



This is a repository copy of *Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest CT*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/154524/>

Version: Published Version

---

**Article:**

Masood, A., Yang, P. [orcid.org/0000-0002-8553-7127](https://orcid.org/0000-0002-8553-7127), Sheng, B. et al. (6 more authors) (2019) Cloud-based automated clinical decision support system for detection and diagnosis of lung cancer in chest CT. *IEEE Journal of Translational Engineering in Health and Medicine*, 8. 4300113. ISSN 2168-2372

<https://doi.org/10.1109/jtehm.2019.2955458>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

Received 26 May 2019; revised 2 September 2019 and 24 October 2019; accepted 8 November 2019.  
Date of publication 4 December 2019; date of current version 26 December 2019.

Digital Object Identifier 10.1109/JTEHM.2019.2955458

# Cloud-Based Automated Clinical Decision Support System for Detection and Diagnosis of Lung Cancer in Chest CT

ANUM MASOOD<sup>1</sup>, PO YANG<sup>2</sup>, BIN SHENG<sup>1</sup>, HUATING LI<sup>3</sup>, PING LI<sup>4</sup>, (Member, IEEE),  
JING QIN<sup>5</sup>, VITAVESKA LANFRANCHI<sup>2</sup>, JINMAN KIM<sup>6</sup>, AND  
DAVID DAGAN FENG<sup>6</sup>, (Fellow, IEEE)

<sup>1</sup>Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup>Department of Computer Science, University of Sheffield, Sheffield S1 4DP, U.K.

<sup>3</sup>Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China

<sup>4</sup>Department of Computing, The Hong Kong Polytechnic University, Hong Kong

<sup>5</sup>Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong

<sup>6</sup>Biomedical and Multimedia Information Technology Research Group, School of Information Technologies, The University of Sydney, Sydney, NSW 2006, Australia

CORRESPONDING AUTHORS: BIN SHENG (shengbin@sjtu.edu.cn) AND HUATING LI (huating99@sjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61872241, Grant 61572316, and Grant 81870598, in part by the National Key Research and Development Program of China under Grant 2017YFE0104000 and Grant 2016YFC1300302, in part by The Hong Kong Polytechnic University under Grant P0030419 and Grant P0030929, and in part by the Science and Technology Commission of Shanghai Municipality under Grant 18410750700, Grant 17411952600, and Grant 16DZ0501100.

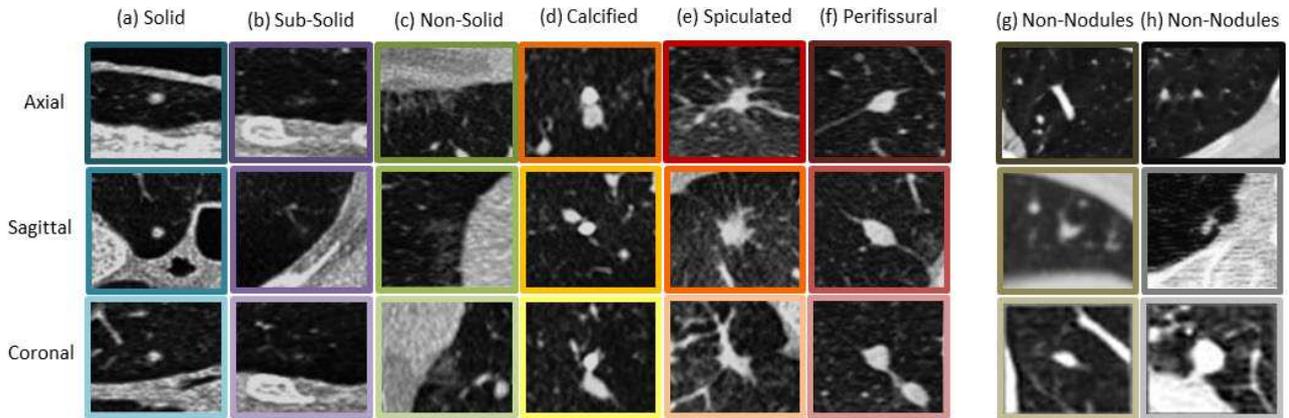
**ABSTRACT** Lung cancer is a major cause for cancer-related deaths. The detection of pulmonary cancer in the early stages can highly increase survival rate. Manual delineation of lung nodules by radiologists is a tedious task. We developed a novel computer-aided decision support system for lung nodule detection based on a 3D Deep Convolutional Neural Network (3DDCNN) for assisting the radiologists. Our decision support system provides a second opinion to the radiologists in lung cancer diagnostic decision making. In order to leverage 3-dimensional information from Computed Tomography (CT) scans, we applied median intensity projection and multi-Region Proposal Network (mRPN) for automatic selection of potential region-of-interests. Our Computer Aided Diagnosis (CAD) system has been trained and validated using LUNA16, ANODE09, and LIDC-IDR datasets; the experiments demonstrate the superior performance of our system, attaining sensitivity, specificity, AUROC, accuracy, of 98.4%, 92%, 96% and 98.51% with 2.1 FPs per scan. We integrated cloud computing, trained and validated our Cloud-Based 3DDCNN on the datasets provided by Shanghai Sixth People's Hospital, as well as LUNA16, ANODE09, and LIDC-IDR. Our system outperformed the state-of-the-art systems and obtained an impressive 98.7% sensitivity at 1.97 FPs per scan. This shows the potentials of deep learning, in combination with cloud computing, for accurate and efficient lung nodule detection via CT imaging, which could help doctors and radiologists in treating lung cancer patients.

