Scientific
Research
Publishing

Attention Based Multi-Patched 3D-CNNs with Hybrid Fusion Architecture for Reducing False Positives during Lung Nodule Detection

Vamsi Krishna Vipparla[1], Premith Kumar Chilukuri[2], Giri Babu Kande[3]

[1]Emerging Technologies Department, Mahindra & Mahindra IT, Mumbai, Maharashtra, India
[2]Computer Vision Division, Supervue AI, Visakhapatnam, Andhra Pradesh, India
[3]Department of Electronics & Communication Engineering, Vasireddy Venkatadri Institute of Technology, Nambur, Andhra Pradesh, India
Email: vipparla.vamsikrishna@gmail.com, premithchilukuri@gmail.com, giribabukande@gmail.com

Abstract

In lung nodules there is a huge variation in structural properties like Shape, Surface Texture. Even the spatial properties vary, where they can be found attached to lung walls, blood vessels in complex non-homogenous lung structures. Moreover, the nodules are of small size at their early stage of development. This poses a serious challenge to develop a Computer aided diagnosis (CAD) system with better false positive reduction. Hence, to reduce the false positives per scan and to deal with the challenges mentioned, this paper proposes a set of three diverse 3D Attention based CNN architectures (3D ACNN) whose predictions on given low dose Volumetric Computed Tomography (CT) scans are fused to achieve more effective and reliable results. Attention mechanism is employed to selectively concentrate/weigh more on nodule specific features and less weight age over other irrelevant features. By using this attention based mechanism in CNN unlike traditional methods there was a significant gain in the classification performance. Contextual dependencies are also taken into account by giving three patches of different sizes surrounding the nodule as input to the ACNN architectures. The system is trained and validated using a publicly available LUNA16 dataset in a 10 fold cross validation approach where a competition performance metric (CPM) score of 0.931 is achieved. The experimental results demonstrate that either a single patch or a single architecture in a one-to-one fashion that is adopted in earlier methods cannot achieve a better performance and signifies the necessity of fusing different multi patched architectures. Though the proposed system is mainly designed for pulmonary nodule detection it can be easily extended to classification tasks of any other 3D medical

diagnostic computed tomography images where there is a huge variation and uncertainty in classification.

1. Introduction

Lung Nodule is also known as lung tumor which is characterized by uncontrolled cell growth in the lung tissues. According to American Cancer Society, Lung cancer occurred in 222,500 people and resulted in 155,870 deaths worldwide in 2017 [1], which makes it the most common cause of cancer related death in men and second most common in women after breast cancer [2]. The 5-year survival rate is only 17% for lung cancer [3], but if detected early on, survival increases to 54% [1]. As it is very difficult to diagnose cancer at a very early stage, a timely detection is necessary to monitor the growth pattern in the affected region. According to National Lung Screening Trail (NLST) [4] apart from traditional two dimensional X-Ray scans, low dose three dimensional CT scans have proved to be most useful for diagnosis and reduced the mortality rate up-to 20%. However, due to their huge volumetric data it becomes increasingly difficult for a CAD system to analyze these scans and is susceptible to more number of false positives per scan resulting in repercussions and financial problems. Hence, the goal of any CAD system for assisting doctors is to reduce the false positives as much as possible. The true positives are difficult to be detected and to get them accurately differentiated from True negatives in first place due to the following reasons.

- Huge variations in the contextual information surrounding the nodules as well as the structural intensity of nodules that vary like solid, part solid and non-solid/Ground Glass opacity and can be visualized in slices of **Figure 1**. Which are extracted from the center of a nodule with patch size of $64 \times 64 \times 64$.
- Small size of nodules ranging from 3 mm and going up to 30 mm and above.
- Various types of nodules with different spatial location and shapes namely Isolated Nodule, Juxta-Pleural, Pleural tail, Cavitary and Calcific nodule that can be visualized in **Figure 2**.
- Some false positives/True Negatives carry similar shapes and structures to true positives which can be coined by the term hard mimics and can be visualized in **Figure 3**.

It is due to these challenges radiologists find it difficult to exactly localize the lung nodules and it is also even more challenging for making a CAD system which can tackle the aforementioned challenges. Hence, a novel approach is required to deal with these significant challenges.
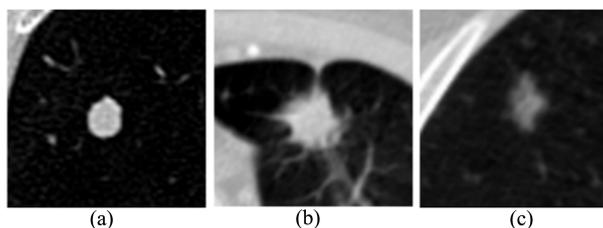
Figure 1. Different types of nodules in terms of variation in their intensity. (a) Solid; (b) Part Solid; (c) Non-Solid.
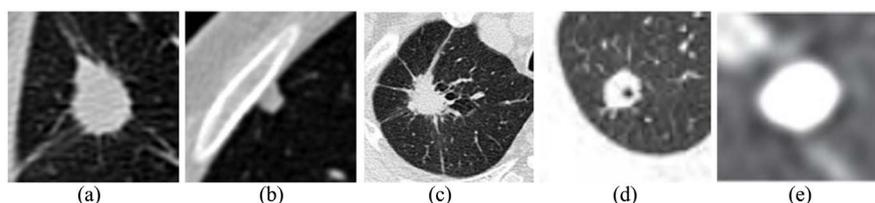
Figure 2. Different types of nodules in terms of their external attachment and shapes (a) Common Isolated Nodule; (b) Juxta-Pleural Nodule; (c) Pleural Tail Nodule; (d) Cavitary Nodule; (e) Calcific Nodule.
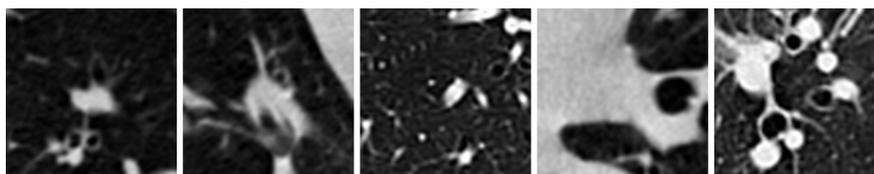
Figure 3. False positives which are similar in appearance to the true positives that make the task even more challenging.

Current CAD approaches can be divided into two categories: classification approaches based on hand-crafted features [5] [6] and deep learning approaches with automated feature extraction [7]-[12]. In the first category, approaches usually measure radiological characteristics like texture, shape, nodule size, location and employ a classifier to find the malignancy status. These methods lead to measurement errors as the collection and selection of suitable set of features for lung nodule diagnosis is trivial.

With the active involvement of research in medical imaging by deep learning community, methods using 3D and 2D CNN's that fall into second category are proposed for lung nodule detection with better false positive reduction which outperformed the methods based on hand crafted features. Though 2D CNN's achieved a greater performance than the methods based on hand crafted features they cannot completely utilize the 3D information of lung CT scans. In order to mimic the 3D information several approaches were proposed which were mostly based on adjusting the 2D CNN's to capture the information from various cross-sectional slices in different orientations or even varying the convolutional filter size in a multi-path fashion to concatenate the feature maps for final result [7] [8]. Similar 2D variants were also employed for the segmentation of lung nodules [9] [10] where [9] utilized scans from different orientations like axial,

sagittal and coronal for voxel level classification and in [10] multi-path fashion is followed with both 3D and 2D slices from center of the nodule as input for each path. But with the recent advancement in the computational power of the systems, methods using 3D CNN were also implemented whose performance surpassed the 2D Variants where 3D patches of different sizes are fed as input to the model architectures in one to one or many to one fashion for averaging each individual result or concatenating all the feature maps before final classification respectively [11] [12]. [13] used different patch sizes as input to the different CNN architectures in a one to one fashion for ensembling all their predictions for a more reliable result where a sensitivity of 94% is achieved at an average of 5 false positives per scan. The methods discussed above suffered from huge imbalance in classification between the positive and negative classes though some methods have performed augmentation. Moreover, the range of variations in position, size, intensity and surrounding contextual information of nodules in Lung CT scans cannot be learned by a single 3D CNN architecture alone which draws a requirement for establishing consensus from diverse multi patch 3D CNN architectures.

Also, attention mechanism has been widely in various tasks recently wherever selective features of most importance at the task are to be weighed more suppressing other complex irrelevant features. For example in [14] attention mechanism was used in U-Net++ for automatic segmentation of liver by merging only task oriented features at different levels in the encoder-decoder architecture. This works due to the ability of attention mechanism in increasing the weight of the focus regions while suppressing the regions in background that are unrelated to the segmentation task at hand.

In this paper, to address the aforementioned problems we propose a unique CAD system for false positive reduction based on diverse multi-patch 3D ACNN architectures. The architectures comprise of a newly developed 3D ACNN architecture, two others inspired from dense-net [15] and res-net [16] namely MP-ACNN1, MP-ACNN2 and MP-ACNN3 respectively. Attention mechanism was incorporated to minimize the effect of background irrelevant non-nodule features on model performances. To deal with the large variations in the shape, size and surroundings of lung nodules and to classify them accurately from their very similar false positives multiple patches are considered for concatenating all their feature maps before the final classification. To further boost the sensitivity at even very low false positive rates, an ensembling technique on predictions of diverse 3D ACNN architectures is followed. Also to overcome the imbalance between true positives and false positives a new iterative training approach is employed where equal number of positive and negative samples is considered in each iteration. We validated our proposed system on LUNA16 [17] dataset and have achieved a state of the art performance which outperformed several methods in the challenge as well as some of the approaches stated above.

