IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 6, December, 2024, pp. 3063-3073
Journal homepage: http://iieta.org/journals/ts

RCED-UNet3+: Unleashing Residual Connections in Encoder-Decoder Architecture for Precise Lung Nodule Segmentation

Sadaf Raza[1*], Razia Zia[1], Irfan Ahmed Usmani[2], Abeer Waheed[3]

[1] Faculty of Electrical and Computer Engineering, Sir Syed University of Engineering & Technology, Karachi 75300, Pakistan
[2] Department of Biomedical Engineering, Salim Habib University (Formerly Barrett Hodgson University), Karachi 74900, Pakistan
[3] Department of Computing Science, University of Alberta, Edmonton AB T6G 1A6, Canada

Corresponding Author Email: sminhaj@ssuet.edu.pk

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ts.410623

Received: 2 May 2024
Revised: 29 September 2024
Accepted: 20 November 2024
Available online: 31 December 2024

Keywords:
deep learning, lung nodule, segmentation, UNet3+, LIDC-IDRI

ABSTRACT

Lung cancer remains a serious public health concern characterize by its alarming mortality rates. Timely detection and precise segmentation of lung nodules are pivotal in patient survival rates. However, lung nodule segmentation presents challenges due to its small size and resemblance to surrounding tissues in computed tomography (CT) scans. Manual segmentation by expert radiologists is prohibitively expensive and time-consuming. This study introduces a residual connection based encoder decoder (RCED) structure into the UNet3+ architecture for lung nodule segmentation. The main contribution includes integrating elements such as residual connections and deep supervision to enhance learning efficiency and overcome vanishing gradient issues in segmenting small, complex lung nodules. To evaluate the effectiveness of the proposed architecture the training and testing were conducted on the publicly available Lung Image Database Consortium - Image Database Resource Initiative (LIDC-IDRI) dataset. The results are highly promising, with the model achieving a Dice similarity coefficient (DSC) of 0.980 and an Intersection over Union (IoU) of 0.963. Augmentation techniques were employed not only to expand the dataset size but also to optimize model performance, further elevating the DSC and IoU scores to 0.984 and 0.979, respectively. The proposed RCED-UNet3+ architecture showcases remarkable potential for diagnosing and treatment of lung cancer.

1. INTRODUCTION

In recent years, there has been a noticeable increase in lung cancer incidence, as reported by both the Global Cancer Observatory (GCO) [1] and the American Cancer Society (ACS). ACS data for 2024 reveals approximately 236,740 new cases of lung cancer, with 116,310 occurring in men and 118,270 in women. Furthermore, lung cancer was responsible for approximately 125,070 fatalities, with 65,790 men and 59,280 women affected [2]. The primary cause of lung cancer is the abnormal and uncontrolled growth of cells in lung tissue, commonly referred to as pulmonary or lung nodules. These nodules can be either benign or malignant [3] and may serve as indicators of cancer. Physicians often classify solitary pulmonary nodules (SPNs) as benign based on specific anatomical features, such as being well-defined, vascularized, adjacent to pleura, or possessing a pleural tail [4-6]. In contrast, malignant nodules tend to grow quickly and can potentially affect nearby organs. Thus, early detection of lung nodules and careful treatment planning may improve the five-year survival rate for patients [7]. Standard diagnostic techniques, such as computed tomography (CT) scans and chest X-rays, are employed to identify lung cancer. Radiologists typically review numerous CT slices to manually identify lung nodules, which may lead to overlooking small, low-density nodules due to the fatigue associated with this demanding task.

Computer-aided diagnosis (CAD) systems have been created to help radiologists detect nodules through advanced image analysis techniques. The goal of these systems is to accurately identify and outline nodules, thereby decreasing the time needed for detection. Their influence on the timely and precise diagnosis of lung cancer is considerable. Enhancements in CAD systems have significantly improved the classification, detection, and segmentation of nodules, which are vital components in the assessment of lung malignancy.

Both traditional feature-based techniques and deep learning models have proven effective in identifying and segmenting nodules. However, feature-based techniques frequently depend on manual feature extraction, which can present difficulties, especially when segmenting nodules that are connected to the lung wall [8]. Recent advances in medical image processing using deep learning have transformed clinical applications. These methods automatically extract vital information from medical images, overcoming limitations of manual feature engineering [9].

Fully Convolutional Networks (FCNs), based on deep

learning, are being increasingly applied in biomedical image segmentation, with U-Net models gaining particular attention [10]. These models consist of an encoder, a central bottleneck component, and a decoder. The U-shaped structure, where the encoder extracts features, the bottleneck ensures information flow, and the decoder reconstructs segmented images, has proven effective for medical image segmentation [11]. Researchers continue to refine the U-Net architecture to enhance feature extraction capabilities, addressing specific challenges in Medical Image Analysis. These adjustments aim to boost model performance by capturing high-level features from complex images. The UNet framework and its advanced versions, such as UNet++ and UNet3+, have proven highly effective in medical image segmentation, particularly for lung nodules. These models utilize skip connections to preserve spatial information and facilitate better gradient flow during the training process. However, as the network deepens, optimization issue arises such as vanishing gradient problems, especially in the encoder, making it harder to accurately segment smaller lung nodules.

This research addresses the issue by incorporating residual connections within the encoder to enhance the learning capability, particularly for small and complex nodules. The inclusion of residual blocks helps mitigate the vanishing gradient issue by ensuring continuous feature propagation and improving feature extraction. This adjustment enhances segmentation accuracy and accelerates network convergence while minimizing information loss during downsampling, outperforming the original UNet3+, which depends solely on skip connections. Furthermore, integrating the model into clinical processes could enhance medical imaging analysis, minimizing the time radiologists spend on manual segmentation. This automation could result in quicker, more uniform outcomes, supporting earlier diagnosis and timely treatment.

This study makes the following contributions:

1. *An RCED-UNet3+ architecture* is introduced for lung nodule segmentation, improving both the encoder and decoder components by incorporating advanced features like residual and skip connections, along with deep supervision.

2. The utilization of *data augmentation techniques* is explored to enhance the model's performance in lung nodule segmentation, aiming to evaluate its effectiveness in improving results.

3. A *comprehensive comparative analysis* is undertaken to evaluate the performance and efficiency of the proposed architecture against other state-of-the-art methods in the field of lung nodule segmentation.

The paper is organized into six sections. Section 2 provides a literature review of lung nodule segmentation. Section 3 outlines the methodology for implementing the proposed RCED-UNet3+ architecture. Section 4 presents the experimental setup, and Section 5 discusses the results along with a comparative analysis of other architectures. Finally, the conclusion is presented in Section 6.