**INDEX TERMS** Computer-aided diagnosis, nodule detection, cloud computing, computed tomography, lung cancer.

## I. INTRODUCTION

Among different types of cancer, pulmonary cancer also refer to as lung cancer is considered to be one of the most deadly cancers. In 2018, there were approximately 2.2 million new pulmonary cancer cases and about 1.8 million deaths in U.S. within a year. Pulmonary cancer is an uncontrollable abnormal lung cells growth, referred to as nodules, whose detection in early stages is highly crucial to the effective control of

disease progression and thus potentially increase the survival rate of the patient. Commonly used manual lung nodule delineation by radiologists on high-resolution and high-quality chest Computed Tomography (CT) is complex, time consuming and extremely tedious [1]. Automation of pulmonary nodule detection with effective and efficient Computer-Assisted Diagnosis (CAD) tools facilitates radiologists in fast diagnosis and improves the diagnostic confidence. Among these



**FIGURE 1.** Nodules and non-nodules in coronal, sagittal and axial view (nodules/non-nodules positioned at the center of the box (40 × 40 mm). Left images set are various types of nodules: (a) Solid (b) Sub-Solid (c) Non-Solid (d) Calcified (e) Spiculated (f) Perifissural while right images set are non-nodules.

approaches, one key challenge in CAD systems for lung cancer is dealing with the morphological variations in the nodules in CT images. Usually such variations are evident in images with different image modalities such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), X-ray or CT but in case of lung nodules numerous morphological variations are present even if the same image modality is used [2], [3]. Fig. 1 shows various nodules and non-nodules examples which depict the variety of morphological features, resulting in complexity in data used for nodule detection and diagnosis systems. Recently, deep learning methods [4], [5] have merged both the hand-designed feature extraction process and nodules classification process into a combined automated training process. Deep learning techniques have demonstrated great performance (i.e. reduced number of False Positive (FP) results) when compared with typical results reported by deploying traditional segmentation techniques [6], [7]. This paper presents a novel computer-aided decision support system for lung nodule detection. The contributions of this paper are threefold.

- A novel automated clinical decision support system for lung detection based on a 3D Deep Convolutional Neural Network (3DDCNN) architecture. In order to leverage 3-dimensional information from CT scans, we applied novel median intensity projection and introduced a novel multi-Region Proposal Network (mRPN) in our architecture for automatic selection of potential region-of-interest.
- To further improve the efficiency and performance of our proposed model, we integrated cloud computing into our CAD system. Proposed computer-aided decision support system is used for nodule detection and for assistance of radiologists in clinical diagnosis at Shanghai Sixth Peoples Hospital.
- A comprehensive experimental evaluation of our CAD system done on four different datasets with varying CT imaging parameters with existing state-of-the-art CAD systems for lung cancer detection demonstrated that our

system outperformed the existing systems and obtained an impressive 98.7% sensitivity at 1.97 FPs per scan.

