



An efficient U-Net framework for lung nodule detection using densely connected dilated convolutions

Zeeshan Ali¹ · Aun Irtaza¹ · Muazzam Maqsood²

Accepted: 24 April 2021 / Published online: 15 June 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Remote health monitoring is an important aspect especially for remote locations where standard medical facilities are not available. Smart cities use a similar concept to provide health facilities even when physicians are unavailable. Lung cancer remains to be one of the most critical types of cancer with a 5-year survival rate of only 18%. Efficient computer-aided diagnostic systems are required to diagnose lung cancer before time for better treatment planning. The variety of lung nodules and their visual similarity with surrounding regions make their detection difficult. Traditional image processing and machine learning methods usually lack the ability to handle all types of nodules with a single method. In this study, we propose an efficient end-to-end segmentation algorithm with an improved feature learning mechanism based on densely connected dilated convolutions. We applied dense feature extraction and incorporated multi-dilated context learning by using dilated convolutions at different rates for better nodule segmentation. First, lung ROIs are extracted from the CT scans using k-mean clustering and morphological operators to reduce the model's search space instead of using full CT scan images or nodule patches. These ROIs are then used by our proposed architecture for nodule segmentation and efficiently handles different types of lung nodules. The performance of the proposed algorithm is evaluated on a publicly available dataset LIDC-IDRI and achieved a dice score of 81.1% and a Jaccard score of 72.5%.

Keywords Lung nodules segmentation · Densely connected convolutions · Dilated convolutions · Remote health monitoring

✉ Muazzam Maqsood
muazzam.maqsood@cuiatk.edu.pk

¹ Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan

² Department of Computer Science, COMSATS University Islamabad, Attock Campus, Islamabad, Pakistan

1 Introduction

Smart cities utilize a concept to collect data using sensors and cameras. It also helps to provide a framework for smart health monitoring. It is usually difficult to handle critical diseases in the absence of hardware or physicians. Lung nodule detection is one of those critical and life-threatening types of cancer with a five-year survival rate of only around 18% [1]. In the last decade, there has been a lot of attention given to design efficient CAD systems with the advancements in computer vision and deep learning techniques [2–4]. However, CAD systems for lung nodule detection still need a lot of attention for better results. The computed tomography (CT) images are generally used for early diagnosis, monitoring, and treatment planning of lung nodules [5]. A generic CAD system for lung nodule detection consists of lung nodule classification, localization, and segmentation. Therefore, accurate lung nodule segmentation remains to be one of the important aspects of these systems. There is a growing need to design fully automated CAD systems for lung nodules as manual identification and nodule segmentation take time and expert knowledge [6].

Automated nodule detection and segmentation have a complex problem due to their heterogeneous nature [7–9]. The most challenging issue in nodule segmentation is the visual and shape similarity of nodules with their surroundings. However, the main challenge here is to identify and detect different types of nodules with a single method for efficient lung nodule detection. There are nodule types like juxtaleural which are visually very similar to lung walls and affect the segmentation performance of conventional methods. The other types of nodules also face similar issues like insignificant contrast difference with surrounding blood vessels or low contrast with the background. These problems make it difficult for traditional threshold or morphological-based algorithms to segment nodules efficiently. The juxta-vascular nodules are also difficult to segment due to their low contrast as compared to their surroundings. The major cause of low contrast is lung parenchyma in which juxta-vascular nodules are attached to blood vessels. Ground-glass nodules are also difficult to segment out due to the exhibition of the same behavior of juxta-vascular nodules. Simple thresholding-based techniques are not useful because of the very low contrast among surrounding regions. To solve these issues, OTSU-based segmentation technique is employed to segment the nodules, but this technique faces limitations to handles adhesion-type nodules which include juxta-vascular and juxtaleural. The cavitory lung nodules have a cavity inside them which is an important indication for their detection. Malignant nodules are greater than 5 mm in size with a shape similar to round air-like low-density shadows. Further, the size of these small nodules is comparable to surrounding noise. Such small nodules affect the network's downsampling capabilities which means deeper semantic features are difficult to detect. This creates a major problem for feature extraction of large-sized nodules. The traditional segmentation algorithms may suit any particular nodule but lack the generalizability for other nodule types [10].

Previously, many traditional approaches have been used for lung nodule detection which includes intensity-based morphological operations [9, 11, 12] and

region growing. Further, energy optimization-based level set [8] and graph cut methods [13] have also been used for this purpose. However, these methods are not much useful for small nodules with less than 6 mm diameter. Generally, it is difficult to design adaptive templates in morphological operations [9]. Some semi-supervised methods proved effective but these methods require human intervention [14]. The rule-based heuristic approaches also proved efficient but fail to handle irregular shapes due to rule violations [15]. The limitations of existing segmentation methods emphasize to design of better algorithms for nodule detection. The recent advancements in deep learning and its use in computer vision make it a better choice for CAD systems [16–21]. The performance of deep learning algorithms generally improves with an increase in data. The most important aspect of deep learning is automatic feature learning where manual and hand-picked feature extraction is not required. The deep learning architecture automatically learns features and patterns from the data and uses it for further processing. The deep learning architectures have also been widely used in medical imaging and proved effective [22–24]. However, the complex nature of lung nodules is problematic because of intensity, irregular shape, contrast, and visual similarity. Due to all these factors, already existing models require a lot of further improvements. These problems make lung nodules detection an open research problem that needs attention.

From this line of research, we propose an efficient semantic segmentation-based algorithm for lung nodule segmentation. The proposed work uses an efficient architecture for nodule localization and segmentation. As in most previous researches [25, 26], algorithms are trained on whole CT scan images or different patches of nodules extracted by patch-based techniques. In this research, ROI-based technique is used to train the algorithm for improved performance. In the first step, lung ROIs are extracted from a complete CT scan with the help of preprocessing steps. These preprocessing steps include some standard operations, noise removal filters, and k-means clustering. The segmented images returned by k-means clustering are further improved by morphological operations which proved helpful in removing similar objects from surroundings. After that, the lung ROIs and their ground truth images are used to train the algorithm. The main aim of extracting lung ROI is to remove extra organs and objects from the whole CT scan. It also makes the training of algorithms easier due to reduced search space. The proposed algorithm is based on densely connected dilated (DCD) convolution blocks. The use of dense connections helps extract dense features by concatenating the extracted features from the previous layer to every next layer and also to improve feature usability. The convolution layers in DCD blocks are based on dilated convolutions which are capable of extracting wider information of context due to their dilated filters. Different convolutions consist of different dilation rates to empower the multi-dilated context learning of the algorithm. Moreover, the proposed work is evaluated over a publicly available dataset LIDC-IDRI. The results are computed by using sensitivity, dice score, and Jaccard score. The published work shows improvement in the segmentation of lung nodules over existing methods and also on standard U-Net proposed by Ronneberger et al. [27]. This study has the following contributions:

- We propose an end-to-end efficient automated method for lung nodule segmentation.
- We incorporated DCD blocks to improve the feature learning process in our architecture which are based on densely connected dilated convolutions
- Dilated convolutions of different rates enable multi-dilated context learning in algorithm to capture wider information for different types of nodules

The rest of the paper is organized as follows; Sect. 2 presents related work, the methodology is explained in Sect. 3, and results are presented in Sect. 4 followed by a conclusion and future work.

