STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. 15, March 2026, pp 1868–1888.
Published online in International Academic Press (www.IAPress.org)

Lung Cancer Segmentation and Classification with Multi-Dataset Integration

Hozan Akram Abdulqader [1,2,*], Adnan Mohsin Abdulazeez [2]

[1]Department of Information Technology, Technical College of Duhok, Duhok Polytechnic University
[2]Department of Information Technology, Technical College of Informatics - Akre, Akre University for Applied Sciences

Abstract Accurate computer-aided lung cancer diagnosis is based on two sequential tasks: precise nodule segmentation and reliable malignancy classification. To this end, we curated the largest open-source CT benchmark to date by unifying five public repositories, resulting in 7,061 annotated slices from 571 patients for segmentation and 17,351 slices from 1,208 patients for classification. A standardized pre-processing pipeline was developed to harmonize voxel spacing, intensity windows, and label conventions. For segmentation, six encoder–decoder architectures were evaluated, with the hybrid UNet++ achieving the highest validation performance (Dice coefficient = 98.5%), demonstrating that attention-augmented dense skip pathways enable more accurate boundary detection of lung nodules. These masks were then used to drive a two-phase classification strategy: models were initially trained using ground-truth masks, followed by fine-tuning on predicted masks to emulate real-world deployment scenarios. Our proposed NoduleHyperFusionNet a dual-stream EfficientNetV2-S architecture, achieved the best overall discrimination (Accuracy = 92%, F1-score = 89%, AUC = 91%). The EfficientNet-B3 model also performed strongly, reaching an AUC of 94%.
Overall, this study demonstrates that the combination of attention-enhanced segmentation and lightweight multichannel fusion architectures can significantly improve automated lung cancer workflows, reducing diagnostic error rates without incurring prohibitive computational costs.

Keywords Segmentation, Classification, Deep Learning, Lung Cancer, Multi-Dataset

AMS 2010 subject Segmentation and classifications 68T07, 68T45, 92C55

DOI: 10.19139/soic-2310-5070-2814

1. Introduction

Lung cancer remains one of the most devastating malignancies worldwide, accounting for a substantial portion of cancer-related deaths [1]. Its aggressive nature and often asymptomatic early stages underscore the profound clinical urgency for precise and timely diagnostic interventions. Early detection is critically correlated with improved patient survival rates and enables the application of less invasive and more effective treatment modalities, highlighting the indispensable need for advanced, reliable diagnostic tools within medical imaging [2]. The rise of deep learning (DL) has profoundly reshaped medical image analysis. Convolutional Neural Networks (CNNs), in particular, have become essential tools for automating and improving the accuracy of tasks such as detection, segmentation, and interpretation of complex patterns in medical scans like Computed Tomography (CT) [1]. Their ability to learn complex hierarchical features directly from raw image data has positioned them as central to next-generation diagnostic systems.

*Correspondence to: Hozan Akram Abdulqader (Email: Hozan.akram@auas.edu.krd, hozan.abdulqader@dpu.edu.krd). Department of Information Technology, Technical College of Duhok, Duhok Polytechnic University, Department of Information Technology, Technical College of Informatics - Akre, Akre University for Applied Sciences.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

In the context of lung cancer, deep learning plays a pivotal role in two primary tasks: nodule segmentation and nodule classification. Segmentation involves the precise delineation of regions of interest, such as lung nodules or tumors, providing crucial volumetric and morphological information for clinical assessment. Early deep learning endeavors in this field prominently featured models like U-Nets [3] and 3D-CNNs [1], which consistently demonstrated robust capabilities in accurately identifying and outlining lung nodules. Subsequent advancements have led to the development of more sophisticated architectures, including hybrid models and the incorporation of attention mechanisms, as seen in architectures like UNet++ [4], [36], SegFormer [35], [43] and DeepLabV3Plus [8] further refining the segmentation process by directing the model's focus to diagnostically relevant features within complex medical imagery [1].

Following accurate segmentation, the subsequent classification of lung nodules as either malignant or benign is paramount for guiding patient management. Deep learning models, frequently integrated with transfer learning paradigms, have achieved substantial improvements in classification accuracy, with various architectures demonstrating high proficiency in distinguishing malignancies [1]. This study investigates a wide array of deep learning models for this task, including advanced CNNs (e.g., EfficientNet-B3 [15], [32], ConvNeXt-Tiny [12], [13], EfficientNetV2-S [14], MobileNetV3-Large [16], [17]), Vision Transformers (e.g., ViT-Small [18], MaxViT, [20]), and hybrid architectures (e.g., CoAtNet-0 [22]]). Crucially, we propose NoduleHyperFusionNet, a novel and specialized architecture designed for multi-scale feature fusion [24] from a comprehensive seven-channel input. This model leverages dual processing streams with an EfficientNetV2-S backbone [14], [15] and lightweight attention for efficient and effective information integration [25], representing a core contribution of this work.

To ensure the robustness and generalizability of our findings, this study uniquely leverages a comprehensive combination of several publicly available CT scan datasets: Luna (LUng Nodule Analysis 2016) [26], MSD (The Medical Segmentation Decathlon) [27], LNDb [28], LungCt [30], and Lung-PET-CT-Dx [51]. This deliberate aggregation of diverse datasets is an academically sound practice that aids in mitigating potential biases from single-source data, significantly enhancing the models' ability to generalize and perform reliably across varied clinical contexts, different imaging protocols, and diverse patient populations. Despite these advancements and diverse data sources, medical image analysis, particularly for lung cancer datasets, frequently encounters inherent challenges such as the number of benign cases significantly outweighing malignant ones. This can lead to models biased towards the majority class, thereby diminishing their sensitivity to critical, but rare, malignant instances [41]. Furthermore, computational demands and ensuring generalizability across diverse clinical settings remain significant hurdles for widespread clinical integration. This paper comprehensively investigates various deep learning techniques, including robust data augmentation strategies cite:Widodo2024-xz, the application of Focal Loss [2], [31] to mitigate class imbalance, and a two-phase training process, to address these challenges and enhance the reliability and efficiency of lung cancer diagnostics.

The remainder of this paper is structured as follows: Section 2 provides an overview of related works in lung cancer segmentation and classification, detailing key deep learning architectures and methodologies. Section 3 outlines the research methods, including the experimental setup, data preparation and preprocessing, model architectures, training and optimization strategies, and evaluation metrics. Section 4 presents and discusses the empirical results obtained for both segmentation and classification tasks. Finally, Section 5 concludes the paper and offers insights into future research directions.

Contributions The main contributions of this work are three-fold:

1. We introduce, to our knowledge, the largest curated public benchmark for this task, created by harmonizing five diverse CT datasets, which enhances model generalizability.
2. We propose a novel dual-stream classification architecture, NoduleHyperFusionNet, which efficiently fuses multi-scale features for highly accurate malignancy assessment.
3. We validate a robust two-phase training strategy, where models are first trained on ground-truth masks and then fine-tuned on predicted masks, effectively simulating a real-world clinical workflow and improving model robustness.

2. Related Works

The application of deep learning (DL) has profoundly impacted medical image analysis, particularly in the realm of lung cancer diagnostics. Convolutional Neural Networks (CNNs) have been instrumental in automating and enhancing the precision of detecting and interpreting intricate imaging patterns derived from medical scans, such as Computed Tomography (CT) [1]. The early and accurate identification of lung cancer is critical, as it significantly correlates with improved patient survival rates and facilitates less invasive treatment modalities [2].

2.1. Lung Cancer Segmentation

Segmentation, which involves precisely delineating regions of interest, is a foundational step in the computational analysis of medical images. Early deep learning endeavours in this field prominently featured models like U-Nets [3] and CNNs [1], which consistently demonstrated robust capabilities in identifying and outlining lung nodules. Subsequent advancements have led to the development of hybrid architectures and the incorporation of attention mechanisms, further refining the segmentation process by directing the model's focus to diagnostically relevant features within complex medical imagery [1]. Various deep learning architectures have been developed and extensively evaluated for their efficacy in lung nodule segmentation.

Res34UNetV2 leverages a ResNet34 backbone [33] integrated within a U-Net V2 architecture. This combination allows for enhanced feature extraction through the deep residual connections of ResNet, which mitigate vanishing gradients and improve information flow, thereby boosting the U-Net's capacity for precise contextual understanding and spatial localization during segmentation. While a specific paper on "Res34UNetV2 for lung cancer segmentation" was not directly found, similar U-Net based architectures with ResNet backbones have been widely applied for medical image segmentation tasks, including lung tumour segmentation Jayaram et al. 2025 used U-Net based architectures achieving dice score of 0.853 and F-score of 0.905 on the MSD dataset [34].

SegFormer introduces a transformer-based approach to semantic segmentation. Unlike traditional CNNs, SegFormer utilizes a hierarchical transformer encoder to capture global contextual information more effectively, followed by a lightweight all-MLP (Multi-Layer Perceptron) decoder. This design allows for robust feature representation across various scales, which is crucial for accurately segmenting objects of diverse sizes, such as lung nodules. Research has explored transformer-based models with attention mechanisms for lung nodule segmentation to address challenges in extracting edge and semantic information and capturing long-range dependencies[35], as in Hu et al. 2024 on LIDC-IDRI and LNDb datasets achieved dic score of 88.29% and 78.51% respectively [7].