The rest of the paper is organized as follows. Section 2 briefly introduces the related work, detailed description of our method is presented in Section 3. In Section 4, we discussed the experimental results on different datasets. Our paper is concluded with relevant future work in Section 5.

## II. RELATED WORK

CAD system is one of the most common means to improve the accuracy of cancer diagnosis done by the radiologists and decrease the time required for interpretation of the CT images. CAD systems are further categorised as: Computer Aided Detection (CADe) systems and Computer Aided Diagnosis (CADx) systems. CADe systems assist in finding the locality of nodules in CT images acquired from different imaging modalities while on the other hand the CADx systems characterize and classify these detected lesions as malignant or benign tumors. In general, a CAD system designed for the detection of pulmonary lesions (nodules) has two steps namely candidate nodule detection and FP Elimination. Firstly, the Regions Of Interest (ROIs) are selected in the input CT image, then the lung nodule candidates are extracted. Teramoto and Fujita [8] used Active Contour Model (ACM) filter for enhancement of contrast then used thresholding of the resultant images for the screening of candidate nodules. Supervised learning methods namely linear discriminant analysis (LDA), gray-scale distance transform, clustering (*k*-means clustering), connected component analysis, and patient-specific priori model have been used in conventional approaches [9]. FP reduction step classifies the lung nodules and non-nodules using machine learning techniques. The main objective is to eliminate the FP results which are considered as candidate in the previous step. Hierarchical Vector Quantization (HVQ), Rule-based filter, LDA, Artificial Neural Networks (ANN), and Support Vector Machine (SVM) are few supervised reduction methods which are used for FP reduction. Random Forest (RF) is reported to surpass

SVM in FPs reduction in lung CAD system. Regression tree-based classifiers have shown efficient discrimination ability in reduction of FPs for improved detection results [10]. Spatial and metabolic features in combination with SVM [8] are other approaches used for FPs elimination.

In the past few years, researchers have presented deep learning based CAD systems for cancer detection with promising results [11]. Convolutional Neural Network (CNN) framework is used for FPs reduction [12]. Nodules were accurately classified by using the fully connected layers (Fc) of CNN integrated with SVM classifier in [13]. Shen *et al.* [14] proposed a Multi-Crop CNN (MC-CNN) comprising of training by cropped convolutional feature maps and max-pooling layers recursively. Multi-View CNN proposed by Setio *et al.* [15] combines three candidate detectors each for sub-solid, solid, and large nodule category and then utilizes a fusion method to classify the input CT image. A 3-dimensional Fully Convolutional Network (FCN) based on Volumes Of Interest (VOI) was employed for classification [16]. This proposed work produces a score map with respect to the input VOI in single pass which is used for training of CNN used for classification. Deep learning based models have also been proposed for the candidate nodule detection [17], [18]. Multi-scale Laplace of Gaussian (LoG) filters and shape priors based multi-scale 3D-CNN model is proposed in [19]. In the past few years both CADE and CADx systems have been researched independently. CADE's major shortcoming for detecting lung cancer is their lack of ability to characterize them. CADE systems assist the radiologists in detection of lung nodules but do not provide detailed radiological characteristics of the lesion, consequently missing the information which is crucial for radiologists, while on the other hand CADx systems do not automatically identify lesions thus they do not possess high automation levels, making it not suitable for clinical use. Therefore, a new and advanced CAD system is needed, that incorporates the benefits of detection from CADE and diagnosis from CADx into a single system for better performance. The CADx systems performance evaluation is conducted in terms of computational efficiency, accuracy, sensitivity and specificity.

### III. MATERIALS AND METHODS

#### A. TRAINING DATASETS

For the training of our proposed method for nodule detection, we used LUNG Nodule Analysis (LUNA16) dataset [20] which comprises of 888 annotated CT scans. In these CT scans, four radiologists marked the lesions as nodule  $< 3mm$ , nodule  $\geq 3mm$ , or non-nodule in a two-phase annotation process. We used 55 CT-scans from ANODE09 dataset [21], among which only 5 have annotations done by three radiologists containing 39 nodules and 31 non-nodules. We used the remaining 50 as testing datasets which contained 433 non-nodules and 207 nodules along with LIDC-IDR dataset [22] to validate the nodule detection and classification of our proposed method. Since we were using two heterogeneous

datasets having varied image resolution therefore we resampled CT scans by the help of spline interpolation by 0.5mm per voxel along  $x$ ,  $y$  and  $z$ -axis to have constant resolution and we further reconstructed all the images by sharp kernel (Siemens B50 kernel).