2 Related work

In the last decade, there has been a lot of research done in the field of lung nodule cancer detection. The methods used for this purpose include region growing, morphological methods, energy-based optimization, and machine learning-based methods. In morphological operations, lung nodules are highlighted, and blood vessels are removed using different morphological operations for nodule detection and isolation [28]. To improve this technique and remove the wall from juxtaleural nodules, the shape hypothesis was included with morphological operators [29, 30]. However, lung nodule detection is a difficult task to be handled by morphological operators. The region-growing techniques lack the ability to handle all nodule types with a single method, especially the small nodules. This limitation was handled by introducing different rules based on intensity, distance, and fuzzy connectivity. Even though this method used various rules but still faced issues to handle irregular nodules because it is difficult to design rules for such nodules. Researchers [31–33] also converted segmentation tasks into energy optimization tasks for processing. Level set functions were used to process the image, and optimization energy reaches a minimum value when segmentation contours match nodule boundary. Farag et al. [8] proposed a similar approach for lung segmentation shape prior hypotheses. Furthermore, the lung nodule detection problem was also formulated as a maximum flow problem [34] using the graph cut method. However, these methods lack the ability to handle all nodule types simultaneously. The machine learning methods require manual feature extraction and selection for nodule segmentation and voxel classification [35, 36]. Lu et al. [37] proposed a method based on translational and rotational invariant features for segmentation. Wu et al. proposed a method based on texture and shape features extraction using conditional random fields. Hu et al. [38] proposed a method and first performed lung extraction. Later, they have used the Hessian matrix for vascular feature extraction. They used artificial neural networks for the classification of nodules. Jung et al. [39] proposed deformable asymmetric multi-phase models for ground-glass nodule detection. A 3D multiscale lung nodule segmentation was also proposed for lung nodule detection by Gonçalves et al. [40].

CNN is a deep learning architecture and extracts features directly from the data and learns underlying patterns between raw data and labels. This category of algorithms also performs voxel classification for segmentation just like machine

learning techniques. Wang et al. proposed a multi-view CNN which consists of three branches, one for each view axial, coronal and sagittal. Later, Wang et al. proposed a center-focused semi-automatic CNN, however, this architecture lacks the ability to detect small nodules. Zhao et al. [41] proposed a pyramid deconvolution neural network for lung nodule detection. The architecture efficiently extracted high-level and low-level features and combined them for classification. Huang et al. [42] designed a fully automated architecture that consisted of four steps: lung candidate detection, merging, false-positive reduction, and segmentation. A fully connected convolution neural network has also been used for this purpose like 2D and 3D U-Net architectures. Moreover, in recent work, Cao et al. [43] proposed a dual branch residual network for the segmentation of lung nodules. In their work, they combined the intensity and deep features. They trained their algorithm by first extracting the patches of nodules using a weighted sampling strategy. Later on, Ali et al. [44] proposed CNN based on transferable texture to increase the performance of classifying pulmonary nodules. The texture features in CNN were extracted by introducing energy layer (EL) in the network instead of using pooling layers. Jiang et al. [45] proposed a multiple resolution residually connected network (MRRN) for the segmentation of lung nodules. The feature across multiple resolutions of images was combined with the features extracted through residual connections. Liu et al. [46] used an object detection-based approach of Mask-RCNN for the detection of lung nodules. They fine-tuned the model trained on the COCO dataset for lung nodule detection. Keetha et al. [26] modified the original U-Net by incorporating a bidirectional feature network (Bi-FPN) for the segmentation of lung nodules. They trained their algorithm on full CT scan images with data augmentation. During feature fusion, the additional weight is incorporated by Bi-FPN with each input. This helped and allowed the network to learn the importance of a particular input feature. Pezzano et al. [25] proposed context-learning-based CNN combined with an adaptive loss for accurate segmentation of nodules. They also extracted the patches from the CT scan and used them for training their network. Furthermore, Tang et al. [47] proposed an algorithm named Nodule-Net which was based on 3D deep CNN. It performed nodule detection, segmentation, and false-positive reduction jointly in single stage as multi-task fashion.

3 Proposed methodology

The main architecture of the proposed methodology is presented in Fig. 1. The first step is dataset labeling, and then, preprocessing is done on extracted images to get lungs ROI. The lungs ROI images and their corresponding annotations are used to train the model and acquired the results of lungs nodules segmentation.

3.1 Dataset labeling

This dataset contains the DICOM files for all patients with their corresponding XML annotations files. A Pylidc Python package is used for labeling the dataset as

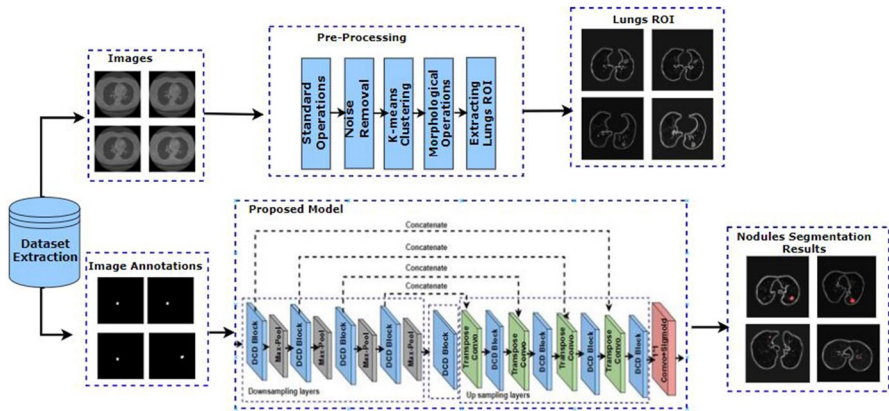


Fig. 1 Schematic overview of the proposed methodology

recommended. This package helps in extracting the ground truth masks images of lung nodules.

3.2 Dataset preprocessing

Generally, medical imaging datasets contain unnecessary objects which are not in the region of interest. Usually, the data need preprocessing for further experimentation and better quality. So, in order to reduce the search space of our model, we extract a particular region of interest from the whole CT scan image where nodules are present. We follow the following sequence of steps to extract a lung's ROI. The complete preprocessing procedure is also shown in Fig. 2.

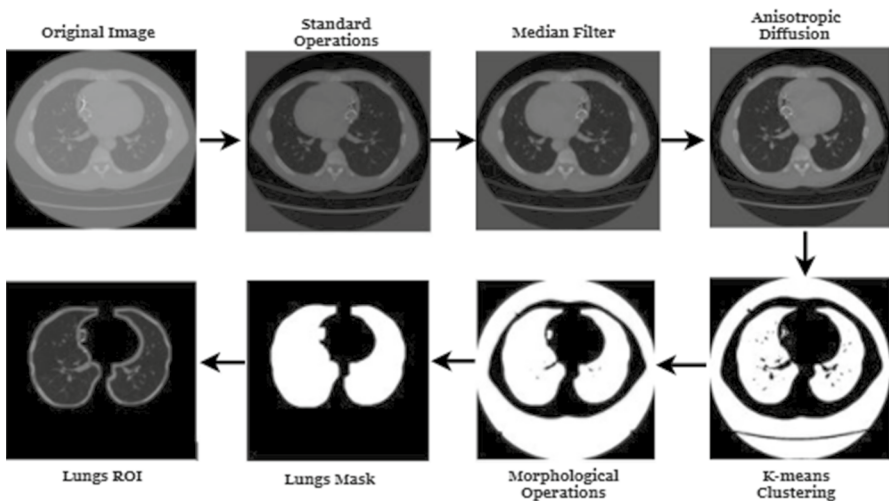


Fig. 2 Steps of preprocessing on LIDC dataset

3.2.1 Standard operations

The basics of image processing operations can be beneficial in achieving simple tasks. Here, we first applied standard mean and deviation on the original image and then subtracted the image from standard mean and deviation. The image cropping is done followed by taking a mean of that cropped region. Further, the maximum and minimum values of the original slice image are replaced by the mean of the cropped region.