DeepLabV3Plus is a robust semantic segmentation model that integrates atrous convolutions (dilated convolutions) and an encoder-decoder structure. The atrous convolutions enable the model to capture multi-scale contextual information without increasing the number of parameters, making it particularly effective for segmenting objects with varying sizes and shapes within medical images. This architecture has been utilized in models for accurate lung nodule segmentation, often combined with attention mechanisms and transformers to enhance global information integration [8].

LightUNet is a streamlined variant of the U-Net architecture, optimized for lower computational cost and memory usage while maintaining strong segmentation performance [37]. Its efficiency makes it ideal for resource-constrained settings and real-time medical imaging applications, such as early lung cancer detection in CT scans. Said et al. (2023) demonstrated its effectiveness on the MSD dataset, achieving a Dice score of 96.42% [38].

UNetPlusPlus extends the original U-Net by incorporating nested and dense skip pathways between the encoder and decoder As shown in Figure 1. These redesigned connections facilitate richer feature aggregation and more efficient information propagation, allowing for finer-grained segmentation results and more precise boundary delineation compared to its predecessor[4]. UNet++ has been successfully applied to various medical image segmentation tasks, including lung CT image segmentation to improve

the efficiency of treating lung cancer, Dong et al. 2021 used U-Net++ on the LIDC-IDRI dataset achieving an 83.43% Dice score value and 72.51% IOU score[36].

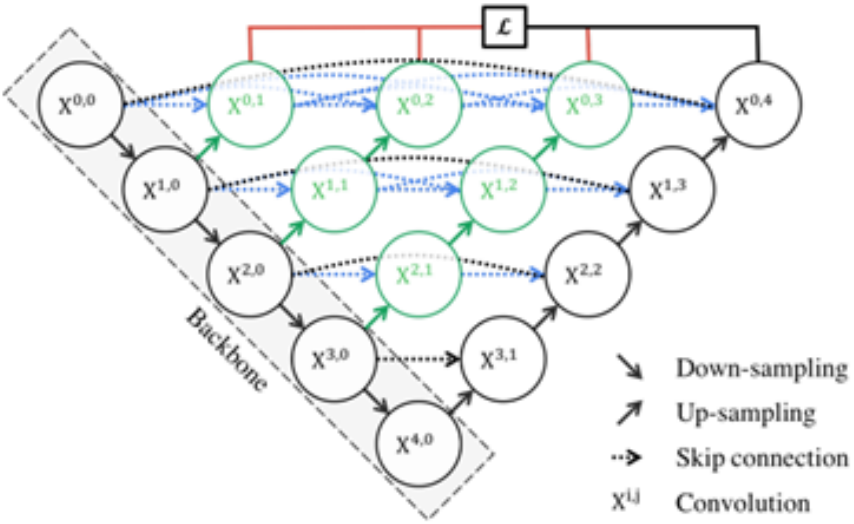

Figure 1. baseline UNet++ architecture

UNAR is a hybrid architecture that integrates UNetPlusPlus [4] with attention gates and a ResNet34d backbone.The attention gates are crucial for improving segmentation performance. They work by automatically learning to focus on target structures of varying shapes and sizes while suppressing irrelevant information from background regions. For each skip connection, the attention gate takes a gating signal from a deeper, more contextually-rich layer of the network and uses it to rescale the feature map from the shallower, more spatially-detailed encoder layer. This process effectively filters the feature flow, ensuring that only the most relevant activations are propagated to the decoder, which significantly enhances the model's precision in delineating complex nodule boundaries. Jayaram et al. (2025) reported an 85.3% Dice score on the MSD dataset [34].

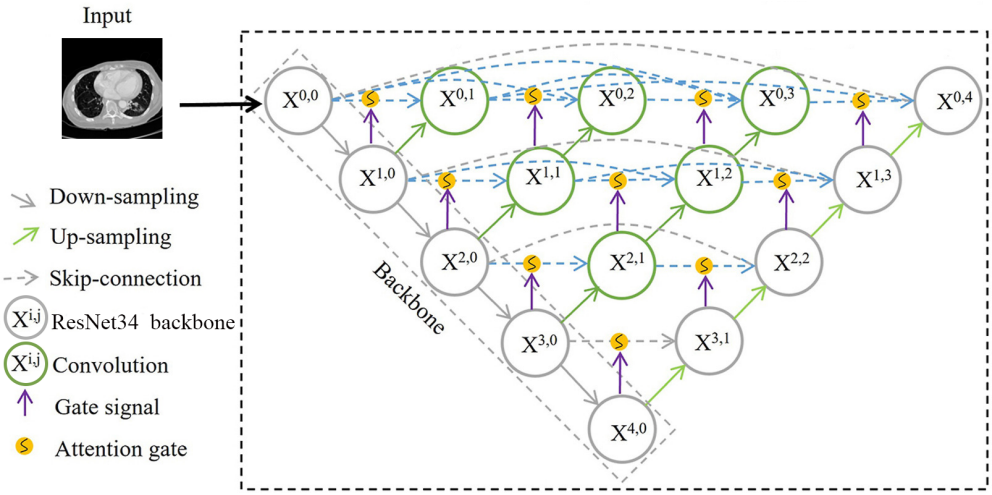

Figure 2. Proposed UNAR Architecture with Attention Gates

furthermore, the term ResNet34d, which serves as the UNAR model's backbone, refers to an improved variant of the standard ResNet-34 architecture. While maintaining the 34-layer structure, it incorporates impactful modifications in its downsampling blocks to preserve more spatial information [44], as illustrated in Figure 3. The key modifications are:

1. Modified Convolution in the Main Path (Path B): In a standard ResNet, the stride of 2 is applied in the first 1x1 convolution of a downsampling block, which discards significant spatial information. In ResNet-D, this stride is moved to the subsequent 3x3 convolution, allowing the model to learn richer features before downsampling.
2. Average Pooling in the Skip Connection (Path A): The strided 1x1 convolution in the skip connection is replaced with a 2x2 Average Pooling layer followed by a non-strided 1x1 convolution. This change avoids information loss in the identity path.

These tweaks ensure that less spatial information is lost during downsampling, leading to better feature representation and ultimately improving the segmentation performance of the UNAR model.

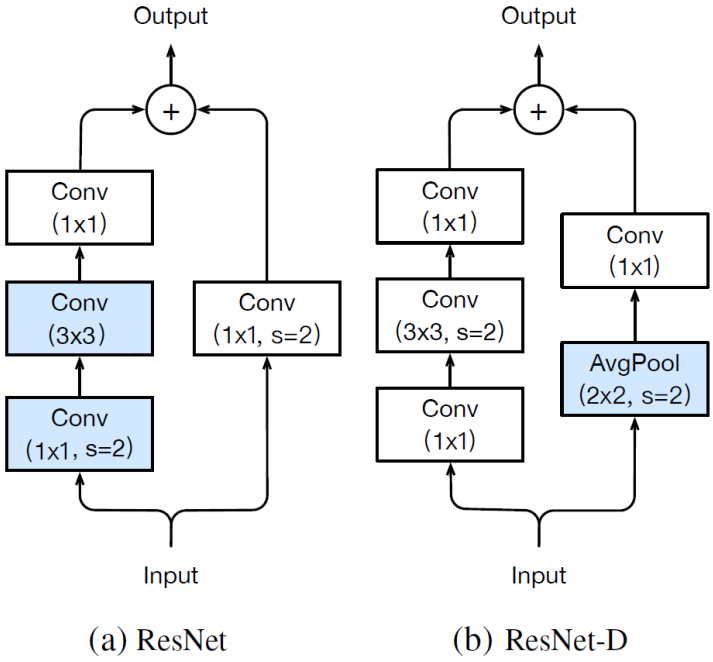

Figure 3. Architectural comparison of a block in a standard ResNet (a) versus the ResNet-D (b). The key changes, involve replacing the strided convolution in the skip connection (A) with an average pooling layer.

2.2. Lung Cancer Classifications

For the classification of lung nodules, distinguishing between malignant and benign cases, CNNs integrated with transfer learning paradigms have yielded substantial improvements in accuracy, with certain models achieving high performance in identifying malignancies [1]. A persistent challenge in medical image classification, particularly pronounced in lung cancer datasets, is class imbalance. This phenomenon occurs when clinically significant but rare instances are underrepresented, leading to models biased towards predominant classes and diminished sensitivity to critical minority classes [1],[30]. Such biases can compromise diagnostic reliability in clinical settings [30]. To address this challenge, traditional strategies like oversampling and under sampling, alongside synthetic data generation techniques such

as SMOTE and ADASYN, have been explored [30]. A wide array of deep learning models has been investigated for lung nodule classification, each offering unique architectural advantages. EfficientNet B3, part of the EfficientNet family, employs compound scaling to balance network depth, width, and resolution, optimizing both accuracy and efficiency. Its performance makes it a strong backbone for image classification tasks, including accurate lung cancer diagnosis from histopathological images, highlighting its potential in reliable medical diagnostic systems [9, 10, 11].

MaxViT-Tiny and MaxViT-Large are hybrid architectures that effectively combine the strengths of Vision Transformers (ViT) with conventional CNNs. Max-ViT models introduce a multi-axis attention mechanism that allows for both local and global feature extraction, leading to a more comprehensive understanding of image context, which is beneficial for complex classification tasks like nodule malignancy assessment. Hybrid transformer models, including those based on MaxViT, have been explored for multi-classification of lung diseases using chest X-rays, showing robust results in detecting multiple lung lesions [20, 21].

