International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

# Novel Algorithm for Pulmonary Nodule Classification using CNN on CT Scans

**Drishti[1], Jaspreet Singh[2]**

**Submitted**: 17/09/2023 **Revised**: 16/11/2023 **Accepted**: 27/11/2023

**Abstract:** For the purpose of making a preliminary diagnosis of lung cancer, computed tomography, or CT, is frequently utilized to find pulmonary nodules. However, as a result visual similarities among non-cancerous and cancerous nodules, identifying malignant from cancer nodules is not easy for doctors to accomplish. Here, a novel Convolution Neural Network architecture known as ConvNet is suggested to classify lung nodules as malignant or benign. A multi-scale, multi-path architecture is developed and utilized to increase the classification performance. This is done since there is a large variance in the nodule characteristics that are displayed in CT scan images, like Shape and Size. The multiple scale method makes use of filters of varying sizes to extract nodule features from local regions in a more efficient manner, and the multiple path architecture combines features extracted from various Convolution Network layers in order to improve the nodule features in relation to global regions. Both of these methods are part of the multi-path architecture. The LUNGx Challenge database is used to train and assess the proposed ConvNet, and it obtains specificity of 0.924, sensitivity of 0.887, and AUC of 0.948. The suggested Convolution Network is able to obtain an AUC improvement that is 14 percent higher than the current state-of-the-art unsupervised learning technique. The proposed Convolution Network also performs better than the previous state-of-the-art Convolution Networks that were specifically created for the categorization of pulmonary nodules. The suggested Convolution Networks has the potential to aid radiologists in making diagnostic judgments during CT screening when it is utilized in clinical settings.

***Keywords***: *CT-Scans, CNN, Deep Learning, Lung Nodule Detection*

## 1. Introduction

The most common type of cancer that results in mortality is lung cancer, as indicated by the statistics that were compiled and released by the American Cancer Society (2017). The ability to detect and diagnose cancer at an earlier stage is one of the most crucial factors in determining whether or not a patient will survive their illness. Computed tomography, sometimes known as CT, is frequently utilized by radiologists in the process of diagnosing lung cancer.

CAD tools have been created to help radiologists differentiate between cancerous and non- cancerous nodules in patients' chests. In general, these CADx systems use a classifier to transform the properties of the nodule into conclusions regarding whether or not the nodule is benign or cancerous. As a result of the fact that the CADx system helps radiologists achieve a higher level of diagnostic accuracy, it is an excellent option for performing preliminary diagnoses. Low-level features of medical imaging are used by early established CADx systems. Contour, size, and form are features image structures in the images. Because pulmonary nodules come in a wide variety of sizes and forms, as shown in Fig. 1, our techniques demonstrate that low-level traits can be beneficial for classifying pulmonary nodules.

Unsupervised learning strategies, on the other hand, call for the manual extraction of features, which makes it difficult to determine which combination of features is best (Anthimopoulos et al.).

DL has seen a significant deal of success in recent years, particularly in the areas of picture objective detection, classification, NLP and many more. Deep neural networks, often known as DNNs, are capable of approaching human levels of performance in several of these areas (Silver et al. 2016).

In a traditional ConvNet, the early layers are responsible for the generation of global structures. When one reaches deeper layers, one begins to encounter increasingly complex local systems. A conventional ConvNet typically consists of a single path and is limited to making use of the local structures while performing classification. On the other hand, in the classification of pulmonary nodules, merging local and global structures by way of a skip connection is anticipated to increase the classification performance. Additionally, the traditional Convolution Network only uses a single-scale filter when it comes to the feature extraction process. On the other hand, because the sizes of the nodules might vary, it may be possible to obtain more accurate nodule features by conducting nodule image analysis using multi-scale filters.

An unique CNN architecture is suggested with the goal of satisfying the demand for better pulmonary nodule classification. The suggested ConvNet makes use of a

[1] *GD Goenka University, Gurugran – 122103, India*
[2] *Sharda University, Greater Noida – 201310, India*

multiple path feature extraction strategy in order to maintain the integrity of both the local and the global structures. In addition, the suggested ConvNet makes use of multi-scale convolutional layers, which include the application of numerous filters of varying sizes, in order to cover a greater number of useful nodule properties. The suggested ConvNet achieves a 14 percent improvement in an area under the ROC curves when compared to the approach (Nishio and Nagashima 2017). This is in comparison to the previous approach (AUC).