3.2.2 Noise removal filters

In this study, we used two different noise removal filters for noise reduction. The first one is the median filter which performs excellently on images to reduce noise by working on its spatial domain and preserves image edges. A small matrix called kernel is used to scan an image, and the value of the central pixel is recalculated by taking the median of matrix values. Here, we used a kernel size of 3*3 to apply a median filter on the resulting image. After applying the median filter, an anisotropic diffusion filter has been applied. It is a non-linear filter known as Perona–Malik diffusion which removes noise without blurring the edges and corners of the nodule boundary. The value of the gamma coefficient is 0.1 which controls the speed of diffusion, while the value kappa coefficient is 50. This advantage of anisotropic diffusion filter is that it removes noise without blurring the edges of nodules.

3.2.3 K-means clustering

There are many different types of research conducted for image segmentation using different clustering methods. K-means algorithm is one of the popular clustering methods based on unsupervised learning and segment out the required region of interest from the image background. After preprocessing, K-means clustering is performed with $k = 2$ on resulting images to segment the lung regions from the image followed by thresholding. At first, the algorithm picks a random point which is the center point for grouping followed by calculating distance among all image pixels. After this, a new mean or centroid point is calculated and updated. This algorithm continuously performs iterations until there is no change found in the mean or centroid point. The similarity and dissimilarity measures in clustering methods are based upon the Euclidian distance as shown in Eq. (1) in which $D(x, y)$ represents the image and x and y are pixel coordinates of images.

$$D(x, y) = \sqrt{\sum x_i y_i} \quad (1)$$

3.2.4 Morphological operations

Morphological operations are used to remove the imperfections and distortions present in an image using different structuring elements. These structuring elements are simply pre-defined kernels that are applied to an image. The image we get from

the previous step is a binary image and this binary image contains different types of imperfections. So, to remove these, we first apply an erosion operation on images followed by dilation on an image. The object boundaries are smooth, and all small objects in the foreground are removed by erosion while dilation repairs intrusions by enlarging objects and reducing gaps. This erosion followed by dilation is also called opening an image $I(x, y)$ which is a binary image. Here, we used a box structuring element of (4, 4) in erosion operation and (10, 10) in dilation. The equation of morphological opening operation is defined below.

$$I^{\circ}s = (I\ominus s)\oplus s \quad (2)$$

In the above equation, structuring element s is applied on image I to form an opening operation denoted by $I\ominus s$ in which erosion and dilation are denoted by \ominus and \oplus , respectively.

3.2.5 Extracting lungs ROI

In this step, the ROIs of the lungs from the CT scan image are extracted. The results of the morphological opening operation are further used for labeling of resulting image based on pixel intensities. The connection between two pixels exists when they are neighboring pixels and have a similar value. In this way, all connected regions are assigned the same integer value. Then, the properties and attributes of each labeled region are accessed using bounding boxes. The bounding box covers all pixels belonging to one region. With the help of these bounding box parameters, a lung mask is extracted. We further perform another dilation operation on the resulting lung mask with a structuring element of size (10, 10). In the end, a lung mask is multiplied with a slice image to get our required ROI.

3.3 Lung nodule segmentation

U-Net architecture is a widely used framework that has emerged in deep learning over the last few years especially in biomedical image segmentation. The U-Net framework was first developed by Ronneberger et al.[27] in the year 2015. Since then, many researchers have exploited this architecture in the field of medical imaging. We use the concept of the U-Net model to design our model for lung nodules segmentation and later on, we also made a comparison between our model and U-Net. Figure 3 shows our proposed architecture.

3.3.1 Downsampling Layers

The U-Net model generally consists of an encoder path, and this path is also referred to as the contracting and downsampling path which consists of downsampling layers. In our proposed architecture, the downsampling layers consist of densely connected dilated convolution blocks (DCD) as shown in Fig. 3. After every DCD block, a max-pool layer of size $2*2$ is applied on an input image to downscale it and it can be defined in Eq. (3):

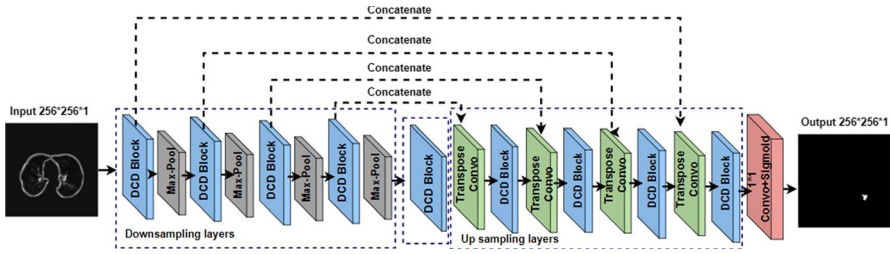


Fig. 3 Architecture of proposed algorithm

$$y_{k,w}^i = \max_{0 \leq a,b \leq p} (x_{i_{k \times p + a, w \times p + b}}) \tag{3}$$

In above Eq. (3), on downsampling layer at a position of (k, w) , a $y_{k,w}^i$ neuron is present in i th output map, whereas x_i denotes the i th input map, in which region $p \times p$ is assigned with neuron value of $y_{k,w}^i$. After that, the dropout layer of rate 0.05 is used after every max-pool layer. This step is performed to avoid overfitting issues due to the random deactivation of neurons found inside hidden layers. In the DCD block, the input is first given to 3×3 convolution with a dilation rate of 1×1 . After that, the batch normalization [48] and ReLu [49] activation function are applied to the input image. In DCD blocks, the dilated convolutions of different rates are strong enough to capture wider information at different scales. It can be defined as to first consider a 2D signal at the input which is lung ROI in our case and the feature map which is represented by x . The kernel that is applied on an input image is denoted by w to get output y on each location i which is calculated as Eq. (4).