ViT-Small is a transformer-based model that processes images as patch sequences using self-attention to capture long-range dependencies. Unlike CNNs, it models global image relationships, providing an effective alternative for feature extraction. ViT and its variants have shown promise in classifying pulmonary nodules as benign or malignant by capturing both local and global contextual information [18, 19].

ConvNeXt-Tiny is a modern convolutional network designed to bridge the performance gap with recent transformer models while retaining the simplicity and inductive biases of CNNs. It incorporates various architectural innovations inspired by transformers, such as large kernel sizes and inverted bottleneck blocks, to achieve state-of-the-art performance with a purely convolutional backbone. While a direct reference for "ConvNeXt-Tiny for lung nodule classification" was not found, CNN-based structures are widely used for lung nodule classification to improve diagnostic accuracy [12, 13].

EfficientNetV2-S is an evolution of the EfficientNet series, designed for faster training and improved performance. It introduces a combination of Fused-MBConv blocks and smaller image sizes for initial layers, leading to quicker training times and better accuracy, particularly for diverse image classification challenges. EfficientNetV2 models have been investigated for classifying lung cancer images using histopathology, demonstrating high accuracy in distinguishing different types of lung cancer [14, 15].

CoAtNet-0 is a novel hybrid model that strategically integrates convolution and attention mechanisms, aiming to leverage the best of both worlds. It combines depth-wise convolutions for local feature extraction with self-attention for global context modeling, making it effective across various data sizes and types, including medical images [22]. While a specific paper on "CoAtNet-0 for lung nodule classification" was not directly found, 3D attention-gated convolutional networks have been used for lung nodule malignancy classification, highlighting the importance of incorporating contextual information [22, 42, 50]

MobileNetV3-Large is part of the MobileNet family, specifically designed for mobile and embedded vision applications. It focuses on achieving a balance between high accuracy and computational efficiency through automated search and novel architectural blocks, making it suitable for real-time or resource-constrained diagnostic tools [16]. MobileNetV3, along with other transfer learning models, has been explored for enhancing lung cancer classification from CT Scan images [17].

NoduleFusionNet is a specialized architecture often designed for medical image analysis tasks, particularly for nodule characterization. While specific architectural details can vary, such networks typically focus on multi-scale feature fusion and potentially incorporate domain-specific knowledge to enhance the diagnostic performance for lung nodules, leading to more robust and accurate classifications. Feature fusion techniques are critical for lung nodule classification, combining various descriptors to improve the distinction between malignant and benign nodules [24].

2.3. Optimization Algorithms

Deep learning models rely on optimization algorithms to minimize loss functions and improve performance. Adam (Adaptive Moment Estimation) combines the advantages of AdaGrad and RMSProp by computing adaptive learning rates for each parameter. It achieves this by using exponentially weighted moving

averages of both the gradient (the first moment, similar to momentum) and the squared gradient (the second moment, similar to RMSProp). Its efficiency, fast convergence, and minimal hyperparameter tuning make it ideal for non-convex problems in deep learning. In medical image analysis, particularly for lung cancer segmentation and classification, Adam is widely used for its robust performance and ability to generalize across diverse datasets [39].

2.4. Loss Functions

Loss functions are essential in deep learning, guiding model optimization by quantifying prediction errors. In medical image segmentation, Dice Loss based on the Dice coefficient measures overlap between predicted and ground truth masks, making it well-suited for imbalanced tasks like small lesion or lung nodule segmentation [40, 41].

Dice Loss was applied for segmentation tasks, effective for measuring spatial overlap. The loss is defined as:

$L_{\text{Dice}} = 1 - \frac{2\sum_i^N p_i g_i + \epsilon}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + \epsilon}$

where p_i is the predicted probability for pixel i, g_i is the ground truth label, and ϵ is a smoothing factor to prevent division by zero.

For classification, Cross-Entropy Loss is widely used, especially in multi-class settings, though it may struggle with class imbalance by favoring dominant classes [42, 23]. To mitigate this, Focal Loss modifies cross-entropy to emphasize hard, misclassified examples,

$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$

where p_t is the model's estimated probability for the ground truth class, and α and γ are focusing parameters. We selected $\alpha = 0.6$ and $\gamma = 0.4$ based on preliminary experiments to give more weight to the underrepresented malignant class. enhancing performance in cases with rare but critical positive samples, such as malignant lung cancer detection [31, 30].

2.5. Data Augmentation

Data augmentation is a crucial technique in deep learning to enhance the diversity of training datasets without actually collecting new data. This strategy helps to prevent overfitting, improve model generalization, and boost robustness, especially when dealing with limited medical imaging datasets. Common augmentation techniques include geometric transformations such as rotations, shifts, flips, and scaling, which simulate variations in image acquisition and patient positioning [45, 46]. These transformations expose the model to a wider range of plausible inputs, making it more invariant to minor distortions. Additionally, photometric augmentations like brightness adjustments, contrast changes, and noise injection can help the model become more robust to variations in image quality. Libraries like Albumentations provide a wide range of fast and flexible augmentation techniques commonly used in medical imaging pipelines [47].

3. Research Methods

This section outlines the methodologies employed for the lung cancer segmentation and classification tasks, detailing the experimental setup, data handling, model training procedures, and evaluation protocols.

3.1. Experimental Setup and Tools

All experiments were conducted in Python 3.10. Deep-learning workflows were implemented with PyTorch 2.2 and the following companion libraries, as shown in Table 1:

3.1.1. Compute infrastructure training and inference were conducted on Kaggle notebooks using dual NVIDIA T4 GPUs (16 GB VRAM each), with 30 GB RAM and four vCPUs. Mixed-precision training

Table 1. libraries utilized

Purpose	Library (version)	Notes
Core tensor & autograd	PyTorch 2.2.0	CUDA 11.8 backend
Data frames & numerics	pandas 2.2.2, NumPy 1.26.4	—
CV utilities & metrics	scikit-learn 1.5.0	Train/val/test splits, AUC, F1, DeLong stats
Image augmentation	Albumentations 1.4.6	3D-compatible transforms enabled
Pre-trained backbones	timm 0.9.16	EfficientNet-B3, ConvNeXt-Tiny, MaxViT-Large MaxV-Tiny
Visualization	Matplotlib 3.9, Seaborn 0.13	Learning-curve & heat-map plots

(via autocast and GradScaler) was enabled to reduce GPU memory usage and training time. To ensure reproducibility, deterministic behavior was enforced and all random seeds were set to 42.

3.2. Data Preparation and Preprocessing

Data preparation is essential for ensuring high-quality, model-ready inputs in deep learning workflows.

3.2.1. Datasets The study utilized Computed Tomography (CT) scan data specifically tailored for lung cancer diagnostics. Most of the original datasets were in DICOM format and were subsequently converted to PNG format for compatibility with the deep learning pipeline. For segmentation tasks, corresponding mask images were generated according to the annotations available in each dataset or directly utilized if already provided, figure 4 showing dataset preprocessing.

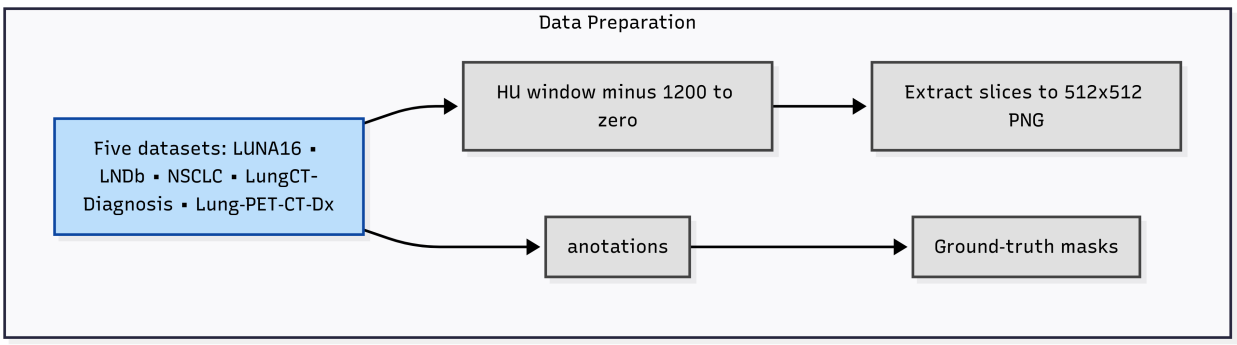

Figure 4. Data Pre-Processing workflow

3.2.2. Data Harmonization Pipeline To create a unified benchmark, a standardized pre-processing pipeline was applied to harmonize the diverse datasets. a standard lung window (-1200 to 0 Hounsfield Units) was applied to the image intensities to maximize the contrast of pulmonary nodules and surrounding parenchyma. These windowed values were then linearly scaled to an 8-bit grayscale range and saved as PNG files.