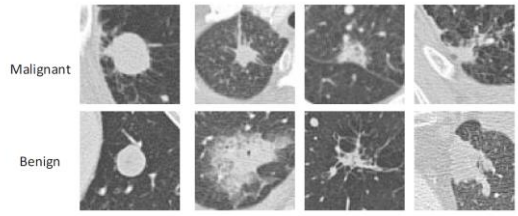


**Fig 1.** The definition of pulmonary nodules with examples. The nodules can be found in the middle of the image boxes that are 80mm by 80mm. These examples show that classifying pulmonary nodules can be difficult due to the wide range of sizes, forms, and comparable visual representations shared by malignant and benign nodules.

## 2. Datasets

[3] contains the CT scans of seventy patients, and the overall collection has 83 pulmonary nodules, of which 41 are malignant and 42 are benign. Each CT scan included in the LUNGx Challenge database was acquired with a tube peak potential energy of either 120kV or 140kV, with a tube current-exposure time product of between 200 and 325 mA.

The Computed Tomography scans are reconstructed DICOM files, which contain a variety of 2D slice images. The size of each slice image is 512 pixels by 512 pixels, and the slice thickness is 1 millimeter.

In Figure 2, a selection of the bounding boxes are depicted as red rectangles for your viewing pleasure. (3) The nodule patches are retrieved by cropping the slice images in a way that respects the bounding boxes. Different patches obtained the same nodules had distinct visual manifestations, as seen in Figure 2 (see also Fig. 1).

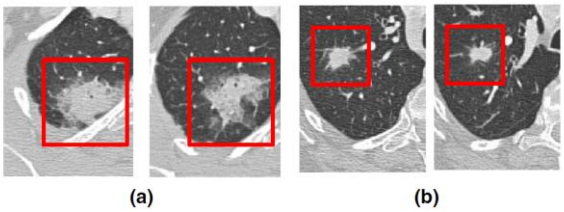


**Fig.2.** The nodule patches were extracted using bounding boxes. **A.** 2 different benign nodule patches photographed after being cut from the same non-cancerous nodule twice. **B**. 2 different malignant nodule patches taken from different slices.

In order for ConvNets to attain good classification performance, they require a substantial dataset on which to train their vast number of learnable parameters. For instance, there are 1.2 million training photos in the ImageNet database, which was used to train AlexNet. According to what was covered in Section 2, the LUNGx Challenge database only has 1131 nodule patches retrieved from it. When there is a finite amount of data, the trained network is unable to model the training data accurately nor can it generalise its performance to make accurate predictions for new data. In most cases, a data augmentation strategy is utilised in order to accomplish the goal of increasing the total number of training photos.

In this paper, the data augmentation involves several techniques by randomly selecting image flipping, image rotation and image translation. The goal of this is to achieve a unique spatial variance for each enhanced nodule patch. The rotation is accomplished by rotating the nodule patches at an arbitrary angle that ranges from 0 to 359 degrees.

The orientation of the nodule patches in either the vertical or horizontal plane is arbitrary. The translation causes a haphazard movement of the nodules closer to the boundaries of the bounding boxes. As can be seen in Figure 3b, the initial bounding box must be expanded by twenty percent in order for translation to be performed.

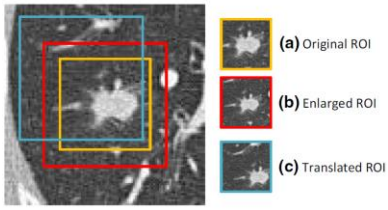


**Fig.3.** A visual representation of the translating process. **A.** Nodule patch that was removed according to the original bounding box. **B.** The extracted nodule patch twice the original. **C.** Nodule patch translation

### 2.2. Image Normalization and Contrasting

When applying the contrast normalization technique, the global mean is subtracted from each pixel in each image before applying the technique. The intensities of all pixels are averaged in order to calculate the global mean. This is accomplished by first adding up the intensities of all pixels in all of the photos, then dividing this total by the total number of images. The performance of the ConvNet may be negatively impacted when a big gradient is present since this may force the network to converge toward a local minimum.

## 3. Proposed Convolution Network

The suggested ConvNet utilizes two strategies that have been developed in order to increase classification performance: (1) The Multiple path Feature Extraction creates one extra path to combine features from other layers and later layers

preserve from loss of features. (2) In order to provide multi-scale features, the traditional conv. layer is removed and replaced with the proposed multiple scale Conv. layer. The multi-scale conv. layer makes use of numerous filters of varying sizes.

