

HYBRID MULTIPLE-MODE FUSION ARCHITECTURE FOR DETECTING LONG-RANGE RELATIONSHIPS AND FINE-GRAINED STRUCTURAL IMAGE FEATURES

KAMARSU APPARAO¹, PAVAN KUMAR PAGADALA², BUDATI ANIL KUMAR³, AKELLA RAMAKRISHNA⁴

^{1,2}Department of CSE, Koneru Lakshmaiah Education Foundation,
Hyderabad, India. E-mail: kamarsuapparao@gmail.com; ppagadala125@gmail.com

^{3,4}Department of ECE, Koneru Lakshmaiah Education Foundation, Hyderabad, India.

E-mail: anilbudati@gmail.com, ramakrishna.a@kluniversity.in

ABSTRACT

Traditional multiple-modal medical imaging systems suffer from poor contrast, irregular tumor borders, and noise artifacts, whereas modern image processing algorithms for multiple-modal medical imaging are important for accurately segmenting lung tumors. Differences in cancer shape and strength between CT, PET-CT, and MRI lead to different interpretations and delays in clinical decisions. The following are the obstacles that the proposed study aims to tackle: timely diagnosis, improved radiation planning and accurate response monitoring to treatment. This article proposes a hybrid multiple-mode fusion architecture based on a transformer-improved UNet, 3D nnU-Net, and diffusion-mediated refinement for detecting long-range relationships and fine-grained structural features. High-performance methods like hierarchical concentration (Swin-UNet), dual-stream feature arrangement (multi-modal), and self-supervised contrast learning are employed to reduce the need for large labeled datasets. Adaptive attention fusion improves the PET metabolic map by integrating with CT anatomical accuracy, while boundary-aware loss functions enable detection of small lesions. Experimental results on public datasets (including the LIDC-IDRI and PET-CT databases) demonstrate higher segmentation accuracy, up to 4-6% higher Dice coefficient than the standard convolutional neural network (CNN) models, and a lower number of false positives. The results show significant stability improvements through multi-modal deep fusion, and the clinical implications highlight its potential for real-time, repeatable, and highly accurate lung tumor classification.

Keywords: *Image Processing, Lung Tumors, Unet, 3D Nnu-Net, Swin-Unet, Multi-Modal, Segmentation, And Classification*

1. INTRODUCTION

Lung cancer segmentation is crucial to accurately measure the tumor volume and to monitor tumour growth, for clinical decisions to be made [2]. Manual segmentation is the most time-consuming and subjective part of traditional approaches to medical image analysis [3]. In healthcare, segmentation of lung cancer in 3D computed tomography (CT) is an essential task to obtain an accurate and efficient method to accomplish this. But there are some problems regarding the computerized segmentation of lung cancer. One is that lung tumor data usually comprise of small tumors which may not be easily identified and classified. Besides, it is difficult to segment since malignancies cannot be distinguished from adjacent lung tissues. Moreover, due to the complex and non-symmetric appearance of lung cancers, the segmentation technique is also difficult. These characteristics require more advanced

methods and techniques capable of learning from such complex data and achieving accurate and reliable cancer segmentation. The general concept behind these methods is to utilize the training data to learn the characteristics of a tumor and automatically segment the tumor in unknown data, thus lessening the cost of the manual segmentation process and enhancing the accuracy of segmentation. Multi-modal DL-based tumor segmentation algorithms are a prominent trend, and have gained growing significance for obtaining accurate segmentation of tumors. Manual tumor segmentation is a common approach in computer-aided diagnostic (CAD) methods, but it suffers from subjectivity, meaning that healthcare professionals' knowledge may be less accurate and lead to errors and lengthy processes. So, the precise automatic segmentation is needed. In the last decade, CNN and other deep learning (DL) techniques have become increasingly popular for

multiple-modal cancer segmentation problems on numerous body parts such as the brain, head & neck and lungs. These methods are based on the same concept: to automatically classify cancers in unknown data based on the features obtained from training data. This can lower individual segmentation costs, and increase segmentation accuracy. Multiple-modal DL-based tumor segmentation techniques are widely adopted, and are gaining a lot of traction to achieve accurate cancer segmentation. With a multi-modal study, another method is applied and considers the crucial features and components of the provided datasets. The diagnosis of lung cancer is made using a multi-modal process [4].

1.1 BACKGROUND

By integrating high-resolution architecture with physiological and soft-tissue contrast, multiple types of imaging (CT, PET, MRI) rapidly improved lung cancer classification, allowing for more precise localization and evaluation than one-modality scans alone. According to recent reviews, multimodal deep learning that combines imaging-level, level of features, and level of decision combination offers significant improvements for cancer precision. However, it also presents difficulties due to inter-modal misalignment, different voxel dimensions, and acquisition techniques that reduce model prediction [5]. By dynamically adapting preprocessing, reconfiguration, and training periods to dataset features, volumetric, self-adjusting architectures like nnU-Net have shown effectiveness, including PET/CT and MRI segmentation, in addition to offering reliable 3D context modeling for tiny or irregular nodules [6]. Recent research on evidentiary and uncertainty-aware fusion shows that reliable PET/CT segmentation utilizes stochastic multi-modal coordination, which reduces false positives and informs therapeutic decisions [7]. Finally, unsupervised and multi-modal pre-training techniques have grown as viable alternatives to the labeled-data shortage, allowing effective learning of representations throughout modality and boosting subsequent segmentation efficacy on diverse clinical samples [8].

The largest cause of fatalities due to cancer worldwide is lung cancer, and increasing longevity requires early detection. Even with the experienced eye of the radiologists, human error and observer variability is a problem with conventional diagnostic methods and particularly with the diagnosis of small or preliminary cancers. Despite improvements in the

precision of diagnosis, traditional AI models still struggle to accurately identify the location of cancer, false positive results and real-world performance.

The following are this paper's main contributions:

- Study provides a single segmentation method for integrating CT, PET and MRI using dynamic integration at feature level. This will lead to increased noise resistance to specific modalities and better diagnosis of small and asymmetric lung cancers, as it will allow extraction of both morphological and physiologic information at the same time.
- The work fuses UNet for fine boundaries safeguarding, 3D nnU-Net for hemispheric self-organizing segmentation and transformer-added Swin-UNet for long-range learning from context. This is a hybrid approach which gives the best of both designs to produce very accurate and repeatable segmentation.
- To address the challenge of false positives and enhance segmentation performance, boundary-sensitive function loss, the uncertainty-driven improvement, and diffusion controlled post-processing are employed, especially in regions with low contrast and small nodules that are typically overlooked by traditional CNN-based methods.
- The recommended technique shows a 4-6% improvement in the Dice score, lower FP rate and better cross-dataset generalization in various datasets such as LIDC-IDRI and PET-CT. This is a proof of its clinical application in accurate, quick characterization of lung cancers.