Label conventions were also unified to a binary system (0: benign, 1: malignant). For the LIDC-IDRI dataset, we followed the LUNA16 challenge protocol, defining a nodule as malignant if at least three of four radiologists assigned it a malignancy rating of 3 or higher. For the LNDb dataset, a 'finding' score greater than 3 was mapped to a malignant label. Critically, to ensure a single, unambiguous ground truth per individual and prevent data leakage, a final **patient-level unification** was performed. If any slice from a patient contained a malignant nodule, the entire patient was assigned the malignant label for all classification tasks. The datasets used in this study include:

Luna : (Luna16) is a subset of The LIDC/IDRI data set which is publicly available, including the annotations of nodules by four radiologists TU2025 [26].

MSD: The medical Segmentation decathlon lung task [27]dataset corresponds to the NSCLC-Radiogenomics collection from The Cancer Imaging Archive (TCIA) [48].

LNDb (Lung Nodule Database) comprises 294 CT scans retrospectively collected at Centro Hospitalar e Universitário de São João (CHUSJ), Porto, Portugal, between 2016 and 2018 [28].

LungCt : LungCT-Diagnosis Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity [29, 49]

Lung-PET-CT-Dx includes CT and PET-CT DICOM images of lung cancer patients, accompanied by tumor location annotations in PASCAL VOC (XML) format [51].

Each dataset was meticulously organized into a CSV file, with each row detailing crucial information for individual image slices. This CSV structure included: image-png paths: Direct links to the raw CT scan image files in PNG format. mask-png paths: Direct links to the corresponding binary mask images (also in PNG format) that delineate lung nodules, essential for segmentation training and evaluation. patient number: A unique identifier for each patient, allowing for patient-level aggregation and analysis. malignant or benign label: A categorical label indicating the ground truth malignancy status of the lung nodule (malignant or benign), derived from existing labels or generated from a 'finding' column (where finding > 3 indicated malignancy for datasets like LNDb).

3.2.3. Dataset Statistics The classification dataset, compiled from Luna, MSD, LNDb, LungCt, and Lung-PET-CT-Dx, includes 17,351 slices from 1,208 unique patients. The segmentation dataset, sourced from Luna and MSD, contains 7,061 slices. As part of preprocessing, patient-level labels were unified by assigning the maximum malignancy label (e.g., 1 for malignant) across all slices per patient, ensuring label consistency. The distribution of patients, slices, and malignancy labels across training, validation, and test sets for classification is summarized in Table 2:

Table 2. final dataset split

Split	Patients	Slices	Benign (0)	Malignant (1)
train	772	11425	2486	8939
val	194	2553	567	1986
test	242	3373	786	2587

Table 2 highlights the distribution of samples and the class imbalance within the dataset, particularly the higher number of malignant cases, which was addressed during model training.

3.2.4. Dataset Splitting and Class Imbalance Handling To ensure a fair and robust evaluation, free from data leakage, specific protocols were established for dataset splitting and handling class imbalance.

Patient-Level Dataset Splitting As recommended for medical imaging tasks, a strict patient-level splitting protocol was enforced for all classification experiments. This protocol ensures that all image slices from a single patient belong exclusively to one data partition (training, validation, or test), preventing the model from being evaluated on data from a patient it has already seen during training.

The splitting process was as follows:

1. A list of all unique patient identifiers was extracted from the complete classification dataset.
2. A corresponding list of patient-level labels was created.
3. This list of patients was first split into a training set (80%) and a test set (20%). The split was stratified by the patient-level malignancy labels to maintain a similar distribution of benign and malignant cases across the partitions.
4. The 80% training set was further split into a final training set (80% of the initial 80%, i.e., 64% of the total) and a validation set (20% of the initial 80%, i.e., 16% of the total), again using stratification.

5. The final slice-level dataframes were then constructed by selecting all slices belonging to the patient IDs in each respective set.

For the segmentation task, which did not involve patient-level labels, a standard 80/20 stratified split on the available slices was performed.

Class Imbalance Handling The compiled classification dataset exhibited a notable class imbalance, as detailed in Table 2. To address this, a WeightedRandomSampler was employed for the training data loader in all classification experiments. The weight for each individual sample in the training set was calculated as the **inverse of its class frequency**. This procedure assigns a higher sampling probability to instances from the underrepresented class (benign nodules), ensuring that the model is exposed to a more balanced distribution of classes within each training batch. This method mitigates the risk of the model developing a bias towards the majority (malignant) class and significantly improves its diagnostic sensitivity to the minority class.

3.2.5. Feature Engineering for Classification For the classification task, seven distinct features were extracted and combined to form a multi-channel input for each image slice. This involved several transformations and derivations.

Global Image: The resized global CT scan slice.

Global Mask: The resized corresponding predicted segmentation mask of the nodule.

Contextual Image Crop: A resized crop of the image around the nodule, providing local context.

Distance Transform: A resized distance map from the nodule boundary, capturing shape information.

Edge Map: A resized representation of the edges within the image or around the nodule, highlighting structural boundaries.

Zoomed Nodule Texture: A magnified and resized view focusing on the texture within the nodule.

Zoomed Nodule Shape : A magnified and resized view of the nodule's shape from its mask.

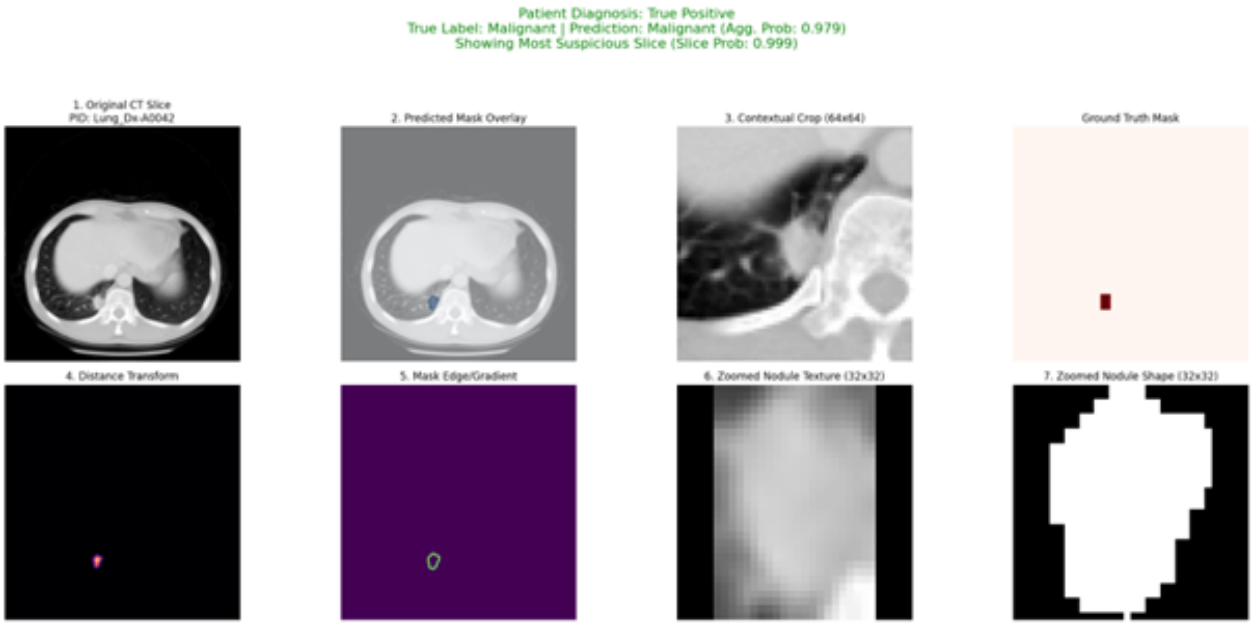

Figure 5. example of features extracted and used

This comprehensive multi-channel input, created by resizing and cropping around the segmented nodule using a defined 64 x 64 nodule crop size, aimed to provide a focused and enriched representation for the classification models.

3.3. Data Augmentation

To enhance model robustness and prevent overfitting, an extensive data augmentation pipeline was applied on-the-fly during training using the Albumentations library. Augmentations were applied only to the training data. The pipeline included a combination of geometric, non-linear, photometric, and regularization transformations, with specific parameters detailed in Table 3.

Table 3. Data Augmentation Pipeline and Parameters

Category	Transformation	Parameters
Geometric	RandomResizedCrop	Target size: (256, 256), Scale range: (0.7, 1.0), Ratio range: (0.75, 1.33)
	HorizontalFlip	Probability: 0.5
	VerticalFlip	Probability: 0.5
	RandomRotate90	Probability: 0.5
	ShiftScaleRotate	Shift limit: 0.0625, Scale limit: 0.15, Rotate limit: 30°
Non-Linear	ElasticTransform	Alpha: 120, Sigma: 6
	GridDistortion	Number of steps: 5, Distort limit: 0.3
Photometric	RandomGamma	Gamma range: (70, 130)
	RandomBrightnessContrast	Brightness/Contrast limit: 0.25
Regularization	CoarseDropout	Max holes: 8, Max height/width: 32x32

3.4. Model Architectures

The study used a diverse set of deep learning architecture training tailored for both segmentation and classification tasks following the steps presented in figure 6.

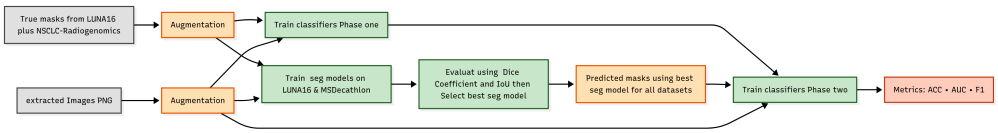

Figure 6. training steps and flow

3.4.1. Segmentation : models selected for evaluation:

Res34UNetV2: A U-Net variant incorporating a ResNet34 backbone to enhance feature extraction and support precise segmentation.