$$y[i] = \sum_k x[i + r \cdot k]w[k] \tag{4}$$

In Eq. (4), the given 2D signal at the input is sampled by r , while r shows the value of stride. This sampling operation is similar to conventional convolution operation on input x with kernels w . These kernels are useful to up-sampled the image along each spatial dimension. But in dilated convolutions, this can be done by adding $r - 1$ zeros between consecutive kernel values. The dilated convolutions assist in getting a large receptive field [50]. These dilated convolutions are able to extract larger contexts due to dilated filters of different rates. In this way, we get multi-dilated context information from the image. Further, there are four different DCD blocks in the downsampling path, and each DCD block consists of convolution of kernel size 3×3 . The dilation rate of each of four convolutions in DCD blocks is $1 \times 1, 2 \times 2, 3 \times 3,$ and $4 \times 4,$ respectively. All these convolutions in each DCD block are densely connected [51]. Here, the feature maps of the previous convolution layer are concatenated to the next convolution layer and increase the chance of getting dense features as shown in Fig. 4. Therefore, the feature maps of lungs ROI in all previous layers y_0 to y_{n-1} are given to every next n^{th} layer as input:

$$y_n = H_n([y_0, y_1, \dots, y_{n-1}]) \tag{5}$$

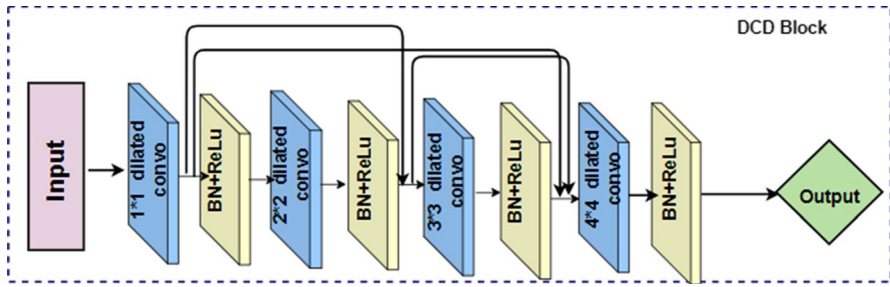


Fig. 4 Internal architecture of DCD blocks

The concatenation of feature map’s generated from layers $0 \dots n - 1$ is referred by $([y_0, y_1, \dots, y_{n-1}])$ in Eq. (5). For example, if we have three layers named as y_0, y_1 and y_2 , then the input of layer y_2 is the concatenation of features maps extracted by layer the y_0 and y_1 . From Fig. 4, it is illustrated that how the input of different dilated convolutions is extracted which results in dense, wider, and multi-dilated features of lung nodules. Furthermore, there is the use of batch normalization [48] and Relu [49] activation after every convolution layer in DCD blocks as shown in Fig. 4. Batch normalization after each convolutional layer over the downsampling path is very important. This is because it enhances the process of training as inputs to each layer are normalized by obtaining zero mean and variance. All values being passed into activation get regulated by batch normalization which eventually speeds up the calculations during the forward pass. Weight initialization also becomes convenient while designing deeper networks which improves the performance of the model. Therefore, it can be said that a batch having N separate examples, where every example is of D -dimensional vector is passed to the batch of normalization layer. All inputs of lungs ROI are given by a matrix as $X \in R^{N \times D}$ where each of the examples is described by row x_i . Every individual example x_i of lungs ROI is normalized with the help of the equation given below in Eq. (6):

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\mu^2 + \epsilon}} \tag{6}$$

where ϵ is the small constant for numerical stability, and μ , and σ^2 represent the mean and variance, respectively, and can be calculated by Eq. (7) and Eq. (8):

$$\mu = \frac{1}{N} \sum_i x_i \tag{7}$$

$$\sigma^2 = \frac{1}{N} \sum_i (x_i - \mu)^2 \tag{8}$$

In above Eqs. (7) and (8), N represents the total number of lungs ROI images in the current batch, while x_i represents the one single ROI example in the batch.

Initially, the total number of filters present over the first DCD block is 16. This number is found to be doubled over further blocks which are 32, 64, and eventually 128 over the last DCD block. More specifically, on the downsampling path, the lung ROI is passed as an input to the first DCD block, which produced feature maps of dimensions $256 \times 256 \times 16$. Afterward, the input is passed through max-pool operation to reduce the spatial dimensions of feature maps. This process is repeated three more times as shown in Fig. 4, and the dimension of feature maps produces each time is $128 \times 128 \times 32$, $64 \times 64 \times 64$, and $32 \times 32 \times 128$, respectively. Moreover, all the filters are initialized by “he normal” [52] weight initialization and it is defined by Eq. (9):

$$W \sim G(0, \sqrt{\frac{2}{n}}) \quad (9)$$

In Eq. (9), n represents the total number of inputs in the node, while G denotes a random number with Gaussian probability distribution. Furthermore, the standard deviation is calculated by $\sqrt{\frac{2}{n}}$, while the mean is 0.0. These filters are applied in lungs ROI to extract nodule features. The main aim of the downsampling layers is to extract features and semantics out of an image as well as to depict image context in an efficient way. By the end of this step, the proposed model can learn the different kinds of information found in the image by using extracted features. This is done by downsampling the image with the help of convolution and pooling layers.

3.3.2 Bottleneck layer

The bottleneck layer consists of a DCD block which consists of densely connected dilated convolution shown in Fig. 3. The max-pool layer after the last DCD block in the downsampling layer is given as an input to the bottleneck layer which is the first layer of the DCD block. Then, the output produced by the bottleneck is given to the first upsampling layer which is a transposed convolution layer. The dimensions of the feature map produce in the bottleneck layer are $16 \times 16 \times 256$.

3.3.3 Upsampling layers

The upsampling layers structurally comprised of four DCD blocks after every transposed convolution of size 3×3 and stride value of 2 as shown in Fig. 3. A transposed convolution is used at this stage as a deconvolution layer. The transposed convolution layer (Conv2DTranspose) is complicated as compared to traditional upsampling and refers to the inverse of the convolution operation. This implies that in the training phase, it works by the upsampling image with suitable learning of how to fill up the details. Meanwhile, the traditional upsampling layer has no weights, and it just doubles the dimensions of the input image. Transposed convolution is also referred to as fractionally stride convolutions. Suppose if a convolution given by kernel w is applied with unit stride and without padding, while inputs and outputs from left to right are unrolled into vectors then this convolution can be represented as a matrix called sparse matrix C . In this matrix, $w_{i,j}$ denotes the non-zero elements of the kernel. By making use of this matrix,

a backward pass is conveniently attained if transpose of C matrix is obtained. Furthermore, the loss is multiplied by transpose of C , and then, error is back-propagated. A kernel w that defines a convolution consists of the forward and backward pass is computed by multiplying a sparse matrix C and the transpose of sparse matrix C^T . Similarly, a kernel w also defines a fractionally strided convolution (transposed convolution) in which both backward and forward passes are computed by multiplying C and $(C^T)^T$, respectively. Followed by each transposed convolution layer, concatenation is carried out from the contracting path with consequent feature maps. The filter size over four DCD blocks in upsampling layers is 128, 64, 32, and 16. In the end, a convolution of size 1*1 followed by a sigmoid activation function is employed as shown in Fig. 3. More specifically, the input from the DCD block of the bottleneck layer is given as an input to first transpose the convolution layer. After that, the output that comes from transposed convolution layer is concatenated with the last DCD block of downsampling layers. The dimension of output feature maps of the first DCD block in the upsampling layer is 32*32*128. Similarly, the output that comes from the second convolution transposed convolution layer is concatenated with the second last DCD block in the downsampling path. The dimension of output feature maps of the second DCD block in the upsampling path is 64*64*64. This process is repeated two more times in the same pattern. So, the output feature maps in last two DCD block are 128*128*32 and 256*256*16.