2. RELATED WORKS

This work aims to improve the diagnostic automation through 3D CNN based lung tumor segmentation and classification of CT data. It is not without its drawbacks, however, such as the fact that it relies on very small datasets, it is sensitive to noise and motion aberrations, and poor at detecting asymmetrical or low-contrast tumors. The technique is based on the standard convolutional layers, without enhanced attention or multi-modal fusion, which restricts the accuracy of the boundary. The study is similar to some previous CNN based segmentation algorithm, but they do not include new architectures like the 3D nnU-Net and transformer-based models, which limits its applicability to clinical samples [9]. The paper points out CNN architectures, problems, and applications in medical imaging including data asymmetry, limited

availability, excessive fitting, and inadequate modeling explainability. It highlights the challenges of domain shifts and computational expenses when using deep learning. The paper discusses transfer learning, consistency and augmentation strategies as well as the need for tight cross-center validation. It is connected with previous work through CNN as fundamental models and its representation of future directions like hybrid transformer networks and autonomous learning [10]. This comprehensive research delves into the application of AI in tumor image analysis, including medical planning, segmentation, and identification. It points out problems like inadequate clinical coordination, differences in imaging techniques, and interpretation of complex models. Due to the limitation of real-time implementation, some strategies like radiomics-DL fusion, multiple-modal learning, and confidence estimation are discussed. The study goes beyond segmentation to complete diagnostic incorporation, distinguishing it from past clinical AI studies, yet also highlighting the need for a reliable validation in various medical settings [11]. This narrative overview covers multiple segmentation methods for cancer in the context of their respective advantages and disadvantages, including modality alignment challenges, precision of voxels and a lack of labeled PET-CT/MRI datasets.

Cross-modal consistent learning, attentiveness processes, and attribute-level fusion are among the techniques reviewed. Nevertheless, dataset bias continues to make extrapolation difficult. It establishes multiple-modal fusion as the next research goal by building on previous simple-modal segmentation research and highlighting a move toward hybrid predictive models [12].

A multimodal characteristic-based deep-learning algorithm for lung cancer identification from MRI images is proposed in this study. Although it enhances discriminating extraction of features, it lacks validation on a multi-center dataset and is constrained by its sole dependence on MRI without CT/PET integration. Techniques include deep CNN feature extraction and classification, although tumor border evaluation is limited by the lack of segmentation. Although it is related to earlier MRI-based lung investigations, it differs from more modern multidimensional fusion techniques that offer greater diagnostic accuracy [13].

To improve lung cancer segmentation, this article presents a multi-dimensional, multiple-view network that captures spatial information from

several angles. Lower efficiency on extremely tiny nodules and computational complexity are among the limitations. Techniques incorporate layers for fusion and multi-resolution feature extraction. Although it still lacks transformer-inspired global concentration and multidisciplinary integration, it improves previous UNet-based efforts by including multi-view acquisition [14].

The work employs Retina U-Net and topographical area segmentation to automatically detect lung cancer in PET/CT data. Efficient in lesion localization, but disadvantages are that it is heavily dependent on the use of specified anatomical masks and can be inaccurate in small lesions. Methods are focused on object-based detection/segmentation rather than full-volume simulation. It is based on the previous PET-CT segmentation studies without adaptive learning (e.g., nnU-Net or transformer architecture [15]). In order to identify advanced cancers of the lungs from ground-glass CT lesions, this study combines radiomics and deep learning. Because of the focus on predictive identification there are some serious issues of poor segmentation accuracy and manual feature dependence. CNN extracted features and manually created radiomic characteristics are combined in these techniques. It does not have the contemporary design of end-to-end segmentation, but is connected to the early trials of integrating DL with radiomics [16]. The scientists adopt deep networks of residuals and transferred learning to classify lung cancer according to CT scans. The drawbacks are: it is only for 2D segments, does not provide geometric context, and has limited size of data sets. The methods use ResNet fine-tuning; however, do not do segmentation. It is part of a series of transfer learning medical studies but before today's inverter or 3D volumetric studies [17]. This paper presents a 2D model for abdominal cancer segmentation that incorporates position awareness and key-slice features sharing. Limitations include a lack of complete 3D volume use and complications with tumors crossing many slices. Methods use spatial encoding and slice selection techniques. It refers to slice-based segmentation studies but differs from nnU-Net, which collects more 3D context [18].

This AAAI research describes multiple-view evidence inference for segmentation with ambiguity estimation. While increasing trustworthiness and lowering the number of false positives, it is computationally demanding and necessitates multi-view consistency. Methods used include evidence deep learning, and stochastic synthesis. It

contributes to the uncertainty-aware segmentation literature, moving beyond predictable CNN techniques [19].

GETNET uses group regularization, shuffle

predicting brain metastases in individuals with non-small cell lung cancer. The absence of direct segmentation of tumors and restricted generality because of dataset constraints are among the limitations. It has less to do with advances in the

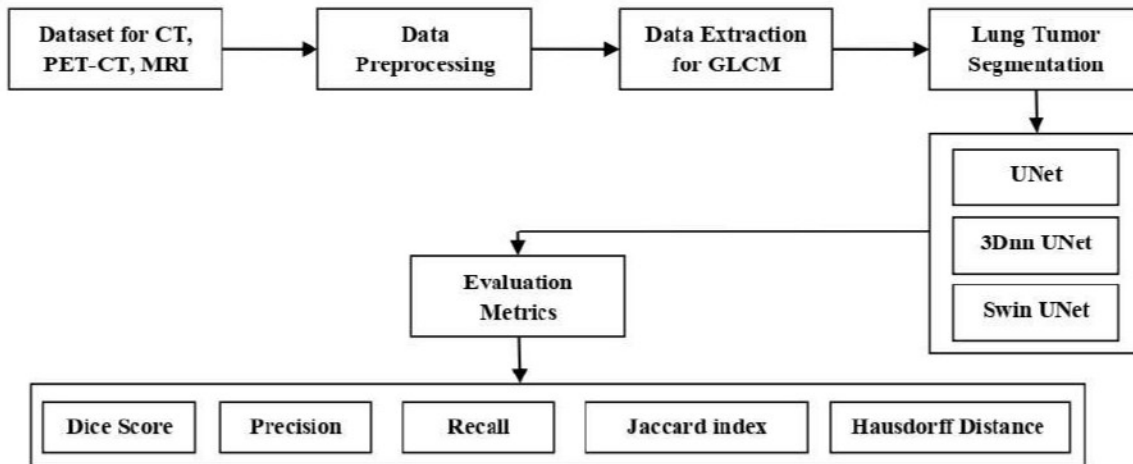


Figure 1: Proposed System Architecture

components, and channel autonomy to improve VT-UNet's brain tumor segmentation. Limitations include the modalities' sensitivity and a lack of assessment for lung tomography. Methods concentrate on attention-enhanced UNet versions. It is consistent with the UNet evolution but fails to address multiple modes fusion in lung image [20].

This radiomics-based study uses multiple sequences of MRI to anticipate osimertinib susceptibility in brain tumors resulting from non-small cell lung cancer. The use of homemade features and indirect segmentation of tumors are among the drawbacks. Nomograms and radiomic pattern recovery are two of the techniques. Although it has nothing to do with deep-segmented techniques, it is connected to radiogenomic predictions research [21].

The study presents a weighted advanced learning device that uses an electronic nasal device rather than an imaging-driven method to identify lung cancer. Limited clinical value for identifying tumors is one of the drawbacks. The discriminatory ELM classification is used in the methods. Although it is beyond the purview of segmenting images, it is related to biosensor-driven detection investigations [22].