SegFormer: A transformer-based semantic segmentation model that captures long-range dependencies and global context.

UNetPlusPlus: An improved U-Net with nested, dense skip connections designed for finer feature refinement and improved boundary delineation.

DeepLabV3Plus: Integrates atrous spatial pyramid pooling and encoder-decoder structure to capture multi-scale contextual information.

LightUNet: A lightweight and computationally efficient U-Net variant optimized for faster inference, suitable for resource-constrained environments.

UNAR: A hybrid architecture that fuses UNet++ with attention mechanisms and a ResNet34d backbone, enhancing feature relevance and segmentation performance.

3.4.2. Classification: models evaluated for classification are:

EfficientNet-B3: A convolutional network that balances accuracy and computational efficiency using compound scaling.

MaxViT-Tiny / MaxViT-Large: Hybrid CNN-transformer models leveraging multi-axis attention for robust feature extraction across scales.

ViT-Small: A Vision Transformer model that treats images as patch sequences, enabling global feature learning via self-attention.

ConvNeXt-Tiny: A modernized convolutional architecture inspired by transformers, combining the strengths of both paradigms.

EfficientNetV2-S: An advanced version of EfficientNet optimized for improved training speed and accuracy.

CoAtNet-0: A hybrid architecture blending convolution and attention mechanisms for adaptability across diverse visual tasks.

MobileNetV3-Large: A compact and efficient model tailored for mobile and embedded systems, offering fast inference with reasonable accuracy.

NoduleFusionNet: A domain-specific architecture tailored for medical imaging, particularly lung nodule classification. It emphasizes multi-scale feature fusion from diverse input channels to form a rich representation. NoduleHyperFusionNet: This model represents a lighter and faster version of "HyperFusionNet" architecture, specifically optimized for lung nodule classification. Its design prioritizes efficiency while maintaining high performance. Key architectural features demonstrated in figure 7 include:

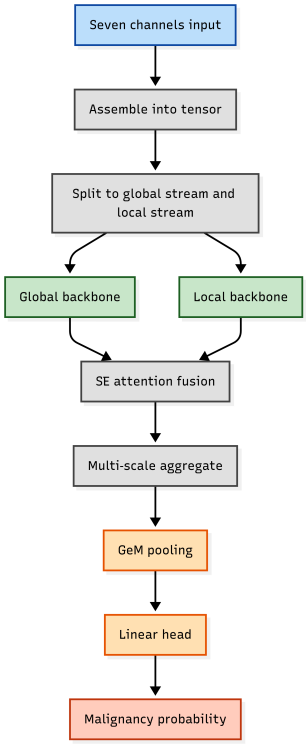

Figure 7. Nodule Hyper Fusion Net layers

- Dual-Stream Processing: It utilizes two distinct streams for feature extraction: a "Global" stream and a "Local" stream, reducing the parameter count compared to more complex multi-stream designs. Both streams leverage an EfficientNetV2 backbone, a powerful and efficient convolutional neural network architecture [13, 24]pre-trained for robust feature representation.

- Hierarchical Feature Extraction: Features are extracted from multiple stages of the backbones, allowing for the capture of information at various scales.
- Lightweight Fusion: Instead of computationally intensive Cross-Attention mechanisms, NoduleHyperFusionNet employs lightweight channel attention for fusion across the two streams at each hierarchical stage. This significantly enhances speed without sacrificing critical feature integration.
- Input Channel Handling: The model is designed to process a multi-channel input. The global stream receives the global CT image (replicated to 3 channels), while the local stream processes a combination of contextual image crop, distance transform, and edge map channels (totaling 3 channels). This specialized input assembly, drawing from critical feature fusion techniques [34, 33], allows the model to leverage both broad contextual information and fine-grained nodule-specific details.
- Final Aggregation and Classifier Head: After hierarchical fusion, features from different stages are aggregated, passed through a final convolutional block, pooled using GeM pooling, and then fed into a linear classifier head to produce the final malignancy prediction.

3.5. Training and Optimization

Model training involved iterative optimization to minimize the defined loss functions:

3.5.1. Optimizer: The Adam (Adaptive Moment Estimation) optimizer was used, known for its adaptive learning rates and efficient convergence.

3.5.2. Loss Functions: Dice Loss was applied for segmentation tasks, effective for measuring spatial overlap and handling class imbalance. Cross-Entropy Loss and Focal Loss were employed for classification. Focal Loss specifically addressed class imbalance by down-weighting well-classified examples, enabling the model to focus on harder, misclassified samples.

3.5.3. Learning Rate Schedulers: Various scheduling strategies from LRscheduler were utilized to dynamically adjust the learning rate during training, including SequentialLR, LinearLR (for warm-up), CosineAnnealingLR, and ReduceLROnPlateau, to facilitate faster convergence and prevent overfitting.

3.6. Hyperparameters and Implementation Details

To ensure full reproducibility, as recommended by the reviewers, this section details the specific hyperparameters used for training the segmentation and classification models. All random seeds were set to 42 for deterministic behavior.

The training parameters for the segmentation models are summarized in Table 4. All segmentation models were trained using a combined loss function consisting of 52% Binary Cross-Entropy and 48% Dice Loss to balance pixel-wise accuracy and spatial overlap. The AdamW optimizer was used, and the learning rate was dynamically adjusted by a ReduceLROnPlateau scheduler, which monitored the validation Dice score.

For the classification task, a two-phase training strategy was employed with distinct hyperparameter configurations for each phase, as detailed in Table 5. Phase 1 (training on ground-truth masks) utilized a standard training recipe with Focal Loss and a sequential scheduler. Phase 2 (fine-tuning on predicted masks) adopted a more advanced recipe with a combined loss function, the OneCycleLR scheduler, gradient accumulation, and Mixup for regularization. The AdamW optimizer was used in both phases, but with a stronger weight decay in Phase 2.

Data augmentation was performed using the Albumentations library. The training pipeline included geometric transformations (RandomResizedCrop, HorizontalFlip, VerticalFlip, RandomRotate90, ShiftScaleRotate), non-linear distortions (ElasticTransform, GridDistortion), photometric adjustments (RandomGamma, RandomBrightnessContrast), and regularization techniques (CoarseDropout).

Table 4. Hyperparameters for Segmentation Models

Parameter	Value
Data & Batches	
Training Batch Size	8
Validation Batch Size	16
Input Image Size	(256, 256)
Optimizer (AdamW)	
Learning Rates (LR)	Model-specific (e.g., 4e-4 for UNAR, 1e-4 for Res34UNetV2)
Weight Decay	2e-5
Betas (β_1, β_2)	(0.9, 0.999) (PyTorch default)
Epsilon (ϵ)	1e-8 (PyTorch default)
LR Scheduler	
Scheduler Type	ReduceLROnPlateau
Monitored Metric	Validation Dice Score
Patience	7 epochs
Decay Factor	0.5
Training Details	
Epochs	70 (with early stopping patience of 13)
Loss Function	$0.52 \times$ BCEWithLogitsLoss $+ 0.48 \times$ SoftDiceLoss

Table 5. Hyperparameters for Two-Phase Classification Training

Parameter	Phase 1 (GT Masks)	Phase 2 (Predicted Masks)
Data & Batches		
Training Batch Size	8	8
Input Size (Multi-channel)	(224, 224)	(224, 224)
Nodule Context Crop Size	64x64	64x64
Optimizer (AdamW)		
Base Learning Rates (LR)	Model-specific (e.g., 3e-4)	Model-specific (e.g., 2e-4)
Weight Decay	2e-3	2e-3
Betas (β_1, β_2)	(0.9, 0.999)	(0.9, 0.999)
LR Scheduler		
Scheduler Type	SequentialLR (Linear warm-up + Cosine Annealing)	OneCycleLR
Warm-up	5 epochs	N/A (handled by OneCycleLR)
Pct_start (OneCycleLR)	N/A	0.1
Loss Function		
Type	FocalLossWithLabelSmoothing	ComboLoss (70% Focal, 30% BCE)
Alpha (α) / Gamma (γ)	0.56 / 2.0	0.25 (default) / 2.0
Label Smoothing	0.1	0.05
Training Details		
Epochs per Phase	90	90
Encoder Freeze	First 7 epochs	First 10 epochs
Gradient Accumulation	1	4
Mixup Alpha	None	0.2

3.7. Evaluation Metrics

The performance of the models was rigorously evaluated using a comprehensive set of metrics:

3.7.1. Segmentation: Performance metrics typically include Dice Coefficient and IoU (Intersection over Union), which are implicitly optimized by Dice Loss.

3.7.2. Classification: For classification, slice-level predictions were aggregated to the patient level by averaging probabilities, ensuring consistent patient-level diagnosis when multiple slices were available.

- Accuracy: Measures the overall correctness of predictions.
- F1-Score: Harmonic mean of precision and recall, providing a balanced assessment of classification performance.
- AUC (Area Under the Curve): Evaluates the model's ability to distinguish between classes across different thresholds.
- Confusion Matrix: Offers a detailed view of classification outcomes, including true/false positives and negatives.