#### B. DATA AUGMENTATION

CNN models have a tendency to overfit data in case of limited labeled training dataset [12], therefore to ensure that the training of our model does not overfit, we trained our model with data-augmented training dataset. Since benign nodules are more in number as compared to malignant nodules, we choose to augment the malignant training samples by cropping, duplicating, random translation within the range of [1, 0, or  $-1$  pixels in each dimensions (3D)] voxels, flipping, scaling, swapping in three dimensions axes and then rotating on the angle of [0, 90, 180, 270 degree] in training dataset. Specifically, among input batch, random translation as well as rotation are performed for up-sampling and down-sampling. These data-augmentation methods assisted our model in capturing nodule attributes invariant to image-level affine transformations.

#### C. PRE-PROCESSING

##### 1) MULTI-SCALE ROI PATCHES

Multi-scale ROI patches were generated by zooming in or out of the CT image in coronal, axial and sagittal views (see Fig. 1). The motivation for the multi-scale ROI patches comes from the real life scenario when the radiologists detects cancer patterns in a patient's CT-scans. In this scenario, the suspected regions were explored on pixel level in the follow-up check-up thus making these regions more scrutinized than the rest. Therefore, the training dataset comprising of nodules was used in different multi-scale patches as shown in Fig. 2. Using multi-scale ROI patches also upsampled the labeled dataset.

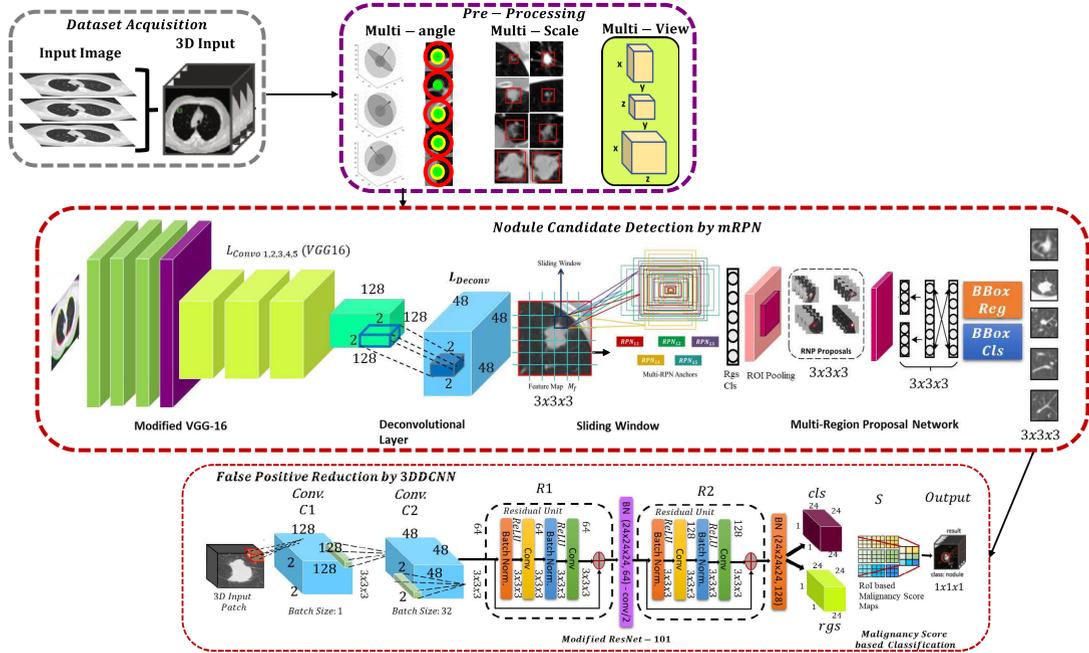
##### 2) MULTI-ANGLE ROI PATCHES

Multi-angle ROI patches were generated by rotation of obtained multi-scale patches by small angle  $\theta$  in orthogonal coordinate systems to obtain labeled data. The deep learning based methods can easily process these multi-scale and multi-view ROI patched to their properties such as shift invariance [23].