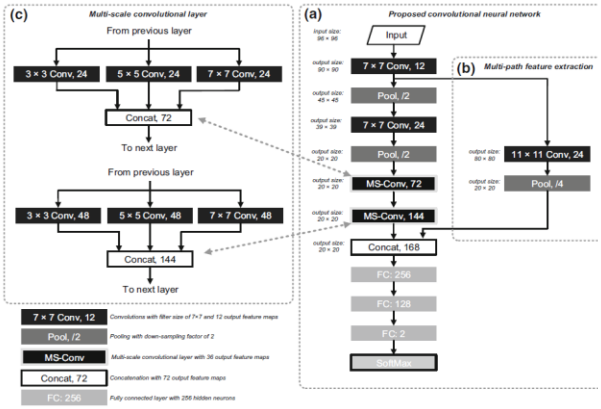


**Fig.4. A.** General outline of the planned architecture for the ConvNet. **B.** An illustration showing how the MP feature extraction works. **C.** A more in-depth look at multiple scale conv. layer.

### 3.1. Architecture

The Figure 4a depicts architecture that would be used for the proposed ConvNet. The suggested ConvNet has a size of input that is 96*96. Two feed-forward routes are included in the proposed ConvNet design. The main path is responsible for performing multiple scale for extraction of features, and its fundamental structure is founded on the hierarchical neural network. The main path's fundamental structure is made up of four consecutive single-scale convolutional layers, which are then followed by three fully connected layers. On the other hand, the second path is the one that is utilized for multi-path feature extraction. This is accomplished by establishing a direct connection between the first single-scale conv. layer and the first FC layer along the main path. In relation to the primary path, the multiple scale extraction of features occurs after the two conv. layers that are only of the single-scale variety. In the 1st single-scale convolutional layer, a convolution operation with a size of 7*7 is carried out in order to obtain 12 output feature maps.

In a similar manner, the second single-scale convolutional layer generates 24 output layer feature maps by utilizing the same filter size. Then, after this, there are two multi-scale convolutional layers that come after this. In the end, the two different routes will be combined and linked to the initial layer that will be entirely connected.

Following the completion of three layers that are FC, the final output features are analyzed in order to divide pulmonary nodules into two categories: benign and malignant. There are 256 neurons that are concealed in the first layer that is fully linked. In order to match the number of output classes, the last fully connected layer has only two hidden neurons, whereas the second fully connected layer contains 128 hidden neurons. After the last layer that is completely connected, the results that were predicted are normalized using SoftMax (Bishop 2006).

The activation function of each convolutional layer is the ReLU layer, which stands for the "rectified linear unit." Positive values are maintained while negative ones are truncated to zero by the ReLU. In comparison to the sigmoid function and the tanh function, the ReLU exhibits significantly more favorable gradient changes (Krizhevsky et al. 2012). In addition, it is simple to apply, which makes it an effective tool for enhancing the speed performance of the network during both training and inference.

Following each layer of convolutional processing, a max-pooling layer is added. In order to preserve more pertinent local structures, the max-pooling layers of the network are utilized to lower the spatial resolution of the layers that immediately follow them in the hierarchy. In addition, the pooling process expands the extent of the receptive field, which enables the network to acquire a greater understanding of intricate regional patterns based on the input.

#### 3.1.1. Multiple path extraction of features

The The traditional ConvNet has a direct route for passing extracted features into the fully connected levels of the network. The earlier layers of a typical ConvNet are responsible for the feature extraction in relation to global structure. When the ConvNet is allowed to propagate to deeper layers, the previously extracted features become less dense and more locally concentrated, while the previously extracted global structures become less prominent. The performance of the traditional ConvNet may be limited in the study of the pulmonary nodules, despite the fact that it is well-suited to extract features that describe local structures. Local structures have a lower capacity to represent such low-level information as compared to global structures. This is due to the fact that the visual representations of pulmonary nodules come in a variety of different shapes and sizes.

The multi-path feature extraction that has been offered is a short-cut solution to the problem of lost global structures. The output of the first conv. layer is split in two by the shortcut, which also establishes a skip connection to the first FC layer, as can be seen in Figure 4b. A maximum PL is put into the shortcut in order to blend the characteristics that come from two different branches. Combination of both local and global structures because they were created by passing the features of the early layer into the first FC layer. As a result, it is anticipated that the final combined features will provide more nodule features.

### 3.1.2. Multiple Scale convolution layer

A simplistic approach to enhancing the performance of a ConvNet is to add more convolutional layers to the network, hence increasing its depth. Due to the limited size of the dataset, this strategy would result in overfitting. Convolution requires careful consideration when selecting an appropriate filter size. This can be challenging. Choosing the appropriate filter size can considerably increase diagnostic accuracy when pulmonary nodule categorization is being performed. A multiple scale conv. layer is offered as a means of improving the extraction of local sparse structures with regard to various receptive fields in this particular piece of research. The filter bank in the multiple scale conv. layer is constructed with 3 different filter sizes as opposed to utilizing a single-scale filter. Figure 4c illustrates the configurational specifics of the multiple scale conv. layer in its entirety.

