Journal of Scientific & Industrial Research
Vol. 85, February 2026, pp. 150-163
DOI: 10.56042/jsir.v85i2.26548

Efficient Lung Cancer Identification through Integrated Deep Learning on Residual U-Net Segmented CT Images and Swin Transformer Features

Sunil Kumar[1,2], Harish Kumar[1], Himani[3], Birendra Kumar Saraswat[4] & Shivaji Sinha[5,*]

[1]Department of Computer Engineering, J.C. Bose University of Science and Technology, YMCA, Faridabad 121 006, Haryana, India

[2]School of Engineering and Technology (UIET), CSJM University, Kalyanpur, Kanpur 208 024, Uttar Pradesh, India

[3]Department of Electronics and Communication Engineering, Ajay Kumar Garg Engineering College, Ghaziabad 201 015, India

[4]Department of Computer Science and Information Technology, GL Bajaj Institute of Technology and Management, Greater Noida 201 306, India

[5]Departmentof Electronics and Communication Engineering, JSS University, Noida 201 301, India

Received 13 December 2025; revised 30 December 2025; accepted 03 February 2026

Late detection of lung cancer continues to be the primary reason for its high mortality rate worldwide, which is responsible for almost 85% of all cases. Computed Tomography (CT) imaging is a powerful tool for locating lung nodules and abnormalities in a short time, thus refining the medical diagnostic process. Advances in Artificial Intelligence (AI), especially Deep Learning (DL), have significantly improved the identification of lung nodules in CT imaging. This research aims to develop a diagnostic system capable of accurately predicting lung nodules. The study harnessed two major CT imaging datasets, NSCLC Radiomics and LUNA16, for pattern analysis related to lung cancer. The residual U-Net model was found to be highly effective in segmenting lung cancer regions, with 96.21% accuracy and a Dice Coefficient ($Dice_{co}$) of 0.934, demonstrating its capability to accurately capture the complex features of lung cancer areas. The Swin Transformer and Principal Component Analysis (PCA) are combined to optimize feature engineering for the segmented CT scan images. The Swin Transformer generates feature vectors of a very high-dimensional space, and PCA takes these as inputs, reducing the dimensionality of the feature space and discarding redundant features to facilitate the selection of the most relevant ones. While investigating lung nodule patterns in the NSCLC radiomics dataset, the residual U-Net model in combination with DenseNet169, ResNet50, and ResNet101 models was able to achieve a very high accuracy level and thus perform better than LUNA16. The integrated residual U-Net and ResNet101 demonstrated outstanding accuracy of 98.97%, an F1 score of 96.21%, and a $Dice_{co}$ of 0.946, highlighting its exceptional ability to accurately detect lung nodules.

Keywords: Convolutional neural networks, Lung nodule, Machine learning, Medical imaging, Transformers

Introduction

On a global scale, lung cancer is a major cause of most deaths as a result of all types of cancer. It is highly triggered by changes in the genes in the pulmonary epithelial cells, which induce uncontrolled mitogenic activity and lead to the production of nodular formations or tumor. With the increase of neoplasm size, there is a risk of intrusive growth into the surrounding tissues and the spread of the tumor through both the humoral and lymphatic pathways.[1] Non-Small Cell Lung Cancer (NSCLC) generally exhibits slow progression with localized metastasis, whereas Small-Cell Lung Cancer (SCLC) is marked by rapid growth and early widespread dissemination. Among other subtypes of NSCLC, about 85% of all the incidences of lung cancer.[2,3] CT scans provide the thoracic images with a greater level of resolution, which makes it easier to identify small pulmonary nodules, which may indicate early malignancy. They also offer accurate evaluation of nodule size and structure and the ability to identify temporal variability, therefore supporting prompt diagnosis. In addition, Low-Dose CT (LDCT) is particularly effective in screening patients at high risk, thus improving the chances of detecting lung cancer at a more treatable stage. High-Resolution CT (HRCT) imaging cannot be replaced by other methods due to the accuracy of identification and evaluation of small nodules in the lungs, including their size, shape, and strict location.[4,5]

Traditional radiology workflows have a lot of differences, between the readers and Traditional radiology workflows cannot keep up with the growing amount of the imaging data. Current Computer Assisted Detection (CAD) approaches based on

*Author for Correspondence
E-mail: shivaji2006@gmail.com

Convolutional Neural Networks (CNNs) do a job at recognizing patterns.

Current CAD approaches based on CNNs have trouble capturing range relationships. Transformer models effectively capture complex patterns in data, they typically require extensive datasets and often lack inherent mechanisms for interpretability. These limitations together make models that either do not work well on to the imaging datasets or are not clear enough, for the clinicians to trust models.

The biases and research gaps are addressed by the proposed framework. In order to improve lung cancer identification, the work presents a novel integrated framework that combines the complementary strengths of the residual U-Net, Swin Transformer architecture, PCA, and CNNs. This leads to in better lung segmentation by the residual U-Net, better feature extraction by the Swin Transformer, feature transformation and reduction by PCA, and classification by the CNNs. The proposed approach enhances the detection of lung cancer by automated diagnostic systems through the use of CNNs' robust local feature learning and the hierarchical feature extraction of the Swin Transformer.

Pre-processing involves noise reduction of CT images through median and Gaussian median, which results in an easier way of finding conspicuous pulmonary structures. Greyscale conversion is then used to streamline the data and bring out nodular appearances. Besides, the use of data augmentation strategies is involved to boost the model performance through the inclusion of various and exaggerated expressions of essential pulmonary characteristics.[6–8]

State-of-the-art AI-based systems[9], especially CNNs have shown considerable potential in automatizing the process of lung cancer detection by properly differentiating benign and malignant nodules in large-scale CT groups, and thus add additional value to radiological decision-making with higher accuracy and on-time and timeliness. The integration of CAD systems in the clinical practice enhances diagnostic accuracy and reduces the clinicians' workload thereby increasing patient care efficiency.[10] By use of constant adaptation using new haunches of information, AI-oriented systems will keep in time with the current medical knowledge base, thus making the diagnosis of lung cancer easier earlier, and eventually benefiting patient-care outcomes.[11]

Segmentation isolates the relevant areas of the lungs, and the background noise repression. The U-Net architecture has shown a great significance in improving the accuracy of the lung and nodule delineation process among the plethora of methods.[12] To increase the learning of deep networks, explicit residual connections are placed in the residual U-Net, thus improving the problem of the vanishing-gradient.[13] This approach contributes moderately to the ability of the model to divide the intricate pulmonary structures at a fine scale. The model ensures that oncogenic lung tissue is segmented accurately and consistently by taking granular information at both micro and macro levels. The correct division is essential in order to narrow down on their identification by making sure that the neoplastic areas are well delineated. Further development and use of the advanced methods of segmentation are recommended to increase the effectiveness and individualization of treatment protocols in patients with lung cancer.[14]

The Swin Transformer also known as Shifted Window Transformer is used to promote higher computational efficiency as it limits the self-attention mechanisms to non-overlapping local windows and maintains cross-window connectivity. Its hierarchical design can be used to derive feature maps at various spatial scales, which is beneficial to many tasks such as: object detection, image segmentation and scan classification.[15,16] There are also feature-level hierarchical feature extraction, advanced attention, and global environment modelling that help the detection of lung nodules.[17,18] The Swin Transformer produces both global features and local features explicit through the modelling of contextual relationships at different scales, making it especially appropriate to find more complex patterns of CT images, including the variability seen in lung nodules. When given self-attention, diagnostically salient regions are selectively strengthened improving the fidelity of features and adaptability.[19] Another feature that makes the Swin Transformer popular is its large scale and high-resolution CT image processing with the ability of extracting reliable features used in segmentation and classification processes.[20] However, Swin Transformers do not avoid the following issues like its computational complexity, overfitting, and redundancy of features when used to process high-dimensional CT data. The feature representations generated by Swin Transformer are subjected to PCA to alleviate these problems, which reduces the dimensionality. In integrative models, PCA complements Swin transformer and CNN hybrid

models[21,22], optimizing the feature vectors accordingly, and, as a result, increasing the interpretability.