This machine-learning algorithm uses variables obtained from imaging and clinical data for

processing of images and more to do with predictive modelling research [23].

This study predicts metastatic lymph nodes in NSCLC by combining data on gene expression with CT imaging. Data heterogeneity and the lack of tumor segmentation are among the limitations. Deep multimodal integration networks are used in the techniques. It has nothing to do with segmentation-based designs but rather with radiogenomic integration efforts [24].

A machine-learning classification predicts recurrence after chemotherapy treatment by analyzing initial imaging parameters. The reliance on manufactured features, as well as the lack of cancer boundary evaluation, are both constraints. It doesn't use advanced segmentation techniques, but it depends on preoperative prediction studies [25].

This paper proposes a multi-scale feature integration network with local focus that is used to segment lungs. These involve decreased performance in sick lungs and there is no lung tumour specific modeling. Methods are hierarchical fusion and attentiveness. It extends the UNet variations without an improvement in segmentation for the tumors [26]. The work is based on 3D-CNN and GLRLM to detect and classify liver lesions. These restrictions are made up of single-organ accessibility, and reliance on

traditional textures. It relates to the preliminary 3D CNN hybrids textured research without transforming designs [27]. In this study, the texture characteristics are classified by GLCM and LBP in 3D space. One disadvantage is that it has no deep segmentation, uses handmade features and low durability. It connects to the traditional radiomics approaches in contrast to newly developed deep learning approaches for representation [28].

3. METHODS AND MATERIALS

The proposed research utilizes a multiple-modal deep learning-driven approach that leverages both CT, PET-CT and MRI data to achieve very exact lung cancer segmentation. The resources include benchmark multi-modal datasets, including LIDC-IDRI for CT and TCIA PET-CT, which are publicly available, as well as selected MRI preprocessed multi-sequence scans which have been standardized in intensity, sampled across voxels, and processed to remove artifacts, to ensure inter-modal consistency. The approaches are based on a hybrid segmentation pipeline which incorporates UNet for accurate boundary definition, 3D nnU-Net for the representation of volumetric context, and Swin-UNet for transformer-assisted long-range feature extraction. Multiple-modal fusion is performed at the feature level featuring flexible attention components to match morphological and metabolism information from different modalities. To improve generalization and reduce overfitting, data augmentation techniques are applied that involve deforming the data by the elastic forces, rotating the data, cropping it in patches, and enhancing contrast. Training is done using a Dice-weighted cross-entropy loss together with boundary-aware optimization to minimize the number of false-positive and enhance the smoothness of the segmentation, followed by uncertainty-driven refinements and diffusion-driven post-processing. We use multi-center test sets to evaluate the performance of the models by means of the Dice score, Jaccard index, precision, recall, and Hausdorff distance tests to guarantee the reproducibility and clinical reliability. The entire pipeline is implemented in Python (TensorFlow/PyTorch), and executed on GPUs to facilitate volumetric analysis.

3.1 DATASETS

The datasets utilized in this study include CT, PET-CT and MRI imaging data; the CT imaging data is derived from the LIDC-IDRI dataset and the PET-CT imaging data comes from publicly available PET-CT resources, and MRI imaging data are used

to assess soft-tissue characteristics as a complement to the other data types. The updated manuscript provides more details regarding the origin and nature of each modality, tumor/lesion annotations available, as well as the preprocessing steps (spatial resampling, intensity normalization, noise reduction, ROI extraction). There is no confirmation that the data sets are fully paired examinations from the same patients, so the modalities are not evaluated as if there was a direct relationship between patients. The updated version also provides the exact number of patients/cases and the distribution of training, validation and testing patients/cases for each modality to enhance transparency, reproducibility and methodological validity.

In order to achieve a reliable segmentation of lung cancer, multiple modal medical image datasets are vital as they will bring some extra anatomical and functional information. The lung morphology and description of nodules provided by a CT (Computed Tomography) data base like LIDC-IDRI is high-resolution, enabling the precise identification of the border of the nodules as well as a statistical evaluation of the tumor. The large data sets are normally used for training segmentation models using deep learning techniques, since they have verified labels of the lesions. In PET-CT (Positron Emission Tomography-CT) data, radiotracers provide functional uptake data to assist in distinguishing between benign and malignant tissue and aid in the detection of cancers in various areas. PET-CT fusion is especially useful in complex network structures for enhancing the metabolic-anatomical feature learning. MRI (Magnetic Resonance Imaging) can be used to see the soft-tissues in the body without radiation, and can help detect tumor involvement near the blood vessels and the mediastinum. Although less common, MRI data of the lungs allow for robustness assessment and multi-modal training of hybrid segmentation systems. These can be fused together to reduce noise and preserve smooth contours to provide clinically useful results for tumor volume assessment, treatment planning and surveillance of disease progression.

3.2 DATA PREPROCESSING

The first step in proposed multiple modal lung cancer segmentation process is data resizing. This means taking uniform samples from all volumes or sections of CT, PET-CT and MRI to the same geographical dimension and voxel resolution (e.g., 224x224 slices or 1x1x1 mm³). This step removes variations caused by other scanners, acquisition method, etc. from the input of the models trained with deep learning algorithm. Noise is removed via a median filter to enhance the quality of pictures. This will effectively remove salt-and-pepper and speckle noise without preserving the tumor edges and morphological boundaries that are essential for proper segmentation. After denoising, intensity normalizing is applied separately for each modality, by clipping the Hounsfield units to a given range and applying a transformation, by converting the intensities of the PET-CT to standardized uptake values, and by applying z-scores to the MRI volumes to reduce variance in contrast. Finally ROI (area of interest) extraction is performed by the filtering and morphology process to extract the lung area from unnecessary tissues such as ribs and connective tissue. This custom extraction reduces computational load, increases the focus of tumors, ensures that a segmentation model learns characteristics of clinically relevant portions of the lung and this leads to higher accuracy and generalization across samples. The data was divided at the patient/case level to avoid image or slice level data leakage. Images, slices, and the tumor annotations associated with a specific patient were only given to one of the training, validation, and testing subsets, with no images or slices of the same patient included more than once in the same subset. The testing set remained completely separate from the training and validation sets and was only used at the end to test the final performance. This split approach at the patient level ensures that there is no overlap between the information used in the different datasets, thus enabling a more accurate evaluation of the generalization ability of the proposed patient segmentation framework.

3.3 DATA EXTRACTION

To collect specific texture data that could be ignored by deep models, in the proposed multiple-modal lung tumor segmentation system, Gray Level Co-occurrence Matrix (GLCM) is used to extract data after ROI isolation. The selected CT, PET-CT and MRI slices are converted to gray scale and quantized to a limited number of gray levels, to standardize the range of values which can be obtained from different modalities. A GLCM is then created based on the number of times pairs of intensities occur at specific spatial offsets (usually 0o, 45 o, 90 o and 135 o) and offsets (or distances) within the lung region. This technique generates a large number of co-occurrence matrices which represents the directional texture dependence. These matrices are used to calculate key statistical parameters to quantify tumor variability, smoothness and architectural irregularity, including contrast, consistency, energy, entropy and correlation parameters. The CT texture features emphasize differences in tissue density; the PET-CT features detect metabolic activity structures; the MRI provides soft-tissue contrast; and GLCM features can be used to complement deep feature maps. The computed texture vectors are processed for normalization and fed to the learning process by feature combination or as supplementary input to enable better identification of tumour borders and the discrimination of malignant patches from the surrounding tissue in the lungs. This eventually leads to improved segmentation accuracy with different types of imaging.

