT fusion, underpinned by the proposed Fusion FormerU-Net architecture, significantly enhances the precision of tumor delineation. The optimal fusion weight, which balances PET's metabolic detail with CT's anatomical clarity, has been identified through rigorous quantitative analysis using metrics such as the Dice Similarity Coefficient, Jaccard Index, Average Surface Distance, Precision, and Recall. The investigation demonstrates that a higher weighting towards PET data, specifically in the 80-90% range, yields superior segmentation performance for primary tumors. However, our study also highlights the nuanced complexity of lymph node segmentation, where a more balanced PET/CT weighting emerges as advantageous, underscoring the unique challenges presented by different tumor characteristics and locations. The variability in segmentation performance, indicated by the broader distribution of DSC scores at certain fusion weights, underscores the importance of personalized approaches in the fusion process, tailored to the specific imaging properties of individual tumors. The study, however, is not without its limitations. The CT/PET/Fusion strategy, while effective, may require further optimization to tailor its application to specific types of tumors or individual patient characteristics. Additionally, the computational requirements and technical complexities involved in effectively fusing CT and PET data warrant further investigation.

Based on our findings, adopting the CT/PET/Fusion strategy could revolutionize clinical workflows, offering enhanced decision-making capabilities for oncological treatment planning. Future research should aim to refine this segmentation approach further, exploring its applicability to a broader range of tumor types and clinical scenarios. Therefore, our future research direction also predicting the clinical H&N outcome based on our proposed segmentation approach.

DECLARATION OF COMPETING INTEREST

The authors indicate that they have no known conflicting financial interests or personal affiliations that may appear to have impacted the work revealed in this study.

DATA AVAILABILITY

The code used in this study is available at:
https://github.com/Alqushaibi/Weighted-Fusion-Transformer-for-Dual-PET-CT-Head-and-Neck-Tumor-Segmentation.

REFERENCES

[1] G. R. Bhat, R. G. Hyole, and J. Li, "Head and neck cancer: Current challenges and future perspectives," in *Advances in Cancer Research*, vol. 152. Amsterdam, The Netherlands: Elsevier, 2021, pp. 67–102.

[2] Q. Al-Tashi, M. B. Saad, A. Muneer, R. Qureshi, S. Mirjalili, A. Sheshadri, X. Le, N. I. Vokes, J. Zhang, and J. Wu, "Machine learning models for the identification of prognostic and predictive cancer biomarkers: A systematic review," *Int. J. Mol. Sci.*, vol. 24, no. 9, p. 7781, Apr. 2023.

[3] G. Sharp, K. D. Fritscher, V. Pekar, M. Peroni, N. Shusharina, H. Veeraraghavan, and J. Yang, "Vision 20/20: Perspectives on automated image segmentation for radiotherapy," *Med. Phys.*, vol. 41, no. 5, Apr. 2014, Art. no. 050902.

[4] C. Sjöberg, M. Lundmark, C. Granberg, S. Johansson, A. Ahnesjö, and A. Montelius, "Clinical evaluation of multi-atlas based segmentation of lymph node regions in head and neck and prostate cancer patients," *Radiat. Oncol.*, vol. 8, no. 1, pp. 1–7, Dec. 2013.

[5] S. Mercieca, J. S. A. Belderbos, and M. van Herk, "Challenges in the target volume definition of lung cancer radiotherapy," *Transl. Lung Cancer Res.*, vol. 10, no. 4, pp. 1983–1998, Apr. 2021.

[6] H. Bollen, S. Willems, M. Wegge, F. Maes, and S. Nuyts, "Benefits of automated gross tumor volume segmentation in head and neck cancer using multi-modality information," *Radiotherapy Oncol.*, vol. 182, May 2023, Art. no. 109574.

[7] N.-N. Zhong, H.-Q. Wang, X.-Y. Huang, Z.-Z. Li, L.-M. Cao, F.-Y. Huo, B. Liu, and L.-L. Bu, "Enhancing head and neck tumor management with artificial intelligence: Integration and perspectives," *Seminars Cancer Biol.*, vol. 95, pp. 52–74, Oct. 2023.

[8] M. Kosmin, J. Ledsam, B. Romera-Paredes, R. Mendes, S. Moinuddin, D. de Souza, L. Gunn, C. Kelly, C. O. Hughes, A. Karthikesalingam, C. Nutting, and R. A. Sharma, "Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer," *Radiotherapy Oncol.*, vol. 135, pp. 130–140, Jun. 2019.