The main aim of the upsampling layers is to up-sample the image to restore as well as capture spatial information. This also restores location information out of feature map size which was lost at the encoder path. All the contextual data from the downsampling layers are transferred to the upsampling layers with the help of skip connections. Skip connections comprised of concatenating intermediate encoder outputs to decoder layers at the required position. This step serves in merging localization data from the decoder path with contextual data from the encoder path.

3.3.4 Training details and hyperparameters

The proposed model is trained on the LIDC-IDRI dataset. For training the model, input images with their ground truth mask are used. The loss function used here is binary cross-entropy and the model runs for 150 epochs with an input batch size of 4. The model is trained using an adaptive learning optimization algorithm called “Adam” optimizer with a learning rate of 0.001. Adam optimizer merges the momentum term along with stochastic gradient descent and RMSprop. The equation used for weight update of Adam optimizer is mentioned as under:

$$W_t = W_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (10)$$

In the above-mentioned equation, W shows proposed model weights and η shows step size. The value of η directly depends on iteration. The values of \hat{m}_t and \hat{v}_t are estimated using equations mentioned below:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \text{ and } \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (11)$$

In the above-mentioned equations, β_1 and β_2 are hyperparameters of the algorithm having default values 0.9 and 0.999, respectively. While the network is being trained, errors among actual values and predicted values are calculated with the help of the binary cross-entropy loss function. It is mentioned below:

$$BCE = \frac{-1}{N} \sum_{i=1}^N y_i * \log (P(y_i)) + (1 - y_i) * \log (1 - p(y_i)) \quad (12)$$

In the above-mentioned equation, BCE stands for binary cross-entropy. y_i refers to the predicted class of pixel in model output. $P(y_i)$ represents probability predicted by the trained model for all pixels being background or belong to nodules.

4 Experimentation and results

4.1 Dataset details

The dataset used for this research is a public dataset of lung nodules from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [53]. In this dataset, there are 986 nodules samples labeled by four different expert radiologists. In this research study, we have randomly partitioned 986 nodules into subsets of training and testing with an 80–20 ratio of split.

4.2 Evaluation metrics

For the assessment of the proposed algorithm in segmenting lung nodules, different types of evaluation metrics are used which include dice score, Jaccard score, the symmetric volumetric difference (SVD), and sensitivity [54]. The dice score gives an overlap score between actual and predicted results of the model, while the difference between actual and predicted mask is measured by SVD. Moreover, the Jaccard analysis helps in measuring similarity and diversity among two samples. Equations from (13) to (16) describe the mathematical formulation of these scores:

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (13)$$

$$SVD = 1 - DSC \quad (14)$$

$$SVN = \frac{TP}{TP + FN} \quad (15)$$

$$\text{JACCARD} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (16)$$

where “TP” represents the true positive, FP represents the false positive, and FN represents the false negative in the results of the model.

4.3 Results and discussions

In this study, we present an efficient and improved feature learning-based segmentation algorithm for the segmentation of different types of nodules. Our methodology starts with preprocessing of the data that is used to train the algorithm. So, due to the different natures of lung nodules, shapes and similarity with surrounding regions require preprocessing step. Therefore, images are enhanced for easy extraction of ROIs and the required objects. The results are presented in the second row of Fig. 5 where final ROI regions are presented against each corresponding actual CT scan image. As mentioned before, the original CT scan contains many objects and artifacts which are visually similar to nodules. Therefore, segmenting nodules directly from the CT scan without any preprocessing are very challenging. Noise removal and image enhancement filters improve the image quality, and then, these improved images are used for ROI extraction. Then, we used k-mean clustering to extract the ROI from the CT scan which is our requirement so that further working can be done to extract lung nodules. However, the extracted lungs also contain various other similar objects which are detected as false positives. These false positives are considered as one of the main problems faced in medical imaging because these objects need to be removed by a domain expert as a post-processing step. This false-positive reduction step requires manual intervention and makes this process of the human dependent. In this work, we used morphological operators, erosion, and dilation to remove such objects. The quality of the extracted ROIs is very good as shown in Fig. 5 and is then used for further analysis.

After preparing the data, the lung's ROI and their corresponding ground truth images are used to train the algorithm. The proposed method segments out lung

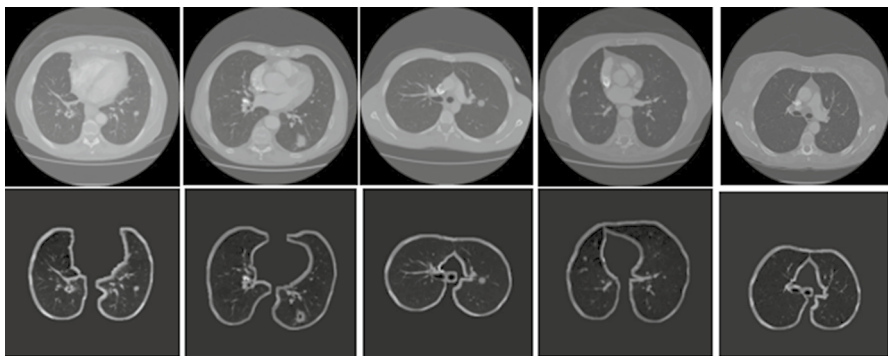


Fig. 5 CT scans from the dataset and their corresponding lungs ROI

nodules very efficiently. The segmentation and localization of the lungs nodule are an important part of semantic segmentation. The results for nodule segmentation are presented in Fig. 6 where column (A) shows ROI images containing lungs, the second column (B) shows its corresponding actual mask provided by radiologists, the third column (C) contains actual overlay image, the fourth column (D) shows the predicted mask by the model and the last column (E) presents predicted overlay image. From Fig. 6, it is observed that the model efficiently handles different types of nodules. The results for juxta-vascular nodules, partially solid nodules, solid nodules attached with the lung wall, irregular shape nodules, and results for tiny nodules attached to lung walls are presented in first, second, third, fourth, fifth, and sixth rows, respectively. It is observed that our model fails to segment very tiny nodules because the model is unable to extract features of tiny nodules as they are just 1–3 pixels (Table 1).

On the other hands, the model is able to segment all other nodule types. The results presented in terms of images are also very promising. Moreover, the CAD systems are required to both identify the nodules and then segment out these nodules