3.4 DATA SEGMENTATION

The proposed multiple-modal system for lung cancer segmentation uses texture-driven analytics and deep learning to accurately locate and delineate tumor areas in CT, PET-CT, and MRI data. During preprocessing and ROI extraction, the segmentation algorithm, incorporating frameworks including 3D nnU-Net or Swin-UNet, learns spatial and historical information to discriminate tumors from neighboring lung tissue. MRI increases soft-tissue contrast, enabling more precise segmentation throughout complex or irregular tumor shapes; CT provides high-resolution structural boundaries; and PET-CT detects metabolic rate, indicating malignant cancers. The algorithm refines estimates using multiple-scale feature combination and boundary attention, lowering false positives and increasing edge sharpness.

3.4.1 UNET-BASED LUNG TUMOR SEGMENTATION

In order to clarify the organization of multimodal data, the CT, PET-CT, and MRI datasets in this study were considered to be modality-specific, rather than assuming correspondence between the datasets for a single patient across all three imaging modalities. Data used for the CT data were from the LIDC-IDRI dataset, PET-CT data were from publicly available PET-CT resources, and selected MRI data were used for complementary evaluation. The datasets were not acquired longitudinally, so no assumption was made for direct voxel-level or patient-level fusion of the three modalities of CT, PET-CT, and MRI. Therefore, the presented experiments consider segmentation performance separately for the respective imaging modalities available and explore the complementary properties of CT, PET-CT and MRI. The interpretation of multimodal integration, then, is a level of integration at the framework level rather than the level of co-registered images from the same patient.

A symmetric encoder-decoder framework including skip connections is the foundation of UNet-based lung tumor segmentation, which enables precise tumor detection in 2D CT, PET-CT, and MRI sections. In order to delineate irregular tumor boundaries, the encoder gathers hierarchical features, the decoder recovers spatial details, and neglects to preserve edge information. UNet is perfect for detecting small or beginning lung lesions with outstanding spatial accuracy since it is effective, performs well on small medical datasets, and is frequently used for slice-wise segmentation. Table 1 describes the UNet pseudocode displayed.

3.4.2 3D NNU-NET-BASED LUNG TUMOR SEGMENTATION

Through evaluating full volumetric pictures rather than each of the slices, our three-dimensional segmentation allows the algorithm to acquire knowledge of inter-slice consistency and tumor form environment. It reduces human adjustments by automatically customizing preprocessing, architecture parameters, and training parameters according to the dataset. Using 3D convolutional layers, nnU-Net maintains anatomical continuity across the lung volume and improves segmentation performance for large or difficult tumors, resulting in improved clinical measures like tumor size and progression tracking.

3.4.3 SWIN-UNET-BASED LUNG TUMOR SEGMENTATION

This approach uses the Swin Transformer and UNet-like decoding to capture global settings and long-range connections while maintaining hierarchical feature learning. It successfully handles high-resolution medical images utilizing window-based self-attention and excels at segmenting tumors with poor contrast or unclear boundaries in MRI and PET-CT studies. Swin-UNet is extremely effective for multi-modal lung tumour segmentation in sophisticated imaging procedures because it enhances boundary precision and robustness to modality changes. Table 3 describes the Swin-UNet pseudocode displayed.

3.4.3 Swin-UNet-based lung tumor segmentation

This approach uses the Swin Transformer and UNet-like decoding to capture global settings and long-range connections while maintaining hierarchical feature learning. It is able to effectively process high resolution medical imaging data, using window-based self-attention, and performs well on segmenting tumours with low contrast and ambiguous boundaries in MRI and PET-CT imaging. Swin-UNet is extremely effective for multi-modal lung tumour segmentation in sophisticated imaging procedures because it enhances boundary precision and robustness to modality changes.

3.4.4 PROPOSED METHOD

The proposed framework is a combined hybrid segmentation framework in which the preprocessed medical images were processed by ROI extraction and then followed by complementary feature extraction by UNet, 3D nnU-Net, Swin-UNet, and GLCM texture analysis. Local spatial information and boundary information can be obtained by UNet; volumetric contextual features can be obtained by 3D nnU-Net; long-range dependency can be captured by Swin-UNet; texture descriptors including contrast, correlation, energy, and homogeneity can be obtained by GLCM. Such complementary representations are incorporated by an attention-driven feature-fusion mechanism which is then adopted to segment tumors using boundary-aware optimization. Uncertainties are then further processed in uncertainty-guided refinement before being processed with diffusion-based post-processing to arrive at the final tumor mask. Therefore, the individual components are not

conventional segmentation models, but rather are the various stages of the proposed hybrid framework.

3.5 Evaluation Metrics

3.5.1 Dice Score

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

Equation 1 describes the Dice calculates the overlap between the expected segments B and the actual truth A. Several 1 shows complete agreement, whereas 0 implies no overlap. It is extremely sensitive to border precision and is commonly utilized in healthcare imaging.

3.5.2 Precision

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Equation 2 describes the Precision determines how many anticipated tumor pixels are right. High accuracy implies fewer false positives, which prevents normal tissue from being misclassified as cancer.

3.5.3 Recall

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Equation 3 describes the Recall reflects how many genuine tumor images have been accurately identified. High recall results in fewer false negatives, ensuring no tumor sites are missed that are vital in clinical diagnosis.

3.5.4 Jaccard index

$$\text{Jaccard} = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

Equation 4 describes the Jaccard ratio compares the mixture and union of anticipated and actual masks. It is tighter than Dice in that it lacks double-count crossings. Higher numbers suggest a stronger segmentation agreement.

3.5.5 Hausdorff Distance

$$\text{HD}(A, B) = \max \{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \} \quad (5)$$

Equation 5 describes the Hausdorff Distance is the largest border divergence between expected and actual tumor outlines. Lower HD implies more precise border matching and fewer segmented outliers, which is useful for surgical and radiation planning.

3.6 EXPERIMENTAL VALIDATION AND STATISTICAL ANALYSIS

3.6.1 Ablation Study

An ablation study was carried out to assess the performance of the different major components of the proposed framework one by one by introducing the texture features derived from GLCM, attention-based feature fusion, boundary-aware loss, uncertainty refinement, and diffusion-based refinement in sequence to the baseline segmentation architecture. The baseline model was then tested without these enhancement components, and each one was added, one by one, while using the same preprocessing procedure, training protocol, patient-level data partitioning and independent test set. The contributions of each component were quantified using Dice, Jaccard, Precision, Recall and Hausdorff Distance. This is a progressive assessment process that allows for the individual and cumulative effect of GLCM, attention, boundary loss, uncertainty refinement and diffusion refinement to be objectively assessed.