The main contributions of our proposed system can be summarized as follows:

- 3D attention based CNN architectures named MP-ACNN1, MP-ACNN2 and MP-ACNN3 was developed which works well with the 3D volumetric data or sequence of 2D frames like CT scans that have more 2D spatial representations along with temporal connections in z-axis similar to video frames that are stacked from a video for action-recognition.
- A new technique is followed to deal with the huge variations in different characteristics of nodules like size, shape and its surroundings where in different sized patches which can encapsulate diverse set of features are used as input to the proposed model architectures in a multi path fashion for final classification.
- Inspired from the way ground truths were formulated in LIDC/IDRI [18] dataset where a consensus is established among at least 3 out of 4 radiologists, reliable result is obtained by fusing the results of diverse model architectures namely MP-ACNN1, MP-ACNN2 and MP-ACNN3.
- To deal with the huge imbalance between positive and negative samples a new iterative training approach is followed in which equal number of positive and negative samples are taken.

2. Materials and Methods

Figure 4 below gives an overview of our methodology. First a Lung CT scan is given as input to a pre-processing module which outputs only the region of interest from the entire scan, retaining all the information of nodules. In the next step the processed scan goes as input to a training samples generator which outputs positive and negative patches of different sizes using the ground truth containing the locations of all nodules. Augmentation is performed on positive samples for up-sampling. The final step is to feed all the samples generated as input to the proposed diverse multi-patch architectures followed by a fusion of their predictions for final classification.
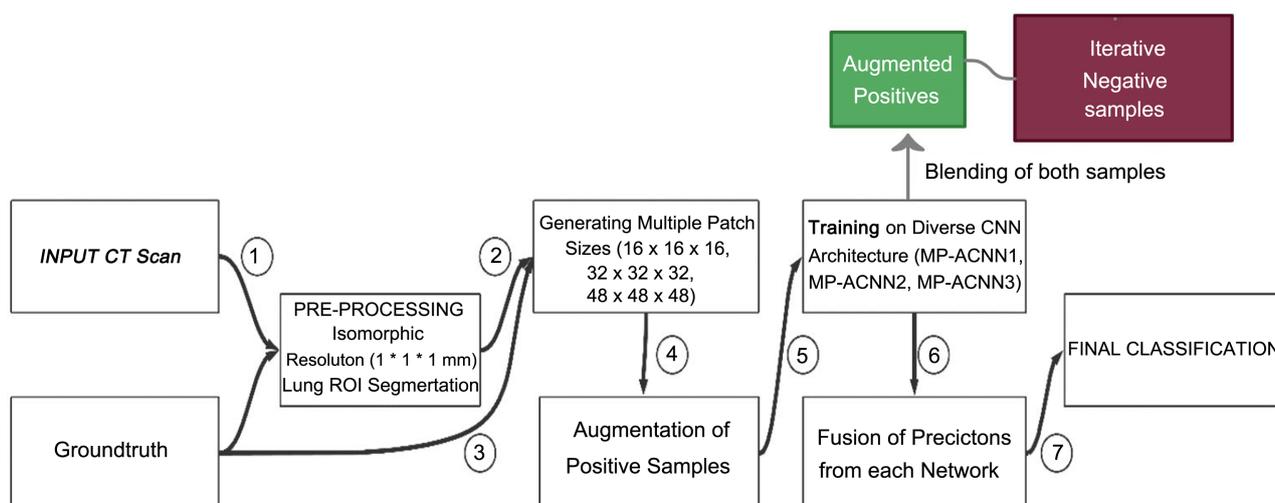
Figure 4. Overview of the detailed flow of the proposed system starting from the input of raw CT scan to that of the final prediction.
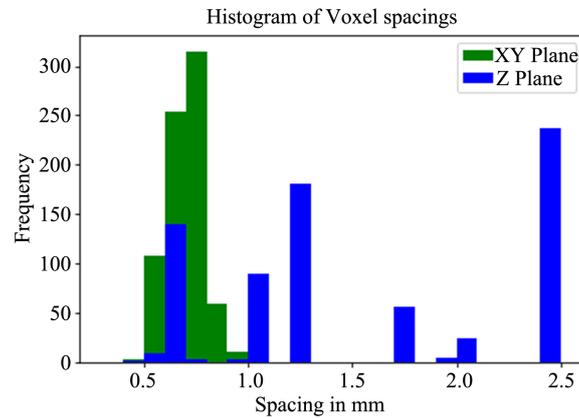
2.1. Dataset

For training our model, we have used LUNA16 challenge dataset [17] that is extracted from LIDC/IDRI dataset [18] which consists of 1018 patient scans. All scans that are having a slice thickness of greater than 2.5 mm are excluded resulting in a total of 888 volumetric thoracic CT scans in LUNA16 dataset [17]. A two phase annotation procedure is followed to collect the ground truths in this dataset [17] by four experienced radiologists. After each radiologist annotated all the candidate nodules of the CT scans, each candidate nodule with an agreement of at least three radiologists was approved as ground truth. All the 888 scans are divided into 10 different subsets for the use of 10 fold cross validation. The CT scans provided are in Meta Image file format (MHD/raw). Each CT scan has around 100 - 400 slices in the z-axis depending upon different patients and each two dimensional slice is of 512×512 pixels. In each CT scan voxel dimensions can vary in all the x, y, and z directions according to the configurations of different CT machines. The nodule sizes in this dataset vary in the range of 3 - 33 mm. The dataset provides respective patient's ID and x, y, z coordinates of the nodule centers along with diameter for positive annotations as given in Table 1.

2.2. Pre-Processing

In LUNA16 dataset [17] depending upon the configuration of CT machines used, different scans have varying voxel spacing along x, y, z directions (Figure 5). This condition affects performance of the ACNN model because of the non-homogeneity in the resolution of input scan. Therefore an isomorphic resolution is necessary so that the ACNN models can generalize equally for all the

Table 1. Sample data of positive annotations ground truths in LUNA16 dataset.

| Seriesuid	coord X	coord Y	coord Z	diameter mm
1.3.6.1.4.1.14519.5.2.1. 6279.6001.1002252872 22365663678666836860	−128.69	−175.31	−298.38	5.65
1.3.6.1.4.1.14519.5.2.1. 6279.6001.1002252872 22365663678666836860	103.78	−211.92	−227.12	4.22
1.3.6.1.4.1.14519.5.2.1. 6279.6001.1003981387 93540579077826395208	69.63	−140.94	876.37	5.78
1.3.6.1.4.1.14519.5.2.1. 6279.6001.1006213830 16233746780170740405	−24.01	192.10	−391.08	8.14
1.3.6.1.4.1.14519.5.2.1. 6279.6001.1006213830 16233746780170740405	2.44	172.46	−405.49	18.54
1.3.6.1.4.1.14519.5.2.1. 6279.6001.1006213830 16233746780170740405	90.93	149.02	−426.54	18.20

Figure 5. Histogram of voxel spacing's of all the scans present in the LUNA16 dataset.

scans. In order to arrive at a best voxel spacing, all the scans present in the dataset are analyzed (**Figure 5**). A mean of 0.68 mm in x-y axis and 1.56 mm in z-axis is achieved upon which on experimenting $1 \times 1 \times 1$ mm is the best spacing which resulted in better accuracy along with less computational load. Nearest interpolation is performed on all the scans for bringing all the voxels to same resolution by using a resize factor which is computed using original spacing and new spacing.

A CT scan of a lung comprises of several irrelevant information like bones, tissues, blood, water, air apart from lungs and have to be removed for the better detection of nodules by the model. Firstly, to eliminate air noise the CT scans are convolved with gaussian kernel that has a standard deviation of 1.0 mm. In CT images, the intensities of ribs, examination bed and fat is generally above −100 HU and the tissue of the lung is in the range of −400 to −600 HU. To segment the lung region iterative thresholding [19] is used. The trachea and bronchi will be preserved even after the coarse segmentation. We have used 3D connected component labeling and 3D region-growing to eliminate the trachea and bronchi. Erosion is performed to separate the nodules attached to the blood vessels. **Figure 6** illustrates the results of preprocessing.