There are three different branches that make up the multiple scale conv. layer. The first branch carries out an operation known as a 3*3 convolution. By utilizing a 5*5 convolutional layer, the second branch of the tree is able to accomplish its goal of extracting comparatively larger local structures. The final branch also consists of one conv. layer, but the size of the filter has been increased to 7*7 in order to take into account a greater number of local structures. A depth-wise concatenation is used to combine the feature maps that were produced by three different branches. In order to carry out an efficient concatenation, the input of the 3*3 conv. layer is padded with a 1*1 border of zeros, while the inputs of the 5*5 and 7*7 convolutional layers, respectively, employ zero paddings of 2 or 3, depending on which layer they are.

The ConvNet that has been proposed features an extension of a single-scale conv. layer to a multiple scale conv. layer. This provides the network with the ability to produce more local structures of varied sizes.

### 3.2. Training of model and result evaluation

The suggested Convolution Network is evaluated and trained Network using 5-fold cross validation, in order to pick a trained model with strong generalization. Each fold has benign and malignant nodules with 10% each, with a total of 41 and 42, respectively. In every iteration of the cross-validation procedure, 3 folds are set aside for training, whereas only one-fold is allocated for validation and one-fold is used for testing. The augmentation is done while the data set is still being used for training. When calculating overall performance, an average is taken of the accuracies, sensitivities, and specificities obtained from each of the 5 cross-validation folds.

Class imbalance may have a negative impact on training convergence and the generalization of the trained model due to the fact that the enhanced nodule patches contain a greater number of malignant samples than benign ones (Mazurowski et al. 2008). Over-sampling the minority group is one of the most prevalent solutions to the problem of imbalance, and research has shown that doing so leads to improved performance in deep learning. The proposed training and validation datasets are subjected to over-sampling so that they are more evenly distributed. This is accomplished by randomly selecting and duplicating benign nodule patches.

## 4. Analysis of Results

### 4.1. Experimental Setup

Caffe, a framework for deep learning developed by Jia et al. (2014), is used to implement the suggested ConvNet. The Caffe environment is often used for the implementation of other research involving convolutional neural networks. Training and testing of the network is carried out on a computer running Ubuntu 16.04 Linux

### 4.2. Proposed Convolution Network Algorithm Analysis

### 4.2.1. Hyper-parameter Tunning

A variety of experiments are carried out in order to provide evidence in support of the decisions made regarding the proposed ConvNet input scale, no. of hidden neurons in the FC layers, and training settings. Table 1 provides a concise summary of the findings. The accuracy of the categorization decreased by approximately 1.5% when moving from 96*96 to 128*128 on the input scale. When the input scale is decreased to 64 by 64, there is a 4% decrease in accuracy. Additionally, when the no. of hidden neurons in the FC layers is either doubled or halved, the classification accuracies that are produced are inferior to those that were proposed. If the dropout technique is not implemented, the proposed ConvNet displays an accuracy that is around 1.5% lower. In addition, there is a noticeable gain in accuracy when the batch size is increased from 96 to 128, which is around 0.5 percent.

| Input scale | FC layer factor | Dropout ratio | Batch size | Weight initialization | Accuracy (%) |
|---|---|---|---|---|---|
| *64* | 1 | 0.5 | 128 | 0.01 | 86.60 |
| *128* | 1 | 0.5 | 128 | 0.01 | 88.77 |
| 96 | *1/2* | 0.5 | 128 | 0.01 | 89.17 |
| 96 | 2 | 0.5 | 128 | 0.01 | 87.61 |
| 96 | 1 | *0* | 128 | 0.01 | 88.50 |
| 96 | 1 | 0.5 | *96* | 0.01 | 89.93 |
| 96 | 1 | 0.5 | 128 | *0.005* | 89.28 |
| 96 | 1 | 0.5 | 128 | *0.05* | 88.50 |
| **96** | **1** | **0.5** | **128** | **0.01** | **90.38** |

Italics represents the changed hyper-parameters.
Bold represents the best configuration of hyper-parameters

**Table 1.** Accuracy of classification achieved by the proposed ConvNet using various settings for its hyper-parameters.

### 4.2.2. Multiple Path Feature Extraction Analysis

The performance of the multiple path feature extraction is examined by comparing the classification performances of the proposed ConvNet with and without the shortcut. This is done in order to evaluate the effectiveness of the multi-

path feature extraction. The short cut raises the accuracy from 86.77 to 90.38 percent, the AUC from 0.914 to 0.948 percent, and adds to a 2 percent increase in sensitivity and a 6 percent increase in specificity. Table 2 presents these findings for your perusal. By establishing numerous branches simultaneously, an investigation is carried out to see how successful the suggested multi-path feature extraction is. Intuitively, each of the branches makes use of the identical input features and extracts those features using an independent convolution operation, which is then followed by maximum pooling.