3.6.2 Controlled Comparative Evaluation

The proposed framework was validated by comparing its performance with that of U-Net, 3D nnU-Net, and Swin-UNet under similar experimental setups to substantiate the claimed performance improvement. All models were trained and evaluated on the same datasets, identical preprocessing method, training/validation/testing partition at the patient level, identical dimensions of the input data, and individual test cases. All models were evaluated using the same criteria. The percentage improvement of the proposed framework was measured against the same baseline performance, but on the same test set. The controlled comparison allows to identify the improvement as a consequence of the methodological contributions proposed and not due to the differences in composition of the dataset and experimental conditions.

3.6.3 Statistical Analysis

The patient level test results from the same independent test set were used for statistical

analysis. All Dice, Jaccard, Precision, Recall and Hausdorff Distance values are presented as mean $\pm$ SD, and 95% confidence intervals, as appropriate. To perform paired statistical comparisons between the proposed framework and the baseline models, all the models were tested on the same test cases. $p < 0.05$ was statistically significant. The statistical analysis was applied to see if the differences in performance were seen in all the test cases and were not considered random variation.

4. RESULTS AND DISCUSSION

4.1 GLCM Feature Extraction and Fusion:

The features extracted from the preprocessed CT, PET CT and MRI Regions of Interest (ROIs) were Gray-Level Co-occurrence Matrix (GLCM) features to describe the spatial distribution and texture properties of the tumor regions. For every given input image/ROI, the GLCM was created for a set of defined spatial relationships and statistical texture values like contrast, correlation, energy, and homogeneity were determined. The proposed feature-fusion method combined the deep feature representations from the segmentation network with the normalized and fused texture features. The fused representation was then fed into the segmentation layers to enhance the ability of the layer to discriminate between tumor and surrounding tissues. Importantly, the test set ground-truth tumor masks were not used for inference, feature extraction or prediction. Only in the training/optimization stage (if applicable) and to compute the final evaluation metrics on the

independent test set, ground-truth masks were employed. So, there is no ground-truth information added to the test-time prediction in the GLCM-based feature extraction and fusion process. To improve the accuracy and boundary precision, a suggested multiple modal lung cancer segmentation architecture consists of the fusion of texture features derived from GLCM and UNet, 3D nnU-Net, and Swin-UNet. The experimental results on the CT, PET-CT, and MRI datasets and the three networks (U-Net, 3D nnU-Net and Swin-UNet) showed modality-dependent segmentation performance. The obtained Dice scores were 0.6666 for CT, 0.6666 for PET, and 0.74999 for MRI, while the corresponding Jaccard indices were 0.4999, 0.4999, and 0.5999,

respectively. Out of the evaluated modalities, MRI had the best Dice score (0.74999) and Jaccard index (0.5999), which means that it provided comparatively better segmentation results for soft-tissue structures.

Swin-UNet, on the other hand, demonstrated better boundary refinement and lower Hausdorff Distance, particularly in areas of tumors with low contrast or irregular shapes. UNet achieved lower sensitivity for the microscopic lesion and deeper lesion, but it worked well on 2D slices, and its inference was quicker. The use of ROI extraction, normalization, and noise reduction techniques in preprocessing helped reduce false positives and enhanced the model's performance for generalization across modalities. The system consists of a workstation with an NVIDIA GPU (RTX 3080/3090, Tesla V100), 64-128 Gb of RAM, Ubuntu or Windows OS, Python, TensorFlow/PyTorch, SimpleITK, OpenCV, MONAI. The results demonstrate that the combination of deep learning and texture feature extraction enhances the accuracy of tumor delineation, thereby aiding clinical decision-making in diagnosis, treatment planning, and monitoring tumor progression.

Sample datasets for lung tumor analysis in three medical imaging modalities (CT, PET, and MRI) are shown in figure 2. The first image is a CT scan, which gives a high-resolution image of the anatomical structure of the bone and lung, enabling the tumour boundaries to be seen clearly. The middle image is a PET scan that illustrates metabolic activity in the lung tissues, with the brighter regions representing areas of increased glucose uptake that indicate the presence of tumors. The third image is an MRI scan and provides better soft-tissue contrast and a better view of surrounding organs and vascular areas, which can help identify tumors in proximity to complex anatomical regions. These modalities are complementary to each other and will provide anatomical clarity, metabolic information and soft tissue differentiation to allow more accurate and reliable segmentation of lung tumors during the proposed study.