2.3. Training

Different sized patches are used in the training and the motive behind this was the variation in the size of nodules which generally ranges from 3 mm to 33 mm in this dataset. Different patch sizes reflect on different contextual information and hence can generally consider various distance dependencies which results in a more reliable and accurate result. If the patch dimension is small, smaller lung nodules are detected better but larger nodules are disregarded. If the patch dimension is too big, irrelevant structures surrounding the nodule are also considered which hinders the performance. Therefore a proper analysis is required before considering patch sizes for which we have examined the histogram of the nodule sizes present in the dataset as given in **Figure 7** and came to an optimum set of patch sizes which are $16 \times 16 \times 16$, $32 \times 32 \times 32$ and $48 \times 48 \times 48$.
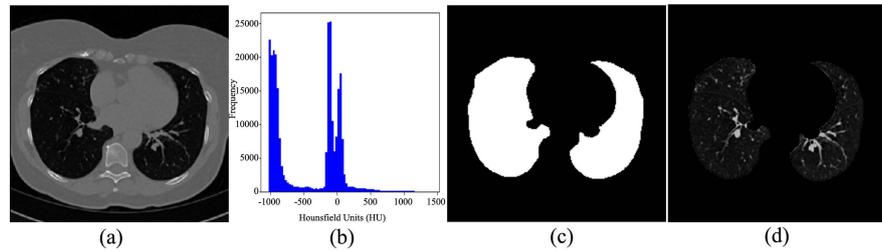
**Figure 6.** (a) Original Slice (b) Histogram of its HU (c) Generated Mask (d) Final segmented slice.
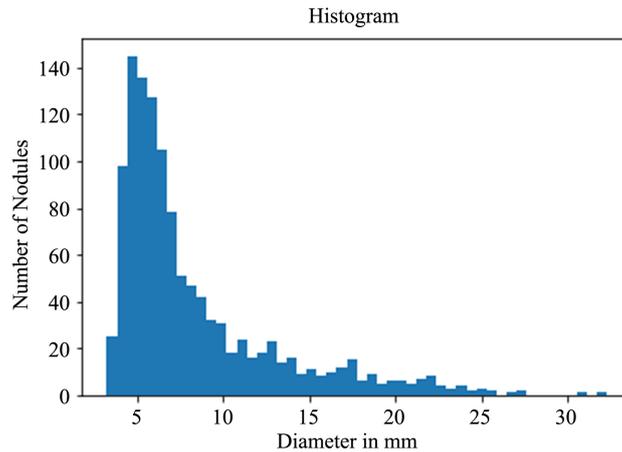
Figure 7. Histogram of nodule sizes present in LUNA16 dataset for the analysis in considering different patch sizes.

Majority of the small sized nodules are in the range of 3 - 10 mm. Hence the first patch considered is of size $16 \times 16 \times 16$ which can enclose the entire small nodule as well as consider certain surrounding contextual information. Medium sized nodules are in the range of 10 - 25 mm and hence the second patch considered is of size $32 \times 32 \times 32$ which can cover the majority of nodules with rich contextual information for smaller nodules along with certain range of contextual information for medium sized nodules. Larger nodules are in the range of 25 - 35 mm and hence the third patch is of size $48 \times 48 \times 48$ for these extreme cases. In **Figure 8** multiple patch sizes that are considered can be visualized on how wide range of contextual information is captured for nodules ranging from smaller size to that of the larger size. After the generation of different patches a two staged approach is followed to tackle the class imbalance problem between positive and negative samples which is generally a common type of problem when dealing with majority of medical images.

First stage:

Data augmentation is performed to up sample the positives from a ratio of 1:483 to 1:12 by rotating them at various angles like 90, 180, 270 and flipping them in all the three directions (x, y and z). In this way the model also becomes invariant to variations in orientation of lung nodules.

Second stage:
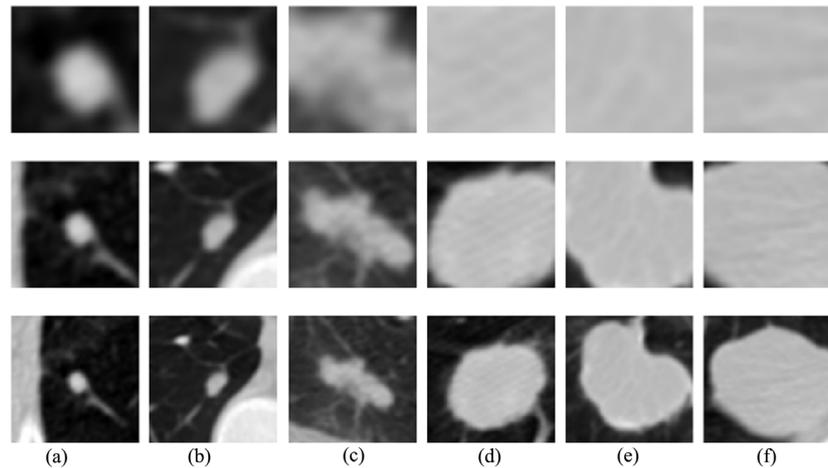
Figure 8. Illustration of the different patch sizes exploiting the different contextual information surrounding the nodule. (a) and (b) are the small sized nodules having diameter less than 10 mm. (c) and (d) are the medium sized nodules having diameter in between 10 - 25 mm. (e) and (f) are the large sized nodules having diameter greater than 25 mm.

Novel iterative training approach is followed which contains the following steps:

- Let's suppose we have N number of positive samples.
- Our model first gets trained on these N positive samples and the first N negative samples as input.
- Similarly in the corresponding iterations the same N positive samples and the next N number of negative samples are used as input for the training of our model.

Using this unique method we got best results as equal emphasis is even given to the positive samples with respect to negative samples while training despite of their huge imbalance. In this way we can use the standard loss functions like cross-entropy without any bias towards negative samples.

Along with the consideration of various patch sizes, to increase the performance we also propose a fusion model of different ACNN architectures to reduce the false positives. A single ACNN architecture has limited learning capability and may not learn all the significant features to differentiate between lung nodules and their very similar false positive/non-nodule structures. This problem draws a requirement for establishing a consensus among diverse Multi Patched Attention based CNN architectures by using a fusion technique.

A total of three architectures are proposed namely Multi-patch Attention based CNN1 (MP-ACNN1), Multi-patch Attention based CNN2 (MP-ACNN2) and Multi-patch Attention based CNN3 (MP-ACNN3) whose architectures can be visualized in a, b and c of Figure 9 respectively. All the three architectures are constructed in such a way that each one is very diverse from the other two whether it is in the number of convolutional layers, number of filters or in the connections made in between the layers thereby creating different hierarchal features. Different patch sizes of $16 \times 16 \times 16$, $32 \times 32 \times 32$ and $48 \times 48 \times 48$ are
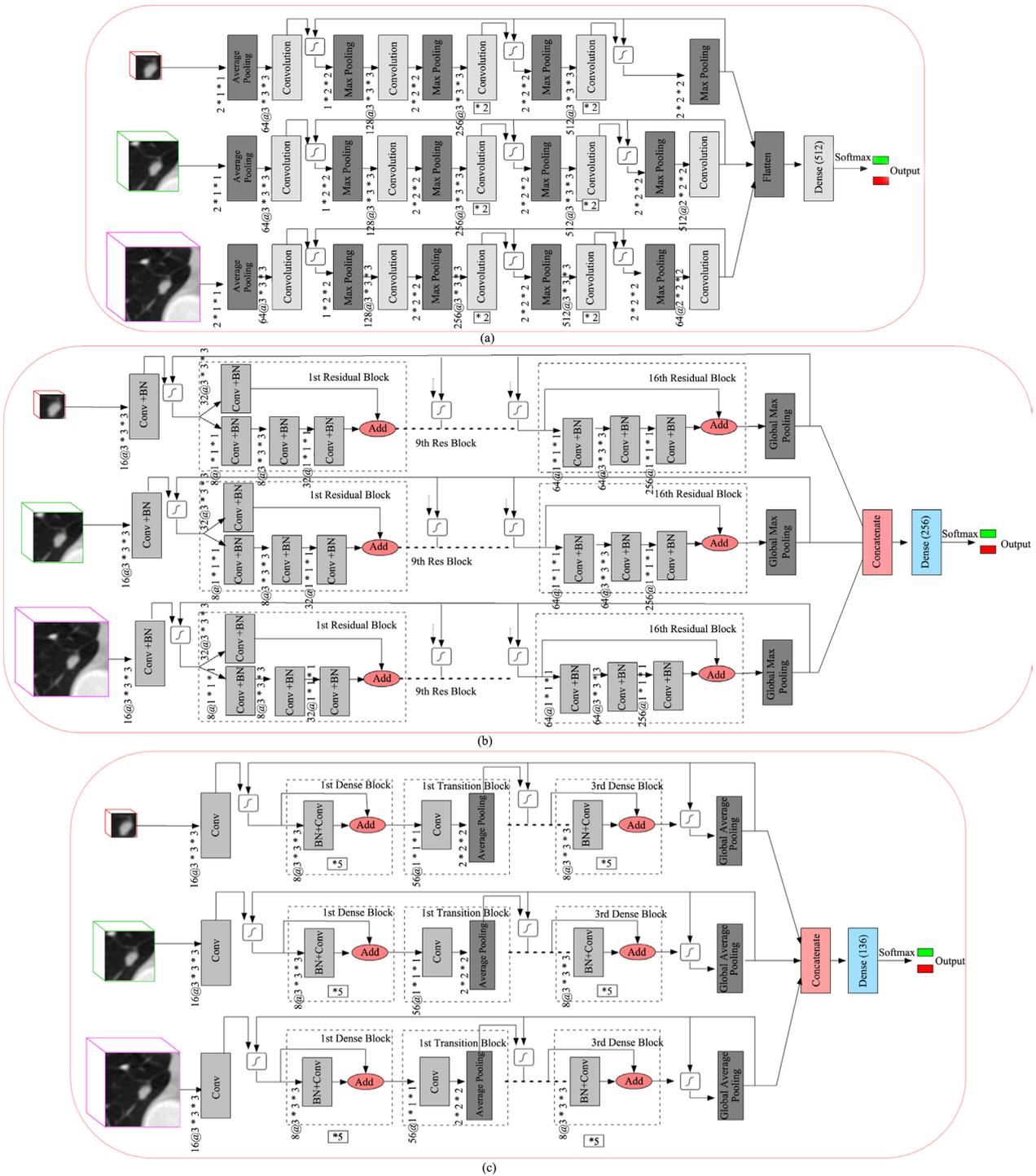
Figure 9. Visualization of the proposed diverse multi-patch architectures. The convolution kernel size is represented as number of filters@ filter dimensions (*i.e.*, 8@3 × 3 × 3 represents 8 filters of kernel size 3 × 3 × 3), (a) MP-ACNN1; (b) MP-ACNN2; (c) MP-ACNN3.

used as input for the proposed architectures in a multi path fashion for the three paths namely P1, P2 and P3 for concatenating their feature maps at the ending layers before final classification. The important units in the construction of proposed 3D ACNN's are given below.