2. LITERATURE REVIEW

Many State-of-the-art frameworks for precise nodule identification, segmentation, and classification have been proposed by numerous researchers. Most extensive survey works on nodule detection and classification are available in studies [12, 13]. Before the rise of deep learning, researchers put forth a collection of conventional machine learning techniques for the segmentation of nodules. These encompass methodologies such as rule-based methods, intensity-driven methods, morphological operations, level-set methods, graph-cut algorithms, threshold-based techniques, and region-growing methods [14-18]. For nodule segmentation, an approach utilizing a rolling ball filter and rule-based analysis is present in study by Messay et al. [15]. Yuan and He [16] provides information on energy optimization techniques including level sets and graph cuts. Demeshki et al. [18] suggests a contrast-based region-growing approach. Kostis et al. [19] utilizing the morphological opening operator in conjunction with connected components serves to effectively delineate the lung nodule and separate it from the associated blood vessels. To facilitate the segmentation of the nodule through region-growing-based methods, users are required to provide an initial seed point. de Carvalho Filho et al. [20] process the lung parenchyma region with Gaussian and median filters. The segmentation of lung nodules and the extraction of shape and texture information are carried out using the quality threshold technique. Afterward, a Support Vector Machine accurately eliminates incorrectly segmented area. Jacobs et al. [21] designed twenty-one context features, which can greatly enhance classification performance, drawing from the grayscale characteristics, as well as from the structure of lung nodules.

The drawback of the above-mentioned algorithms is that they depend on the manual feature extraction technique for the effective segmentation of the area of interest. Additionally, most methods cannot separate nodules connected to the lung wall. Features that are manually crafted are derived from the input image data, leading to a reduction in dimensionality through a process of summarizing the input. Deep learning, a subset of machine learning, has demonstrated greater effectiveness compared to traditional approaches. State-of-the-art deep learning (DL) techniques deliver cutting-edge performance in image classification, segmentation, and object detection. In image classification, DL models are capable of autonomously learning complex patterns and features within the data, allowing for highly accurate categorization of images into various classes. For segmentation tasks, DL algorithms excel at identifying and labeling specific regions or objects within an image, enabling detailed and precise analysis of individual segments. Moreover, in object detection, DL approaches are highly effective at recognizing and locating multiple objects in an image, providing both class labels and highly accurate bounding box coordinates for each detected object. These methods significantly surpass the limitations of manual feature extraction, marking a major advancement in the field.

Deep learning-based lung nodule segmentation and classification algorithms can be implemented on both 2D and 3D images [22]. Among several deep learning architectures, convolutional neural network (CNN) is the most effective and commonly used method for analyzing medical images [23]. CNN is capable of classifying each pixel in an image independently by providing the surrounding regions of a specific pixel as input. For example, Ciresan et al. [24] proposed a method for segmenting neuronal membranes in microscopic images based on light patches and sliding windows. There are two issues because of the sliding window one is redundant calculation and the other is substantial input patch overlap with neighboring pixels. Long et al. [25] address these challenges by employing a fully convolutional network,

where the final fully connected layers of the CNN are replaced with transposed convolutional layers. Ronneberger et al. [26] introduced the U-Net architecture, a U-shaped network integrated with an end-to-end fully convolutional network (FCN) for medical image segmentation. U-Net has become a widely adopted FCN for segmentation applications. Ding et al. [27] present a lung nodule detection approach leveraging Faster Region-based Convolutional Neural Network (R-CNN) alongside deep convolutional neural networks (DCNN). The technique initially extracts features using Visual Geometry Group (VGG16), followed by deconvolution to recover the feature map's dimensions, and ultimately employs Faster R-CNN and DCNN to identify lung nodules and eliminate false positives. Setio et al. [28] propose a multi-view CNN-based method for detecting lung nodules. After the initial detection stage, this method processes potential lung nodules by utilizing axial, sagittal, and coronal plane images. These images are then stored in a substantial residual network for further analysis. The predicted result is obtained by combining the outputs from the multi-view networks. Gong et al. [29] proposed an enhanced U-Net model for pulmonary nodule segmentation by incorporating squeeze-and-excite modules. This improvement embeds SE-ResNet modules into both the encoder and decoder of U-Net, allowing for a more effective fusion of high-level and low-level semantic features, thereby increasing the network's representational capability.

Zhou et al. [30] introduced three distinct UNet network architectures. The first is UNete, which consists of an overarching architecture built of UNets of various depths that share some codifiers but have unique decoders for each. This suggests that the subsequent networks do not monitor the decoders of the preceding ones, and the skip connections only combine the decoder characteristic maps at the same resolution scale, which is too limited. The second is UNet+ which is a direct link between two neighboring nodes that solves the skip connection issue. Additionally, this resolves the issue of shallow and deep decoders not receiving a supervision signal. The third approach is UNet++, which introduces a notably useful fusion of capabilities within the decoder nodes. In this method, each decoder node undertakes two specific tasks: vertically integrating multi-scale functions from prior nodes with differing resolutions, and horizontally amalgamating features from all preceding nodes operating at the same resolution. Maps of the aggregate characteristics are created in this manner, leading to a more accurate training procedure with less loss of semantic information.

A variant stemming from UNet is UNet3+, notable for its utilization of deep supervision extended across multiple scales during training. It further enhances the skip connections by incorporating multiscale features. Another innovation by Aversano et al. [31] known as GUNet3++, combines the strengths of UNet++ and UNet3+ networks. This architecture effectively conveys information across a wide range of sizes, spanning from shallower to deeper nodes, through the dense pyramidal transducer block. Furthermore, the multiscale skip connections enable the network to gather knowledge from various sources. Agnes et al. [32] introduce a novel approach known as Wavelet U-Net++ for precise lung nodule segmentation. This method combines the U-Net++ architecture with wavelet pooling to effectively capture both high and low-frequency details in images, thereby improving segmentation accuracy. The Haar wavelet transform is employed during the encoding phase's down-sampling process, ensuring the preservation of fine-grained details in the image. Gite et al. [33] evaluate the performance of four well-known neural network architecture: U-Net, FCN, SegNet, and UNet++ using the Shenzhen and Montgomery datasets. The results reveal that the UNet++ architecture significantly outperforms the others, attaining a remarkable accuracy of 98%. The U-Net model demonstrates flexibility in overcoming various challenges in medical image segmentation through its numerous extensions and enhancements. Table 1 highlights various improved models based on the U-Net architecture, aimed at achieving precise segmentation in medical images.

Table 1. Summary of improved network structure for a few UNet variants.

Model Structure	Year	Improved Structure	Highlights
UNet [34]	2020	Fully connected layer	Fully connected layer changed to an up sampling layer.
UNet++ [35]	2019	Skip connection	Use dense blocks and in-depth Supervision.
UNet 3+ [36]	2020	Skip connection	Full-scale jump connection and deep Supervision.
Context Encoder Network [37]	2019	Bottleneck between encoder and decoder	DAC and RMP structure.
nnU-Net [38]	2021	Network organization	Multiple ordinary U-Nets form a network pool.
Attention Net [39]	2018	Skip connection	Add the attention module to the skip Connection.

3. METHODOLOGY

This section explains the proposed lung nodule segmentation model shown in Figure 1. The architecture modifies UNet3+ by adding residual connections to enhance the encoder's abstraction of high-level features, effectively addressing network degradation. Additionally, the hybrid loss function penalizes incorrect background and foreground pixel classification, ensuring accurate segmentation. The following subsections detail the step-by-step implementation of the proposed methodology.