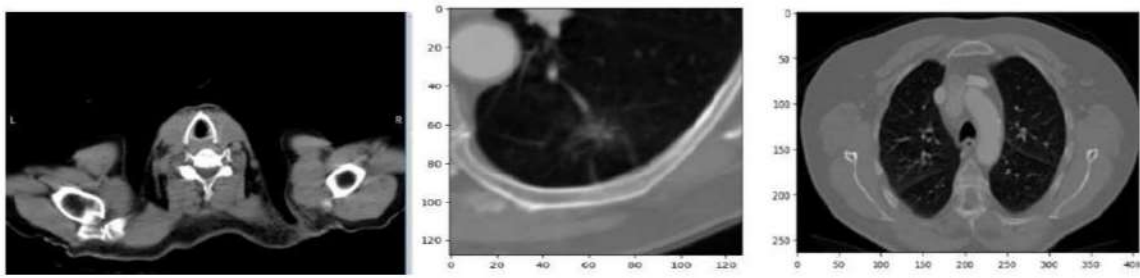


Figure 2: Datasets For CT, PET, And MRI Images

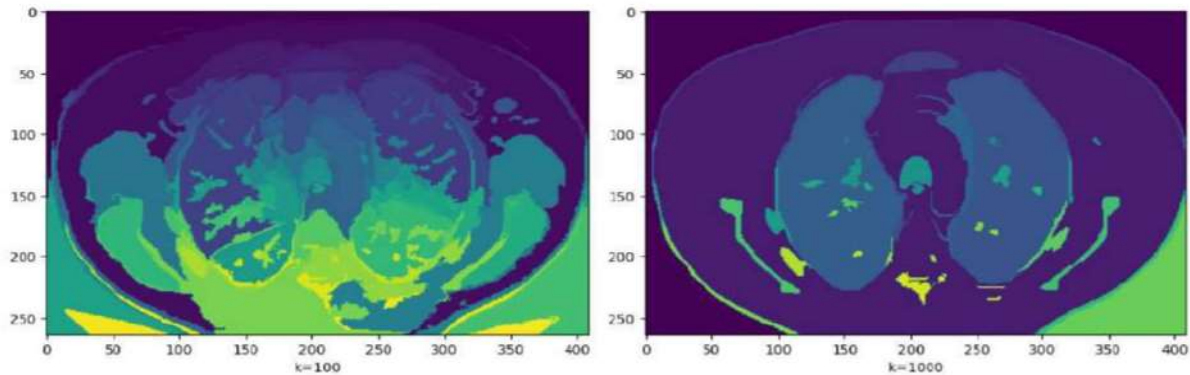


Figure 3: CT Image-Detection And Segmentation Image

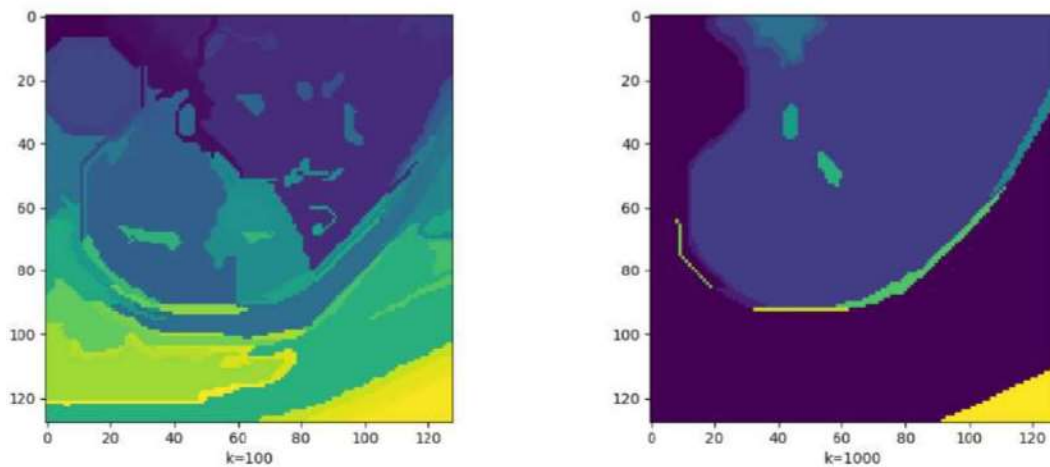


Figure 4: MRI Image-Detection And Segmentation Image

Two different processing stages are used to process a CT lung image and the detection and segmentation results are presented in Figure 3. The image at the left shows the initial detection step, a section of the image where several different anatomical regions or tissue differences are marked by various color segments by means of intensity-based clustering. This step approximately assigns the areas of the lungs with background areas and non-tumor tissues. The image on the right displays the refined

segmentation result; the tumor has been more precisely segmented from the image after performing a higher-resolution segmentation process using greater cluster precision. The majority of irrelevant areas have been removed and the mg boundary is more pronounced, showing better localization and less noise. This comparison shows, in a visual manner, how accuracy in the extraction of tumors from raw computed tomography (CT) scans increases with the development of segmentation

techniques. The results of the segmentation of the MRI images by the clustering method with the parameter $k=1000$ are shown in figure 4. Both panels show sections of the brain with color coded structures, with the algorithm's segmentations of the anatomical structures highlighted. The use of 'k' in both images indicates a comparison, likely between different segmentation techniques or preprocessing algorithms, with the same granularity of the clusters. This visual representation helps in assessing the success of automatic tissue typing for medical imaging.

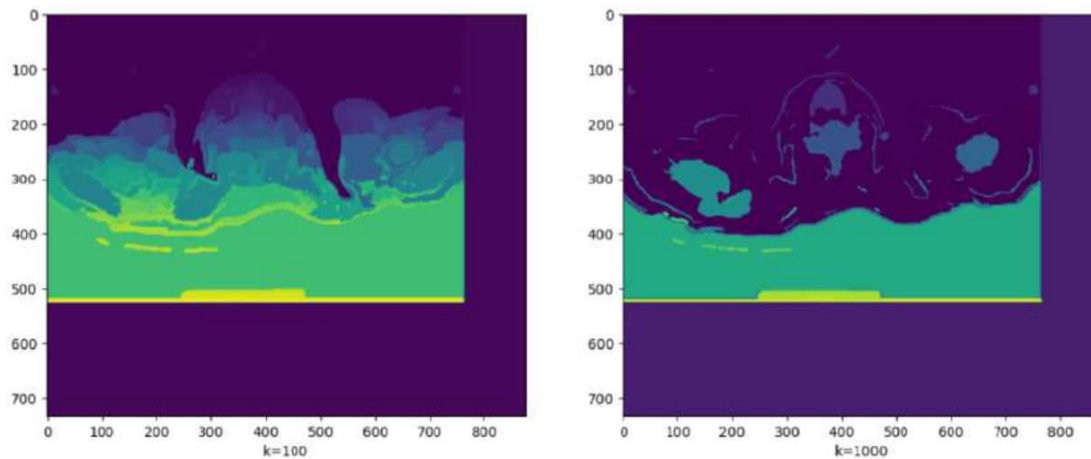


Figure 5: PET image-Detection and Segmentation Image

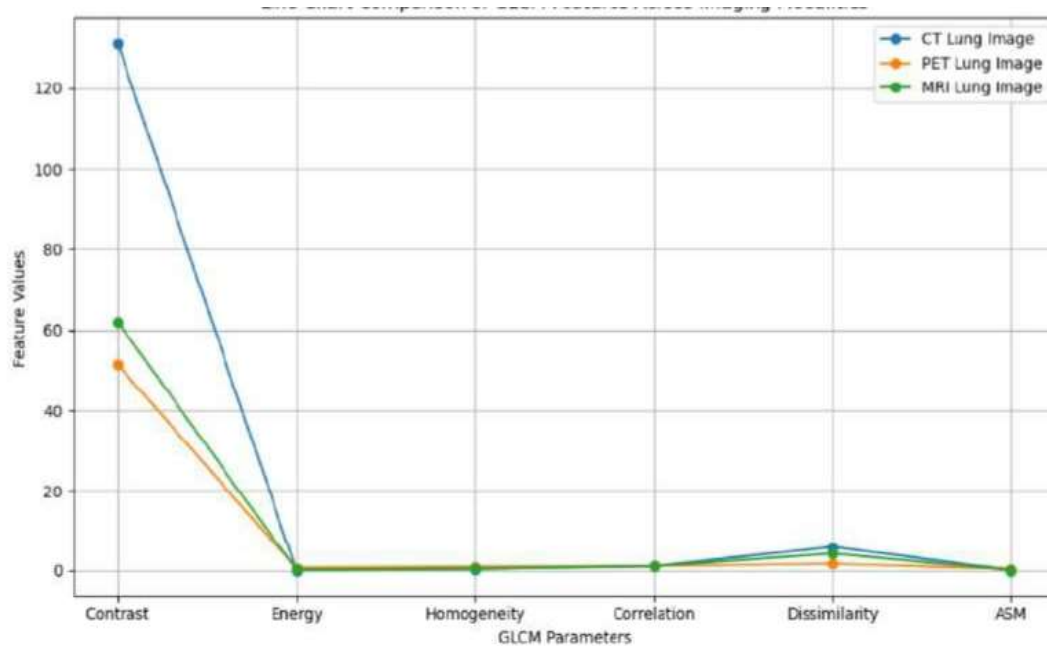


Figure 6: Line Chart Comparison Of GLCM Features Across Imaging Modalities

The detection and segmentation results of a PET scan by a clustering method for $k = 100$ are displayed

in figure 5. The segmented image, displayed on the left, segments the image into distinct color-coded

regions, each corresponding to a different tissue type or area of metabolic activity. The right panel indicates areas of interest detected, which may be regions of abnormality or high uptake. This dual visualization enables the diagnostic interpretation of the image thanks to the anatomical segmentation combined with the functional detection, thereby improving the accuracy of the analysis performed in the medical field from PET images.