The study employs the NSCLC Radiomics[23] and LUNA16[24] datasets for the proposed approach. Using these datasets allowed for rigorous validation of the proposed diagnostic method, demonstrating its reliability and accuracy in detecting lung cancer. The proposed architecture combines DenseNet169, ResNet50, and ResNet101 classifiers with a residual U-Net for segmented CT images, followed by a Swin Transformer for feature extraction, which refines these representations to enhance lung nodule diagnosis.[25,26] Furthermore, Grad-CAM increases the visual explanations for forecasts. This study proposes an integrated system which integrates Swin Transformer features, utilizes PCA to reduce dimensionality, and employs CNN architectures for classification, resulting in compact yet relevant representations that capture both local and global patterns for accurate diagnosis. Below is a list of the investigation's significant contributions:

- Employment of the LUNA16 and NSCLC radiomics, because of their varying instances.
- To detect lung cancer in the segmented CT images precisely, a couple of modern tools were used: the CNN architecture was used as a classification approach, and the Swin Transformer was used for feature extraction.
- The application of the residual U-Net enhances the segmentation process.
- The combined approach that integrates Swin Transformers, PCA, and CNNs for the task.
- The research evaluates the integrational/ensemble approach, with a focus on evaluating its effectiveness and enhancements to the diagnostic process.

The work is structured as follows: Section II outlines the studies that support the study. Section III expands on the suggested system approach, including information on datasets, segmentation techniques, the integrational network, and identification procedures. Section IV offers the findings and explains the experimental conditions, while Section V summarizes the research and suggests future directions.

Related Work

Recent DL development can be divided into three categories for the issue: CNN and transformer-based architectures, and their hybridization. This subsection thoroughly explores these methods, taking into account their strengths and weaknesses, and seeing how well they perform in different scenarios.

CNN-based Architectures

CNNs have been widely used in the issue due to its hierarchical feature extraction capabilities.

Investigators used nine unconventional models pre-trained to classify lung cancer from a large dataset of CT scans.[27,28] ResNet50[(29)] solved vanishing gradient issues with a residual connection and obtained the highest accuracy (97.7%), sensitivity (100%), F1 score (97.7%) and AUC (0.999). InceptionV3[(30)] used multi-scales convolutional kernels which span the diverse range of spatial information and excelled above other CNNs, with a precision of 97.9% and a specificity of 98.0%. The analysis demonstrated the role of Grad-CAM to visualize and guide the training process.[31,32] ACNN trained on over 42,000 CT images achieved high sensitivity (94.4%) and specificity (93.9%) in the detection, demonstrating its potential in identifying lesions with potential to be missed in normal radiology.[33] When analyzing manually segmented CT images, the CNN indicated that both DeepLab- v3[(34)] and VGG-19[(35)] achieved better performance than segmentation in terms of accuracy but their computational costs made them unsuitable for real-time applications. In contrast, SegNet provided similar performance to segmentation without being inefficient.[36] Researchers used the idea of developing a combined architecture technique by using MobileNetV2[(37)] as a feature extraction backbone and a Capsule Network[38] to deal with the orientation variance problem, which is a major drawback of traditional CNN classification problems.[39] Architectures like VGG-19 and InceptionV3 struggle with their receptive fields and computational demands, which make it tough for them to capture long-range dependencies and make them less efficient in resource-constrained scenarios.

Transformer-based Architectures

Transformer-based models help to address major limitations of the CNN, which is the lack of global context. This capability allows for the more reliable classification of ambiguous nodules and can generalize more reliably, as it reduces the inductive bias.

The hybridization makes traditional DL methods more efficient, and it focuses on features that are more relevant, resulting in more robust and efficient diagnostics. In one of the review study, LASSO was shown to be the best feature selection method for nodules based on thermography images; in particular, it was combined with the Random Forest and

XGBoost classifiers.[22] Research has also found that ViT[40] and CNNs were able to distinguish lung tumours. The architecture was created to extract key information in an encoder-decoder framework. It used convolutional and matching blocks to better extract features. Transformers possess self-attention mechanisms which effectively capture complex global features. The model obtained average $Dice_{co}$ of 0.7468 and 0.6847 and Hausdorff distances of 15.336 and 17.435, which showed its decent performance on both public NSCLC-Radiomics dataset[23] and local hospital dataset.[16] The Swin Transformer[41] model is extremely useful for lung cancer classification and segmentation, and it demonstrates versatility between tasks. After pre-training, the Swin-B variant achieved a classification accuracy of 82.26%, which is 2.529% ahead of ViT in top ranked prediction. Additionally, the Swin-S variant outperformed other segmentation methods, which proved that pre-training increases the accuracy of Swin Transformer. The study heavily relied on a large, 888 low-dimensional figures for CT images including annotations as the foundation of study called the LUNA16 dataset. Both the models, Swin-B and Swin-S, have shown an improvement over ViT, with the former achieving an accuracy of 82.26% and the latter showing significant advances in overlapping measurements, in this case measured as the mean Intersection over Union (mIoU). For the future, it would be worthwhile to investigate 3D medical image classification.[20] To automate the system, a systematic way to create transformer-based models was presented in the work. Experimental findings indicated that the pre-trained ViT feature extractor performed better than the CNN-based encoder, i.e., DenseNet121. Moreover, the utilization of both frontal and lateral images as dual-view input was more effective compared to single-input approaches.[32] Researchers have developed three interdependent deep-fusion learning algorithms specifically to identify lung nodules in CT images. These include Multi-Perspective Fusion (MPF), Single-Feature Multi-Perspective Fusion (SFMPF), and Multi-Feature Multi-Perspective Fusion (MFMPF), with each reflecting a distinct hierarchical structure. The models were tested on bilateral, trilateral, Gabor, and LOG-filtered pictures, which resulted in a complete multi-feature, multi-perspective hierarchical deep fusion learning model.[42] Despite their strengths, transformers require large-scale data and incur high computational costs, limiting scalability and practicality for high-dimensional imaging in data-constrained settings.

Hybrid Architectures

Hybrid neural networks like CDC-HNN used 3D-CNNs on datasets such as LIDC-IDRI and LUNA16 to extract features from CT scans. This method identified cancer at an early stage and accurately, the diagnostic accuracy rate is up to 95%.[43] Guided by the awful noise present in the CT images, a technique called guided bilateral filtering is used to reduce noise in the image, and prepare it for analysis. A transformer aided generative adversarial network (T-GAN) detected various types of lung cancer and DyLF-CO fine tunes a trained model. The results of the python implementation were 99.70% accuracy; 99.60% precision, 99.80% specificity, 0.104 RMSE and runtime 120 seconds.[31] SwinResNet was a hybrid model that combines the Swin Transformer and ResNet models. By leveraging their strengths, it becomes even better feature extraction than many traditional implementations. The architecture made use of the multi-head attention of the transformer and the residue learning of the ResNet to capture strong representations. Both the Swin Transformer and ResNet were encoders, extracted features at multiple levels. These features were then enhanced by receptive field blocks and aggregation modules that facilitates a synergy between them to achieve superior performance.[44] The combination of Support Vector Machine (SVM) and CNN exhibited good ability of classification. On the LUNA16 dataset, this hybrid approach was able to achieve readable accuracy rates of 94.00% and 94.5%, respectively, and hence its effectiveness.[45] Hybrid architectures are also known to have increased training complexity and reduced interpretability, which can limit their scalability and transparency in the applications.