3.1 Dataset

The dataset used in this study is the LIDC-IDRI database, which is a publicly available database of lung CT scans. The dataset comprises chest CT images accompanied by XML files, that have been already annotated by four certified medical experts. This dataset encompasses 1018 CT scans belonging to 1010 distinct patients. Each scan is represented in its original form through DICOM images alongside corresponding XML files. The CT scans are uniformly structured in DICOM format, quantified in Hounsfield Units (HU), possess a trichromatic composition, and exhibit dimensions of 512 × 512 pixels. The dataset is often used in the design and testing of deep learning algorithms for the segmentation and detection of lung cancer and is currently considered a benchmark dataset in this domain [40].

3.2 Preprocessing

The preprocessing of CT scans involves isolating the lung

parenchyma from the surrounding tissues, such as bed frames, clothing, muscles, and bones, to eliminate any interference. Since lung nodules are located within the lung parenchyma, accurately segmenting this area from CT images is essential for reducing the chances of false positives and improving the accuracy of segmentation results. In CT images, the lung parenchyma appears as a continuous region with lower grayscale values, contrasted by higher grayscale values representing the chest muscles. The segmentation process for lung parenchyma is complex and comprises of multiple steps. Initially, the CT images undergo binarization, which generates a binary image of the lung region, typically by representing lung tissue as black and the rest as white through thresholding. Morphological dilation techniques are then utilized to fill gaps caused by denser tissues within the lung parenchyma. This method expands the boundaries of the lung regions, allowing small interruptions or breaks to be seamlessly filled. Consequently, this leads to a complete representation of the parenchyma. Figure 2 illustrates the various stages involved in the segmentation of lung parenchyma from CT images.

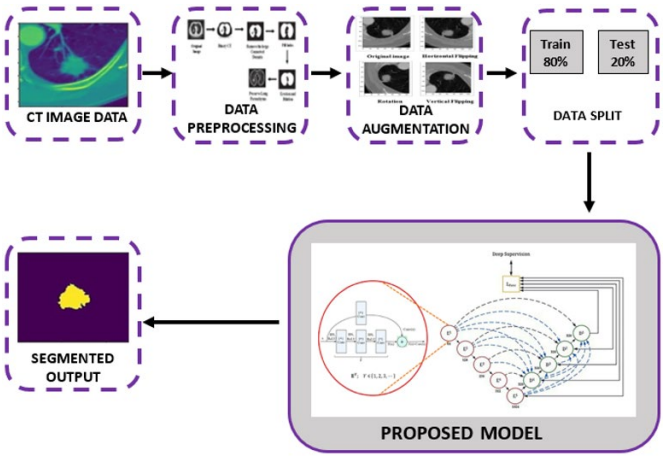


Figure 1. Proposed methodology for lung nodule segmentation

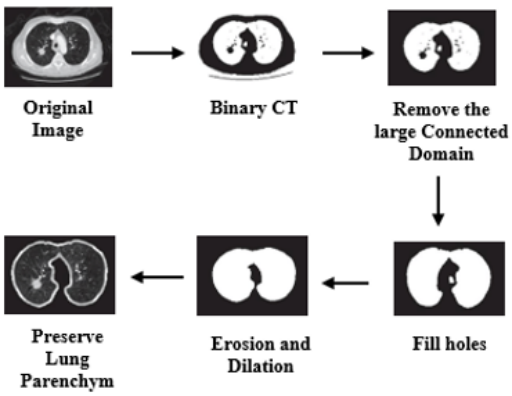


Figure 2. The steps involved in segmenting lung parenchyma from CT images

3.3 Data augmentation

After preprocessing, data augmentation is performed to increase the data size and evaluate the effectiveness of the model with and without data augmentation. A total of three augmentation techniques were implemented, extending each data sample to 4 samples, resulting in a total of 62,200 samples. The three techniques for geometric transformation invariance, including horizontal flipping, vertical flipping, and rotation, are defined below and illustrated in Figure 3.

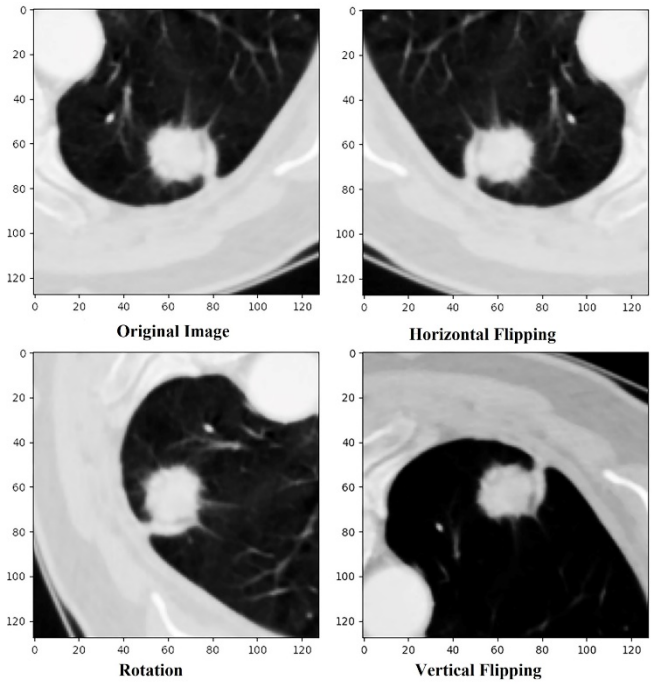


Figure 3. Original image and three augmented variations

3.3.1 Horizontal and vertical flipping

These techniques are used to simulate changes in lung nodule orientation. Nodules can appear on either side or in multiple positions within the lungs, and flipping helps the model recognize consistent features regardless of orientation.

3.3.2 Rotation

It is designed to correct slight positional shifts that may occur during medical scans due to patient posture. By introducing rotations, the model becomes more adaptable to these positional variations, thereby enhancing its segmentation accuracy across different nodule locations.

3.4 Proposed model

UNet++ and UNet3+ are two variants of the UNet architecture that aim to improve the original UNet model's capabilities for segmentation tasks. They achieve this by incorporating more advanced features and design modifications. The UNet++ architecture uses a nested, multi-scale approach, allowing the network to capture more comprehensive information at different scales. On the other hand, the UNet3+ architecture modifies the skip pathways (which connect the encoder sub-network to the decoder sub-network) and deep supervision block, enhancing the model's effectiveness in image segmentation tasks. However, the encoder and decoder subnetworks remain unchanged as compared to UNet. The encoder and decoder in all variants of UNet are made up of two 3 × 3 convolutional layers with a size of 64, and each convolutional layer is followed by a ReLU activation function. This means that there is a ReLU activation at the very end of each residual block and this can negatively affect propagation as large weight updates (especially in the beginning) will cause the activation value to always be 0, regardless of the input.

The encoder captures the abstract features of the input

image which are essential semantic information. If more semantic information can be extracted from the encoder, the segmentation accuracy should improve. To improve feature extraction in encoders and to avoid the problem of network degradation, the proposed architecture has a residual connection network for encoders and decoders.

Residual connections are crucial in deep neural networks for addressing the vanishing gradient problem. By allowing information to flow directly during both forward and backward propagation, these connections improve learning efficiency and optimization, especially for small, complex lung nodule structures. In contrast to standard UNet3+ approaches that often lack residual learning, this method promotes enhanced feature reuse and faster convergence.