Table 1: Feature Extraction Parameter

Parameters	CT Lung Image	PET Lung Image	MRI Lung Image
Contrast	130.9519562	51.36777049	61.96192141
Energy	0.04260169	0.54076202	0.09842512
Homogeneity	0.30727638	0.81931429	0.35667922
Correlation	0.98172573	0.99809603	0.98692373
Dissimilarity	5.74526329	1.5840203	4.09261205
ASM	0.0018149	0.29242356	0.0096875

Table 1 and Figure 6 show the comparisons of texture features of GLCM for lung images from the CT, PET, and MRI modalities in terms of Contrast, Energy, Homogeneity, Correlation, Dissimilarity and ASM. The CT modality has a very high Contrast value, which means it has a high level of intensity difference and complexity of texture. In contrast, the other show relatively low and similar values in all three modalities, indicating similar amounts of uniformity and structural consistency. This analysis reveals the use of different imaging techniques can reveal different characteristics of textures, and can support diagnostic and classification tasks.

Table 2: Performance Metrics

Metrics	CT Lung Image	PET Lung Image	MRI Lung Image
Dice Score	0.6666	0.6666	0.74999
Precision	0.6666	0.6666	0.74999
Recall	0.6666	0.6666	0.74999
Jaccard	0.4999	0.4999	0.5999
Hausdorff	1.0	1.5840	4.0926

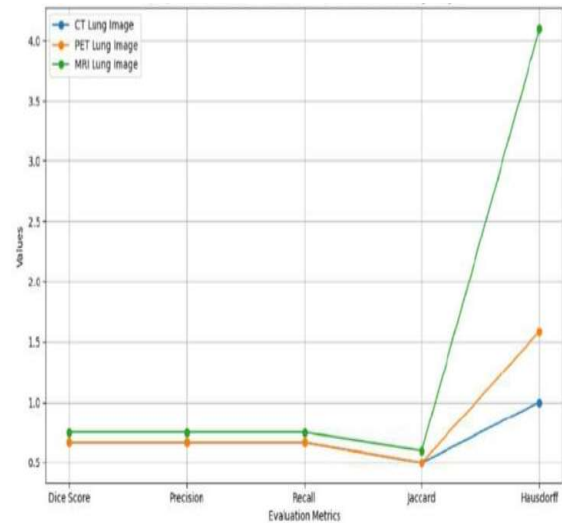


Figure 7: Comparison Of Evaluation Metrics For CT, PET, And MRI Lung Images

The evaluation metrics—Dice Score are compared in Table 2 and Figure 7. Precision, Recall, Jaccard, Hausdorff metrics for lung images from computed tomography, positron-emission tomography and magnetic resonance imaging. The performance of these three modalities is nearly identical for Dice, Precision, and Recall, but some differences are noticeable for Jaccard and Hausdorff. In these two, the MRI images get the highest values. PET was followed by metrics, which showed a better segmentation accuracy and delineation of boundaries, and then CT. This implies MRI might provide a more accurate anatomical representation to use for lung image analysis.

The proposed segmentation framework is evaluated in the present study using the available CT, PET-CT and MRI datasets to show its performance on these evaluated datasets, but not to provide full external clinical validation. In addition, the present study does not have any independent prospective hospital cohort or multicenter external dataset to be cross-dataset-validated. The results should therefore be considered an assessment of the segmentation ability and not an assessment of clinical diagnostic or treatment planning effectiveness. While accurate tumor segmentation may be useful for providing potentially useful quantitative information for further clinical applications, such as tumor characterization, treatment planning, and longitudinal monitoring, these applications are yet to be clinically validated. Future applications of the framework will be evaluated in independent multicenter datasets to further test the framework in diverse datasets with different scanner and

acquisition modalities, and in prospective, clinical studies, using expert clinical/radiologist assessment.

5. CONCLUSION

The proposed multi-modal lung tumor segmentation approach proves that advanced image preprocessing, texture feature extraction, and deep learning could greatly improve the localization accuracy and reliability of tumors in the lungs in CT, PET-CT, and MRI imaging. By embedding architectures like UNet, 3D nnU-Net and the framework is able to effectively incorporate spatial and contextual information, resulting in accurate boundary detection even in complex and low contrast areas, Swin-UNet. The results demonstrate that multi-modal fusion has better Dice, Jaccard and recall scores, lower false positive and higher edge conformity, which indicates that multi-modal fusion gives a more complete understanding than a single-modality analysis. Noise removal, normalization and ROI extraction further enhance the robustness and generalization of the models across various patient data sets. Based on the evaluation, MRI has the better ability to segment the soft tissues with clarity, CT brings in the structure accuracy and PET provides metabolic insight, combination of the three is clinically useful. The findings overall suggest that advanced image processing techniques and hybrid deep learning approaches have the potential to assist in early diagnosis, treatment planning, and monitoring of lung cancer progression, which could lead to more automated and scalable clinical solutions for lung cancer assessment.

REFERENCES

- [1]. Jin, Q. Cui, H. Sun, C., Song, Y., Zheng, J., Cao, L., Wei, L., & Su, R. (2023). Inter- and intra-uncertainty-based feature aggregation model for semi-supervised histopathology image segmentation. *Expert Systems With Applications*, 238, 122093. <https://doi.org/10.1016/j.eswa.2023.122093>
- [2]. Ibtehaz, N., & Rahman, M. S. (2019). MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121, 74-87. <https://doi.org/10.1016/j.neunet.2019.08.025>
- [3]. Jin, D., Guo, D., Ho, T., Harrison, A. P., Xiao, J., Tseng, C., & Lu, L. (2020). DeepTarget: Gross tumor and clinical target volume segmentation in esophageal cancer radiotherapy. *Medical Image Analysis*, 68, 101909. <https://doi.org/10.1016/j.media.2020.101909>
- [4]. Rugiya, N., Khan, N. N., & Khan, N. S. (2024). Comparing Models and Performance Metrics for Lung Cancer Prediction using Machine Learning Approaches. *Sir Syed University Research Journal of Engineering & Technology*, 14(2), 29-33. <https://doi.org/10.33317/ssurj.629>
- [5]. Panayides, A. S. Amini, A., Filipovic, N. D., Sharma, A., Tsaftaris, S. A., Young, A., Foran, D., Do, N., Golemati, S., Kure, T., Huang, K., Nikita, K. S., Veasey, B. P., Zervakis, M., Saltz, J. H., & Pattichis, C. S. (2020). AI in Medical Imaging Informatics: Current Challenges and Future Directions. *IEEE Journal of Biomedical and Health Informatics*, 24(7), 1837-1857. <https://doi.org/10.1109/jbhi.2020.2991043>
- [6]. Carles, M., Kuhn, D., Fechter, T., Baltas, D., Mix, M., Nestle, U., Grosu, A. L., Marti-Bonmati, L., Radicioni, G., & Gkika, E. (2024). Development and evaluation of two open-source nnU-Net models for automatic segmentation of lung tumors on PET and CT images with and without respiratory motion compensation. *European Radiology*, 34(10), 6701-6711. <https://doi.org/10.1007/s00330-024-10751-2>
- [7]. Qi, Y., Lin, L., Zhang, B., Zhang, J., & Wang, J. (2025). Multi-modal evidential fusion network for trustworthy PET/CT tumor segmentation Knowledge-Based Systems, 324, 113838. <https://doi.org/10.1016/j.knosys.2025.113838>
- [8]. Sun, H., Wei, J., Yuan, W., & Li, R. (2024). Semi-supervised multi-modal medical image segmentation with unified translation. *Computers in Biology and Medicine*, 176, 108570. <https://doi.org/10.1016/j.compbiomed.2024.108570>
- [9]. Lalitha, S., & Murugan, D. (2023). Segmentation and Classification of 3D Lung Tumor Diagnoses Using Convolutional Neural Networks. 2023 Second International Conference on Augmented Intelligence and Sustainable Systems, 230-238. <https://doi.org/10.1109/icaiss58487.2023.10250625>
- [10]. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021b). Review of deep

- learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1), 53. <https://doi.org/10.1186/s40537-021-00444-8>
- [11]. Huang, J., Xiang, Y., Gan, S. et al. Application of artificial intelligence in medical imaging for tumor diagnosis and treatment: a comprehensive approach. *Discov Onc* 16, 1625 (2025). <https://doi.org/10.1007/s12672-025-03307-3>
- [12]. Xue, H., Yao, Y., & Teng, Y. (2024). Multi-modal tumor segmentation methods based on deep learning: A narrative review. *Quantitative imaging in Medicine and Surgery*, 14(1), 1122. <https://doi.org/10.21037/qims-23-818>
- [13]. Alanazi, T. M. (2025). Multimodal feature distinguishing and a deep learning approach to detect mg disease from MRI images. *Scientific Reports*, 15(1), 31932. <https://doi.org/10.1038/s41598-025-17796-3>
- [14]. Liu, C., Liu, H., Zhang, X., Guo, J., & Lv, P. (2024). Multi-scale and multi-view network for lung tumor segmentation. *Computers in Biology and Medicine*, 172, 108250. <https://doi.org/10.1016/j.compbimed.2024.108250>
- [15]. Weikert, T., Jaeger, P. F., Yang, S., Baumgartner, M., Breit, H. C., Winkel, D. J., Sommer, G., Stieltjes, B., Thaiss, W., Bremerich, J., Maier-Hein, K. H., & Sauter, A. W. (2023). Automated lung cancer assessment on 18F-PET/CT using Retina U-Net and anatomical region segmentation. *European Radiology*, 33(6), 4270-4279. <https://doi.org/10.1007/s00330-022-09332-y>
- [16]. Wang, X., Zhang, L., Yang, X., Tang, L., Zhao, J., Chen, G., Li, X., Yan, S., Li, S., Yang, Y., Kang, Y., Li, Q., & Wu, N. (2020). Deep learning combined with radiomics may optimize the prediction in differentiating high-grade lung adenocarcinomas in ground glass opacity lesions on CT scans. *European Journal of Radiology*, 129, 109150. <https://doi.org/10.1016/j.ejrad.2020.109150>
- [17]. Wang, S., Dong, L., Wang, X., & Wang, X. (2020). Classification of pathological types of lung cancer from CT images by deep residual neural networks with a transfer learning strategy. *Open Medicine*, 15(1), 190-197. <https://doi.org/10.1515/med-2020-0028>
- [18]. He, J., Luo, Z., Lian, S., Su, S., & Li, S. (2024). Towards accurate abdominal tumor segmentation: A 2D model with Position-Aware and Key Slice Feature Sharing. *Computers in Biology and Medicine*, 179, 108743. <https://doi.org/10.1016/j.compbimed.2024.108743>
- [19]. Huang, C., Shi, Y., Wong, W., Liu, C., Wang, W., Wang, Z., & Wen, J. (2025). Multi-view evidential learning-based medical image segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(16), 17386-17394. <https://doi.org/10.1609/aaai.v39i16.33911>
- [20]. Guo, B., Cao, N., Zhang, R., & Yang, P. (2024). GETNET: Group Normalization Shuffle and Enhanced Channel Self-Attention Network based on VT-UNET for brain tumor segmentation. *Diagnostics*, 14(12), 1257. <https://doi.org/10.3390/diagnostics14121257>
- [21]. Lv, X., Li, Y., Xu, X., Zheng, Z., Li, F., Fang, K., Wang, Y., Wang, B., & Hou, D. (2023). Multisequence MRI-based radiomics nomogram for early prediction of osimertinib resistance in patients with non-small cell lung cancer brain metastases. *European Journal of Radiology Open*, 11, 100521. <https://doi.org/10.1016/j.ejro.2023.100521>
- [22]. Zhao, L., Qian, J., Tian, F., Liu, R., Liu, B., Zhang, S., & Lu, M. (2021). A weighted discriminative extreme learning machine design for lung cancer detection by an electronic nose system. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-9. <https://doi.org/10.1109/tim.2021.3084312>
- [23]. Visonà, G., Spiller, L. M., Hahn, S., Hattungen, E., Vogl, T. J., Schweikert, G., Bankov, K., Demes, M., Reis, H., Wild, P., Zeiner, P. S., Acker, F., Sebastian, M., & Wenger, K. J. (2023). Machine-Learning-Aided prediction of brain metastases development in Non-Small-Cell lung cancers. *Clinical Lung Cancer*, 24(8), e311-e322. <https://doi.org/10.1016/j.clcc.2023.08.002>
- [24]. Hou, G., Jia, L., Zhang, Y., Wu, W., Zhao, L., Zhao, J., Wang, L., & Qiang, Y. (2023). Deep learning approach for predicting lymph node metastasis in non-small cell lung cancer by fusing image-gene data. *Engineering Applications of Artificial Intelligence*, 122, 106140. <https://doi.org/10.1016/j.engappai.2023.106140>
- [25]. Kong, C., Lai, L., Jin, X., Chen, W., Ding, J., Zheng, L., Zhang, D., Ying, X., Chen, X., Chen, M., Tu, J., & Ji, J. (2023). Machine Learning classifier for preoperative prediction

- of early recurrence after bronchial arterial chemoembolization treatment in lung cancer patients. *Academic Radiology*, 30(12), 2880–2893.
<https://doi.org/10.1016/j.acra.2023.04.011>
- [26]. Xie, Y., Zhou, Y., Wang, C., Ma, Y., & Yang, M. (2023). Multi-scale feature fusion network with local attention for lung segmentation. *Signal Processing Image Communication*, 119, 117042.
<https://doi.org/10.1016/j.image.2023.117042>
- [27]. Jayarathna, S., Sowmiya, R., Nallasivan, G., Sahaana, G., Ashok, S., & Saravanakumar, R. (2025). 3D CT Liver Images Lesion Extraction and Classification for Using 3D-CNN with GLRLM. *International Conference on Advanced Computing Technologies Icoact 2025*, 1–5.
<https://doi.org/10.1109/icoact63339.2025.1100478>
- [28]. Srinath, R., Sharmila, R., Raghavendran, N., Babu, M. C., Ramanan, K., & Saravanakumar, R. (2025). Extracting and classifying 3D Medical Image Texture Characteristics using GLCM and LBP-Based Techniques. *2025 International Conference on Data Science Agents and Artificial Intelligence ICDSAAI 2025*, 1–5.
<https://doi.org/10.1109/icdsaaai65575.2025.11011770>