2.3.1. 3D Convolution

In 3D convolution layer, a group of 3D kernels convolve with the output of preceding layer to extract some high level feature maps. Unlike a 2D convolution where a single patch is given as input, 3D convolution takes an input of stacked patches and outputs a stack of feature maps based on the number of kernels used. Kernel size also effects the variation in feature volumes. After the convolution is done a bias value is added to it and an activation function like ReLU, Softmax is applied on the whole resulting value. This process can be formulated as the following equation

$$f_i^l(x, y, z) = \sigma\left(b_i^l + \sum_k \sum_{u,v,w} f_k^{l-1}(x-u, y-v, z-w) W_{ki}^l(u, v, w)\right) \tag{1}$$

where f_i^l and f_k^{l-1} represent the ith and k^{th} 3D feature volume in l^{th} and the previous $(l-1)^{th}$ layer respectively. W_{ki}^l is the 3D convolutional weight kernel connecting f_i^l, f_k^{l-1} and $f_i^l(x, y, z)$, $f_k^{l-1}(x-u, y-v, z-w)$ and $W_{ki}^l(u, v, w)$ are their corresponding element wise values where x, y, z and u, v, w are the co-ordinates of f_i^l, W_{ki}^l respectively. b_i^l is the bias term and σ is the activation function which in our case is ReLU.

2.3.2. Pooling Layer

3D max-pooling layers are used in between 3D convolutional layers to down sample the dimensions of input 3D feature volumes in all the three directions to gain invariance to translations in local 3D space. If l^{th} layer is a convolutional layer and $(l+1)^{th}$ layer is a 3D max-pooling or 3D average pooling layer, then the pooling layer will receive a four dimensional tensor $T = \left[f_1^l, f_2^l, \cdots, f_N^l\right] \in \mathbb{R}^{X \times Y \times Z \times N}$ as input. Max-pooling operation selects the maximum activation and average pooling selects the average of all the activations within a neighborhood and gives an abstracted output $T' \in \mathbb{R}^{X' \times Y' \times Z' \times N}$, where $(X; Y; Z)$ and $(X'; Y'; Z')$ are the 3D feature volumes sizes before and after the corresponding pooling operation, N represents the total number of 3D feature volumes and it is same throughout the pooling operation. If the pooling window size is W and the stride followed is S, the reduced feature volume size along X-axis can be computed using Equation (2) which remains the same for Y' and Z'.

$$X' = \frac{X - W}{S} + 1, \tag{2}$$

2.3.3. Fully Connected Layer

The nodes have more dense connections in fully connected layers compared to convolutional layers (*i.e.* each node in dense layer is connected to all the nodes of adjacent layers unlike in convolutional layers where local connections are made). The dense layers are very useful in the better representation of the extracted features. Fully connected layers are implemented by flattening the volumetric features into a vector for matrix multiplication of weights, then adding a bias value to it followed by a non-linear activation function. This process can be

formulated as the following equation:

$$f^l = \left(b^l + W^l f^{l-1} \right) \tag{3}$$

where f^{l-1} is the input feature vector obtained after flattening the volumetric features in $(l-1)^{th}$ layer, f^l is the output feature vector of the lth layer which is fully connected, W^l is the weight matrix and b^l is the bias.

2.3.4. Soft-Max Layer

The ending layer of the 3D ACNN architecture before output is soft-max. If the nodes at the ending layer are denoted by a vector f^L with C number of output classes, final prediction probability for each class can be determined using

$$p_c \left(f^L \right) = \frac{\exp \left(f_c^L \right)}{\sum_{c=0}^{C-1} \exp \left(f_c^L \right)}, \tag{4}$$

where f_c^L is the cth node in feature vector of last layer. All the resulting activations from the soft-max layer are positive and lie in between 0 to 1 with their summations resulting to 1. As a result, they can be interpreted as the estimated probability distribution predicted by the network.

2.3.5. Cost Function/Loss Function

Binary cross entropy loss/log loss function is used to optimize the parameters of 3D ACNN by minimizing the loss function $H_p(q)$ or until the point of convergence as follows:

$$H_p(q) = \frac{-1}{N} \sum_{i=1}^{N} \left[l_i \cdot \log \left(p(l_i) \right) + \left(1 - l_i \right) \cdot \log \left(p \left(1 - p(l_i) \right) \right) \right] \tag{5}$$

where l_i is the label and $p(l_i)$ is the corresponding probability. Similarly $(1-l_i)$ and $p(1-l_i)$ are for other class and N represent the total number of samples.

2.3.6. Batch Normalization

Batch normalization decreases the internal co-variance shift in the values of hidden layers due to the continuous update of weights during backpropagation. Batch normalization can be expressed as,

$$b^* = \left(b - M(b) \right) / std(b), \tag{6}$$

where b^* the new value for a single element within a batch is, $M(b)$ is the mean for a batch and $std(b)$ is standard deviation in a batch.

The Equation (6) is further extended to identity function $b^{**} = \gamma \times b^* + \beta$, where b^{**} is the final value after normalization. γ and β are the values learned for each layer.

2.3.7. 3D Attention Gate

Human vision system is the basis in the design of Attention gate which gives more importance to the object features in context and less weight age to other irrelevant features. The Attention gate usage can be expressed using below equa-

tions,

$$x_{out} = x_l \cdot \alpha_i, \tag{7}$$

where x_{out} is the element by element wise multiplication between the input features map x_l and the attention coefficient α. The attention coefficient belongs to a set of [0 - 1] and prunes the networks irrelevant features in classification. We have used additive attention instead of multiplicative attention due to its accuracy despite having a tradeoff in computational complexity [20]. The 3D multi-dimensional multiplicative attention coefficient can be computed as:

$$\alpha_i = \sigma_2 \left(\Psi^{\mathrm{T}} \left(\sigma_1 \left(W_x^{\mathrm{T}} x_l + W_g^{\mathrm{T}} g_i + b_g \right) \right) + b_\Psi \right), \tag{8}$$

where σ_1 and σ_2 are the ReLU and Sigmoid functions respectively. W_x, W_g and Ψ are the linear transformations. For linear transformations on the input feature vector X_l and gating feature vector g_i $1 \times 1 \times 1$ kernels are used. b_Ψ and b_g are the bias terms. The architecture of our 3D attention mechanism can be visualized in (Figure 10).

2.3.8. MP-ACNN1

The structure of proposed MP-ACNN1 can be visualized in **Figure 9(a)** MP-ACNN1 has less number of convolutional layers compared to other two proposed architectures MP-ACNN2 and MP-ACNN3. But the number of trainable parameters are high due to the number of filters which can learn more primitive features at starting layers. In this architecture $2 \times 1 \times 1$ average pooling is applied at the starting of each path followed by four convolutions each with a kernel size of $3 \times 3 \times 3$ and a total of 64, 128, 256 and 512 kernels respectively followed by a max-pooling. For the other two paths P2 and P3 an extra convolution is applied of kernel size $2 \times 2 \times 2$ with total kernels of 512 and 64 respectively. The ending layers of all the three paths are flattened for concatenation before passing them to dense layer having 512 nodes with a dropout of 50% for final classification using soft-max.

2.3.9. MP-ACNN2

The proposed MP-ACNN2 architecture is given in **Figure 9(b)**. This architecture is inspired from res-net [16] and has a total of 16 residual blocks for all the
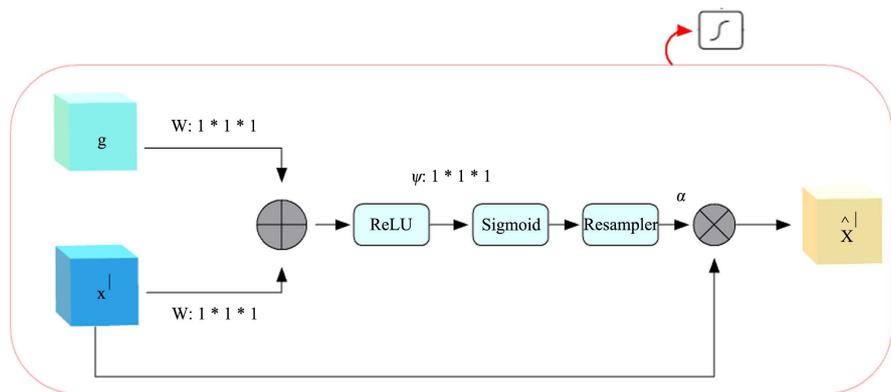

Figure 10. 3D Attention gate architecture.

three paths namely P1, P2 and P3. Before the start of residual block, a convolution of 16 kernels each having a size $3 \times 3 \times 3$ and batch normalization is applied at the input. Each residual block consists of 3 convolutions with kernel sizes of $1 \times 1 \times 1$, $3 \times 3 \times 3$ and $1 \times 1 \times 1$ respectively and varying number of kernels according to corresponding residual block. Each convolution in the residual block is followed by a batch normalization. A residual connection of output from previous block is also included in each residual block with an exception in the first residual block where an output from a parallel convolution of 32 kernels each having a size $3 \times 3 \times 3$ is taken as residual. After 16 residual blocks global max pooling is applied and the resulting nodes in each path are concatenated before passing them to dense layer having 256 nodes with a dropout of 50% for final classification using soft-max.