Additionally, residual connections are essential for handling the variability in the shapes and sizes of lung nodules, ensuring that the model preserves critical spatial and contextual information across layers. Without residual learning, deep networks might struggle to capture the fine details necessary for precise segmentation. A residual connection network consists of many residual blocks and allows information to propagate more easily, both forward and backward. In our proposed structure, each encoder and decoder block consist of a residual connection network as shown in Figure 4, hence improving the encoder and decoder sub-networks.

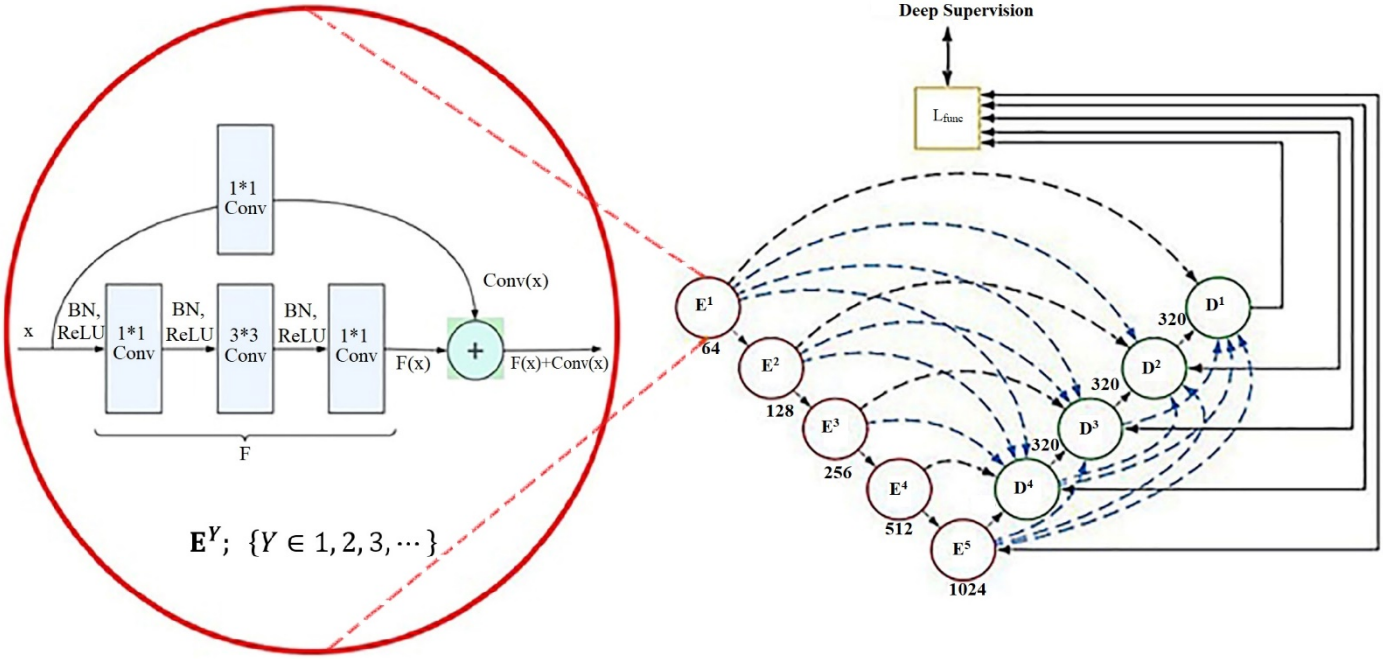


Figure 4. Proposed encoder structure of RCED-UNet3+ architecture

The proposed residual connection network has 3 convolutional layers of, sizes 1 × 1, 3 × 3, and 1 × 1, and before each of these layers, a batch normalization (BN) and ReLU activation are performed where batch normalization is a technique used to address the vanishing gradient problem. It helps to maintain a stable distribution of inputs by normalizing the activations within each layer. It acts as a regularizer, preventing overfitting, improving generalization, and facilitates the use of higher learning rates, leading to faster convergence and reduced training time in these deep neural networks. The output from these layers combined with the original input form a residual connection (RC). If x is the original input to the block and $F(x)$ represents the output through all the layers then RC can be expressed as:

$$RC = x + F(x) \tag{1}$$

However, the dimensions of $F(x)$ and x need to be the same for element-wise addition; to ensure this we introduce another 1 × 1 convolutional layer with 0 padding and apply it to x. Then we get:

$$RC = Conv(x) + F(x) \tag{2}$$

This residual network should improve the accuracy as it reduces the number of channels and solves the problem of having an activation of a value of 0 regardless of input.

The architecture proposed in this paper modifies the UNet3+ architecture by incorporating residual connections to solve the problem of vanishing gradient. The proposed architecture not only surpasses the performance of the original UNet architecture by incorporating advanced skip pathways and deep supervision, but it also ensures higher feature abstraction at the encoder level and mitigates network degradation. The proposed RCED-UNet3+ model is illustrated in Figure 1 which significantly improves overall segmentation performance, achieving a Dice score of 0.984. In contrast, the baseline UNet3+ model exhibits a Dice score of 0.965. This highlights the impact of residual connections in improving feature learning, especially for the segmentation of irregularly shaped lung nodules.

Table 2 describes the parameters of the proposed architecture, encompassing layer names, operations executed at each layer, and the resulting output size. The model is structured with four primary layer types: input layer, encoder layers, decoder layers, and output layer. Each layer undertakes operations such as convolution, batch normalization, and rectified linear unit (ReLU) activation. The table also describes the number of output channels and the size of the convolution kernels for each layer.