Materials and Methods

The proposed pipeline performs lung nodule segmentation using a residual U-net, and then extracts discriminative representations using a Swin Transformer. After that, it uses PCA for efficient dimensionality reduction and application of DenseNet169, ResNet101 and ResNet50for classification. This provides for accurate and reliable lung cancer detection. All the steps of the integrational approach are illustrated in Fig. 1.

The combination of the methods, utilize a systematic methodology to get the precise delineation

of lung cancer regions and malignant nodules. Different DL models and feature extraction methods for precise lung nodules segmentation and identification step by step with good details about each phase in the algorithm illustrated through Algorithm 1.

Algorithm 1: LungNoduleDetection(CT_images)

$\mathcal{M} \leftarrow$ *LoadImages*(CT_images) // Load Input Images from dataset
$\mathbb{R}C \leftarrow \Re U.Segment(\mathcal{M})$ // Segment Lung Cancer Regions
$\mathcal{F} \leftarrow$ *ST.ExtractFeatures*($\mathbb{R}C$) // Extract Features from Segmented Regions
PS ← *PCA.Transform*($\mathcal{F}$) // Reduce and Select Features Using PCA
$\mathbb{R}\mathcal{F} \leftarrow$ *PCA.Select*(*PS*)
$\mathbb{R}\_1.Result \leftarrow \mathbb{R}\_1.Predict(\mathbb{R}\mathcal{F})$ // Identify Lung Nodule Patterns Using integrational Models
$\mathbb{R}\_2.Result \leftarrow \mathbb{R}\_2.Predict(\mathbb{R}\mathcal{F})$
$\mathbb{R}\_3.Result \leftarrow \mathbb{R}\_3.Predict(\mathbb{R}\mathcal{F})$
$\mathcal{E} \leftarrow$ *Choose_best*($\mathbb{R}$_1.Result, $\mathbb{R}$_2.Result, $\mathbb{R}$_3.Result) // Choose the best Performer
$\mathcal{S} \leftarrow \mathcal{E}.GetFinalResult()$ // Determine Final Identification Result
Return $\mathcal{S}$// Return Final Results

Abbreviation:$\Re U$: Residual U-Net model, $\mathcal{M}$: Input image CT dataset (lung CT scans), $\mathbb{R}C$: Segmented lung cancer regions (output from $\Re U$), *ST*: Swin Transformer, $\mathcal{F}$: Feature set extracted using Swin Transformer, *PS*: Principal Components obtained from PCA, $\mathbb{R}\mathcal{F}$: Reduced feature set after PCA, *R*_1: DenseNet-169 model, *R*_2: ResNet-50 model, *R*_3: ResNet-101 model, $\mathcal{E}$: Choose the best Performer, $\mathcal{S}$: Final results.

Dataset

The NSCLC Radiomics[23] and LUNA16[24] datasets, that provide expert-annotated CT scans for lung cancer analysis, were used in the research. LUNA16 has 888 images from 630 patients with a variety of nodule annotations, whereas NSCLC-Radiomics contains nodule-containing slices from 1,010 patients with radiologist-validated tumor masks. In order to ensure objective assessment, all patient-level data were divided into training (70%), validation (15%), and testing (15%) sets. The number of CT scan instances used in the research is summarized in Table 1.

To ensure there was no data leakage, CT slices were split up to get the training and test sets. Augmentation was applied to the training data. It is actually increasing the size of the supplied images, reaching 10,316 images for NSCLC-Radiomics. For LUNA16, it went up to 1,244 images. The test sets were unchanged, though. It ensures unbiased evaluation when checking the results at the end.

Preprocessing

CT images were processed through a unified preprocessing pipeline comprising noise suppression via median and Gaussian filtering, intensity normalization, grayscale conversion, and Hounsfield Units (HU) windowing (−1000 to 1000) to enhance

Table 1 — CT instances in the investigation

NSCLC Radiomics		LUNA16	
Dataset ratios	CT images	Dataset ratios	CT images
Training	5,158	Training	622
Validating	1,105	Validating	133
Testing	1,105	Testing	133
Total	7,368	Total	888

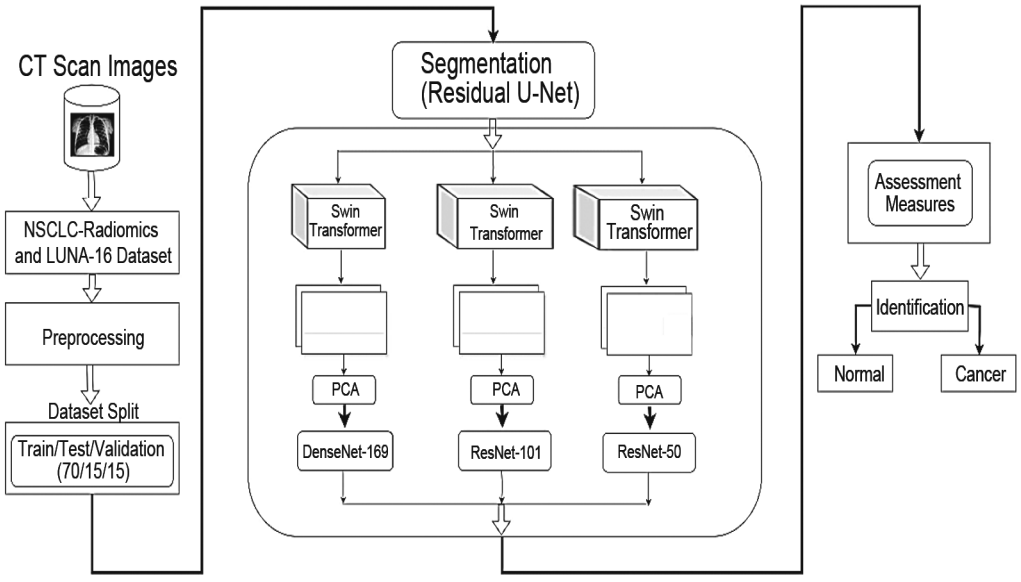

Fig. 1 — Hybridization methodology for the identification of the lung cancer

lung structure visibility and ensure input consistency. Controlled data augmentation is applied during preprocessing to create a broader, clinically plausible set of training examples, this helped the model learn more stable, generalizable features, reduced overfitting to scarce classes, and improved overall reliability in a way that reflects realistic variation in the imaging.

Segmentation

Prior to segmenting the lung, the CT images are preprocessed for better quality data. Pulmonary segmentation separates the lungs from other structures including ribs, arteries and veins.[46] Precise segmentations of lungs and lungs nodules is very important to be able to identify their location in CT scans, which would allow lung cancer to be detected and diagnosed earlier. Comprehensive segmentation is useful to differentiate between the cancerous nodules and non-cancerous structures, thereby improving diagnostic accuracy as well as medical treatment plan.[13,14]

Residual U-Net

The residual U-net enhances the training process and improves the segmentation accuracy by combining both the U-net encoder-decoder and the residual connections. This combination of capturing context efficiently and localizing features precisely is useful for detailed segmentation tasks.[13,46]

During the encoder phase, each layer will process its input using a number of convolution operations. After that, Batch Normalization (BN) is applied and Rectified Linear Unit (ReLU) activation helps to increase the feature extraction. The process could be represented as a mathematical equation**:**

$$O_i = \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}(I_i - 1)\right)\right) \quad \ldots (1)$$

where, O_i is the output, *Conv* is the convolution and I_i is the input.