2.3.10. MP-ACNN3

The proposed MP-ACNN3 architecture is as shown in **Figure 9(c)**. This architecture is inspired from dense-net [15] and has a total of three dense blocks with transition blocks in between them for all the three paths namely P1, P2 and P3. Before the start of dense block a convolution of 16 kernels each having a size $3 \times 3 \times 3$ and batch normalization is applied at the input. Dense block consists of 5 convolutions and each having 8 kernels of size $3 \times 3 \times 3$. All the convolutions in dense block are followed by a batch normalization. Dense connections are also included in the dense block with the output from preceding layer concatenated to the outputs of current layer.

In the transition block a convolution of 56 kernels each having a size of $1 \times 1 \times 1$ followed by an average pooling of window size $2 \times 2 \times 2$ are applied. After the end of three dense blocks global average pooling is applied and the resulting nodes in each path are concatenated before passing them to dense layer having 136 nodes with a dropout of 50% for final classification using soft-max.

2.3.11. MP-AFNet

As visualized in **Figure 11**, in proposed MP-AFNet predictions from all the
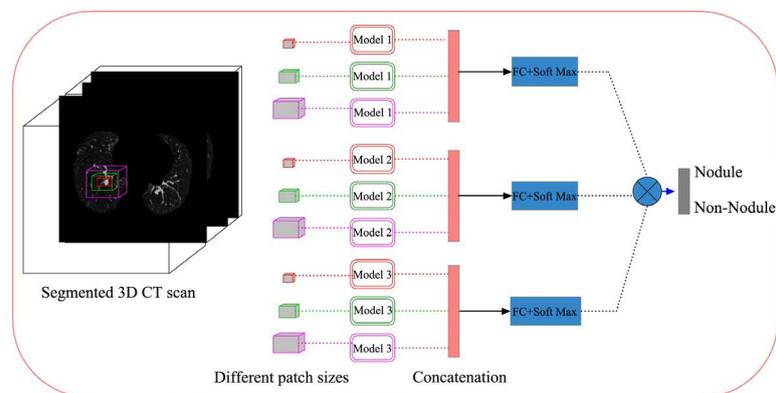
Figure 11. Visualization of the proposed MP-AFNet that fuse the results from diverse Multi patched ACNN's proposed MP-ACNN1, MP-ACNN2 and MP-ACNN3 for final classification of either nodule or non-nodule.

architectures MP-ACNN1, MP-ACNN2 and MP-ACNN3 after training are fused in a hybrid approach. First majority voting is conducted to get the initial decision of either nodule or non-nodule. Second the prediction probabilities of majority decision are averaged to get the final probability of being a nodule. (*i.e.* if the predicted probabilities are 0.482, 0.997, 0.993, from initial majority voting it is evident that it is a nodule with 2 votes and the values of these votes are averaged to get a final probability of 0.995). In this way if a single model fails at establishing correct result other two models can curb that mistake leading to a more reliable result with less number of false positives.

3. Results and Discussion

3.1. Experimental Setup

The proposed system is trained on GTX1080 Ti GPU and Keras [21] is used as deep learning framework which is built on Tensor flow [22] as backend. To deal with different medical image formats, SimpleITK library is used. A set of 3D patches at resolutions of $48 \times 48 \times 48$, $32 \times 32 \times 32$, and $16 \times 16 \times 16$ are extracted from CT scans by using the x, y and z center coordinates of the nodule candidates present in the dataset. The patch sizes in [11] are taken by analyzing the distribution of voxels covered by the nodules separately in both x-y plane and z-plane. But lung nodules are known for their huge variations in the growth patterns. Hence the patch sizes cannot be generalized by experimenting only on this particular dataset [18] and is the reason for taking uniform voxels along x, y, z directions for all the patch sizes considered. All these patches of size $48 \times 48 \times 48$, $32 \times 32 \times 32$, and $16 \times 16 \times 16$ are considered after a proper analysis which can cover all the nodules of varying sizes along with different views of contextual information. They covered 100%, 99% and 90% of the nodules in the positive annotations of the dataset.

For faster convergence, we applied a min-max normalization to patches in the range of [−1000, 400] Hounsfield units (HU). For nonlinear transformation in convolution and fully-connected layers, we used a ReLU function. To make our network robust, we also applied a dropout technique to fully connected layers with a rate of 0.5. During the training phase, we set the learning rate to 0.001, the momentum to 0.9, the batch size to 64 and completed the training through an iterative approach of 12 iterations. For each iteration three epochs are taken so that the model doesn't get over-fitted over the completion of all the iterations, yet can well generalize between the positive and negative samples.

To evaluate the performance of the proposed system, free receiver operation characteristics (FROC) analysis [23] is employed. In the FROC curve, sensitivity is plotted as a function of the average number of false positives per scan (FPs/scan). Competitive Performance Metric (CPM) [24] score is obtained by calculating average sensitivity at seven predefined false positive rates: 1/8, 1/4, 1/2, 1, 2, 4 and 8 FPs per scan.

The proposed system is evaluated with 10-fold cross-validation. That is, after

dividing all data into 10 disjoint subsets, 9 subsets are added to the training set, and the remaining is used for testing.

The dataset comprises of 754,975 candidates which were detected using five different lung nodule detection CAD systems [25] [26] [27] [28] [29] of which only 1557 are positives and the remaining are negatives that indicates a serious imbalance between positives and negatives (1:483). To circumvent this potential bias problem, nodule samples are augmented by 90˚, 180˚, and 270˚ rotation on a transverse plane and 1-pixel shifting along the x, y, and z axes. Thus, the proportion between the numbers of nodules to non-nodules is approximately 1:12. Table 2 presents the details about the number training, validation, and test samples.

3.2. Performance Comparison

The performance of the proposed system is evaluated by comparing CPM score of our system with CPM scores of state-of-the-art methods on LUNA.

16 challenge dataset [7] [11] [30] [31] [32] [33]. Precisely, Setio *et al.*'s method [7] employs multi view (9) 2D patches, Xie *et al.*'s method [32] used a boosting architecture with three 2D slices. Zou *et al.* method [31] used multi resolution 2D patches. Though, the task is of three-dimensional nature, these methods used variants of 2D CNNs. Ding *et al.*'s method [33] takes 3D patches as input, Dou *et al.*'s method and Gorkam Polat *et al.* method [11] [30] used multi-level 3D patches. The CPM scores at seven distinct FPs per scan are summarized in Table 3.

The proposed MP-AFNet surpassed the average CPM scores of state of the art systems stated in Table 3. Particularly when comparing with Ding et.al method, which also uses 3D CNN, our method has an increased average CPM score of

Table 2. The data of the number of training, validation and test samples used for each fold. The values in the parenthesis indicate the number of samples retained for validation and test (#validation samples/#test samples). The value outside the parenthesis is the number of training samples which also include 20% validation samples. # = After-Augmentation.

Dataset	Scans	Nodules	Nodules (#)	Non-Nodules
Fold 1	799 (159/89)	1418 (283/139)	56,720	681,474 (136,294/71,944)
Fold 2	799 (159/89)	1435 (287/122)	57,400	679,607 (135,921/73,811)
Fold 3	799 (159/89)	1380 (276/177)	55,200	679,559 (135,911/73,859)
Fold 4	799 (159/89)	1368 (273/189)	54,720	673,878 (134,775/79,540)
Fold 5	799 (159/89)	1371 (274/186)	54,840	675,342 (135,068/78,076)
Fold 6	799 (159/89)	1455 (291/102)	58,200	679,828 (135,965/73,590)
Fold 7	799 (159/89)	1398 (279/159)	55,920	675,201 (135,040/78,217)
Fold 8	799 (159/89)	1414 (282/143)	56,560	678,505 (135,701/74,913)
Fold 9	800 (160/88)	1406 (281/151)	56,240	678,049 (135,609/75,369)
Fold 10	800 (160/88)	1368 (273/189)	54,720	679,319 (135,863/74,099)

Table 3. CPM scores comparison of proposed system with other state of the art systems on LUNA16 dataset at seven false positives per scan (0.125, 0.25, 0.5, 1, 2, 4 and 8).

Method	CPM								
	CNN	0.125	0.25	0.50	1	2	4	8	Average
Zuo *et al.*	2D	0.672	0.694	0.714	0.739	0.766	0.787	0.822	0.742
Xie *et al.*	2D	0.493	0.688	0.796	0.852	0.864	0.864	0.864	0.775
Setio *et al.*	2D	0.636	0.727	0.792	0.844	0.876	0.905	0.916	0.814
Gorkam *et al.*	3D	0.588	0.669	0.749	0.831	0.863	0.892	0.913	0.786
Dou *et al.*	3D	0.677	0.834	0.927	0.972	0.981	0.983	0.983	0.908
Ding *et al.*	3D	0.797	0.857	0.895	0.938	0.954	0.970	0.981	0.913
MP-AFNet (proposed)	3D	0.821	0.869	0.935	0.968	0.971	0.976	0.982	0.931

0.931 (1.97% increase). It is also important to note that, even though our system has lower sensitivity at 1, 2, 4, and 8 false positives per scan compared to [11], our system still achieved a higher sensitivity of 0.821%, 0.869% and 0.935% at 0.125, 0.25, 0.50 false positives per scan respectively compared to all the methods given in Table 2. That is, even at very low false positive rates which is the main goal of our automated lung nodule detection system.