Table 2. Detailed parameters of proposed RCED-UNet3+

Name	Operations	Output Size
Input		512 × 512 × 1
Encoder 0 (E0)	Conv, c = 64, k = 1 BN, ReLU, Conv, c = 64, k = 3 BN, ReLU, Conv, c = 64, k = 1 RC	512 × 512 × 64
Encoder 1 (E1)	MaxPooling, k = 2 BN, ReLU, Conv, c = 128, k = 1 BN, ReLU, Conv, c = 128, k = 3 BN, ReLU, Conv, c = 128, k = 1 RC	256×256×128
Encoder 2 (E2)	MaxPooling, k = 2 BN, ReLU, Conv, c = 256, k = 1 BN, ReLU, Conv, c = 256, k = 3 BN, ReLU, Conv, c = 256, k = 1 RC	128×128× 256
Encoder 3 (E3)	MaxPooling, k = 2 BN, ReLU, Conv, c = 512, k = 1 BN, ReLU, Conv, c = 512, k = 3 BN, ReLU, Conv, c = 512, k = 1 RC	64 × 64 × 512
Bottleneck	MaxPooling, k = 2 BN, ReLU, Conv, c = 1024, k = 1 BN, ReLU, Conv, c = 1024, k = 3 BN, ReLU, Conv, c = 1024, k = 1	32 × 32 × 1024
Decoder3 (D3)	**E0:** MaxPooling, k = 8 Conv, BN, ReLU, c = 64, k = 3 **E1:** MaxPooling, k = 4 Conv, BN, ReLU, c = 64, k = 3 **E2:** MaxPooling, k = 2 Conv, BN, ReLU, c = 64, k = 3 **E3:** Conv, BN, ReLU, c = 64, k = 3 **Bottleneck:** UpSampling, k = 2 Conv, BN, ReLU, c = 64, k = 3 Concatenate new E0, E1, E2, E3, Bottleneck BN, ReLU, Conv, c = 320, k = 1 BN, ReLU, Conv, c = 320, k = 3 BN, ReLU, Conv, c = 320, k = 1 RC	64 × 64 × 320
Decoder2 (D2)	**E0:** MaxPooling, k = 4 Conv, BN, ReLU, c = 64, k = 3 **E1:** MaxPooling, k = 2 Conv, BN, ReLU, c = 64, k = 3 **E2:** Conv, BN, ReLU, c = 64, k = 3 **D3:** UpSampling, k = 2 Conv, BN, ReLU, c = 64, k = 3 **Bottleneck:** UpSampling, k = 4 Conv, BN, ReLU, c = 64, k = 3 Concatenate new E0, E1, E2, D3, Bottleneck BN, ReLU, Conv, c = 320, k = 1 BN, ReLU, Conv, c = 320, k = 3 BN, ReLU, Conv, c = 320, k = 1 RC	128×128× 320
Decoder 1 (D1)	**E0:** MaxPooling, k = 2 Conv, BN, ReLU, c = 64, k = 3 **E1:** Conv, BN, ReLU, c = 64, k = 3 **D2:** UpSampling, k = 2 Conv, BN, ReLU, c = 64, k = 3 **D3:** UpSampling, k = 4 Conv, BN, ReLU, c = 64, k = 3 **Bottleneck:** UpSampling, k = 8 Conv, BN, ReLU, c = 64, k = 3 Concatenate new E0, E1, D2, D3, Bottleneck BN, ReLU, Conv, c = 320, k = 1 BN, ReLU, Conv, c = 320, k = 3 BN, ReLU, Conv, c = 320, k = 1 RC	256×256×320
Decoder 0 (D0)	**E0:** Conv, BN, ReLU, c = 64, k = 3 **D1:** UpSampling, k = 2	512×512× 320

	Conv, BN, ReLU, c = 64, k = 3 **D2:** UpSampling, k = 4 Conv, BN, ReLU, c = 64, k = 3 **D3:** UpSampling, k = 8 Conv, BN, ReLU, c = 64, k = 3 **Bottleneck:** UpSampling, k = 16 Conv, BN, ReLU, c = 64, k = 3 Concatenate new E0, D1, D2, D3, Bottleneck BN, ReLU, Conv, c = 320, k = 1 BN, ReLU, Conv, c = 320, k = 3 BN, ReLU, Conv, c = 320, k = 1 RC	
Output	Conv, BN, ReLU, c = 1, k = 3	512 × 512 × 1

4. EXPERIMENTAL SETTING

The experiments were carried out using a dataset of CT scans from the LIDC-IDRI database, which had 15550 CT images with annotation of lung nodules by four experienced radiologists. To eliminate interference from other tissues, the dataset has been preprocessed to extract the lungs' parenchyma from the CT image. Data augmentation is done after reprocessing to improve the model efficiency and expand the data size to 62,200 CT images.

Two experiments were conducted: one to assess whether the proposed architecture shows improvement over a UNet3+ network, and the other to evaluate the effectiveness of the proposed architecture with augmented data and without augmented data.

Both architectures, UNet3+ and the proposed, were implemented in Python using the PyTorch library. The hyperparameters for UNet3+ were chosen based on the best-performing hyperparameters in study [31] except for the loss function. Both models utilized a learning rate of 0.01, a dropout rate of 0.15, a batch size of 32, and employed the SGD optimizer.

The model was trained using a hybrid loss function that integrates both dice loss and binary cross-entropy loss. Dice loss is effective in measuring the similarity between two samples, while binary-cross entropy loss is commonly used to penalize false classifications. The model trained with the hybrid loss function showed significant improvements compared to using separate dice loss and binary cross-entropy loss. Let C be the correct mask, S be the predicted mask and then soft dice loss (DLS) is defined as:

$$l_{DLS=}1 - \frac{2CS + 1}{C + S + 1} \tag{3}$$

where, binary-cross entropy loss (BCE) is defined as:

$$l_{BCE=}Clog(s) + [(1 - Clog(1 - S)] \tag{4}$$

Then proposed RCED-UNet 3+ model hybrid loss is defined as:

$$l_H = l_{DLS} + l_{BCE} \tag{5}$$

This hybrid loss function penalizes the model for misclassifying background and foreground pixels and also ensures that the model produces accurate segmentation. The models were evaluated using the Intersection of Union metric (IoU), and the Sørensen-Dice similarity coefficient (Dice). Both metrics measure the similarity between two images based on the presence and absence of data. The dice coefficient is very similar to the Intersection over Union metric. Defining C as the correct mask and S as the generated segmentation, Dice can be defined as:

$$Dice=2*|\mathrm{C} \cap \mathrm{S}|/(|\mathrm{C}| + |\mathrm{S}|) \tag{6}$$

IoU metric can be defined as:

$$\mathrm{IoU} = |\mathrm{C} \cap \mathrm{S}| \,/\, |\mathrm{C} \cup \mathrm{S}| \tag{7}$$

The Hausdorff distance is characterized by the maximum difference in distance between two boundaries. Specifically, it is identified as the greatest minimum distance from any point p located on the boundary of set S_A to the boundary of set S_B. The Hausdorff distance is defined as:

$$h(S_A, S_B) = \max_{p \in S_A}(d_{min}\,(p, S_B)) \tag{8}$$

$$Hausdorffdisatance(S_A, S_B) = \max(h(S_A, S_B), h(S_B, S_B)) \tag{9}$$

The model's performance can be further evaluated based on true positives (correctly identified positives), true negatives (correctly identified negatives), false positives (incorrectly identified positives), and false negatives (incorrectly identified negatives) by computing Accuracy, Precision, and Recall. Accuracy is defined as the ratio of correctly identified pixels to the overall number of pixels in the segmentation output. Precision represents the ratio of correctly identified pixels classified as lung masses to the total number of pixels predicted as such. Conversely, recall assesses the proportion of correctly identified pixels within a lung mass relative to the total number of pixels in the area that the lung mass occupies.:

They are defined as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

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

5. RESULTS AND DISCUSSION

Dice coefficient and Intersection over Union (IoU) are common metrics used to find the accuracy of semantic

segmentation models. In this work, we employed these metrics to evaluate the performance of the proposed RCED-UNet3+ architecture. Table 3 presents a comparison between the IoU and Dice scores of our proposed model and the UNet3+ baseline architecture and Figure 5 shows the evolution of Dice scores and IoU scores across epochs on training. The results indicate a significant improvement in the Dice score when the proposed architecture is used. Additionally, the table shows that both the IoU and Dice scores increase when the proposed architecture is trained with augmented data, as compared to when it is trained without augmented data. Table 4 shows the evaluation of model performance in terms of accuracy, precision, recall and Hausdorff distance with data augmentation and without data augmentation.