MaxPooling-down samples the feature maps without losing spatial resolution of the maps but retaining key elements in the map that helps to code context well. Residual connections help to enhance U-Net by aiding in gradient flow and stabilizing the training process, they do this by reusing features from previous layers. The residual relation is represented as:

$$R_o = O_i + (I_i - 1) \quad \ldots (2)$$

where, R_o is the Residual output.

By retaining information, such connections make it easier to train deeper networks. In the decoder finer detail is recovered by combining up-sampled feature maps with their encoder counterparts. This concatenating is written as:

$$C_i = U_s(O_i) \oplus (O_{i-1}) \quad \ldots (3)$$

where, C_i denotes a feature coming by concatenating, U_s is the up-sampling operation.

After it, the combination of features is refined through convolutional layers, which retain spatial information. The resulting feature maps are then optimized for combination with Swin Transformer.[14]

Swin Transformer

The Swin Transformer is a novel breakthrough in ViTs and a significant step towards better understanding of images. It is a sophisticated ViT that has a hierarchical design with shifted windows. This system works by analyzing images by dividing them into small chunks known as image patches. It then adds certain features of the image to understand the overall context by recognizing important visual elements. The combination of layered and adaptive modules changing positions have shown good interest.[41] Its assessment ability for complex patterns and abnormalities in CT images is due to its excellent focus mechanism and multi-layered feature representation.[17] By analyzing detailed image thoroughly both locally and globally, it is able to recognize the signs of infection like formation of abnormal tissue, which can detect abnormality. With tremendous training datasets, the Swin Transformer can successfully differentiate between normal lung tissue and abnormal growths in the lung[18] by learning the patterns and variation from the dataset.

The Swin Transformer consists of following a set of procedures to extract features well both from specific image information and overall context, respectively. The following describes the main stages, including feature extraction and analysis, involved in the Swin Transformer's image processing:

Stage 1. Input Image Processing: The Swin Transformer first divides the image into patches of a certain size, such as the square patch with dimensions of s × s, which refers to the patch size. Next, it forms a flat vector by combining these patches that best represents the features of the image.

First, the input image I is represented as a tensor of dimensions of H×W×C, where H is the height of the image, W is the width, and C is the number of image channels.[41] These patches are formed as follows: the image is divided into non-overlapping patches of size s × s, for the N patches. The number of patches i.e. N is set by the image dimensions and patch size:

$$P = P_1, P_2, P_3 \dots \dots \dots . P_N \quad \dots (4)$$

where, *P* is individual patch.

Stage 2. Linear Embedding: With a linear layer, for each patch P*i* (where *I* ranges from 1 to N), it is flattened and projected into a D dimensional embedding space.

$$Zi = Linear(Pi) \quad \dots (5)$$

As a consequence, a series of embeddings distinguished with Z are produced.

Stage 3. Hierarchical Feature Extraction: The Swin Transformer adopts a hierarchical representation to analyze the images at different scales. It combines and decreases data resolution as it gathers details on various scales in order to capture local and global contexts well. Each stage consists of multiple Swin Transformer blocks, which consists of the windowed multi-head self-attention and shifted windows:

- *Windowed Multi-Head Self-Attention (W-MSA):* W-MSA calculation is performed over discrete windows, which do not overlap. By utilizing multiple attention heads, the MSA enables the model to simultaneously pay attention to different sections of the image, making it better able to capture complex visual patterns. Each attention head picks out specific aspects of interactions between patches, helping the model pick up on complex patterns and relationships. The following method calculates the attention scores of each window:

$$Attention\ (Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{D_k}}\right) V \quad \dots (6)$$

where, Q, K, and V represent the queries, keys, and values obtained from the input embeddings, respectively. D_k denotes the dimension to which the keys are assigned.[41]

- *Shifted Windows:* An alternative approach to capture cross-window information entails shifting the windows to alternate levels, allowing for a more comprehensive context. The model improves its capacity to aggregate characteristics across regions by moving the input embeddings for the next layer.[19,41]

Stage 4. Layer Normalization: A particular Swin Transformer block consists of layer normalization and a Feed-Forward Network (FFN), which is defined as:

$$FFN(P) = ReLU\left(x * W_1 + b_1\right) * W_2 + b_2 \quad \dots (7)$$

The weight matrices are W_1 and W_2, and the biases are b_1 and b_2.[19]

Stage 5. Down-sampling: The spatial dimensions are effectively reduced while the feature depth is increased by down-sampling the feature maps after a succession of blocks using a patch merging strategy.

$$Z_{out} = Concat(MaxPool(Z_i), AveragPool(Z_i)) \quad \dots (8)$$

After completing the processing stages, a classification system receives the final feature representation.[41]

Fusion of Swin Transformer and PCA for Feature Extraction

The Swin Transformer balances between local and global context by using shifting windows while it processes images. This technique is very effective in recognizing complex features. In addition, the model is able to process high-resolution CT images while keeping the segmentation accurate, which is a must for the detection of small lung nodules. PCA redefines the data in a new coordinate system with the most significant features, starting from the one that explains most of the variation and then the following ones. This method helps in data simplification by lowering its complexity and at the same time keeping the important features.[41,42] PCA was performed on the standardized Swin Transformer features to de-correlate variables preserving dominant variance, thus realizing noise reduction, overfitting alleviation, as well as compact representation, simultaneously, which in turn leads to classification efficiency and interpretability increase.

Leveraging the top features of both Swin Transformer and PCA produces a superior feature extraction pipeline for segmented CT scan images. First, the Swin Transformer extracts rich and complex features. Then PCA delves into these features to lower dimensionality and drop redundant ones. PCA enriches these features by reducing the feature space, removing duplicates, and discarding less important features. Improved effectiveness and precision in feature extraction directly improve the reliability of diagnostic models and enhance the accuracy.

Performance Measures

Performance metrics measure the efficiency of a DL model. The performance indicators include different metrics: recall measures retrieval, accuracy reflects correctness, precision indicates exactness, and the F1 score is a combination of precision and recall.[5] $Dice_{co}$ evaluates the performance of the segmentation task by measuring the similarity between the expected and actual overlaps and results.[10]

Results and Discussion

The proposed model effective in identifying lung cancer in the segmented lung regions of CT images as mentioned stepwise.

Preprocessing

An in-depth analysis of the CT images was performed to ensure the detection of lung nodules was both precise and reliable. To address the problem of class imbalance, a variety of data augmentation techniques were implemented. The preprocessing involved changes in intensity, random flips, zooms, and rotations. After the data were split, augmentation methods including rotation around ±15°, scaling from 0.9 to 1.1, and intensity modification (around ±20% for brightness and ±15% for contrast) were applied to the training set. Thus, the number of images in the NSCLC Radiomics training data was increased from 5,158 to 10,316 and in LUNA16 from 622 to 1,244 images. These steps not only adjusted the class distributions but also increased the model's robustness. Expanding the dataset in this way was successful in increasing the diversity of the training data, which in turn improved the robustness of the proposed models. Moreover, it adjusted the brightness of the CT scan within the range of −1000 to 1000 HU to provide the best image quality for the correct identification and segmentation of lung nodules. Normalization played an important role in the accurate detection and segmentation of lung nodules. It not only helped in keeping uniform brightness and contrast for all the images but also contributed to the results being accurate. These preprocessing steps were the foundation on which the proposed method was largely built, as they allowed models to be trained on a wide range of high, quality data, which in turn ensured their accuracy and performance. The dataset division at the patient level and the augmentation usage after the split were critical steps taken to avoid any data leakage, thereby preserving the model training process's integrity. Moreover, the validation on different datasets supported the model's ability to generalize to unseen data, hence, its reliability and applicability.