4. Results and Discussion

This section presents the empirical results obtained from the proposed deep learning architecture for lung cancer segmentation and classification. The performance of various models was rigorously evaluated using comprehensive metrics for each task.

4.1. Segmentation Results

For the segmentation task, all models were trained on a dataset of 7,061 annotated slices from 571 patients, which was split into training, validation, and held-out test sets. Performance was assessed using the Dice Score on both validation and test data to measure segmentation accuracy and generalizability, respectively. The comprehensive results are presented in Table 6.

Table 6. Final Segmentation Model Leaderboard on Validation and Test Sets

Model	Validation Dice	Test Dice
UNAR	0.9840	0.9827
Res34UNetV2	0.9804	0.9789
SegFormer	0.9783	0.9765
UNetPlusPlus	0.9774	0.9751
DeepLabV3Plus	0.9767	0.9748
LightUNet	0.9754	0.9732

As shown in Table 6, the UNAR model achieved the highest performance on the validation set with a Dice Score of 0.9840. This superiority was confirmed on the unseen test set, where it also attained the top score of 0.9827, demonstrating its strong generalization capabilities. The Res34UNetV2 model also performed robustly, securing the second-highest scores on both validation (0.9804) and test (0.9789) sets. The consistently high scores across all evaluated architectures underscore the effectiveness of modern deep learning approaches for precise lung nodule segmentation, a critical precursor to accurate diagnosis. The small and consistent drop from validation to test scores across all models indicates that our training methodology successfully avoided significant overfitting.

Figure 8 illustrates the training and validation history for these segmentation models, including metrics such as loss, Dice, and IoU over epochs. The plots provide additional insight into the convergence and stability of these models during the learning phase.

4.2. Classification Results

The classification models were evaluated on an extensive dataset, integrating data from Luna, msd, LNDb, LungCt, and Lung-PET-CT-Dx datasets, which encompassed 17351 slices and 1208 patients. Key classification metrics included Accuracy, F1-score, and Area Under the Curve (AUC), with patient-level

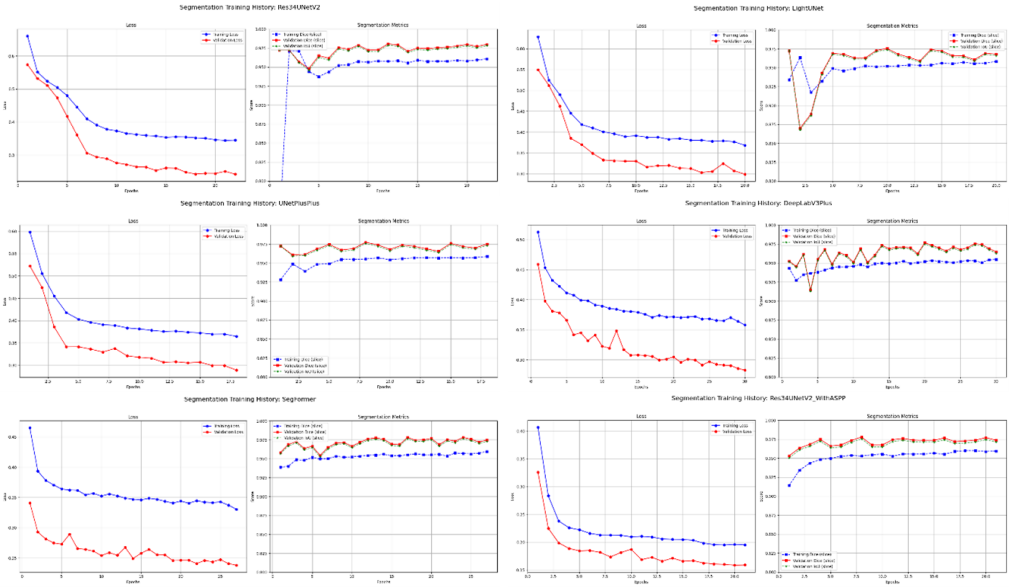

Figure 8. Segmentation Models training and validation history.

aggregation of probabilities to derive final predictions. All classification models underwent a two-phase training process to optimize their performance and leverage both ground truth and predicted segmentation information. The two-phase training process for the classification models involved:

4.2.1. Phase 1 (Training with Ground Truth Masks): In the initial phase, models were trained using ground truth segmentation masks. This allowed the models to learn robust features directly from precisely delineated nodules, minimizing the noise and inaccuracies that might arise from imperfect segmentation. This phase established a strong foundation for the classification task.

4.2.2. Phase 2 (Fine-tuning with Predicted Masks): Following the initial training, the models were fine-tuned using predicted segmentation masks generated by the best-performing segmentation model (UNAR). This crucial step simulated a real-world scenario where ground truth masks are unavailable and the classification model relies on automatically segmented nodules. This fine-tuning process helped the models adapt to potential variations and errors in predicted masks, thereby enhancing their robustness and the generalizability in a practical deployment setting. While a comprehensive set of models was evaluated, the initial discussion highlighted a few specific models (MobileNetV3-Large, EfficientNetV2-S, CoAtNet-0, and EfficientNet-B3) to illustrate the general performance trends and the impact of the two-phase training strategy. The complete results for all tested classification models are presented in Table 7, providing a detailed overview of their performance.

The training and validation history for individual classification models, illustrating the progression of loss, AUC, accuracy, and F1-score over epochs, is provided in separate figures 9 and 10 for NoduleHyperFusionNet Phase 1 and 2. Furthermore, the patient-level confusion matrices for the champion classification models are presented in figure 11, offering a detailed breakdown of true and false predictions

As shown in Table 7, the NoduleHyperFusionNet model demonstrated the highest accuracy among all tested classification models, achieving 0.924341. This model also exhibited a strong F1 Score of 0.891495 and an AUC of 0.918135, indicating a balanced performance in precision and recall, and a high capability to discriminate between classes. The EfficientNet-B3 also performed exceptionally well with an accuracy of 0.880165, an F1 Score of 0.911315, and a remarkable AUC of 0.939582, suggesting its robustness in classification. Models like MaxViT-Large and NoduleFusionNet also showed competitive performance