Table 3. Comparison of mean IoU scores and dice scores of UNet3+ model, RCED-UNet3+ model, and RCED-UNet3+ model trained with augmented data

Model	Mean IoU Score	Mean Dice Score
UNet3+	0.932	0.965
RCED-UNet3+ Without Data Augmented	0.963	0.980
RCED-UNet3+ With Data Augmentation	0.979	0.984

Table 4. Comparison of accuracy, precision, recall and Hausdorff distance of RCED-UNet3+ model (without augmentation) and RCED-UNet3+ model trained with augmented data

Model	Accuracy	Precision	Recall	Hausdorff Distance
RCED-UNet3+ (Without Data Augmented)	0.989	0.983	0.975	5.011
RCED-UNet3+ (With Data Augmentation)	0.998	0.989	0.979	4.640

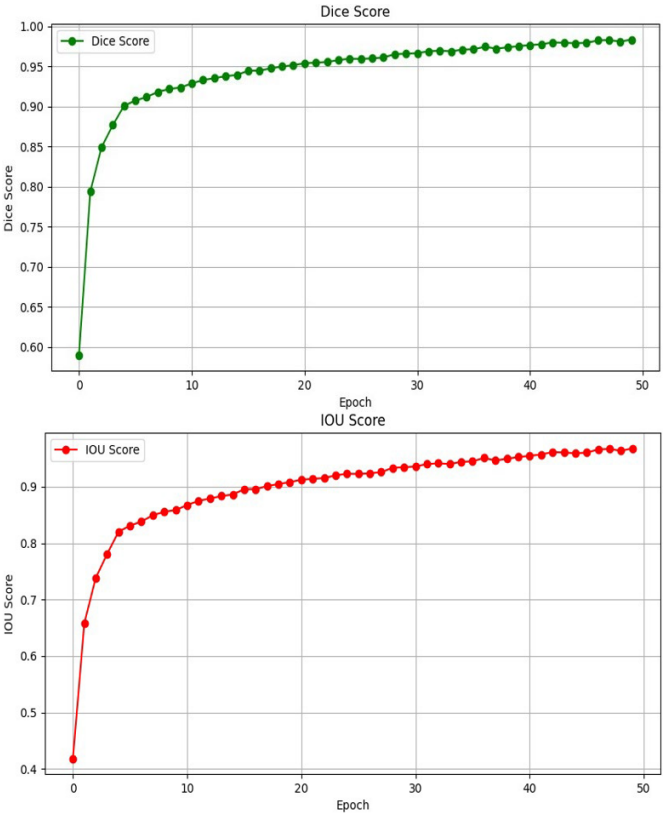


Figure 5. Evolution of dice scores and IoU scores across epochs on training

The hybrid loss function showed a consistent decrease over a larger number of epochs, indicating improved model learning and adaptation to the complexities of lung nodule segmentation as depicted in Figure 6.

Figure 7 presents a box plot, while Figure 8 illustrates a qualitative comparison of the three models. The results indicate that the UNet3+ model performed the worst followed by RCED-UNet3+. This shows that the introduction of residual connection networks in UNet3+ has improved the performance of the model. The box plot also shows that the RCED-UNet3+ trained with augmented data performed the best, moreover, it has a smaller interquartile range than RCED-UNet3+ suggesting that its performance is uniform across the dataset.

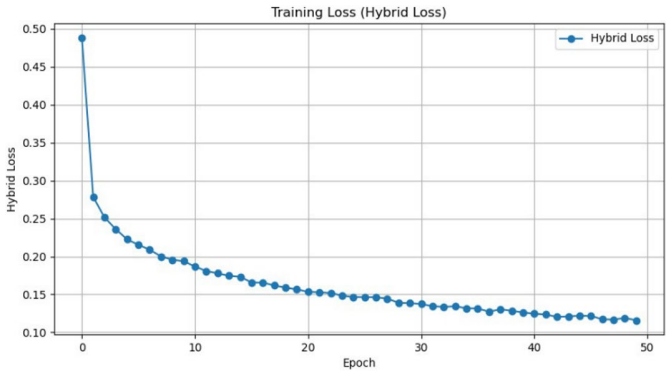


Figure 6. Variation of the hybrid Loss, over epochs on training set

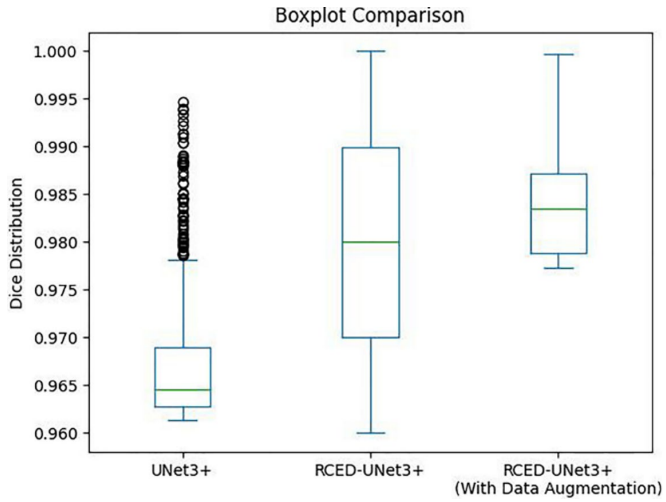


Figure 7. Boxplot comparison of UNet3+, RCED-UNet3+, and RCED-UNet3+ trained with augmented data

The RCED-UNet3+ architecture outperforms seven other advanced models in lung nodule segmentation, as detailed in

Table 5. It achieves a Dice score of 0.9802 ± 0.0986, significantly surpassing other models. ResiUNet, which combines pretrained residual networks with U-Net, records a Dice score of 94.87%, but RCED-UNet3+ excels in handling irregular lung nodules. Additionally, RCED-UNet3+ surpasses the DL-based graph model, which uses a multi-scale PET/CT fusion technique and attains a Dice score of 0.86, as well as the Atrous Convolution-based Convolutional Neural Network (ATCNN), which captures multi-scale HRCT features and scores 0.9715. The Residual UNet, achieving 0.9545 ± 0.1372, also falls behind RCED-UNet3+, which leverages residual connections in both the encoder and decoder for better performance. Furthermore, RCED-UNet3+ outpaces GUNet3++ by 1.92% in Dice score, focusing on enhancing the encoder and decoder units, while GUNet3++ relies more on skip connections. The RCED-UNet3+ also exceeds APU-Net, which requires more training data due to its dual UNet structure, showing that additional PET/CT images aren't crucial for higher segmentation accuracy. Lastly, though DENSE-UNET shares similarities, the incorporation of residual connections into UNet3+ gives RCED-UNet3+ a clear advantage in lung nodule segmentation.