Hyper-Parameters

CNNs such as DenseNet121, ResNet50, and ResNet101 employed certain specific hyperparameters for the enhancement of the identification process. This study deliberately tested various hyperparameter combinations through a grid search to optimize the model's performance, thereby making a compromise between accuracy and generalizability. The hyperparameters, which illustrate the leading configurations for performance maximization, are presented in Table 2. This exhaustive procedure was a cautious consideration of numerous parameter choices and it thus strongly establishes the base for future enhancements.

Segmentation and Classification

The primary objective of the investigation was to assess the model in locating and mapping lung cancer areas for exact diagnosis. In the first column, there are original CT scans, in the second column, the validation ground truth, and in the third column, residual U-Net predictions. The model figures out better boundary delineation, thus producing extremely precise segmentation, and higher precision. Original images are used as a standard to measure the accuracy of the segmentation, which is a vital step in verifying the precision of the segmentation results, as seen in Fig. 2.

Residual U-Net successfully combined the exact segmentation features of U-Net with the more revolves on residual learning aspects. As a result, this mixture gives rise to a model that is suitable for imaging applications because of its strong segmentation accuracy and efficiency. The design improved segmentation accuracy by accurately delineating boundaries, facilitating training through efficient learning processes, and being a versatile tool for different segmentation tasks.

A statistical experiment was performed with four performance metrics: standard deviation (STD),

Table 2 — Hyper-parameters

Terms	Instances
Learning rate	0.001
Decay	0.01
Batch size	128
Optimizer	AdamW
Dropout rate	0.5
Loss function	Binary Cross-Entropy
Number of epochs	100

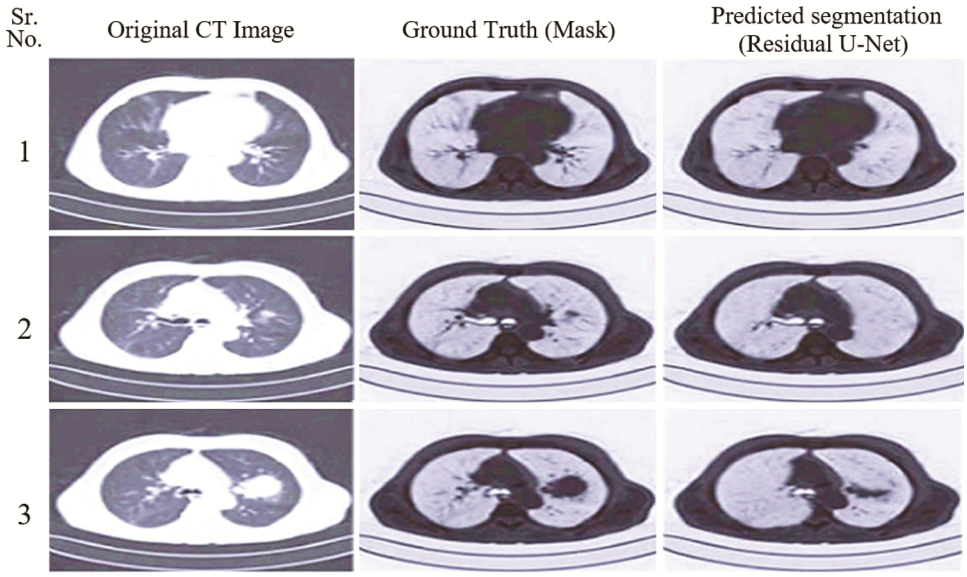

Fig. 2 — Segmentation of the supplied CT images done by the residual U-Net

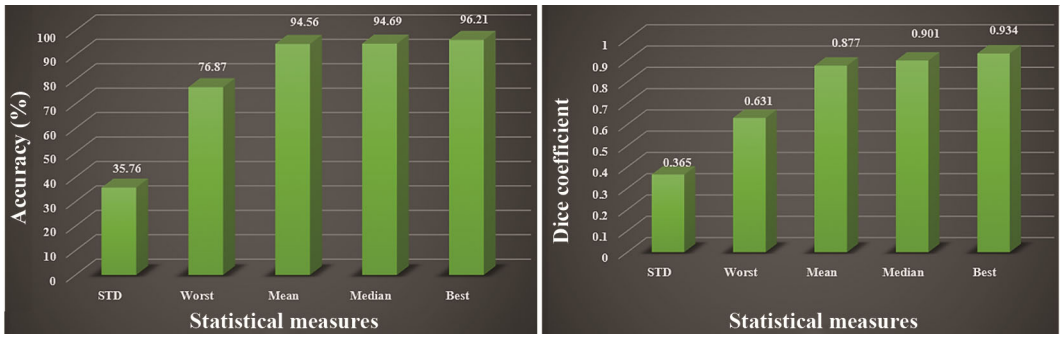

Fig. 3— Segmentation results achieved from residual U-Net

worst, mean, median, and best. The accuracy of lung cancer lesion segmentation by the residual U-Net is illustrated in Fig. 3.

The residual U-Net model reached an accuracy of 96.21% and a $Dice_{co}$ of 0.934 by preserving spatial information via skip connections. The model successfully fused the distinctive features of lung cancer lesions, thus, increasing the accuracy of the area segmentation with dense-flattened neurons. For confirmation of the findings, this study analyzed a Swin Transformer that is a better option for extracting features from data.

The Swin Transformer calculated features from the segmented regions, including shape properties, and texture composition, intensity distribution, edge characteristics, spatial distribution, and some other supplemental features. These features offer a total representation of areas affected by lung cancer, thus, they allow more accurate analysis and classification. The extracted features were spectacular, reflecting the potential of this combined approach to the problem of lung cancer detection and analysis. PCA was employed to increase performance and specify the factor set by selecting the most influential factors that determine data diversity. This step made the model simpler, less computationally complex, and reduced the probability of the overfitting. The principal components, accounting for 95% of the total variance, were able to preserve the maximum possible amount of information from the original feature set.

The initial step of scaling down the image to 224×224 resulted in 3136 patches for each image, which allowed for very detailed feature extraction—each patch, being a part of a feature vector, led to an extremely large number of features per image and the entire dataset. To reduce the number of features, PCA was applied to the feature set, which kept 95% of the variance and thus gave around 249 features. The

transformation greatly simplified the feature space while it still contained the necessary information for an accurate investigation. The study suggests that the residual U-Net with Swin Transformer for feature extraction and PCA for feature reduction and selection used is a very effective method.

To confirm the usefulness of the reduced feature set in locating lung nodules, this study employed ResNet101, DenseNet169, and ResNet50 models for the processed features. The method of using PCA for feature reduction, Swin Transformer for feature extraction, and the U-Net for segmentation is a promising approach in lung cancer detection. The efficiency of these models were tested, which were combined with the integrated architectures that were also used previously: ResNet101, **DenseNet169, and** ResNet50. The association between features and the DenseNet169, ResNet101, and ResNet50 models can be seen in the F1 scores, recall, accuracy, and precision, which were used as metrics for their performance in the detection of lung nodules. This method serves as a compelling diagnostic tool, as it not only enhances computational efficiency but also notably increases the accuracy of lung cancer diagnosis. Both LUNA16 and NSCLC Radiomics datasets were employed to thoroughly assess the diagnostic performance of the proposed method. The framework was efficiently achieved leveraging pre-trained transformers, targeted augmentation, and cross-dataset validation, which can then be a very powerful method even for small datasets such as LUNA16. Identification results on the LUNA16 dataset through different methods are presented in Table 3.