[9] P. Bhatnagar, M. Subesinghe, C. Patel, R. Prestwich, and A. F. Scarsbrook, "Functional imaging for radiation treatment planning, response assessment, and adaptive therapy in head and neck cancer," *RadioGraphics*, vol. 33, no. 7, pp. 1909–1929, Nov. 2013.

[10] D. A. X. Schinagl, P. N. Span, F. J. A. van den Hoogen, M. A. W. Merkx, P. J. Slootweg, W. J. G. Oyen, and J. H. A. M. Kaanders, "Pathology-based validation of FDG PET segmentation tools for volume assessment of lymph node metastases from head and neck cancer," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 40, no. 12, pp. 1828–1835, Dec. 2013.

[11] G. Delouya, L. Igidbashian, A. Houle, M. Bélair, L. Boucher, C. Cohade, S. Beaulieu, É. J. Filion, G. Coulombe, M. Hinse, C. Martel, P. Després, and P. F. Nguyen-Tan, "18F-FDG-PET imaging in radiotherapy tumor volume delineation in treatment of head and neck cancer," *Radiotherapy Oncol.*, vol. 101, no. 3, pp. 362–368, Dec. 2011.

[12] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability*, vol. 13, no. 3, p. 1224, Jan. 2021.

[13] V. Andrearczyk, V. Oreiller, S. Boughdad, C. C. L. Rest, H. Elhalawani, M. Jreige, J. O. Prior, M. Vallières, D. Visvikis, M. Hatt, and A. Depeursinge, "Overview of the HECKTOR challenge at MICCAI 2021: Automatic head and neck tumor segmentation and outcome prediction in PET/CT images," in *Lecture Notes in Computer Science*. Cham, Switzerland: Springer, 2021, pp. 1–37.

[14] V. Andrearczyk, V. Oreiller, M. Hatt, and A. Depeursinge, *Head and Neck Tumor Segmentation and Outcome Prediction*. Cham, Switzerland: Springer, 2022.

[15] A. Alqushaibi, M. H. Hasan, S. J. Abdulkadir, K. U. Danyaro, M. G. Ragab, S. M. Al-Selwi, E. H. Sumiea, and H. Alhussian, "Enhanced colon cancer segmentation and image synthesis through advanced generative adversarial networks based-sine cosine algorithm," *IEEE Access*, vol. 1, pp. 1–18, 2024.

[16] V. Andrearczyk, V. Oreiller, M. Hatt, and A. Depeursinge, *Head and Neck Tumor Segmentation and Outcome Prediction: Third Challenge, HECKTOR*, vol. 13626. Cham, Switzerland: Springer, Sep. 22, 2022.

[17] W. L. Wong, "PET-CT for staging and detection of recurrence of head and neck cancer," *Seminars Nucl. Med.*, vol. 51, no. 1, pp. 13–25, Jan. 2021.

[18] J. J. Vaquero and P. Kinahan, "Positron emission tomography: Current challenges and opportunities for technological advances in clinical and preclinical imaging systems," *Annu. Rev. Biomed. Eng.*, vol. 17, no. 1, pp. 385–414, Dec. 2015.

[19] L. Bi, M. Fulham, N. Li, Q. Liu, S. Song, D. Dagan Feng, and J. Kim, "Recurrent feature fusion learning for multi-modality pet-ct tumor segmentation," *Comput. Methods Programs Biomed.*, vol. 203, May 2021, Art. no. 106043.

[20] Z. Diao, H. Jiang, X.-H. Han, Y.-D. Yao, and T. Shi, "EFNet: Evidence fusion network for tumor segmentation from PET-CT volumes," *Phys. Med. Biol.*, vol. 66, no. 20, Oct. 2021, Art. no. 205005.

[21] C. Lian, S. Ruan, T. Denœux, H. Li, and P. Vera, "Joint tumor segmentation in PET-CT images using co-clustering and fusion based on belief functions," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 755–766, Feb. 2019.

[22] Y. Suh, I. Amelio, T. Guerrero Urbano, and M. Tavassoli, "Clinical update on cancer: Molecular oncology of head and neck cancer," *Cell Death Disease*, vol. 5, no. 1, pp. e1018–e1018, Jan. 2014.