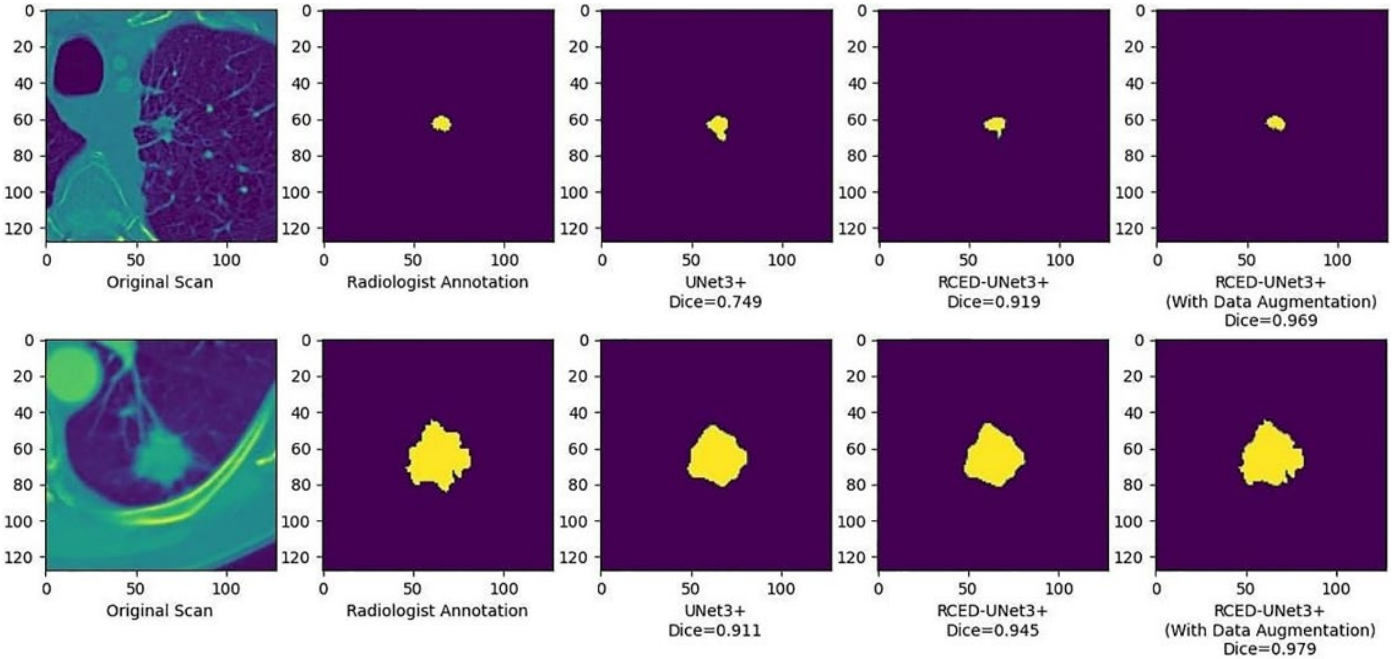


Figure 8. A qualitative comparison of the output of UNet3+, RCED-UNet3+, and RCED-UNet3+ trained with augmented data with radiologist's annotation

Table 5. Comparison of dice score of RCED-UNet architecture with 7 other state-of-the-art lung nodule segmentation architecture

Model	Year	Metric	Value
ResiU-Net [41]	2023	Dice Score	0.948
DL based graph model [42]	2023	Dice Score	0.86
ATCNN2PR framework [43]	2023	Dice Score	0.9715
Residual UNet [44]	2022	Dice Score	0.954
GUNet++ [31]	2022	Dice Score	0.9610
APU-Net [45]	2022	Dice Score	0.9686
Dense-UNET [46]	2022	Dice Score	0.7442
RCED-UNet3+ (Proposed)		Dice Score	0.984

6. CONCLUSION

Recent studies have focused on the challenges associated with detecting and segmenting lung cancer. The need for early lung cancer identification to ensure patients have a longer life expectancy is at the core of the issue. In this study, a modified UNet variant, named RCED-UNet3+ has been proposed to segment pulmonary nodules from CT images. The proposed model was validated by using the LIDC-IDRI dataset. The RCED-UNet3+ model introduced the improved residual connection network for the encoder and decoder structure, skip connection linking the two paths, and deep supervision block. The residual connection network should improve the accuracy as it reduces the number of channels and solves the problem of having an activation of a value of 0 regardless of input. At the same time, the hybrid loss function was introduced to penalize the model for misclassifying background and foreground pixels and also ensure that the model produces accurate segmentation. Furthermore, the RCED-UNet3+ architecture compared with seven other state-of-the-art architectures demonstrates better performance and efficiency. This suggests that the modification introduced in UNet3+ architecture along with the hybrid loss function contributes to improved lung nodule segmentation accuracy in CT images. For future research, we intend to evaluate the model using entirely new datasets from various scanners to further verify and confirm its robustness.

REFERENCES

[1] Chhikara, B.S., Parang, K. (2023). Global Cancer Statistics 2022: The trends projection analysis. Chemical Biology Letters, 10(1): 451-451.

[2] Siegel, R.L., Giaquinto, A.N., Jemal, A. (2024). Cancer statistics, 2024. CA: A Cancer Journal for Clinicians, 74(1): 12-49. https://doi.org/10.3322/caac.21820

[3] Dolejsi, M., Kybic, J., Polovincák, M., Tuma, S. (2009). The lung time: Annotated lung nodule dataset and nodule detection framework. Medical Imaging 2009: Computer-