The combination of residual U-Net + ResNet101 performed better than the combinations of residual U-Net + DenseNet169 and residual U-Net + ResNet50, as well as the ResNet101 network in terms of performance. To further boost the model's performance and also get more dependable outcomes, it would be a noble idea to investigate the use of another extensive dataset, for example, NSCLC radiomics, besides the good results on the LUNA16 dataset. The detailed findings derived from the NSCLC radiomics, which reveal how the models performed, are presented in Table 4.

The evaluation of different models' ability to detect lung cancer in CT scans from the NSCLC, Radiomics dataset led to several notable conclusions. The combination of DenseNet169 and residual U-Net architecture resulted in 97.67% accuracy, 95.31% F1 score, and 0.931 $Dice_{co}$. However, the residual U-Net

Table 3 — Identification outcomes (LUNA16)

Model	Accuracy	Precision	Recall	F1 Score	$Dice_{co}$
Residual U-Net + DenseNet169	91.71	89.99	96.11	92.63	0.828
Residual U-Net + ResNet101	95.81	86.43	89.72	95.05	0.859
Residual U-Net + ResNet50	93.64	92.57	93.71	92.99	0.837

Table 4 — Identification Outcomes (NSCLC Radiomics)

Model	Accuracy	Precision	Recall	F1 Score	Diceco
Residual U-Net + DenseNet169	97.67	94.21	96.54	95.31	0.931
Residual U-Net + ResNet101	98.97	97.87	96.01	96.21	0.946
Residual U-Net + ResNet50	97.21	95.01	96.98	95.09	0.929

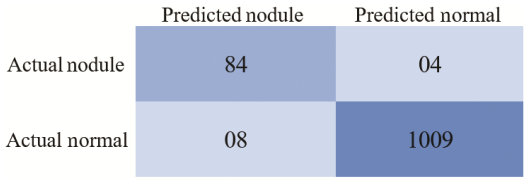

Fig. 4 — Residual U-Net and ResNet-101 confusion matrix

architecture with ResNet50 had somewhat lower performance metrics: 97.21% accuracy, 95.09% F1 score, and 0.929 $Dice_{co}$. The combination of a residual U-Net architecture and ResNet101 performed better than the other models with an accuracy of 98.97%, an F1 score of 96.21%, and a $Dice_{co}$ of 0.946. Although another dataset (LUNA16) is of high quality, it underperformed compared to the NSCLC Radiomics dataset. The larger and more comprehensive dataset of NSCLC radiomics significantly improved the reliability and accuracy.

The presented confusion matrix provides a visual representation of the performance evaluation of the residual U-Net + ResNet-101 model, it shows the model's ability to correctly and incorrectly classify cancerous regions and illustrated through Fig. 4.

The study results show that the residual U-Net + ResNet101 model is the most efficient among different setups in finding cancerous areas in lung tests for NSCLC. The residual U-Net + ResNet101 model can be considered the most reliable and effective in detecting cancerous areas in lung tests for NSCLC, which is corroborated by its higher accuracy rate and $Dice_{co}$, the two most important metrics that reflect its precision in locating the cancerous areas.

The residual U-Net and ResNet101 combined model exhibited enhanced generalization with an

accuracy of 98.97% and a $Dice_{co}$ of 0.946. This performance was achieved through a patient, based division of the data into training (70%), validation (15%), and testing (15%) sets. The minute difference of 0.67% between the validation accuracy of 98.30% and the test accuracy of 98.97% reflects stability of the model, as can be seen in Fig. 5. For the LUNA16 dataset, it obtained an average accuracy of 95.81%, with less variation, thus demonstrating the Swin Transformer, PCA, and CNN model's capacity for small datasets. This improvement is due to the detailed statistical analysis brought in by cross validation, which provides reliable performance evaluation. Depictions of accuracy and loss values for the residual U-Net + ResNet101 model during training, testing, and validation are provided in Fig. 5.

The findings from the research confirm that the proposed combined U-Net and ResNet101 model projections are most similar to the initially identified nodules, thereby evidencing the capability of the proposed fused model.

The outcomes of lung nodules in test images from different CT scans are presented in Fig. 6. It

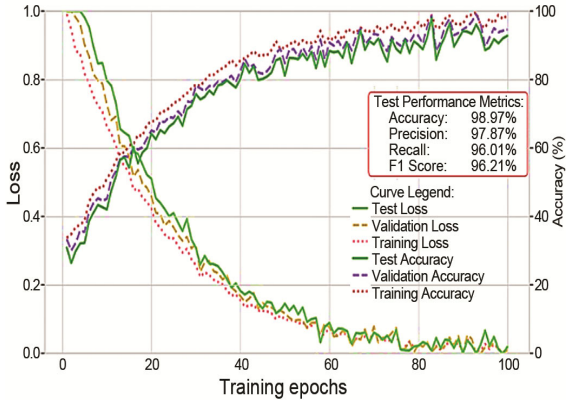

Fig. 5 — Training, validation and test loss and accuracy of the Residual U-Net + ResNet-101 model

comprises the original test image, the Grad-CAM output, the predicted nodule outcomes of the residual U-Net + ResNet101 model, and the ground truth mask. The figure demonstrates the use of Grad-CAM to assist the interpretation of CT scans by visually localizing the most relevant parts. The heat-maps are overlaid on the grayscale CT images to show the important areas that the model has identified. The highlighted areas of this segmentation delineate the parts that are crucial for an accurate evaluation. This visual representation enables you to better comprehend the model's focus, thus, its interpretations become more transparent and reliable.

Ablation Study

An ablation analysis on the NSCLC, Radiomics dataset illustrates that every single element of the proposed framework has a meaningful contribution to the overall performance. Single CNNs (ResNet50, DenseNet169, and ResNet101) trained on raw CT images only achieve limited accuracy, thus directly classifying without structured feature refinement turns out to be an insufficient approach. The use of Residual U-Net segmentation leads to a significant performance increase around 6%, thereby confirming the crucial role of accurate region isolation. The subsequent addition of the Swin Transformer lifts accuracy even further as the model global contextual dependencies; however, it generates high, dimensional representations. Hence, PCA serves as a perfect solution to this problem as it lowers the dimensionality of the features while still keeping the variance, which in turn leads to a more efficient and faster inference without losing accuracy. The combined effect of SwinTransformer, PCA with ResNet101 leads to the highest performance thus, the best single and partial configurations are

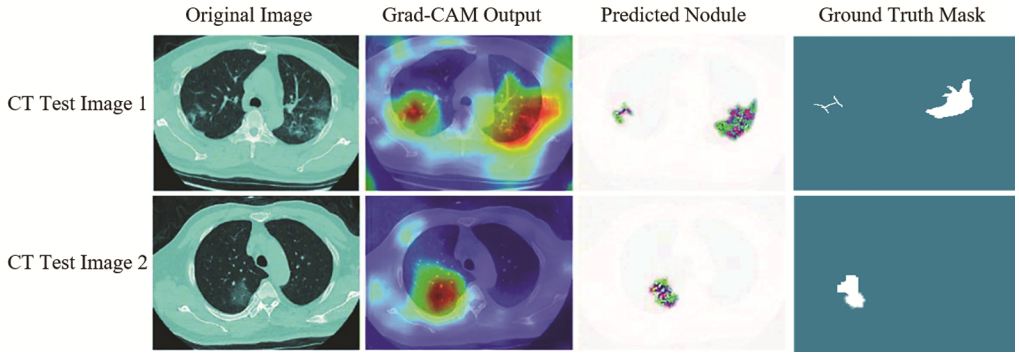

Fig. 6 — Identification results of the lung nodules by Residual U-Net + ResNet101

Table 5 — Comparative analysis

Model[Ref]	Dataset	Acc.	F1	$Dice_{co}$
UPerNet + Swin Transformer[20]	LUNA16	82.26	—	—
SVM+CNN[45]	LUNA16	94.00	94.5	—
3D-convolution neural network (3D-CNN)[42]	LUNA16	95.00	—	—
MobileNet-based multitask learning with attention mechanism (MN-MTL-A)[53]	NSCLC	91.40	—	—
ViT + CNN[16]	NSCLC	—	—	0.746
DSC-L[44]	NSCLC	—	—	0.82
Automated lung tumor detection and 3D volumetric segmentation[48]	NSCLC	—	—	0.82
Residual U-Net + Swin Transformer + PCA + ResNet-101[Proposed]	LUNA16 and NSCLC	95.81 and 98.97	95.05 and 96.21	0.859 and 0.946

outperformed. Testing cross dataset on LUNA16 shows that generalization is still strong even though there is variability in annotations and the sample size is small. Removal of any module results in a drop albeit small, thus the power of the proposed method lies in the segmentation, global context modeling, dimensionality reduction, and deep residual learning interaction which is a coordinated one.