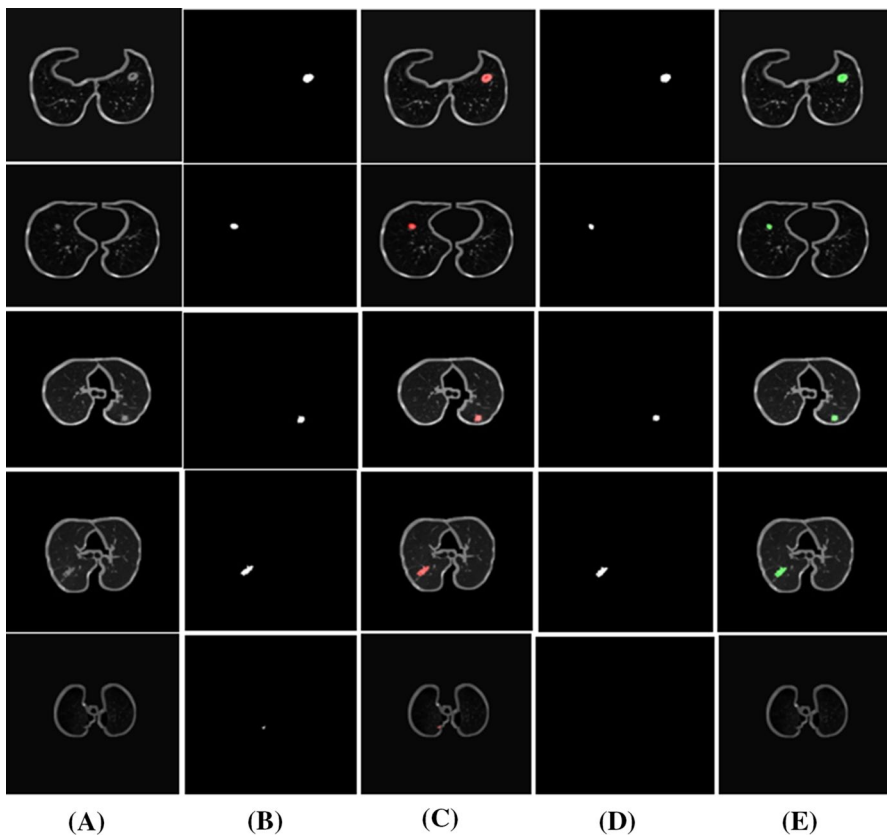


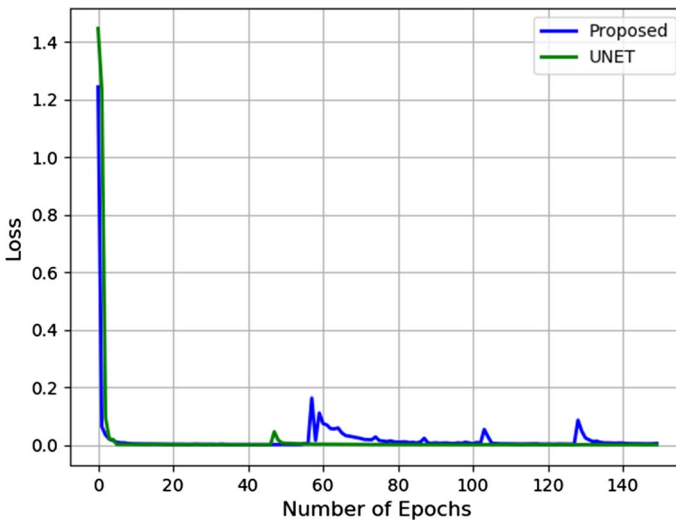
Fig. 6 a Original image b actual mask c actual overlay image d predicted mask e predicted overlay

Table 1 Performance evaluation for lungs nodule detection method

Authors	Dice score (%)	Jaccard (%)	Sensitivity (%)	SVD	Year
U-Net [27]	71.0	62.8	70.2	0.29	2015
Proposed	81.1	72.5	82	0.19	2021

to assist physicians more effectively. The proposed work performed both these tasks efficiently. The test images include different types of nodules which include tiny, solid, partially solid, irregular shapes, and cavity-based nodules. Furthermore, in terms of different evaluation metrics, the proposed system segments out these nodules properly and the dice score achieved for lungs nodules is 81.1% and the Jaccard score of 72.5% is achieved. The overall performance of the proposed work is very good as presented in Table 1. The comparison with the original U-Net proposed by Ronneberger et al. [27] is also drawn, and results are also shown in Table 1. Our proposed algorithm achieves a 10.1% improvement in dice score over U-Net, while the 11.8% improvement was found in the Jaccard score. There is also significant improvement found on other scores which includes sensitivity and SVD. The sensitivity of our proposed algorithm is 82% and SVD of 0.19 which is very small. On the other hands, the sensitivity comes with U-Net is 70.2% and SVD of 0.29. The training and testing division of data are the same for both U-Net and the proposed model.

Further, the accuracy and loss graphs of the proposed model and U-Net are also presented in Figs. 7, 8, respectively. In Figs. 7 and 8, the x -axis shows the number of epochs, while the y -axis shows the accuracy and loss values of both U-Net and the

**Fig. 7** Graph plotted to determine model loss

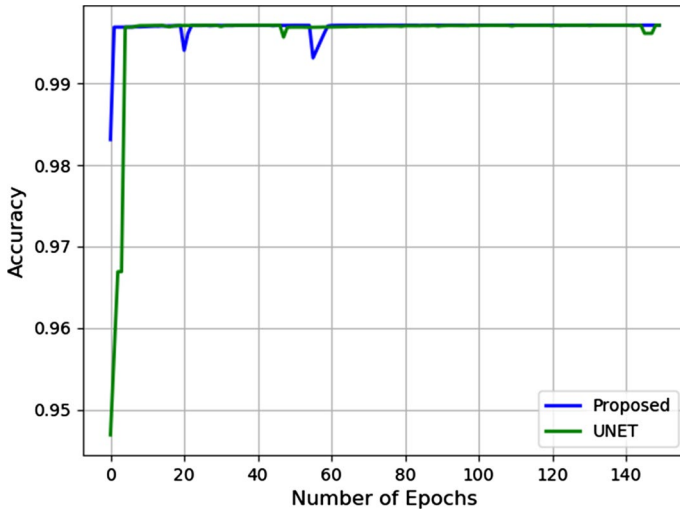


Fig. 8 Graph plotted to determine model accuracy

proposed algorithm. The reason for high accuracy is a class imbalance because the background pixels are more than nodules pixels. The accuracy does not much tell about the segmentation model so we evaluate the model on popular segmentation metric which includes dice and Jaccard score and it is evident from these scores and results that the performance of the proposed method is very good.

Moreover, feature activation maps extracted by our proposed algorithms are shown in Fig. 9. Each layer output different activation maps. These activation maps

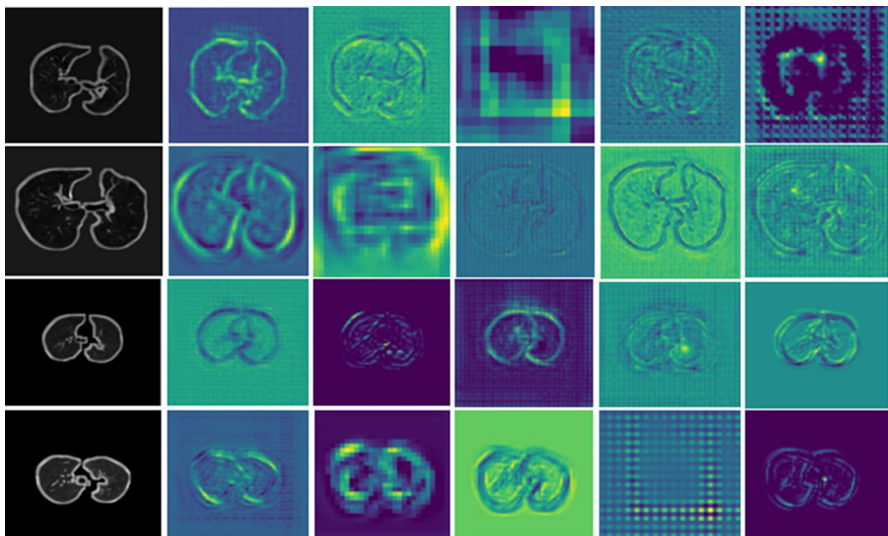


Fig. 9 Activation maps result on intermediate layers

help get information that how the model depicts and encodes the contextual information of the given image. In the above figure, activation maps of different layers are shown. The above activation maps show that at starting layers, the model detects and finds small and fine-grained details and on subsequent layers, more high-level features are extracted. It is observed that the model pays attention to lung areas which indicate and show that the model learns effective features from the image context. For example, the last image present in the last row of above Fig. 9 shows that model is very near to extract nodules as features are represented by yellow color.