Aided Diagnosis, 7260: 538-545. https://doi.org/10.1117/12.811645
[4] Mao, K., Deng, Z. (2016). Lung nodule image classification based on local difference pattern and combined classifier. Computational and Mathematical Methods in Medicine, 2016(1): 1091279. https://doi.org/10.1155/2016/1091279
[5] Wang, X., Mao, K., Wang, L., Yang, P., Lu, D., He, P. (2019). An appraisal of lung nodules automatic classification algorithms for CT images. Sensors, 19(1): 194. https://doi.org/10.3390/s19010194
[6] Shaukat, F., Raja, G., Frangi, A.F. (2019). Computer-aided detection of lung nodules: A review. Journal of Medical Imaging, 6(2): 020901. https://doi.org/10.1117/1.JMI.6.2.020901
[7] Xie, Y., Meng, W.Y., Li, R.Z., Wang, Y.W., et al. (2021). Early lung cancer diagnostic biomarker discovery by machine learning methods. Translational Oncology, 14(1): 100907. https://doi.org/10.1016/j.tranon.2020.100907
[8] Maqsood, M., Yasmin, S., Mehmood, I., Bukhari, M., Kim, M. (2021). An efficient DA-net architecture for lung nodule segmentation. Mathematics, 9(13): 1457. https://doi.org/10.3390/math9131457
[9] Haque, I.R.I., Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked, 18: 100297. https://doi.org/10.1016/j.imu.2020.100297
[10] Uçar, G., Dandıl, E. (2024). Enhanced detection of white matter hyperintensities via deep learning-enabled MR imaging segmentation. Traitement du Signal, 41(1): 1-21. https://doi.org/10.18280/ts.410101
[11] Du, G., Cao, X., Liang, J., Chen, X., Zhan, Y. (2020). Medical Image Segmentation based on U-Net: A Review. Journal of Imaging Science & Technology, 64(2): 020508-1-020508-12. https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508
[12] Valente, I.R.S., Cortez, P.C., Neto, E.C., Soares, J.M., de Albuquerque, V.H.C., Tavares, J.M.R. (2016). Automatic 3D pulmonary nodule detection in CT images: A survey. Computer Methods and Programs in Biomedicine, 124, 91-107. https://doi.org/10.1016/j.cmpb.2015.10.006
[13] Halder, A., Dey, D., Sadhu, A.K. (2020). Lung nodule detection from feature engineering to deep learning in thoracic CT images: a comprehensive review. Journal of Digital Imaging, 33(3): 655-677. https://doi.org/10.1007/s10278-020-00320-6
[14] Diciotti, S., Picozzi, G., Falchini, M., Mascalchi, M., Villari, N., Valli, G. (2008). 3-D segmentation algorithm of small lung nodules in spiral CT images. IEEE Transactions on Information Technology in Biomedicine, 12(1): 7-19. https://doi.org/10.1109/TITB.2007.899504
[15] Messay, T., Hardie, R.C., Rogers, S.K. (2010). A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Medical Image Analysis, 14(3): 390-406. https://doi.org/10.1016/j.media.2010.02.004
[16] Yuan, Y., He, C. (2012). Adaptive active contours without edges. Mathematical and Computer Modelling, 55(5-6): 1705-1721. https://doi.org/10.1016/j.mcm.2011.11.014
[17] Boykov, Y., Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(9): 1124-1137. https://doi.org/10.1109/TPAMI.2004.60
[18] Dehmeshki, J., Amin, H., Valdivieso, M., Ye, X. (2008). Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach. IEEE Transactions on Medical Imaging, 27(4): 467-480. https://doi.org/10.1109/TMI.2007.907555
[19] Kostis, W.J., Reeves, A.P., Yankelevitz, D.F., Henschke, C.I. (2003). Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Transactions on Medical Imaging, 22(10): 1259-1274. https://doi.org/10.1109/TMI.2003.817785
[20] de Carvalho Filho, A.O., de Sampaio, W.B., Silva, A.C., de Paiva, A.C., Nunes, R.A., Gattass, M. (2014). Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. Artificial Intelligence in Medicine, 60(3): 165-177. https://doi.org/10.1016/j.artmed.2013.11.002
[21] Jacobs, C., Van Rikxoort, E.M., Twellmann, T., Scholten, E.T., et al. (2014). Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. Medical Image Analysis, 18(2): 374-384. https://doi.org/10.1016/j.media.2013.12.001
[22] Kumar, M.S., Rao, K.V., Kumar, G.A. (2021). MRI image based classification model for lung tumor detection using convolutional neural networks. Traitement Du Signal, 38(6): 1837-1842. https://doi.org/10.18280/ts.380628
[23] Zhou, K., Gu, Z., Liu, W., Luo, W., Cheng, J., Gao, S., Liu, J. (2018). Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, pp. 2724-2727. https://doi.org/10.1109/EMBC.2018.8512828
[24] Ciresan, D., Giusti, A., Gambardella, L., Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In Proceedings of the 25th International Conference on Neural Information Processing Systems, pp. 2843-2851.
[25] Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965
[26] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th international conference, Munich, Germany, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
[27] Ding, J., Li, A., Hu, Z., Wang, L. (2017). Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. https://doi.org/10.48550/arXiv.1706.04303
[28] Setio, A.A.A., Traverso, A., De Bel, T., Berens, M.S., et al. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16

challenge. Medical Image Analysis, 42: 1-13. https://doi.org/10.1016/j.media.2017.06.015
[29] Gong, L., Jiang, S., Yang, Z., Zhang, G., Wang, L. (2019). Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks. International Journal of Computer Assisted Radiology and Surgery, 14: 1969-1979. https://doi.org/10.1007/s11548-019-01979-1
[30] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J. (2018). UNet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Granada, Spain, pp. 3-11. https://doi.org/10.1007/978-3-030-00889-5_1
[31] Aversano, L., Bernardi, M.L., Cimitile, M., Iammarino, M., Verdone, C. (2022). An enhanced UNet variant for effective lung cancer detection. In 2022 International Joint Conference on Neural Networks (IJCNN): Padua, Italy, pp. 1-8. https://doi.org/10.1109/IJCNN55064.2022.9892757
[32] Agnes, S.A., Solomon, A.A., Karthick, K. (2024). Wavelet U-Net++ for accurate lung nodule segmentation in CT scans: Improving early detection and diagnosis of lung cancer. Biomedical Signal Processing and Control, 87: 105509. https://doi.org/10.1016/j.bspc.2023.105509
[33] Gite, S., Mishra, A., Kotecha, K. (2023). Enhanced lung image segmentation using deep learning. Neural Computing and Applications, 35(31): 22839-22853. https://doi.org/10.1007/s00521-021-06719-8
[34] Beheshti, N., Johnsson, L. (2020). Squeeze U-Net: A memory and energy efficient image segmentation network. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, pp. 1495-1504. https://doi.org/10.1109/CVPRW50498.2020.00190
[35] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J. (2019). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging, 39(6): 1856-1867. https://doi.org/10.1109/TMI.2019.2959609
[36] Huang, H., Lin, L., Tong, R., Hu, H., et al. (2020). UNet 3+: A full-scale connected UNet for medical image segmentation. In ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, pp. 1055-1059. https://doi.org/10.1109/ICASSP40776.2020.9053405
[37] Gu, Z., Cheng, J., Fu, H., Zhou, K., et al. (2019). Ce-net: Context encoder network for 2d medical image segmentation. IEEE transactions on medical imaging, 38(10): 2281-2292. https://doi.org/10.1109/TMI.2019.2903562
[38] Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2): 203-211. https://doi.org/10.1038/s41592-020-01008-z
[39] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., et al. (2018). Attention u-net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999
[40] Dutande, P., Baid, U., Talbar, S. (2021). LNCDS: A 2D-3D cascaded CNN approach for lung nodule classification, detection and segmentation. Biomedical Signal Processing and Control, 67: 102527. https://doi.org/10.1016/j.bspc.2021.102527
[41] Bhattacharjee, A., Murugan, R., Goel, T., Mirjalili, S. (2023). Pulmonary nodule segmentation framework based on fine-tuned and pretrained deep neural network using CT images. IEEE Transactions on Radiation and Plasma Medical Sciences, 7(4): 394-409. https://doi.org/10.1109/TRPMS.2023.3236719
[42] Xia, X., Zhang, R. (2023). A novel lung nodule accurate segmentation of PET-CT images based on Convolutional neural network and Graph Model. IEEE Access, 11: 34015-34031. https://doi.org/10.1109/ACCESS.2023.3262729
[43] Halder, A., Dey, D. (2023). Atrous convolution aided integrated framework for lung nodule segmentation and classification. Biomedical Signal Processing and Control, 82: 104527. https://doi.org/10.1016/j.bspc.2022.104527
[44] Sousa, J., Pereira, T., Silva, F., Silva, M.C., Vilares, A.T., Cunha, A., Oliveira, H.P. (2022). Lung segmentation in CT images: A residual U-Net approach on a cross-cohort dataset. Applied Sciences, 12(4): 1959. https://doi.org/10.3390/app12041959
[45] Zhou, T., Dong, Y., Lu, H., Zheng, X., Qiu, S., Hou, S. (2022). APU-Net: An attention mechanism parallel U-Net for lung tumor segmentation. BioMed Research International, 2022(1): 5303651. https://doi.org/10.1155/2022/5303651
[46] Lu, D., Chu, J., Zhao, R., Zhang, Y., Tian, G. (2022). A novel deep learning network and its application for pulmonary nodule segmentation. Computational Intelligence and Neuroscience, 2022(1): 7124902. https://doi.org/10.1155/2022/7124902