Comparative Analysis

This research accentuated the classification accuracy and F1 scores, with the $Dice_{co}$ being the main performance metric for segmentation. The comparison table illustrated an in-depth evaluation of various methods for CT image classification, primarily focusing on lung cancer diagnoses. A comparative study of the proposed integrational method (Residual U-Net + ResNet101) with the selected alternatives on NSCLC and LUNA16 is presented in Table 5.

The proposed research accomplished both classification and segmentation tasks, whereas existing research[20,45,42] was limited to classification[16,46,48] and focused on segmentation only.

Overall, the Residual U-Net + ResNet101 model's results demonstrate its capability to accurately classify and precisely segment CT images of lung cancer, outperforming other models. The findings indicate that the two models, i.e., the proposed method and UPerNet + Swin Transformer, need more work to further enhance the accuracy of segmentation.[50]

Conclusions

This investigation looks at how machine learning and computer imaging facilitate finding lung malignancies early. Since lung malignancies cause predominant deaths, finding them timely is really important. The framework here leverages a few specific tools, improving accuracy. It leverages residual U-Net for segmentation, Swin Transformer for feature encoding, principal component analysis for dimensionality optimization, and classification using employed CNNs. Basically, these approaches work together to spot lung malignancies better. To validate the work, it was tested on the LUNA16 and NSCLC Radiomics datasets. The results established, it is pretty for cross-dataset robustness, also. A comparative analysis between the proposed approach and ResNet101 identified the ensemble backbone as the superior model for nodule classification and demonstrated efficacy at delineating the tissue. Key limitations, like model generalization and computational overhead, translational barriers. For future steps, the work should really focus on architectural refinement. The best way to do this is by using larger-scale datasets. Also, one should look into things like generative adversarial networks for synthetic data synthesis. The findings clearly demonstrate that the proposed deep neural networks provide substantial utility in this context. The approach enhances diagnostic precision by improving accuracy and facilitates early disease intervention, which is critical for timely detection and effective clinical management.

References

1 World Health Organization, The top 10 causes of death, https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (1 Jan 2026).

2 Clark S B & Alsubait S, Non–Small cell lung cancer, StatPearlsNCBI, https://www.ncbi.nlm.nih.gov/books/NBK562307 (2 Jan 2026).

3 American Lung Association, *Types of Lung Cancer*, https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/basics/lung-cancer-types (6 Jan 2026).

4 Ackah K J, Diab M, Elbrow K, Lewis E & Marchbank A, 203 Should interactive 3D-CT models be used as an alternative to repeat contrast computed tomography in lung cancer screening programme patients for surgical planning?, *Lung Cancer*, **190** (2024) 107764, doi: 10.1016/j.lungcan.2024.107764.

5 Kumar S, Kumar H, Kumar G, Singh S P, Bijalwan A & Diwakar M, A methodical exploration of imaging modalities

from dataset to detection through machine learning paradigms in prominent lung disease diagnosis: A review, *BMC Med Imaging*, **24(1)** (2024) 30, doi: 10.1186/s12880-024-01192-w.

6 Wang J, Sourlos N, Zheng S, Velden N V D, Pelgrim G J, Vliegenthart R & OoijenV P, Preparing CT imaging datasets for deep learning in lung nodule analysis: Insights from four well-known datasets, *Heliyon*, **9(6)** (2023) e17104, doi: 10.1016/j.heliyon.2023.e17104.

7 Mary A A & Thanammal K K, Lung cancer detection via deep learning-based pyramid network with honey badger algorithm, *Meas Sens*, **31** (2023) 100993, doi: 10.1016/j.measen.2023.100993.

8 Zhou X, Wang X, Ma H, Zhang J, Wang X, Bai X, Zhang L, Long J, Chen J, Le H, He W, Zhao S, Xia J & Yang G, Customized T-time inner sampling network with uncertainty-aware data augmentation strategy for multi-annotated lesion segmentation, *Comput Biol Med*, **180** (2024) 108990, doi: 10.1016/j.compbiomed.2024.108990.

9 Alamgeer M, Mengash H A, Marzouk R, Nour M K, Hilal A M, Motwakel A, Zamani A S & Rizwanullah M, Deep learning enabled computer aided diagnosis model for lung cancer using biomedical CT images, *Comput Mater Contin*, **73(1)** (2022) 1437–1448, doi: 10.32604/cmc.2022.027896.

10 Sridevi S & Rajiv Kannan N A, Development of 3DTDUnet++ with novel function and multi-scale dilated-based deep learning model for lung cancer diagnosis using CT images, *Biomed Signal Process Control*, **94** (2024) 106243, doi: 10.1016/j.bspc.2024.106243.

11 Mohandass G, Krishnan G H, Selvaraj D & Sridhathan C, Lung cancer classification using optimized attention-based convolutional neural network with DenseNet-201 transfer learning model on CT image, *Biomed Signal Process Control*, **95** (2024) 106330, doi: 10.1016/j.bspc. 2024.106330.

12 Wang T W, Hong J S, Huang J W, Liao C Y, Lu C F & Wu Y T, Systematic review and meta-analysis of deep learning applications in computed tomography lung cancer segmentation, *Radio Ther Oncol*, **197** (2024) 110344, doi: 10.1016/j.radonc.2024.110344.

13 Bbosa R, Gui H, Luo F, Liu F, Efio-Akolly K & Chen Y P P, MRUNet-3D: A multi-stride residual 3D UNet for lung nodule segmentation, *Methods*, **226** (2024) 89–101, doi: 10.1016/j.ymeth.2024.04.008.

14 Faria F T J, Moin M B, Debnath P, Fahim A I & Shah F M, Explainable convolutional neural networks for retinal fundus classification and cutting-edge segmentation models for retinal blood vessels from fundus images, *arXiv*, (2024), https://arxiv.org/pdf/2405.07338v1.

15 Ali H, Mohsen F & Shah Z, Improving diagnosis and prognosis of lung cancer using vision transformers: a scoping review, *BMC Med Imaging*, **23(1)** (2023) 129, doi: 10.1186/s12880-023-01098-z.

16 Tyagi S, Kushnure D T & Talbar S N, An amalgamation of vision transformer with convolutional neural network for automatic lung tumor segmentation, *Comput Med Imaging Graph*, **108** (2023) 102258, doi: 10.1016/j.compmedimag. 2023.102258.

17 Cui F, Li Y, Luo H, Zhang C & Du H, SF2T: Leveraging swin transformer and two-stream networks for lung nodule detection, *Biomed Signal Process Control*, **95** (2024) 106389, doi: 10.1016/j.bspc.2024.106389.