A detailed comparative analysis is also drawn of our proposed model with existing methods and techniques for segmentation of lungs nodules which is shown in Table 2. In previous years, it is evident from the literature and observed that our proposed achieved the highest dice score value. Shen et al. [55] achieved a dice score of 78.55% in their experiments. In their work, they first extracted nodule patches and given them as an input to the multi-crop convolutional neural network (MC-CNN) in which a multi-crop pooling strategy was used. This dice score value is further improved in the work of Huang et al. [56] in which they proposed 3D convolutional neural networks whose input is nodule candidates generated by the local geometric model filter. Similarly, Wang, S. et al [41] proposed a central focused convolutional neural network (CF-CNN) which is based on a data-driven model and achieved a dice score of 77.67%. Later, the proposed study of Wu et al. [57] produces results that were not encouraging enough. In their work, multi-task CNN based on joint learning technique was proposed. In the same way, Jiang et al. [45] have also proposed another technique that did not perform well with a DSC value of 68%. They proposed multiple resolutions and residual connection-based networks and trained their network on 160*160 patches of nodules. If the Jaccard score is compared, Hancock et al. [58] and Huang et al. [42] achieved 71.85% and 70.24%, respectively. Hancock et al. [58] used the concept of a level set machine learning, while Huang et al. [42] proposed faster-regional CNN (RCNN) for candidate detection of nodules followed by false-positive reduction stage and nodule segmentation. Similarly, Qian et al. [59] proposed pyramid convolutional neural networks and achieved a Jaccard score of 71.93% (Table 2).

Table 2 Comparative analysis with the existing methods

Sr. No	Method	Dice score	Jaccard score	Year
1	Wang, S. et. al [41]	77.67	–	2017
2	Jiang et. al [45]	68%	–	2019
3	Huang et. al [56]	80.52		2017
4	Qian et. al [59]	62.8	71.93	2019
5	Hancock et. al [58]	–	71.85	2019
6	Huang et. al[42]		70.24	2019
7	Wu et. al [57]	74.05	58	2018
8	Shen et. al [55]	78.55	–	2016
	U-Net [27]	71.0	62.8	2015
	Proposed Method	81.1	72.5	2021

It is clear from Table 2 and above discussion that our proposed approach is better than the existing and standard U-Net. The reason behind this improvement over U-Net is the improved feature learning. We extract dense multi-dilated features for nodules segmentation. Our model follows dense connections between dilated convolutions of different dilated filters. The dense connections concatenate the feature maps extracted by the current layer to every next layer. Moreover, the multi-dilated filters extract wider information about context at different rates. The features are propagated more efficiently in dense structures, and it also overcomes the problem of vanishing gradients. Another big advantage of dense connections is feature reusability which reduces the calculation of different number parameters. Furthermore, the other factor of improved results is the reduced search space of the model. In most of the previous approaches, the algorithm is trained on full CT scan images. Most of the researchers extract patches from CT scans and perform training on those patches. The patch-based techniques usually required some time to first extract a lot of patches and usually face a class imbalance problem. In comparison, our approach followed an end-to-end mechanism to segment nodules.

5 Conclusion

Lung cancer remains one of the critical and common types of cancer, and its early diagnosis is required to improve its treatment. In this work, we propose an efficient algorithm to segment lung nodules. The proposed architecture extracts an improved set of features with the help of densely connected dilated convolution blocks. Moreover, to increase the performance, the CT scan image is passed through some basic preprocessing steps to lungs ROI which is further used to train the network. Comparative analysis shows that our proposed method outperformed existing approaches and standard U-Net and achieved a dice score of 81.1% and 72.5% Jaccard score on the LIDC-IRDI. Proposed work significantly addressed and segment different types of nodules present, e.g, juxta-vascular, solid, partially solid, and irregular shape nodules. In the future, we will design a more effective feature learning process combined with attention gates that is able to segment tiny nodules. Moreover, we will also embed another module in the algorithm which sends feedback about diagnosis results in real-time and based on that feedback we optimize our algorithm to improve its performance in segmenting lung nodules.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Siegel RL, Miller KD, Jemal A (2016) Cancer statistics. *CA Cancer J Clin* 66(1):7–30

2. Thomas GAS, Robinson YH, Julie EG, Shanmuganathan V, Nam Y, Rho S (2020) Diabetic retinopathy diagnostics from retinal images based on deep convolutional networks.
3. Nawaz H, Maqsood M, Afzal S, Aadil F, Mehmood I, Rho S (2020) A deep feature-based real-time system for Alzheimer disease stage detection. *Multimedia Tools and Applications*:1–19
4. Jung S, Moon J, Park S, Rho S, Baik SW, Hwang E (2020) Bagging ensemble of multilayer perceptrons for missing electricity consumption data imputation. *Sensors* 20(6):1772
5. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5(1):1–9
6. Reeves AP, Chan AB, Yankelevitz DF, Henschke CI, Kressler B, Kostis WJ (2006) On measuring the change in size of pulmonary nodules. *IEEE Trans Med Imaging* 25(4):435–450
7. Lassen B, Jacobs C, Kuhnigk J, Van Ginneken B, Van Rikxoort E (2015) Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans. *Phys Med Biol* 60(3):1307
8. Farag AA, Abd El Munim HE, Graham JH, Farag AA (2013) A novel approach for lung nodules segmentation in chest CT using level sets. *IEEE Trans Image Process* 22(12):5202–5213
9. Kubota T, Jerebko AK, Dewan M, Salganicoff M, Krishnan A (2011) Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models. *Med Image Anal* 15(1):133–154
10. Jj Z, Ji Gh, Xia Y, Xi Z (2015) Cavitory nodule segmentation in computed tomography images based on self-generating neural networks and particle swarm optimisation. *Int J Bio-Inspir Comput* 7(1):62–67
11. Diciotti S, Lombardo S, Falchini M, Picozzi G, Mascaldi M (2011) Automated segmentation refinement of small lung nodules in CT scans by local shape analysis. *IEEE Trans Biomed Eng* 58(12):3418–3428
12. Messay T, Hardie RC, Rogers SK (2010) A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med Image Anal* 14(3):390–406
13. Ye X, Beddoe G, Slabaugh G (2010) Automatic graph cut segmentation of lesions in CT using mean shift superpixels. *Int J Biomed Imaging*. <https://doi.org/10.1155/2010/983963>
14. Messay T, Hardie RC, Tuinstra TR (2015) Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset. *Med Image Anal* 22(1):48–62
15. Keshani M, Azimifar Z, Tajeripour F, Boostani R (2013) Lung nodule segmentation and recognition using SVM classifier and active contour modeling: a complete intelligent system. *Comput Biol Med* 43(4):287–300
16. Ding J, Li A, Hu Z, Wang L Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017. Springer, pp 559–567
17. Pereira S, Pinto A, Alves V, Silva CA (2016) Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging* 35(5):1240–1251
18. Robinson YH, Vimal S, Julie EG, Narayanan KL, Rho S (2021) 3-Dimensional Manifold and Machine Learning Based Localization Algorithm for Wireless Sensor Networks. *Wireless Personal Communications*:1–19
19. Jifara W, Jiang F, Rho S, Cheng M, Liu S (2019) Medical image denoising using convolutional neural network: a residual learning approach. *J Supercomput* 75(2):704–718
20. Kalsoom A, Maqsood M, Ghazanfar MA, Aadil F, Rho S (2018) A dimensionality reduction-based efficient software fault prediction using Fisher linear discriminant analysis (FLDA). *J Supercomput* 74(9):4568–4602
21. Jiang F, Grigorev A, Rho S, Tian Z, Fu Y, Jifara W, Adil K, Liu S (2018) Medical image semantic segmentation based on deep learning. *Neural Comput Appl* 29(5):1257–1265
22. Valverde S, Oliver A, Roura E, González-Villà S, Pareto D, Vilanova JC, Ramió-Torrentà L, Rovira À, Lladó X (2017) Automated tissue segmentation of MR brain images in the presence of white matter lesions. *Med Image Anal* 35:446–457
23. Moeskops P, Viergever MA, Mendrik AM, De Vries LS, Benders MJ, Išgum I (2016) Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Trans Med Imaging* 35(5):1252–1261
24. Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, Shen D (2015) Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage* 108:214–224