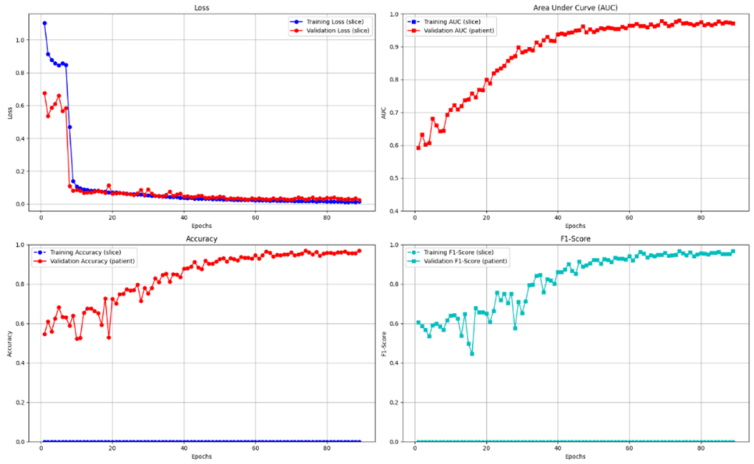

Figure 9. Phase 1 training chart for NoduleHyperFusionNet

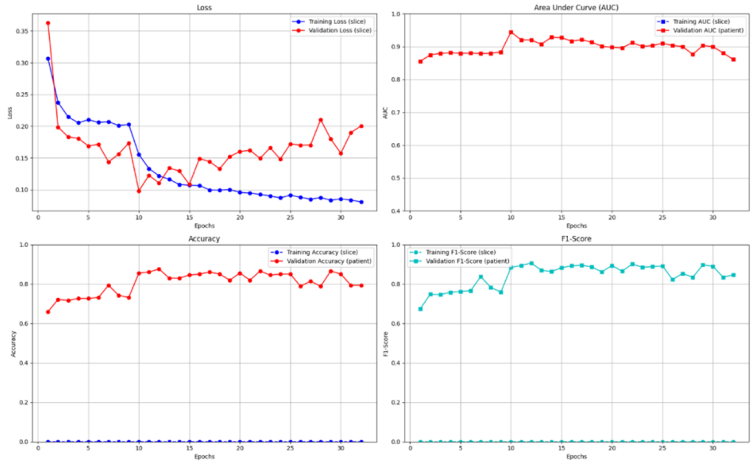

Figure 10. Phase 2 training for NoduleHyperFusionNet

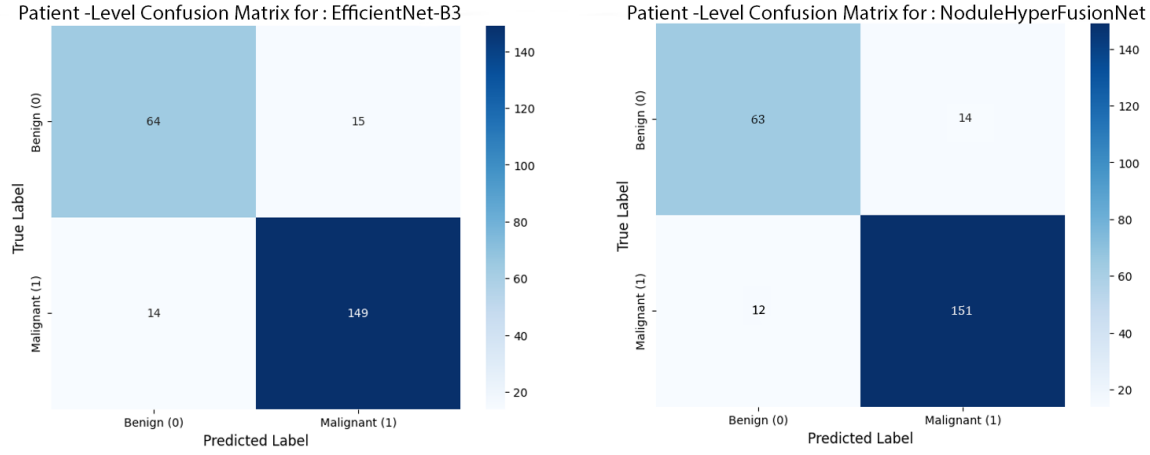

Figure 11. Comparing confusion matrix on test set

Table 7. Final Classification Model Leaderboard

Model	Accuracy	F1 Score	AUC
EfficientNet-B3	0.880165	0.911315	0.939582
MaxViT-Tiny	0.777293	0.825939	0.836118
MaxViT-Large	0.842795	0.880795	0.922785
ViT-Small	0.659389	0.792553	0.793671
ConvNeXt-Tiny	0.655022	0.791557	0.640675
EfficientNetV2-S	0.68559	0.806452	0.635527
CoAtNet-0	0.655022	0.791557	0.816962
MobileNetV3-Large	0.703057	0.805714	0.785907
NoduleFusionNet-B0	0.847162	0.881356	0.916371
NoduleFusionNet-V2	0.834711	0.871795	0.898035
NoduleHyperFusionNet	0.924341	0.891495	0.918135

with AUC values exceeding 0.90, demonstrating their potential for accurate lung cancer classification. Conversely, models such as ViT-Small, ConvNeXt-Tiny, and EfficientNetV2-S showed comparatively lower performance across all metrics, indicating areas for further optimization or that these architectures may be less suited for this specific classification task without significant modifications. It is noteworthy that while NoduleHyperFusionNet achieved the highest overall accuracy (0.924), the EfficientNet-B3 model obtained a slightly higher F1-score (0.911 vs. 0.891). The F1-score is the harmonic mean of precision and recall and is particularly sensitive to the balance between false positives and false negatives for the malignant class. This result suggests that while NoduleHyperFusionNet is better at correctly classifying all cases (especially true negatives), EfficientNet-B3 demonstrates a slightly more balanced performance between precision and recall for identifying malignant nodules, making both models strong contenders depending on the desired clinical application.

The patient-level aggregation strategy proved crucial in handling multiple slices per patient, providing a robust final prediction. The high AUC scores observed in the better-performing models indicate their strong ability to distinguish between benign and malignant cases. The varying performance across models highlights the importance of architectural choice and training methodology for optimal diagnostic accuracy in lung cancer classification.

5. Discussion

The comprehensive evaluation of deep learning models for lung cancer segmentation and classification presented in this study demonstrates significant advancements in automated diagnostic capabilities. Our findings underscore the effectiveness of specialized architectures and multi-phase training strategies in achieving high performance on diverse medical imaging datasets.

For the segmentation task, the superior performance of the UNAR model, evidenced by its Dice Score of 98.40% (Table 6), highlights the critical role of advanced U-Net variants enhanced with attention mechanisms and robust backbones. This result is consistent with the literature indicating that UNet++'s nested and dense skip pathways, combined with attention gates, enable finer-grained segmentation and selective focusing on diagnostically relevant features, which is crucial for accurate nodule delineation [4], [24]. The high Dice Scores across other tested segmentation models like Res34UNetV2 (98.04%) and SegFormer (97.83%) further validate the robustness of deep learning in precisely outlining lung nodules. The visual trends in figure 8, depicting the convergence of loss and stability of Dice and IoU metrics during training, corroborate these quantitative results, demonstrating that these models learn effectively and generalize well to unseen segmentation data.

In the realm of lung nodule classification, the implemented two-phase training process proved instrumental in enhancing model robustness and generalizability. By initially training on ground truth masks (Phase 1) and subsequently fine-tuning on predicted masks (Phase 2), our approach effectively bridges the gap between ideal training conditions and real-world clinical deployment, where segmentation is typically automated. This strategy addresses the inherent inaccuracies of automated segmentation while preparing the classification models to perform optimally with real-time pipeline outputs.

Among the classification models, NoduleHyperFusionNet emerged as the top performer with an accuracy of 92.43%, an F1 Score of 89.15%, and an AUC of 91.81% (Table 7). Its strong performance suggests that specialized fusion architectures, particularly when combined with efficient backbones like EfficientNetV2-S, are highly effective in integrating complex multi-channel features (including global image, masks, contextual crops, distance transforms, and edge maps) for robust malignancy assessment. The EfficientNet-B3 model also demonstrated exceptional discriminative power with an AUC of 93.96%, reinforcing the utility of the EfficientNet family's compound scaling for balancing accuracy and computational efficiency [10], [11]. The training history plots (Figure 9, Figure 10 for NoduleHyperFusionNet) would visually confirm their stable learning dynamics and strong performance progression.

The adoption of Focal Loss was pivotal in addressing class imbalance—common in medical datasets where malignant cases are underrepresented [30], [31]. By emphasizing hard-to-classify examples, it reduced bias toward the dominant benign class and enhanced sensitivity to malignant findings. Confusion matrices (Figure 11) further elucidate model performance by detailing false positives and false negatives, offering clinically valuable insight into error patterns.

While the results are highly promising, certain aspects warrant discussion for future research. The reliance on predicted masks in Phase 2 of classification training, while simulating real-world conditions, means that classification performance is inherently tied to the accuracy of the segmentation model. Further improvements in segmentation could directly translate to enhanced classification accuracy. Additionally, although the dataset size is substantial (17,351 slices, 1,208 patients), further validation on independent, external datasets from diverse clinical settings would strengthen the generalizability and clinical applicability of these models. The computational demands of some larger models, even with optimizations like mixed-precision training on GPUs, might still pose deployment challenges in resource-constrained environments, suggesting a continued need for research into more lightweight yet equally performant architectures.

6. Conclusion

This study successfully established and evaluated a robust deep learning framework for the automated segmentation of lung nodules and the classification of lung cancer, leveraging a diverse set of CT scan datasets. We investigated various state-of-the-art deep learning architectures for both tasks, employing advanced training methodologies, including a two-phase training process for classification models to enhance real-world applicability. For lung nodule segmentation, the UNAR model demonstrated superior performance, achieving an impressive Dice Score of 0.984012. This highlights the effectiveness of integrating attention mechanisms and strong residual backbones within U-Net-based architectures for the precise delineation of complex anatomical structures, a critical step for accurate diagnosis.

In the classification task, our two-phase training strategy, which involved initial training with ground truth masks followed by fine-tuning with predicted masks, proved highly beneficial in bridging the gap between theoretical performance and practical deployment. The NoduleHyperFusionNet model emerged as the top classifier, exhibiting high accuracy 92.43%, F1 Score 89.15%, and AUC 91.81%. The EfficientNet-B3 also showed exceptional discriminative capabilities with a remarkable AUC of 93.96%. The utilization of Focal Loss effectively addressed the class imbalance inherent in medical datasets, further enhancing the models' sensitivity to critical malignant cases. These results collectively underscore the immense

potential of deep learning techniques to improve the accuracy and efficiency of lung cancer diagnostics significantly. The high performance achieved in both segmentation and classification tasks can contribute to earlier detection, leading to improved patient survival rates and facilitating more targeted treatment strategies. Future efforts will aim to improve model generalizability through validation on larger, more diverse external datasets. Research will also explore lightweight architectures to reduce computational overhead for deployment in clinical environments and enhance model interpret-ability to increase trust and adoption among healthcare professionals. Addressing these challenges will further solidify the role of deep learning as a vital tool in lung cancer diagnosis and management.