18 Ren J X, Xiong Y J, Xie X J & Dai Y F, Learning transferable feature representation with swin transformer for object recognition, *Neural Process Lett*, **55(3)** (2022) 2211–2223, doi: 10.1007/s11063-022-11004-3.

19 Kim J H, Kim N & Won C S, Global–local feature learning for fine-grained food classification based on Swin Transformer, *Eng Appl Artif Intell*, **133** (2024) 108248, doi: 10.1016/j.engappai.2024.108248.

20 Sun R, Pang Y & Li W, Efficient lung cancer image classification and segmentation algorithm based on an improved swin transformer, *Electronics*, **12(4)** (2023) 1024, doi: 10.3390/electronics12041024.

21 Etehadtavakol M, Sirati-Amsheh M & Ng EYK, Radiomics feature selection from thyroid thermal images to improve thyroid nodules interpretations, *Lect Notes Comput Sci*, (2023) 121–42, doi: 10.1007/978-3-031-44511-8_10.

22 Etehadtavakol M, Sirati-Amsheh M, Moallem G & Ng E Y K, Enhancing thyroid nodule classification: A comprehensive analysis of feature selection in thermography, *Infrared Phys Technol*, (2025) **105730**, doi: 10.1016/j.infrared.2025.105730.

23 NSCLC-Radiomics, *The Cancer Imaging Arch*, [dataset], *NIH*, (2015), doi: 10.7937/K9/TCIA.2015.PF0M9REI(11 Jan 2026).

24 LUNA16 - Grand Challenge, Grand Challenge [dataset], (2016), https://luna16.grand-challenge.org/Data/(19 Jan 2026).

25 Tan Z, Madzin H, Norafida B, Rahmat R W O, Khalid F & Sulaiman PSS, SwinUNeLCsT: Global-local spatial representation learning with hybrid CNN-transformer for efficient tuberculosis lung cavity weakly supervised semantic segmentation, *J King Saud Univ Comput Inf Sci*, **36(4)** 2024102012, doi:10.1016/j.jksuci.2024.102012.

26 Lin A, Chen B, Xu J, Zhang Z & Lu G, DS-TransUNet: Dual swin transformer U-Net for medical image segmentation, *arXiv*, (2021), doi:10.48550/arxiv.2106.06716.

27 Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D & Batra D, Grad-CAM: Visual explanations from deep networks via gradient-based localization, *Proc IEEE Int Conf Comput Vis*, (2017), doi:10.1109/iccv.2017.74.

28 Chu P T M, Ha T P B, Vu N M, Ha H & Doan T M, The application of deep learning in lung cancerous lesion detection, *medRxiv*, (2024), doi:10.1101/2024.04.12.24305708.

29 He K, Zhang X, Ren S & Sun J, Deep residual learning for image recognition, *arXiv* (2015) doi:10.48550/arxiv.1512.03385.

30 Szegedy C, Vanhoucke V, Ioffe S, Shlens J & Wojna Z, Rethinking the inception architecture for computer vision, *arXiv*, (2015) doi:10.48550/arxiv.1512.00567.

31 Murthy S V S N & Prasad P M K, Adversarial transformer network for classification of lung cancer disease from CT scan images, *Biomed Signal Process Control*, (2023) doi:10.1016/j.bspc.2023.105327.

32 Elaanba A, Ridouani M & Hassouni L, Transformer-based model for radiology text reports generation from frontal and lateral chest X-ray images, *Int J Comput Inf Syst Ind Manag Appl*, (2024), https://cspub-ijcisim.org/index.php/ijcisim/article/view/732.

33 Ardila D, Kiraly A P, Bharadwaj S, Choi B, Reicher J J, Peng L, Tse D, Etemadi M, Ye W, Corrado G, Naidich D P & Shetty S, End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed

tomography, *Nat Med*, **25(6)** (2019) 954–961, doi: 10.1038/s41591-019-0447-x.

34 Chen L C, Papandreou G, Schroff F & Adam H, Rethinking Atrous convolution for semantic image segmentation, *arXiv*, (2017), doi: 10.48550/arxiv.1706.05587.

35 Simonyan K & Zisserman A, Very deep convolutional networks for large-scale image recognition, *arXiv*, (2014), doi: 10.48550/arxiv.1409.1556.

36 Chen X, Duan Q, Wu R & Yang Z, Segmentation of lung computed tomography images based on SegNet in the diagnosis of lung cancer, *J Radiat Res Appl Sci*, **14(1)** (2021) 396–403, doi: 10.1080/16878507.2021.1981753.

37 Sandler M, Howard A, Zhu M, Zhmoginov A & Chen L C, MobileNetV2: Inverted residuals and linear bottlenecks, *arXiv*, (2018), doi: 10.48550/arxiv.1801.04381.

38 Sabour S, Frosst N & Hinton G E, Dynamic routing between capsules, *arXiv*, (2017), doi: 10.48550/arxiv.1710.09829.

39 Nandipati B L & Devarakonda N, Hybrid deep learning model for detection and classification of lung cancer fusion images using MCNet, *J Intell Fuzzy Syst*, **45(2)** (2023) 2235–2252, doi: 10.3233/jifs-231145.

40 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 16×16 words: Transformers for image recognition at scale, *arXiv*, (2020), doi: 10.48550/arxiv.2010.11929.

41 Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S & Guo B, Swin Transformer: Hierarchical vision transformer using shifted windows, *arXiv*, (2021), doi: 10.48550/arxiv.2103.14030.

42 Sekeroglu K & Soysal Ö M, Multi-perspective hierarchical deep-fusion learning framework for lung nodule classification, *Sensors*, **22(22)** (2022) 8949, doi: 10.3390/s22228949.

43 Wankhade S & Vigneshwari S, A novel hybrid deep learning method for early detection of lung cancer using neural networks, *Healthc Anal*, **3** (2023) 100195, doi: 10.1016/j.health.2023.100195.

44 Choi S H, Park K B & Lee J Y, SwinResNet: Volumetric medical image segmentation by fusing swin transformer and ResNet, *Korean J Comput Des Eng*, **28(3)** (2023) 282–293, doi: 10.7315/cde.2023.282.

45 Shafi I, Din S, Khan A, De La Torre Díez I, Del JesúsPalí Casanova R & Pifarre K T, An effective method for lung cancer diagnosis from CT scan using deep learning-based support vector network, *Cancers*, **14** (2022) 5457, doi: 10.3390/cancers14215457.

46 Magdaline P P & Babu T R G, Detection of lung cancer using novel attention gate residual U-Net model and KNN classifier from computer tomography images, *J Intell Fuzzy Syst*, **45(4)** (2023) 6289–6302, doi: 10.3233/jifs-233787.

47 Liao Z, Fan N & Xu K, Swin transformer assisted prior attention network for medical image segmentation, *Appl Sci*, **12(9)** (2022) 4735, doi: 10.3390/app12094735.

48 Zhang F, Wang Q, Fan E, Lu N, Chen D & Jiang H, Enhancing non-small cell lung cancer tumor segmentation with a novel two-step deep learning approach, *J Radiat Res Appl Sci*, **17** (2023) 100775, doi: 10.1016/j.jrras. 2023.100775.

49 Primakov S P, Ibrahim A, Van Timmeren J E, Wu G, Keek S A & Beuque M, Automated detection and segmentation of non-small cell lung cancer computed tomography images, *Nat Commun*, **13** (2022), doi: 10.1038/s41467-022-30841-3.

50 Open AI, Chat GPT: *Large Language Model*, San Francisco, CA, (2023), https://chat.openai.com/