As can be seen in Table 3, when branches are added to the proposed multi-path feature extraction method, the classification performance suffers as a result. In particular, the multi-path feature extraction performed using three branches has a worse classification performance compared to the one performed using two branches. In addition, the classification performance of a system that uses multi-scaled convolution filters (for example, "11*11, 13*13") is superior to that of a system that uses a fixed-scale convolution filter (for example, "11*11, 11*11") in terms of both accuracy and speed.

| Network | Accuracy (%) | Sensitivity | Specificity | AUC |
|---|---|---|---|---|
| Proposed ConvNet without shortcut | 86.77 | 0.869 | 0.863 | 0.914 |
| Proposed ConvNet with shortcut | **90.38** | **0.887** | **0.924** | **0.948** |

Bold represents the best performance among all compared methods

**Table 2.** Comparison of proposed Convolution Network accuracy in classifying data with and without the use of a shortcut

| Filter configuration {11 × 11} | Accuracy (%) **90.38** | Sensitivity **0.887** | Specificity **0.924** | AUC **0.948** |
|---|---|---|---|---|
| {11 × 11, 11 × 11} | 88.40 | 0.870 | 0.892 | 0.920 |
| {11 × 11, 13 × 13} | *88.79* | *0.877* | *0.901* | *0.938* |
| {11 × 11, 15 × 15} | 87.63 | 0.861 | 0.895 | 0.922 |
| {11 × 11, 17 × 17} | 86.34 | 0.847 | 0.884 | 0.916 |
| {11 × 11, 19 × 19} | 86.21 | 0.849 | 0.878 | 0.911 |
| {11 × 11, 21 × 21} | 85.18 | 0.833 | 0.875 | 0.910 |
| {11 × 11, 11 × 11, 11 × 11} | 87.11 | 0.866 | 0.878 | 0.914 |
| {11 × 11, 13 × 13, 15 × 15}* | 87.79 | 0.870 | 0.887 | 0.929 |
| {11 × 11, 13 × 13, 17 × 17} | 85.95 | 0.838 | 0.887 | 0.921 |

{11 × 11, 13 × 13, 15 × 15}*–Three branches' multi-path feature extraction with filter sizes of 11 × 11, 13 × 13 and 15 × 15
Bold represents the top-1 values among all filter configurations.
Italics represents the top-2 values among all filter configurations

**Table 3** Comparison of the classification abilities of the proposed multiple path feature extraction with a variety of filter configurations

### 4.3. Comparison with previous works

#### 4.3.1. Comparison with Unsupervised feature learning approach

The suggested method is contrasted with Nishio's work, which is an approach to nodule-based unsupervised feature learning. The purpose of this comparison is to demonstrate that the proposed method is more effective than an unsupervised learning scheme. The research conducted by Nishio uses a combination of principal component analysis (PCA) and convolution operation in order to extract robust unsupervised features. According to the results of the nodule-based study, the suggested method performs better than Nishio's method (Specificity = 0.889, Sensitivity = 0.875, AUC = 0.958), as shown in Table 4. The unsupervised feature extraction that is detailed in Nishio's work is surpassed by the suggested method in two significant ways, which are as follows:

(1) The covariance matrix serves as the foundation for principal components. Because of the complexity of high-dimensional spaces, it is not possible to accurately evaluate the covariance matrix in a way that would allow it to offer adequate high-level characteristics. (2) The PCA approach does not have a particularly great ability to assess the differences in object placement (Li et al. 2008). During the testing phase of the PCA approach, it is possible for erroneous predictions to be generated when the nodules occur in different positions of the input images. On the other hand, the approaches that are based on ConvNet have a great ability to process the translational variance that is present in objects. The ability to extract features is improved as a result by the method that was proposed.

#### 4.3.2. Comparison with other ConvNet approaches

These ConvNets include those that use single-scale convolutional layers (Li et al. 2014; Zhao et al. 2018), pre-trained AlexNet and GoogLeNet with transfer learning (Shin et al. 2016; Tajbakhsh et al. 2016), and multi-cropped ConvNets (Shen et al. 2017). The classification performance statistics of previous research are created by training and evaluating their ConvNets using the LUNGx Challenge database with identical data partitioning and evaluation scheme as used in the proposed architecture. This ensures that the results are comparable (as described in Sect. 3.2). The findings are presented in a nutshell in Table 5. When compared to other ConvNets, the suggested ConvNet has the highest levels of accuracy (90.38 percent), specificity (0.924), and AUC (0.948). The resultant sensitivity of the suggested ConvNet is 0.887, which is only 0.019 points lower than the sensitivity achieved by the model with the highest sensitivity, which was a pretrained GoogLeNet with transfer learning (Shin et al. 2016).