##### 3) MULTI-VIEW COMBINATION

In a two dimensional lung CT scans, most of the FPs are caused by the trachea and the blood vessels having tubular spatial structure. To reduce the noise caused by trachea and the blood vessels, we used a three dimensional bilateral smoothing filter  $S(x, y, z)$  for the pre-processing of CT-scans  $C(x, y, z)$ .

$$S_i(x, y, z) = n_c^{-1} \sum_{(\theta, \xi, \eta) \in \Omega_w} C(\theta, \xi, \eta) w_c(\theta, \xi, \eta; x, y, z) \quad (1)$$



**FIGURE 2.** Our CAD system comprises of four stages: data acquisition (2D CT scan image to generate MIP projected images), pre-processing (multi-angle, multi-scale and multi-view), candidate screening (mRPN for nodule detection), false positive reduction using 3DDCNN.

where  $S_i(x, y, z)$  is the output smoothed image,  $n_c^{-1}$  is the normalization coefficient and  $\Omega_w$  means the filter window  $w_c(\theta, \xi, \eta; x, y, z)$  denotes the weight coefficient which can be described as:

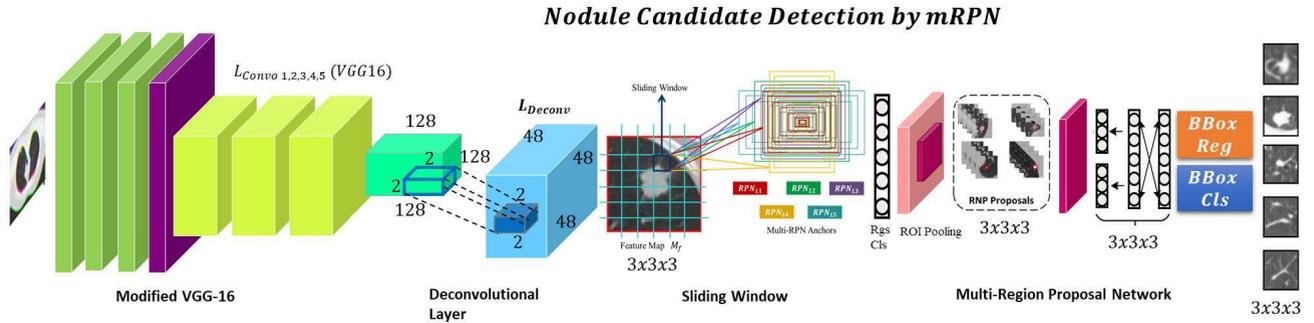
$$w_c(\theta, \xi, \eta; x, y, z) = w_G((\theta, \xi, \eta), (x, y, z)) \cdot w_E(C_{\theta, \xi, \eta}, C_{x, y, z}) \quad (2)$$

where  $w_G((\theta, \xi, \eta), (x, y, z))$  expressed geometric similarity whereas  $w_E(C_{\theta, \xi, \eta}, C_{x, y, z})$  expressed energy similarity. For removal of the effect of trachea and the blood vessels meanwhile enhancing the ROIs, three dimensional isotropic Gauss function was utilized. Thus, the weight coefficient can be described as follows:

$$w_c(\theta, \eta, \xi; x, y, z) = \left[ -\frac{1}{n} \left( \frac{\theta - x}{\sigma_g} \right)^2 + \left( \frac{\xi - y}{\sigma_g} \right)^2 + \left( \frac{\eta - z}{\sigma_g} \right)^2 \right] \cdot \left[ -\frac{(C_{\theta, \xi, \eta} - C_{x, y, z})^2}{2\sigma_E^2} \right] \quad (3)$$

where  $\left[ -\frac{1}{n} \left( \frac{\theta - x}{\sigma_g} \right)^2 + \left( \frac{\xi - y}{\sigma_g} \right)^2 + \left( \frac{\eta - z}{\sigma_g} \right)^2 \right]$  represents the geometric similarity,  $\left[ -\frac{(C_{\theta, \xi, \eta} - C_{x, y, z})^2}{2\sigma_E^2} \right]$  represents the energy similarity and  $\sigma_E$  and  $\sigma_g$  represents the standard deviation ( $\sigma$ ) of gray-level energy values and Gauss function. The terms  $C(\theta, \xi, \eta)$  and  $C(x, y, z)$  are related to CT scan gray-level energy values therefore they are referred to as energy similarity whereas  $\left( \frac{\theta - x}{\sigma_g} \right)^2, \left( \frac{\xi - y}{\sigma_g} \right)^2, \left( \frac{\eta - z}{\sigma_g} \right)^2$  represent geometric similarity since these terms are related to the spatial structure of the anatomical structures. We used  $-\frac{1}{n}$  with the geometric similarity where  $n$  represents the independent planes, in case