[23] W. Bi, A. Hosny, M. Schabath, M. Giger, N. Birkbak, A. Mehrtash, and H. Aerts, "Artificial intelligence in cancer imaging: Clinical challenges and applications," *CA, A cancer J. Clinicians*, vol. 69, no. 2, pp. 127–157, 2019.

[24] S. Vaz, J. Adam, R. D. Bolton, P. Vera, W. van Elmpt, K. Herrmann, and L. de Geus-Oei, "Joint eanm/snmmi/estro practice recommendations for the use of 2-[18 F] FDG pet/CT external beam radiation treatment planning in lung cancer v1. 0," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 1, pp. 1–21, Aug. 2022.

[25] R. Forghani, P. Savadjiev, A. Chatterjee, N. Muthukrishnan, C. Reinhold, and B. Forghani, "Radiomics and artificial intelligence for biomarker and prediction model development in oncology," *Comput. Struct. Biotechnol. J.*, vol. 17, pp. 995–1008, Aug. 2019.

[26] P. Nikulin, S. Zschaeck, J. Maus, P. Cegla, E. Lombardo, C. Furth, and J. van den Hoff, "A convolutional neural network with self-attention for fully automated metabolic tumor volume delineation of head and neck cancer in FDG pet/CT," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 1, pp. 1–16, Oct. 2023.

[27] P. Lambin, J. Zindler, B. Vanneste, L. Van De Voorde, D. Eekers, I. Compter, and S. Walsh, "Decision support systems for personalized and participative radiation oncology," *Adv. drug Del. Rev.*, vol. 109, pp. 131–153, Jun. 2017.

[28] L. G. Marcu, C. Boyd, and E. Bezak, "Feeding the data monster: Data science in head and neck cancer for personalized therapy," *J. Amer. College Radiol.*, vol. 16, no. 12, pp. 1695–1701, Dec. 2019.

[29] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Int. MICCAI Brainlesion Workshop*, 2018, pp. 311–320.

[30] X. Sun, C. An, and L. Wang, "A coarse-to-fine ensembling framework for head and neck tumor and lymph segmentation in CT and PET images," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 38–46.

[31] H. Jiang, J. Haimerl, X. Gu, and W. Lu, "A general web-based platform for automatic delineation of head and neck gross tumor volumes in PET/CT images," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 47–53.

[32] L. Rebaud, T. Escobar, F. Khalid, K. Girum, and I. Buvat, "Simplicity is all you need: Out-of-the-box nnUNet followed by binary-weighted radiomic model for segmentation and outcome prediction in head and neck PET/CT," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 121–134.

[33] Z. Salahuddin, Y. Chen, X. Zhong, N. M. Rad, H. C. Woodruff, and P. Lambin, "HNT-AI: An automatic segmentation framework for head and neck primary tumors and lymph nodes in FDG- PET/CT images," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 212–220.

[34] A. Wang, T. Bai, D. Nguyen, and S. Jiang, "Octree boundary transfiner: Efficient transformers for tumor segmentation refinement," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 54–60.

[35] K. Wang, Y. Li, M. Dohopolski, T. Peng, W. Lu, Y. Zhang, and J. Wang, "Recurrence-free survival prediction under the guidance of automatic gross tumor volume segmentation for head and neck cancers," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 144–153.
[36] A. Jain, J. Huang, Y. Ravipati, G. Cain, A. Boyd, Z. Ye, and B. H. Kann, "Head and neck primary tumor and lymph node auto-segmentation for PET/CT scans," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 61–69.
[37] J. Chen and A. L. Martel, "Head and neck tumor segmentation with 3D UNet and survival prediction with multiple instance neural network," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 221–229.
[38] M. Meng, L. Bi, D. Feng, and J. Kim, "Radiomics-enhanced deep multi-task learning for outcome prediction in head and neck cancer," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*. Cham, Switzerland: Springer, 2023, pp. 135–143.
[39] A. Gandhamal, S. Talbar, S. Gajre, A. F. M. Hani, and D. Kumar, "Local gray level S-curve transformation—A generalized contrast enhancement technique for medical images," *Comput. Biol. Med.*, vol. 83, pp. 120–133, Apr. 2017.

MOHAMMED A. MAHDI received the Ph.D. degree in computer science from the School of Computer Science, University Science Malaysia (USM), in 2016. He is currently an Expert in the field of routing protocols and mobile ad-hoc networks with ten years of research and academic experience. He is an Assistant Professor of computer science and the Deputy Director of quality and development of the College of Computer Science and Engineering, University of Ha'il, Saudi Arabia. He has contributed to publishing more than 20 research papers in journals and conference proceedings. He is participating in academic and institutional development activities at the college and university levels. His research interests include wireless networks, mobile networks, cryptography and network security, artificial intelligence, the IoT, and cloud computing.