In addition to this, the performance of the proposed ConvNet is evaluated and contrasted with that of the multi-crop ConvNet (Shen et al. 2017), which crops feature maps in the direction of multi-scale feature learning. Using multi-crop ConvNet to extract patch-based data should also be legitimate, despite concerns that the multi-crop ConvNet is evaluated under nodule-based data. The aim is to crop the feature maps with varying sizes to boost the core nodule feature, regardless of the input shape. According to the findings presented in Table 5, the suggested ConvNet demonstrates superior accuracy as well as AUC in comparison to the multi-crop ConvNet. The suggested

multi-scale convolutional layer does not crop the feature maps; rather, it uses many filters of varying sizes to produce more local structures. As a result, the central nodule features are also enhanced. On the other hand, the multi-path feature extraction that has been presented utilises the beneficial global structures to further improve the nodule features that correlate to the size or shape of the nodule.

| | Method | Cross validation | ROC analysis | Sensitivity | Specificity | AUC |
|---|---|---|---|---|---|---|
| Nishio and Nagashima (2017) | Hisogram + SVM | 10-fold | nodule-based | 0.867 | 0.488 | 0.640 |
| | LBP-TOP + SVM | | | 0.900 | 0.558 | 0.688 |
| | RLBP + SVM | | | 0.800 | 0.674 | 0.725 |
| | Multi-stages + SVM | | | 0.867 | 0.744 | 0.837 |
| Proposed | ConvNet | 5-fold | nodule-based | 0.875 | 0.889 | 0.958 |
| | | | patch-based | 0.887 | 0.924 | 0.948 |

Histogram, histogram of CT density; LBP-TOP, local binary pattern on the three orthogonal planes; RLBP, LBP with 3D random sampling; Multi-stages, combination of PCA, convolution and pooling operations; SVM, support vector machine

**Table 4.** The suggested ConvNet is compared with an unsupervised feature learning approach.

| | Methods | Accuracy (%) | Sensitivity | Specificity | AUC |
|---|---|---|---|---|---|
| Li et al. (2014) | ConvNet with 1 single-scale convolutional layer | 80.15 | 0.789 | 0.818 | 0.854 |
| Zhao et al. (2018) | ConvNet with 2 single-scale convolutional layers | 84.97 | 0.843 | 0.858 | 0.902 |
| Tajbakhsh et al. (2016) | Transfer learning on pre-trained AlexNet | 80.58 | 0.821 | 0.787 | 0.855 |
| Shin et al. (2016) | Transfer learning on pre-trained GoogLeNet | 86.68 | **0.906** | 0.798 | 0.933 |
| Shen et al. (2017) | Multi-cropped ConvNet | 86.77 | 0.846 | *0.895* | *0.940* |
| Proposed method | Multi-scale + multi-path | **90.38** | *0.887* | **0.924** | **0.948** |

Bold represents the top-1 values among all ConvNets
Italics represents the top-2 values among all ConvNets

**Table 5** The suggested ConvNet is compared to several different ConvNet topologies that are stored in the LUNGx Challenge database.

Figure 5 illustrates the ROCs for a variety of ConvNet designs. A statistical analysis, namely the significance level under the DeLong test (Efron 1993), is carried out in order to examine the statistical significance of the AUC differences that exist between the proposed ConvNet and other ConvNet techniques.

According to the findings of the DeLong test, the suggested ConvNet showed statistical significance ($\rho < 0.05$) in comparison to other ConvNet techniques, with the exception of pre-trained

GoogLeNet (Shin et al. 2016) ($\rho$=0.0509) and multi-crop ConvNet. The suggested multi-scale convolutional layer does not crop the feature maps; rather, it uses many filters of varying sizes to produce more local structures. As a result, the central nodule features are also enhanced. On the other hand, the multi-path feature extraction that has been presented utilises the beneficial global structures to further improve the nodule features that correlate to the size or shape of the nodule.