REFERENCES

1. Abdullahi, K., Ramakrishnan, K. & Ali, A. Deep learning techniques for lung cancer diagnosis with computed tomography imaging: A systematic review for detection, segmentation, and classification. Information (Basel). 16, 451 (2025,5)
2. Gayap, H. & Akhloufi, M. Deep machine learning for medical diagnosis, application to lung cancer detection: A review. BioMedInformatics. 4, 236-284 (2024,1)
3. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes In Computer Science. pp. 234-241 (2015)
4. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2018;11045 LNCS:3-11. doi:10.1007/978-3-030-00889-5_/FIGURES/3
5. Chen, K., Xuan, Y., Lin, A. & Guo, S. Lung computed tomography image segmentation based on U-Net network fused with dilated convolution. Comput. Methods Programs Biomed.. 207, 106170 (2021,8)
6. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. & Luo, P. SegFormer: Simple and efficient design for semantic segmentation with Transformers. (2021,5)
7. Hu, T., Lan, Y., Zhang, Y., Xu, J., Li, S. & Hung, C. A lung nodule segmentation model based on the transformer with multiple thresholds and coordinate attention. Sci. Rep.. 14, 31743 (2024,12)
8. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. Lecture Notes in Computer Science
9. Tan M, Le Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 36th International Conference on Machine Learning, ICML 2019. 2019;2019-June:10691-10700.
10. Devi BR, Ashok KS, Gowda SK, Sumalatha K, Kadiravan G, Painam RK. Efficient NetB3 for Enhanced Lung Cancer Detection: Histopathological Image Study with Augmentation. Journal of Information Technology Management. 2024;16(1):98-117. doi:10.22059/JITM.2024.96377
11. Raza, R., Zulfiqar, F., Khan, M., Arif, M., Alvi, A., Iftikhar, M. & Alam, T. Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. Eng. Appl. Artif. Intell.. 126, 106902 (2023,11)
12. Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T. & Xie, S. A ConvNet for the 2020s. 2022 IEEE/CVF Conference On Computer Vision And Pattern Recognition (CVPR). (2022,6)
13. Cao, H., Liu, H., Song, E., Ma, G., Xu, X., Jin, R., Liu, T. & Hung, C. A two-stage convolutional neural networks for lung nodule detection. IEEE J. Biomed. Health Inform.. 24, 2006-2015 (2020,7)
14. Tan M, Le Q V. EfficientNetV2: Smaller Models and Faster Training. International Conference on Machine Learning. International Conference on Machine Learning 2021 DOI:10.48550/arXiv.2104.00298
15. Hidayah, A. & Wisesty, U. Lung cancer classification based on ensembling EfficientNet using histopathology images. 2024 International Conference On Intelligent Cybernetics Technology & Applications (ICICyTA). pp. 76-81 (2024,12)
16. Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H. & Le, Q. Searching for MobileNetV3. 2019 IEEE/CVF International Conference On Computer Vision (ICCV). (2019,10)
17. Das, S., Salih, A., Bin Sulaiman, R. & Farhan, M. Enhancing lung cancer classification with MobileNetV3 and EfficientNetB7: A transfer learning approach. 2024 International Conference On Computer And Applications (ICCA). pp. 1-8 (2024,12)
18. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. An image is worth 16x16 words: Transformers for image recognition at scale. (2020,10)
19. Zhu, Q. & Fei, L. Cross-ViT based benign and malignant classification of pulmonary nodules. PLoS One. 20, e0318670 (2025,2)
20. Tu Z, Talebi H, Zhang H, et al. MaxViT: Multi-Axis Vision Transformer. Accessed July 19, 2025. https://github.com/google-research/maxvit.
21. Fu, X., Lin, R., Du, W., Tavares, A. & Liang, Y. Explainable hybrid transformer for multi-classification of lung disease using chest X-rays. Sci. Rep.. 15, 6650 (2025,2)
22. Dai, Z., Liu, H., Le, Q. & Tan, M. CoAtNet: Marrying Convolution and Attention for All Data Sizes.
23. Park, I., Kim, W. & Ryu, J. Style-KD: Class-imbalanced medical image classification via style knowledge distillation. Biomed. Signal Process. Control. 91, 105928 (2024,5)

24. Farag, A., Ali, A., Elshazly, S. & Farag, A. Feature fusion for lung nodule classification. Int. J. Comput. Assist. Radiol. Surg.. 12, 1809-1818 (2017,10)
25. Zhang G, Yang Z, Gong L, Jiang S, Wang L, Zhang H. Classification of lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations. Radiologia Medica. 2020;125(4):374-383. doi:10.1007/S11547-019-01130-9/METRICS
26. Armato, S., Mclennan, G., Bidaut, L., Mcnitt-Gray, M., Meyer, C., Reeves, A. & Clarke, L. The Cancer Imaging Archive. The Cancer Imaging Archive. 111 (2015)
27. Antonelli M, Reinke A, Bakas S, et al. The Medical Segmentation Decathlon. Nat Commun. 2022;13(1):1-13. doi:10.1038/S41467-022-30695-9
28. Pedrosa, J., Aresta, G., Ferreira, C., Rodrigues, M., Leitão, P., Carvalho, A., Rebelo, J., Negrão, E., Ramos, I., Cunha, A. & Campilho, A. LNDb: A lung nodule database on computed tomography. (2019,11)
29. Li,Quantitative Computed Tomographic Descriptors Associate Tumor Shape Complexity and Intratumor Heterogeneity with Prognosis in Lung Adenocarcinom Quantitative Computed Tomographic Descriptors Associate Tumor Shape Complexity and Intratumor Heterogeneity with Prognosis in Lung Adenocarcinom
30. Widodo, A., Setiawan, B. & Indraswari, R. Machine learning-based intrusion detection on multi-class imbalanced dataset using SMOTE. Procedia Comput. Sci.. 234 pp. 578-583 (2024)
31. Lin, T., Goyal, P., Girshick, R., He, K. & Dollar, P. Focal Loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell.. 42, 318-327 (2020,2)
32. Alruwaili, M. & Mohamed, M. An integrated deep learning model with EfficientNet and ResNet for accurate multi-class skin disease classification. Diagnostics (Basel). 15 (2025,2)
33. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. 2016 IEEE Conference On Computer Vision And Pattern Recognition (CVPR). pp. 770-778 (2016,6)
34. Jayaram, J., Haw, S., Palanichamy, N., Thillaigovindhan, S. & Al-Tarawneh, M. Lung tumor segmentation in medical imaging using U-NET. Journal Of Informatics And Web Engineering. 4, 140-151 (2025,2)
35. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. & Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers.
36. Dong, L. & Liu, H. Segmentation of pulmonary nodules based on improved UNet++. 2021 14th International Congress On Image And Signal Processing, BioMedical Engineering And Informatics (CISP-BMEI). (2021,10)
37. Zhao, Y. & Lin, L. A Lightweight U-Net for Medical Image Segmentation. 2024 Photonics & Electromagnetics Research Symposium (PIERS). 34 pp. 1-5 (2024,4)
38. Said, Y., Alsheikhy, A., Shawly, T. & Lahza, H. Medical images segmentation for lung cancer diagnosis based on deep learning architectures. Diagnostics (Basel). 13 (2023,2)
39. Sun, H., Zhou, W., Yang, J., Shao, Y., Xing, L., Zhao, Q. & Zhang, L. An improved medical image classification algorithm based on Adam optimizer. Mathematics. 12, 2509 (2024,8)
40. Chi, W., Ma, L., Wu, J., Chen, M., Lu, W. & Gu, X. Deep learning-based medical image segmentation with limited labels. Phys. Med. Biol.. 65, 235001 (2020,11)
41. Peng, J. & Wang, Y. Medical image segmentation with limited supervision: A review of deep network models. IEEE Access. 9 pp. 36827-36851 (2021)
42. Li, W., Cao, P., Zhao, D. & Wang, J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. Comput. Math. Methods Med.. 2016 pp. 6215085 (2016,12)
43. Kljajić, Jovana and Đukić, Nataša and Lazić, Ivan and Lončar-Turukalo, Tatjana and Boban, Jasmina and Rapaić, Milan and Jakovljević, Nikša, "SegFormer Model in Mammography Lesion Segmentation: A Study on the Impact of GLAM Saliency Maps," 2024 11th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), Nis, Serbia, 2024, pp. 1-6, doi: 10.1109/IcETRAN62308.2024.10735436.
44. Q. Li et al., "Performance Evaluation of Deep Learning Classification Network for Image Features," in IEEE Access, vol. 9, pp. 9318-9333, 2021, doi: 10.1109/ACCESS.2020.3048956.
45. Kozah, N., Dornaika, F., Charafeddine, J. & El Jaam, J. Data Augmentation Techniques for Medical Image Segmentation – A Review. 2024 International Conference On Computer And Applications (ICCA). pp. 1-8 (2024,12)
46. Kumar, T., Brennan, R., Mileo, A. & Bendechache, M. Image data augmentation approaches: A comprehensive survey and future directions. IEEE Access. 12 pp. 187536-187571 (2024)
47. Buslaev, A., Iglovikov, V., Khvedchenya, E., Parinov, A., Druzhinin, M. & Kalinin, A. Albumentations: Fast and flexible image augmentations. Information (Basel). 11, 125 (2020,2)
48. Bakr, S., Gevaert, O., Echegaray, S., Ayers, K., Zhou, M., Shafiq, M., Zheng, H., Zhang, W., Leung, A., Kadoch, M., Shrager, J., Quon, A., Rubin, D., Plevritis, S., & Napel, S. (2017). Data for NSCLC Radiogenomics (Version 4) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2017.7hs46erv
49. Grove, O., Berglund, A., Schabath, M., Aerts, H., Dekker, A., Wang, H., Velazquez, E., Lambin, P., Gu, Y., Balagurunathan, Y., Eikman, E., Gatenby, R., Eschrich, S. & Gillies, R. Correction: Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. PLoS One. 16, e0248541 (2021,3)
50. Y. Nishitaki, T. Kamiya and S. Kido, "Identification of Nodular Shadows from CT Images Using Improved CoAtNet Incorporated Clinical Recording," 2023 23rd International Conference on Control, Automation and Systems (ICCAS), Yeosu, Korea, Republic of, 2023, pp. 1727-1732, doi: 10.23919/ICCAS59377.2023.10316787.
51. Li, P., Wang, S., Li, T., Lu, J., HuangFu, Y., and Wang, D. (2020). A Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis (Lung-PET-CT-Dx) [Data set]. The Cancer Imaging Archive. doi: 10.7937/TCIA.2020.NNC2-0461