3.3. Quantitative Analysis of Proposed Architectures

To make the networks more diverse, the architectures are designed to have different number of parameters that are trainable: MP-ACNN3 (801,434), MP-ACNN2 (6,006,338) and MP-ACNN1 (43,634,882) where the number of parameters trainable is in parenthesis along with different number of convolutional layers and filters used. We have quantitatively assessed the performance of the proposed multi patch attention architectures along with that of the final fusion model. The detection sensitives of each model at 7 pre-defined false positives per scan and the final average CPM score are given in Table 4. In order to ensure that the system can deal with the accurate detection of true positives along with a very few number of false positives, very low false positive rates (0.125, 0.25, 0.50 false positives per scan) are included in the evaluation metrics.

The FROC curves of all the models are given in Figure 12(a). First, regarding the approach of Multi-Patch Attention based Feature integration, each of the networks that used multiple patches outperformed the networks that have used single patches. While the MP-ACNN3, MP-ACNN2 and MP-ACNN1 achieved an average CPM of 0.878, 0.886 and 0.900 respectively, the corresponding SP-ACNN3, SP-ACNN2 and SP-ACNN1 achieved an average CPM of only 0.770, 0.777 and 0.799respectively. It is also observed that for each of the multi patch networks the sensitivities can reach above 90% at the rate of 8 false positives per scan. These results show the importance of considering different scales of contextual information from the center of nodules and also the discriminating capability of complex nodule representations in huge volumetric CT scans by 3D
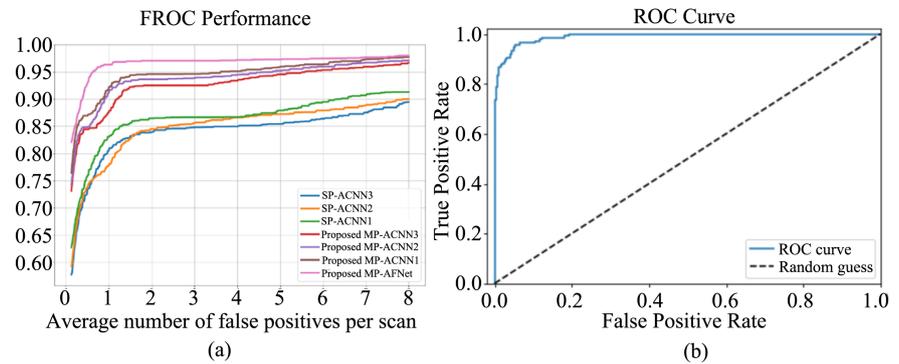
Figure 12. (a) FROC performance curve of all the models given in Table 3 [SP-ACNN1, SP-ACNN2, SP-ACNN3, MP-ACNN1, MP-ACNN2, MP-ACNN3, MP-AFNet]. The curve includes detection sensitivities at 7 pre-defined false positives per scan [0.125, 0.25, 0.50, 1, 2, 4, and 8]; (b) ROC performance curve for the final proposed MP-AFNet (AUC score: 98.98). The curve includes the plotting of False Positive Rates on X-axis to True Positive Rates on Y-axis.

Table 4. The CPM scores of different Attention Based Single, Multi-Patched models and of the proposed fusion model (SP-ACNN3, SP-ACNN2, SP-ACNN1, MP-ACNN3, MP-ACNN2, MP-ACNN1, MP-AFNet).

Model	CPM							
	0.125	0.25	0.50	1	2	4	8	**Average**
SP-ACNN3	0.577	0.663	0.736	0.810	0.843	0.869	0.897	0.770
SP-ACNN2	0.592	0.674	0.743	0.812	0.847	0.871	0.905	0.777
SP-ACNN1	0.627	0.693	0.765	0.836	0.866	0.897	0.914	0.799
Proposed MP-ACNN3	0.731	0.813	0.847	0.908	0.924	0.955	0.968	0.878
Proposed MP-ACNN2	0.743	0.824	0.852	0.916	0.938	0.962	0.973	0.886
Proposed MP-ACNN1	0.764	0.846	0.873	0.925	0.947	0.968	0.979	0.900
Proposed MP-AFNet	0.821	0.869	0.935	0.968	0.971	0.976	0.982	0.931

Attention based CNN's. The MP-ACNN1 has even achieved a detection sensitivity of 0.925% at 1 false positives per scan. It is also worth noting the difference of the models CPM score between Attention and Non-attention based methods as seen in Table 5. There is an average difference of up to 3.7% CPM score between Attention and Non-Attention models. Also in Figure 12(b). ROC curve of the proposed MP-AFNet can be visualized which has an AUC score of 98.98. This shows the importance of suppressing various non nodule features using gating mechanisms.

Second, regarding the effect of using fusion/ensembling, the proposed MP-AFNet outperformed all the individual Multi-Patch attention networks by obtaining an average CPM of 0.931. When considering 0.125 false positives per scan, MP-ACNN3, MP-ACNN2 and MP-ACNN1 achieved a sensitivity of merely 0.731%, 0.743% and 0.764% respectively. Whereas, our fusion model has reached a sensitivity of 0.821% which exceeded the results of MP-ACNN3,

Table 5. The CPM scores of different non-attention single, multi-patched models and of the proposed fusion model (SP-CNN3, SP-CNN2, SP-CNN1, MP-CNN3, MP-CNN2, MP-CNN1 and MP-FNet).

| Model	CPM							
	0.125	0.25	0.50	1	2	4	8	**Average**
SP-CNN3	0.502	0.614	0.683	0.742	0.797	0.810	0.838	0.712
SP-CNN2	0.532	0.594	0.685	0.756	0.798	0.832	0.847	0.720
SP-CNN1	0.561	0.615	0.693	0.776	0.811	0.857	0.868	0.740
Proposed MP-CNN3	0.679	0.762	0.795	0.853	0.896	0.910	0.924	0.831
Proposed MP-CNN2	0.685	0.782	0.820	0.871	0.913	0.927	0.932	0.847
Proposed MP-CNN1	0.712	0.793	0.833	0.895	0.926	0.934	0.948	0.863
Proposed MP-FNet	0.778	0.821	0.893	0.924	0.941	0.948	0.954	0.894

MP-ACNN2 and MP-ACNN1 by 0.09%, 0.078% and 0.057% respectively. These results signify the importance of establishing a consensus among diverse ACNN architectures for the best performance in false positives reduction. It is also noteworthy that the individual multi–patch networks MP-ACNN3, MP-ACNN2 and MP-ACNN1 achieved better results than the methods of [7] [30] [31] [32] at all the 7 false positives per scan as given in Table 3. From these results, it is evident that the methods using 3D ACNN's outperform its 2D variants in learning features of complex lung structures and also signifies the importance of multi patch attention based feature integration in classification of nodules. The training and validation accuracy, loss of all the three different architectures we have proposed are given in Figure 13.

On analyzing the curves in Figure 13, it can be easily noted that the MP-ACNN3 has given the highest training and validation accuracy at the starting iterations due to its less number of trainable parameters. But as the training process converged gradually after a few iterations, MP-ACNN1 having more number of trainable parameters has yielded highest training and validation accuracy. The final performance (accuracy) of the models are of the following order in our case (MP-ACNN1 > MP-ACNN2 > MP-ACNN3). Though one model might outperform others in some cases thereby drawing a requirement of fusing the predictions from diverse architectures for a more reliable result. Some of these cases can be visualized in Figure 14.

Though we cannot completely generalize the reason for varying predictions in Table 6 by different models proposed, an appropriate reason can be established on analyzing the values along with the variation in the parameters of models. In (i) of Figure 14, the nodule is more complex in its size and shape which is why the MP-ACNN1 model failed to detect whereas MP-ACNN2 and MP-ACNN3 models being deeper with more number of layers detected the nodule better with a probability of more than 99% by learning complex features well with more number of convolutions. Correspondingly in (ii) of Figure 14, the nodule is more simple in its nature which is why the MP-ACNN1 having lesser number of
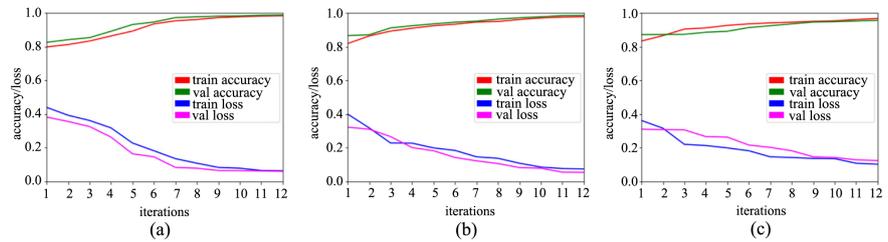
Figure 13. Training, Validation Accuracy and Loss Curves of different models considered with the number of epochs along x-axis, Accuracy/Loss along y-axis (a) MP-ACNN1; (b) MP-ACNN2; (c) MP-ACNN3.
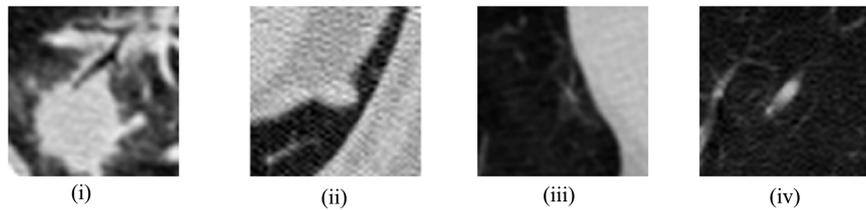

| (i) | (ii) | (iii) | (iv) |

Figure 14. (i), (ii), (iii), (iv) are the 2D transversal slices extracted from the middle of 48 × 48 × 48 cube which is extracted from the center of the nodule. (i), (ii) are nodules and (iii), (iv) are non-nodules.