SHAHANAWAJ AHAMAD (Member, IEEE) received the Ph.D. degree in computer science, specializing in software engineering from Jamia Millia Islamia Central University, New Delhi, India, and the D.Sc. degree in computer science and engineering, specializing in software engineering from Universidad Azteca, Chalco de Dí-az Covarrubias, Mexico. He is currently an Associate Professor with the Department of Software Engineering, College of Computer Science and Engineering, University of Ha'il, Saudi Arabia. He is an Active Academician and a Researcher in the field of software engineering and computer science with 20 years of experience. He served in several academic positions, such as the HoD, the Deputy Director, a Program Coordinator, and the Chair of research groups, college examinations, and academic scheduling and management in the universities of India and Saudi Arabia. He is involved in academic development and scholarly activities. He has published 80 research articles, 15 book chapters, four books, and 20 patents in his field of interest, such as software engineering, AI, the IoT, ML, legacy systems, software aging, service-oriented software engineering, and migration. He is a fellow of the British Computer Society, U.K., a member of ACM, a member of the Institution of Engineering and Technology, U.K., and a Life Member of the Computer Society of India. He is contributing as a member of editorial and reviewer boards in conferences and various computer journals published by Springer, IGI Global, Inderscience, and other universities.

SAWSAN A. SAAD received the B.Sc. degree in electrical engineering from the University of Khartoum, Sudan, in 2003, and the M.Sc. and Ph.D. degree in electrical, electronic and systems engineering from the National University of Malaysia (UKM), Malaysia, in 2008 and 2016, respectively. She was an Assistant Professor with the College of Communication Engineering, The Future University, Sudan. She is currently with the Department of Computer Engineering, College of Computer Science and Engineering, University of Ha'il, Saudi Arabia, in which she has taught many courses. Her current research interests include wireless communication engineering, mobile cellular communication, visible light communications (VLC), mobility management over mobile networks, the Internet of Things (IoT) Technology, artificial intelligence, machine learning, and deep learning technologies in wireless communication networks.

ALAA DAFHALLA received the B.Sc. degree in computer engineering from the Faculty of Engineering and Technology, University of Gezira, Sudan, the M.Sc. degree in computer engineering and networks from the University of Gezira, and the Ph.D. degree in computer engineering from Universiti Malaysia Perlis, in December 2018, with a focus on "Bio-Inspired chameleon technique in MAC protocol for energy-efficient wireless networks." Her research interests include wireless network protocols development, modeling and simulation, software engineering, network programming, and optimization methodologies with robust engineering, programming with open sources that involve network modeling and simulation and nanotechnology in networks.

RIZWAN QURESHI (Senior Member, IEEE) received the M.S. degree from the Institute of Space Technology, Islamabad, Pakistan, in 2015, and the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2021. He was an Assistant Professor with the Fast National University of Computer and Emerging Sciences, Karachi, Pakistan, and a Lecturer with COMSATS University Islamabad, Wah Campus, Pakistan. He was a Researcher with the College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar, a Postdoctoral Research Fellow with the MD Anderson Cancer Research Center, The University of Texas at Austin, Houston, TX, USA, and the AI-Team Lead of the Chinese Academy of Sciences, Hong Kong. He is currently with the Center for Research in Computer Vision (CRCV), University of Central Florida, Orlando, FL, USA. He has published at top venues, including, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, *The Lancet Digital Health*, *Scientific Reports* (Nature), *Expert Systems with Applications*, *Seminars in Cancer Biology*, CVPR, ACCV, and IEEE BIBM. His research interests include AI applications in life sciences, cancer data sciences, computer vision, and machine learning. He was nominated for USERN Prize, in 2024.

ALAWI ALQUSHAIBI received the B.Sc. degree in computer networks and security from Universiti Teknologi Malaysia, in 2012, and the master's degree in research from Universiti Teknologi Petronas, in 2021. He is currently an Academic Researcher. During his academic journey, he has acquired knowledge and skills in conducting independent research, producing academic writing, and teaching computer science courses. His research interests include machine learning, data science, optimization, feature selection, classification, data analytics, and image processing.