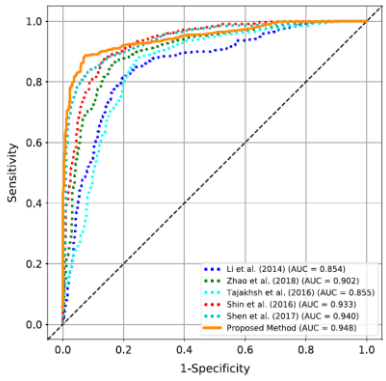


**Fig. 5.** ROC curves for various ConvNet architectural configurations

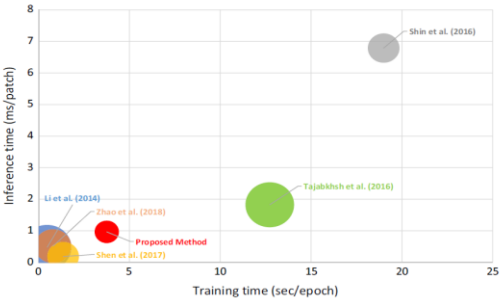


**Fig. 6.** Various ConvNets inference times compared to their training times and their testing errors.

The proposed ConvNet is compared to other ConvNets in terms of its training time, inference time, and testing error. Note that each experiment was carried out using a graphics processing unit (GPU) powered by an NVIDIA K40. The estimated time needed by the proposed ConvNet to finish predicting one nodule patch is 0.966 seconds. The suggested ConvNet requires 3.750 milliseconds for each training session to be completed. As can be shown in Figure 6, the training and inference times for the proposed ConvNet are significantly longer than those reported by Li et al. (2014), Zhao et al. (2018), or Shen et al (2017). It is reasonable to anticipate that the suggested ConvNet will require more time to train or perform inference than the aforementioned three ConvNets due to the fact that it has a greater number of learnable parameters. In spite of this, the suggested ConvNet has an inference time that is only a fraction of a millisecond slower than the other three ConvNets, which is an extremely insignificant difference for actual clinical applications. The proposed ConvNet had the lowest testing error of all the ConvNets that were examined; this is the most important finding.

## 5. Conclusion

A unique Convolution Network architecture is described here for the categorization of lung nodules based on CT scans. In order to blend fine global structures from the early layers with sparse local structures from the later layers, the proposed ConvNet makes use of a multi-path feature extraction approach. This results in an accuracy improvement of 4% when compared to the previously proposed ConvNet, which does not utilise a multi-path feature extraction approach. When compared to the use of single-scale convolutional layers, the proposed ConvNet makes use of a multi-scale convolutional layer to extract features at different scales. This helps the network learn local structures more efficiently and results in an increase in classification accuracy of more than 2 percent. The suggested ConvNet delivers a 14 percent improvement in AUC when compared to an earlier unsupervised feature learning technique. The suggested ConvNet has the potential to produce up to 13 percent greater accuracy and 11 percent higher AUC when compared to earlier versions of ConvNets that were utilised in medical imaging.

## References


[1] American Cancer Society. 2017. Retrieved January, 2019 from Cancer facts & figures 2017. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html

[2] Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Transactions on Medical Imaging*, *35*(5), 1207–1216. https://doi.org/10.1109/TMI.2016.2535865.

[3] Armato, S. G., Drukker, K., Li, F., Hadjiiski, L., Tourassi, G. D., Kirby, J. S., et al. (2016). LUNGx challenge for computerized lung nodule classification. *Journal of Medical Imaging*, *3*, 3–9. https://doi.org/10.1117/ 1.JMI.3.4.044506.

[4] Armato, S. G., Giger, M. L., Moran, C. J., Blackburn, J. T., Doi, K., & MacMahon, H. (1999). Computerized detection of pulmonary nodules on CT scans. *RadioGraphics*, *19*(5), 1303–1311.

[5] Bishop, C. M. (2006). Pattern recognition and machine learning (information science and statistics). Secaucus, NJ: Springer.

[6] Chen, J., & Shen, Y. (2017). The effect of kernel size of CNNs for lung nodule classification. In *2017 9th international conference on advanced infocomm technology (ICAIT)* (pp. 340–344). https://doi.org/10.1109/ICAIT.2017.8388942.

[7] Ciompi, F., Jacobs, C., Scholten, E. T., Wille, M. M. W., de Jong, P. A., Prokop, M., et al. (2015). Bag-of-frequencies: A descriptor of pulmonary nodules in computed tomography images. *IEEE Transactions on Medical Imaging*, *34*(4), 962–973. https://doi.org/10.1109/TMI.2014.2371821.

[8] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255).

[9] Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*, 2121–2159.

[10] Efron, B. (1993). An introduction to the bootstrap. Monographs on statistics and applied probability (Series) (Vol. 57). New York: Chapman & Hall.

[11] Giger, M. L., Bae, K. T., & Macmahon, H. (1994). Computerized detection of pulmonary nodules in computed tomography images. *Investigative Radiology*, *29*(4), 459–465.

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2015a). Deep residual learning for image recognition. CoRR arXiv:1512.03385.

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2015b). Deep residual learning for image recognition. CoRR arXiv:1512.03385.

[14] Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehen- sive tutorial with selected use cases. *Journal of Pathology Informatics*, *7*, 29. https://doi.org/10.4103/ 2153-3539.186902.