Table 6. Nodule predictions of different models proposed (a) MP-ACNN1; (b) MP-ACNN2; (c) MP-ACNN3 which shows the importance of fusing the results from diverse architectures thereby maintaining a consensus decision.

	Predictions			
	MP-ACNN1	MP-ACNN2	MP-ACNN3	MP-AFNet
i	0.495	0.998	0.995	0.997
ii	0.999	0.412	0.794	0.897
iii	0.001	0.0006	0.984	0.001
iv	0.000003	0.000001	0.542	0.000002

convolutional layers detected the nodule with a probability of greater than 99% whereas MP-ACNN2 failed in detecting the nodule. The same is the reason for MP-ACNN1 performing better than other models in terms of accuracy and the final FROC/CPM score as the smaller nodules constitute majority portion in this dataset [17] of up-to 80% in the dataset. In (iii), (iv) cases of Figure 14, the MP-ACNN1 and MP-ACNN2 have correctly classified them as non-nodules whereas MP-ACNN3 failed by miss-classifying them as nodules. MP-ACNN3 failed in many cases apart from (iii), (iv) of Figure 14 in classifying non-nodules which is why it has achieved a low FROC/CPM score. But in automating the nodule detection system, more complex nodules as well as non-nodules have to be discriminated well, which is why in this research 3 diverse architectures are considered to fuse their results for overall increase in the performance of the proposed system. Some of the true positives detected by the proposed system with highest probability along with the nodules which were detected with relatively lowest probability are given in Figure 15 and Figure 16 respectively.

(i) Prob = 1   (ii) Prob = 1   (iii) Prob = 1

(iv) Prob = 0.99   (iv) Prob = 0.98   (iv) Prob = 0.97
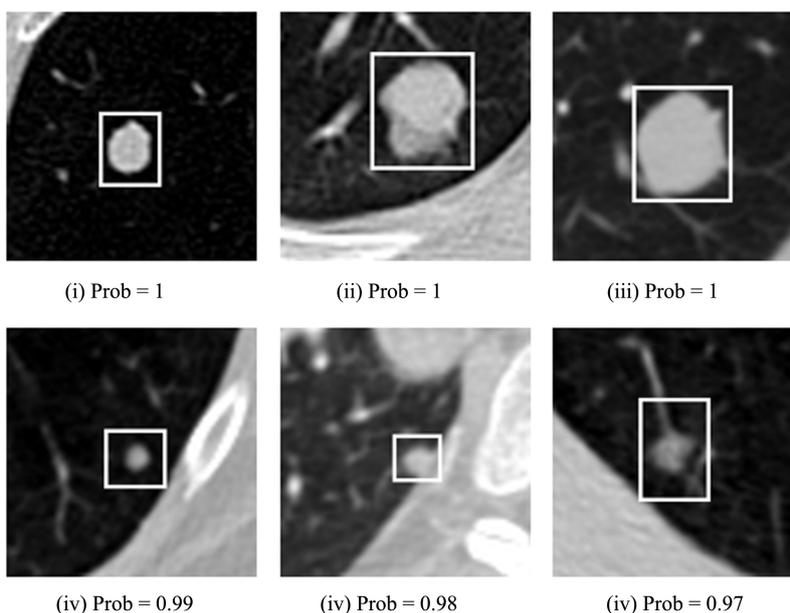
Figure 15. Nodules detected with highest probability by the proposed system. Each patch is a representative 64×64 transverse slice extracted from the center of the nodule. The nodule is enclosed inside the white bounding box.

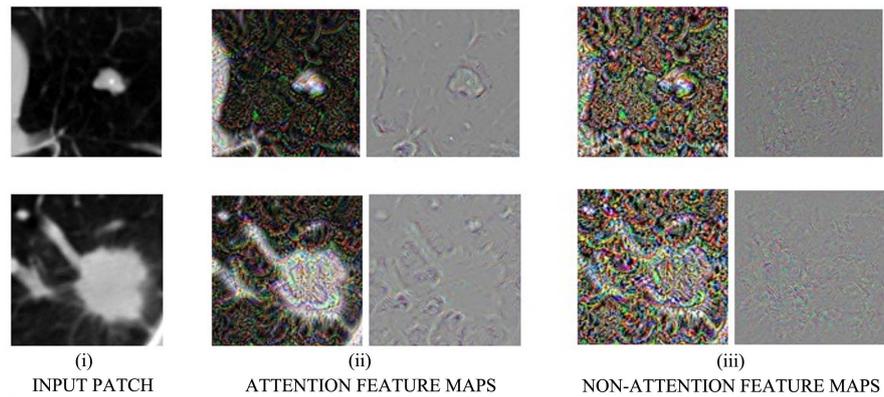(i) Prob = 0.37   (ii) Prob = 0.43   (iii) Prob = 0.47
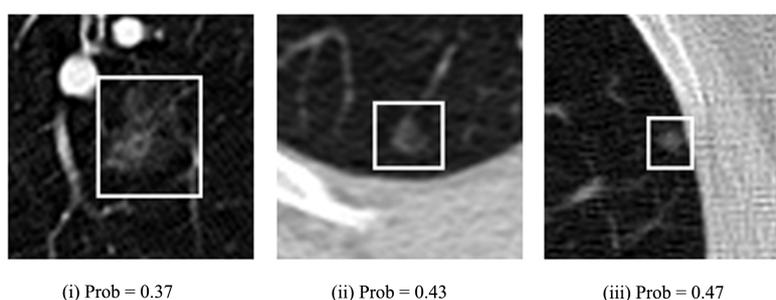
Figure 16. Nodules detected with lowest probability by the proposed system. Each patch is a representative 64×64 transverse slice extracted from the center of the nodule. The nodule is enclosed inside the white bounding box.

As seen in **Figure 15**, the nodules detected with highest probabilities are uniform in shape, bigger in size and are of solid or part solid type whereas the nodules detected with lowest probabilities as seen in **Figure 16** are non-uniform in shape, smaller in size and are of non-solid type which makes our model difficult to predict as they are similar to the background. As majority of the nodules in the dataset are either of solid or semi-solid type, there is a certain bias in training towards the non-solid types and hence to detect them with high probability more number of such samples should be included in the training set.

Our system has also performed well in detecting the nodules attached to the lung wall as seen in (v) of **Figure 14** and the reason is while preprocessing the CT scans emphasis is also given to retain the information of nodules attached to the borders using the techniques discussed in lung segmentation.

As seen in **Figure 17**, it is evident that there is a significant role attention has

<table>
<tr><td>(i)</td><td>(ii)</td><td>(iii)</td></tr>
<tr><td>INPUT PATCH</td><td>ATTENTION FEATURE MAPS</td><td>NON-ATTENTION FEATURE MAPS</td></tr>
</table>

Figure 17. Intermediate feature maps showing the importance of attention based architectures to that of non-attention architectures. (i) Input Patch (ii) Attention based Intermediate features of proposed MP-AFNet, (ii) Intermediate features without attention of proposed MP-AFNet.

on identifying and retaining nodule specific features even after several iterations of convolutions by giving more weightage to region of interests. With attention there is a great improvement in the classification performance of the model.

4. Conclusions

In this work an automated lung nodule detection CAD system for lung CT scans is proposed based on Attention based multi-patch, multi-network strategy for false positive reduction. In this work, we exploited three major approaches: 1) Use of different patch sizes that can cover varying nodule sizes in CT scans along with different views of contextual information; 2) Fusion of three diverse 3D Attention based CNN architectures namely MP-ACNN1, MP-ACNN2 and MP-ACNN3 for false positive reduction in complex structures where there is no homogeneity; and 3) an iterative training procedure to tackle the problem of unbalanced classification. The novel attention approach followed in our 3D CNN's helped to detect even small sized nodules accurately without the need for any nodule localization method. Also, the fusion approach used helped the system in detecting all the nodules with huge variations where in nodules with less detection complexity get easily detected by less deeper models having few convolutions with the highest probability and correspondingly nodules with high detection complexity get easily detected by deeper models having more convolutions with the highest probability. Especially, our system got promising results even at very low false positives per scan which is the main requirement for any CAD system. Our current work is mostly focused on false positive reduction given the center coordinates of nodule candidates but our system can be easily extended to a complete CAD system for detecting positive nodules in low dose CT scans by including an initial candidate screening system before our false positive reduction system. The proposed system is generic as well as modular and it can easily be extended to any other classification tasks of 3D medical diagnosis data. These results also signify that CT scans can be leveraged to bring in prom-

ising automated diagnosis systems using latest technologies. Also with the improvements that are happening in low dose CT scans it has proven to be a safer option in clinical applications utility space with no fear of cancer due to radiation.