25. Pezzano G, Ripoll VR, Radeva P (2021) CoLe-CNN: Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation. *Comput Methods Programs Biomed* 198:105792
26. Keetha NV, Annavarapu CSR (2020) U-Det: A modified U-Net architecture with bidirectional feature network for lung nodule segmentation. *arXiv preprint*
27. Ronneberger O, Fischer P, Brox T U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, 2015. Springer, pp 234–241
28. Kostis WJ, Reeves AP, Yankelevitz DF, Henschke CI (2003) Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. *IEEE Trans Med Imaging* 22(10):1259–1274
29. Sargent D, Park SY Semi-automatic 3D lung nodule segmentation in CT using dynamic programming. In: *Medical Imaging 2017: Image Processing*, 2017. International Society for Optics and Photonics, p 101332R
30. Kuhnigk J-M, Dicken V, Bornemann L, Bakai A, Wormanns D, Krass S, Peitgen H-O (2006) Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Trans Med Imaging* 25(4):417–434
31. Wang J, Guo H (2016) Automatic approach for lung segmentation with juxta-pleural nodules from thoracic CT based on contour tracing and correction. *Comput Math Methods Med*. <https://doi.org/10.1155/2016/2962047>
32. Nithila EE, Kumar S (2016) Segmentation of lung nodule in CT data using active contour model and Fuzzy C-mean clustering. *Alex Eng J* 55(3):2583–2588
33. Chan TF, Vese LA (2001) Active contours without edges. *IEEE Trans Image Process* 10(2):266–277
34. Mukherjee S, Huang X, Bhagalia RR Lung nodule segmentation using deep learned prior based graph cut. In: *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, 2017. IEEE, pp 1205–1208
35. Mukhopadhyay S (2016) A segmentation framework of pulmonary nodules in lung CT images. *J Digit Imaging* 29(1):86–103
36. Shen S, Bui AA, Cong J, Hsu W (2015) An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. *Comput Biol Med* 57:139–149
37. Lu L, Devarakota P, Vikal S, Wu D, Zheng Y, Wolf M Computer aided diagnosis using multi-level image features on large-scale evaluation. In: *International MICCAI Workshop on Medical Computer Vision*, 2013. Springer, pp 161–174
38. Hu Y, Menon PG A neural network approach to lung nodule segmentation. In: *Medical Imaging 2016: Image Processing*, 2016. International Society for Optics and Photonics, p 978420
39. Jung J, Hong H, Goo JM (2018) Ground-glass nodule segmentation in chest CT images using asymmetric multi-phase deformable model and pulmonary vessel removal. *Comput Biol Med* 92:128–138
40. Gonçalves L, Novo J, Campilho A (2016) Hessian based approaches for 3D lung nodule segmentation. *Expert Syst Appl* 61:1–15
41. Wang S, Zhou M, Liu Z, Liu Z, Gu D, Zang Y, Dong D, Gevaert O, Tian J (2017) Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. *Med Image Anal* 40:172–183
42. Huang X, Sun W, Tseng T-LB, Li C, Qian W (2019) Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks. *Comput Med Imaging Graph* 74:25–36
43. Cao H, Liu H, Song E, Hung C-C, Ma G, Xu X, Jin R, Lu J (2020) Dual-branch residual network for lung nodule segmentation. *Appl Soft Comput* 86:105934
44. Ali I, Muzammil M, Haq IU, Khaliq AA, Abdullah S (2020) efficient lung nodule classification using transferable texture convolutional neural network. *IEEE Access* 8:175859–175870
45. Jiang J, Hu Y-C, Liu C-J, Halpenny D, Hellmann MD, Deasy JO, Mageras G, Veeraraghavan H (2018) Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images. *IEEE Trans Med Imaging* 38(1):134–144
46. Liu M, Dong J, Dong X, Yu H, Qi L Segmentation of lung nodule in CT images based on mask R-CNN. In: *2018 9th International Conference on Awareness Science and Technology (ICAST)*, 2018. IEEE, pp 1–6

47. Tang H, Zhang C, Xie X Nodulenet: Decoupled false positive reduction for pulmonary nodule detection and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019. Springer, pp 266-274
48. Ioffe S, Szegedy C Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning, 2015. PMLR, pp 448–456
49. Nair V, Hinton GE Rectified linear units improve restricted boltzmann machines. In: Icm1, 2010.
50. Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv preprint
51. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. pp 4700–4708
52. He K, Zhang X, Ren S, Sun J Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, 2015. pp 1026–1034
53. Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA (2011) The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys* 38(2):915–931
54. Bukhari M, Bajwa KB, Gillani S, Maqsood M, Durrani MY, Mehmood I, Ugail H, Rho S (2020) An efficient gait recognition method for known and unknown covariate conditions. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2020.3047266>
55. Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, Zang Y, Tian J (2017) Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recogn* 61:663–673
56. Huang X, Shan J, Vaidya V Lung nodule detection in CT using 3D convolutional neural networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 2017. IEEE, pp 379–383
57. Wu B, Zhou Z, Wang J, Wang Y Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018. IEEE, pp 1109–1113
58. Hancock MC, Magnan JF (2019) Lung nodule segmentation via level set machine learning. arXiv preprint
59. Zhao X, Sun W, Qian W, Qi S, Sun J, Zhang B, Yang Z Fine-grained lung nodule segmentation with pyramid deconvolutional neural network. In: Medical Imaging 2019: Computer-Aided Diagnosis, 2019. International Society for Optics and Photonics, p 109503S

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.