[15] Jarrett, K., Kavukcuoglu, K., Ranzato, M., & LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? In *2009 IEEE 12th international conference on computer vision* (pp. 2146–2153).

[16] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014) Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093.

[17] Kamiya, A., Murayama, S., Kamiya, H., Yamashiro, T., Oshiro, Y., & Tanaka, N. (2014). Kurtosis and skewness assessments of solid lung nodule density histograms: Differentiating malignant from benign nodules on CT. *Japanese Journal of Radiology*, *32*(1), 14–21.

[18] Kang, G., Liu, K., Hou, B., & Zhang, N. (2017). 3D multi-view convolutional neural networks for lung nodule classification. *PLOS ONE*, *12*, 1–21. https://doi.org/10.1371/journal.pone.0188290.

[19] Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th international conference on neural information processing systems—volume 1, Curran Associates Inc., USA, NIPS'12 (pp. 1097–1105).

[20] Levi, G., & Hassncer, T. (2015). Age and gender classification using convolutional neural networks. In *2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 34–42). https://doi.org/10.1109/CVPRW.2015.7301352.

[21] Li, C., Diao, Y., Ma, H., & Li, Y. (2008). A statistical PCA method for face recognition. In *2008 Second international symposium on intelligent information technology application* (vol. 3, pp. 376–380). https://doi.org/10.1109/IITA.2008.71.

[22] Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., & Chen, M. (2014). Medical image classification with convolutional neural network. In *2014 13th international conference on control automation robotics vision (ICARCV)* (pp. 844–848).

[23] Li, W., Cao, P., Zhao, D., & Wang, J. (2016). Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Computational and Mathematical Methods in Medicine*, *2016*, 6215085. https://doi.org/10.1155/2016/6215085.

[24] Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., & Tourassi, G. D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*, *21*(2), 427–436. (Advances in Neural Networks Research: IJCNN '07).

[25] Monkam, P., Qi, S., Xu, M., Han, F., Zhao, X., & Qian, W. (2018). CNN models discriminating between pulmonary micro-nodules and non-nodules from CT images. *BioMedical Engineering OnLine*, *17*(1), 96. https://doi.org/10.1186/s12938-018-0529-x.

[26] Nishio, M., & Nagashima, C. (2017). Computer-aided diagnosis for lung cancer: Usefulness of nodule heterogeneity. *Academic Radiology*, *24*(3), 328–336.

[27] Shen, W., Zhou, M., Yang, F., Yu, D., Dong, D., Yang, C., et al. (2017). Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*, *61*, 663–673. https://doi.org/10.1016/j.patcog.2016.05.029.

[28] Shin, H., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298. https://doi.org/10.1109/TMI.2016.2528162.

[29] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*, 484–489.

[30] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. CoRR arXiv:1409.1556.

[31] Song, Q., Zhao, L., Luo, X., & Dou, X. (2017). Using deep learning for classification of lung nodules on computed tomography images. *Journal of Healthcare Engineering*, *2017*, 7. https://doi.org/10.1155/2017/8314740.

[32] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

[33] Szegedy, C., Ioffe, S., & Vanhoucke, V. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. CoRR. arXiv:1602.07261.

[34] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015b). Rethinking the inception architecture for computer vision. CoRR arXiv:1512.00567.

[35] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., et al. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on MedicalImaging*, *35*(5), 1299–1312.

[36] The National Lung Screening Trial Research Team. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, *365*(5), 395–409.

[37] Van Beek, E. J., Mirsadraee, S., & Murchison, J. T. (2015). Lung cancer screening: Computed tomography or chest radiographs? *World Journal of Radiology*, *7*(8), 189–193. https://doi.org/10.4329/wjr.v7.i8.189.

[38] van Ginneken, B., Setio, A. A. A., Jacobs, C., & Ciompi, F. (2015). Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: *2015 IEEE 12th international symposium on biomedical imaging (ISBI)* (pp. 286–289).

[39] Zhang, F., Song, Y., Cai, W., Lee, M. Z., Zhou, Y., Huang, H., et al. (2014). Lung nodule classification with multilevel patch-based context analysis. *IEEE*

*Transactions on Biomedical Engineering*, *61*(4), 1155– 1166.

[40] Zhao, X., Liu, L., Qi, S., Teng, Y., Li, J., & Qian, W. (2018). Agile convolutional neural network for pulmonary nodule classification using CT images. *International Journal of Computer Assisted Radiology and Surgery*, *13*(4), 585–595. https://doi.org/10.1007/s11548-017-1696-0.

[41] Zhu, H., Cheng, H., & Fan, Y. (2015). Random local binary pattern based label learning for multi-atlas segmentation. *ProcSPIE*, *9413*, 8.