Acknowledgements

Conflicts of Interest

No conflicts of interest, financial or otherwise, are declared by the authors. The authors state no conflict of interest and have nothing to disclose.

Author Biography

Vamsi Krishna Vipparla is a part-time researcher, and a full-time Assistant Manager in the Emerging Technologies Department of Mahindra. His research interests include deep learning, machine learning, digital image processing. He received the Bachelor of Technology degree in computer science and engineering from BML Munjal University, Gurgaon.

Premith Kumar Chilukuri is a part-time researcher, and full-time Lead machine learning engineer at Supervue Ai. His research interests include Cognitive intelligence, Generative Learning, deep learning, machine learning, digital image processing. He received his Bachelor of Technology degree in computer science and engineering from BML Munjal University, Gurgaon.

Dr. Giri Babu Kande is a professor in Electronics and Communication Department in VVIT, Guntur. He has teaching experience of about 20 years. He is guiding many UG, PG projects, and research scholars. His research interests include digital image processing, VLSI, and communication. He received the PhD degree in digital image processing from Jawaharlal Nehru Technological University, Hyderabad. He is a member of various professional chapters and published many research papers in various SCI journals and national and international conferences.

References

[1] WHO (2018). https://www.who.int/news-room/fact-sheets/detail/cancer

[2] Our World in Data (2018). https://ourworldindata.org/cancer

[3] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018) Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **68**, 394-424. https://doi.org/10.3322/caac.21492

[4] The National Lung Screening Trial Research Team (2011) Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal*

of Medicine, **365**, 395-409. https://doi.org/10.1056/NEJMoa1102873

[5] Su, H., Sankar, R. and Qian, W. (2006) A Knowledge-Based Lung Nodule Detection System for Helical CT Images. *International Journal of Computational Intelligence and Applications*, **6**, 371-387. https://doi.org/10.1142/S146902680600185X

[6] John, J. and Mini, M.G. (2016) Multilevel Thresholding Based Segmentation and Feature Extraction for Pulmonary Nodule Detection. *Procedia Technology*, **24**, 957-963. https://doi.org/10.1016/j.protcy.2016.05.209

[7] El-Regaily, S.A., Salem, M.A.M., Aziz, M.H.A. and Roushdy, M.I. (2020) Multi-View Convolutional Neural Network for Lung Nodule False Positive Reduction. *Expert Systems with Applications*, **162**, 113017.
http://www.sciencedirect.com/science/article/pii/S0957417419307341
https://doi.org/10.1016/j.eswa.2019.113017

[8] Sori, W.J., Feng, J. and Liu, S. (2019) Multi-Path Convolutional Neural Network for Lung Cancer Detection. *Multidimensional Systems and Signal Processing*, **30**, 1749-1768. https://doi.org/10.1007/s11045-018-0626-9

[9] Dong, X.L., Xu, S., Liu, Y., *et al.* (2020) Multi-View Secondary Input Collaborative Deep Learning for Lung Nodule 3D Segmentation. *Cancer Imaging*, **20**, Article No. 53. https://doi.org/10.1186/s40644-020-00331-0

[10] Wang, S., Zhou, M., Liu, Z., Gu, D., Zang, Y. and Dong, D. (2017) Central Focused Convolutional Neural Networks: Developing a Data-Driven Model for Lung Nodule Segmentation. *Medical Image Analysis*, **40**, 172-183.
https://doi.org/10.1016/j.media.2017.06.014

[11] Dou, Q., Chen, H., Yu, L., Qin, J. and Heng, P.A. (2017) Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection. *IEEE Transactions on Biomedical Engineering*, **64**, 1558-1567.
https://doi.org/10.1109/TBME.2016.2613502

[12] Shen, W., Zhou, M., Yang, F., Yang, C. and Tian, J. (2015) Multi-Scale Convolutional Neural Networks for Lung Nodule Classification. *International Conference on Information Processing in Medical Imaging*, **9123**, 588-599.
https://doi.org/10.1007/978-3-319-19992-4_46

[13] Li, C., Zhu, G., Wu, X. and Wang, Y. (2018) False-Positive Reduction on Lung Nodules Detection in Chest Radiographs by Ensemble of Convolutional Neural Networks. *IEEE Access*, **6**, 16060-16067. https://doi.org/10.1109/ACCESS.2018.2817023

[14] Li, C., Tan, Y., Chen, W., *et al.* (2020) Attention Unet++: A Nested Attention-Aware U-Net for Liver CT Image Segmentation. 2020 *IEEE International Conference on Image Processing* (*ICIP*), Abu Dhabi, 25-28 October 2020, 345-349.
https://doi.org/10.1109/ICIP40778.2020.9190761

[15] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2016) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 2261-2269.
https://doi.org/10.1109/CVPR.2017.243

[16] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778.
https://doi.org/10.1109/CVPR.2016.90

[17] Lung Nodule Analysis (2016). http://luna.grand-challenge.org/

[18] Armato III, S.G., McLennan, G., Bidaut, L., Mcnitt-Gray, M.F., Meyer, C.R., Reeves, A.P., *et al.* (2011) The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules

on CT Scans. *Medical Physics*, **38**, 915-931. https://doi.org/10.1118/1.3528204

[19] Pulagam, A.R., Kande, G.B., Ede, V.K.R. and Inampudi, R.B. (2016) Automated Lung Segmentation from HRCT Scans with Diffuse Parenchymal Lung Diseases. *Journal of Digital Imaging*, **29**, 507-519. https://doi.org/10.1007/s10278-016-9875-z

[20] Luong, M.T., Pham, H. and Manning, C.D. (2015) Effective Approaches to Attention-Based Neural Machine Translation. In: *Proceedings of the* 2015 *Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Lisbon, 1412-1421. https://doi.org/10.18653/v1/D15-1166

[21] Chollet F. Keras (2015). https://keras.io/

[22] Abadi, M., Agarwal, A., Barham, P., *et al.* (2016) TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. In: *Computer Vision and Pattern Recognition*. arXiv preprint arXiv:1603.04467.

[23] DeLuca Jr., P.M., Wambersie, A. and Whitmore, G.F. (2008) Receiver Operating Characteristic Analysis in Medical Imaging. *Journal of the International Commission on Radiation Units and Measurements*, **8**, No. 1.

[24] Niemeijer, M., Loog, M., Abràmoff, M.D., Viergever, M.A., Prokop, M. and van Ginneken, B. (2011) On Combining Computer-Aided Detection Systems. *IEEE Transactions on Medical Imaging*, **30**, 215-223.
https://doi.org/10.1109/TMI.2010.2072789

[25] Jacobs, C., van Rikxoort, E.M., Twellmann, T., Scholten, E.T., de Jong, P.A., Kuhnigk, J.M., *et al.* (2014) Automatic Detection of Subsolid Pulmonary Nodules in Thoracic Computed Tomography Images. *Medical Image Analysis*, **18**, 374-384.
https://doi.org/10.1016/j.media.2013.12.001

[26] Murphy, K., van Ginneken, B., Schilham, A.M.R., de Hoop, B.J., Gietema, H.A. and Prokop, M. (2009) A Large-Scale Evaluation of Automatic Pulmonary Nodule Detection in Chest CT Using Local Image Features and k-Nearest-Neighbour Classification. *Medical Image Analysis*, **13**, 757-770.
https://doi.org/10.1016/j.media.2009.07.001

[27] Setio, A.A.A., Jacobs, C., Gelderblom, J. and van Ginneken, B. (2015) Automatic Detection of Large Pulmonary Solid Nodules in Thoracic CT Images. *Medical Physics*, **42**, 5642-5653. https://doi.org/10.1118/1.4929562

[28] Tan, M., Deklerck, R., Jansen, B., Bister, M. and Cornelis, J. (2011) A Novel Computer-Aided Lung Nodule Detection System for CT Images. *Medical Physics*, **38**, 5630-5645. https://doi.org/10.1118/1.3633941

[29] Traverso, A., Torres, E.L., Fantacci, M.E. and Cerello, P. (2017) Computer-Aided Detection Systems to Improve Lung Cancer Early Diagnosis: State-of-the-Art and Challenges. *Journal of Physics: Conference Series*, **841**, 012013.
https://doi.org/10.1088/1742-6596/841/1/012013

[30] Polat, G., Dogrusöz, Y.S. and Halici, U. (2018) Effect of Input Size on the Classification of Lung Nodules Using Convolutional Neural Networks. 26*th Signal Processing and Communications Applications Conference* (*SIU*), Izmir, 2-5 May 2018, 1-4. https://doi.org/10.1109/SIU.2018.8404659

[31] Zuo, W., Zhou, F., Li, Z. and Wang, L. (2019) Multi-Resolution CNN and Knowledge Transfer for Candidate Classification in Lung Nodule Detection. *IEEE Access*, **7**, 32510-32521. https://doi.org/10.1109/ACCESS.2019.2903587

[32] Xie, H., Yang, D., Sun, N., Chen, Z. and Zhang, Y. (2019) Automated Pulmonary Nodule Detection in CT Images Using Deep Convolutional Neural Networks. *Pattern Recognition*, **14**, 1969-1979.

[33] Ding, J., Li, A., Hu, Z. and Wang, L. (2017) Accurate Pulmonary Nodule Detection

in Computed Tomography Images Using Deep Convolutional Neural Networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D. and Duchesne, S., Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI* 2017. *Lecture Notes in Computer Science*, Vol. 10435, Springer, Cham, 559-567. https://doi.org/10.1007/978-3-319-66179-7_64