of energy similarities there is no clear gray level boundaries therefore we used  $-$  with the energy similarity. The distance between  $(\theta, \xi, \eta)$  and  $(x, y, z)$  is the distance between the spatial nodule localization and the gray level energy difference between nodule and other anatomical structure. This similarity metric is the first decision making step to reduce the undesirable feature redundancies which tend to cause false positive results. This metric ensures that the nodule class similarity is maximized on the other hand the non-nodule to nodule class similarity is minimized. Traditionally the Gaussian function refers to mean and co-variance matrix but when the number of independent parameters increase as the number of dimension increase then the multi-dimensional isotropic Gaussian distribution is considered where the variance of each dimensional is the same. In our case, we have three dimensional MIP projected images with multi-view combination therefore the 3D isotropic Gauss function was used. Afterwards an enhancement filter is used which suppresses the spatial tubular structure but enhances the nodule structures. Although commonly the classification phase has a problem when the input channels have multiple color channels, yet in our CT-scans dataset we only have gray-level images. CT input training images are 2-D whereas the locality of the nodule is presented by  $z$ -axis leveraging the inter-slice dependencies through memory units are three dimensional. Thus, to combine the multi-views of three dimensions of CT-scans, each voxel of the 3-D CT scans input was processed by the dot-enhancement filter which is inspired by the  $3 \times 3$  Hessian matrix. We define  $\theta$  as the image projected by Maximum Intensity Projection (MIP). With input image



**FIGURE 3.** Overview of the Proposed mRPN architecture for Nodule Candidate Detection using modified VGG-16 baseline for  $L_{conv}(1,2,3,4,5)$  integrated with deconvolutional layer  $L_{deconv}$ , feature map  $M_f$ , multi-RPN anchor with Classification (cls) and Regression (rgs) layers.

patch  $I$ ,  $\vartheta$  for three dimensions can be presented as:

$$\begin{aligned} \vartheta(y_i, z_i) &= \underset{x_i}{\text{med}} I(x_i, y_i, z_i) \\ \vartheta(x_i, z_i) &= \underset{y_i}{\text{med}} I(x_i, y_i, z_i) \\ \vartheta(x_i, y_i) &= \underset{z_i}{\text{med}} I(x_i, y_i, z_i) \end{aligned} \quad (4)$$

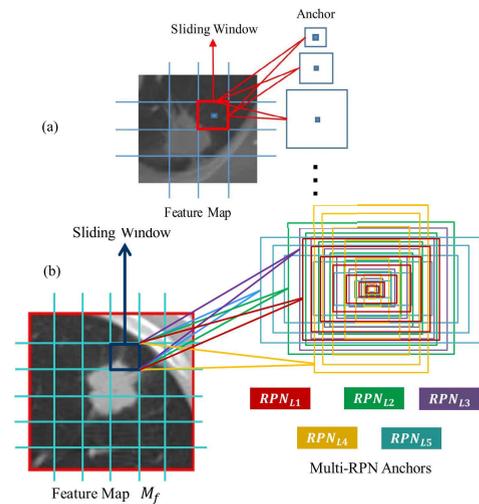
where *med* denotes median operator. Different views can provide different plane information, while patches with combination of different dimensions can provide the space distribution of tumor tissues. In order to construct input image sets with three channels, we connect three MIP projected images together:  $\vartheta = [\vartheta(y_i, z_i), \vartheta(x_i, z_i), \vartheta(x_i, y_i)]$ .

#### D. PROPOSED MODEL ARCHITECTURE

Our proposed architecture uses the basic framework of the Faster R-CNN [24]. Candidate detection is done by the proposed mRPN while the FP reduction is done by novel 3D DCNN.

##### 1) CANDIDATE DETECTION BY mRPN

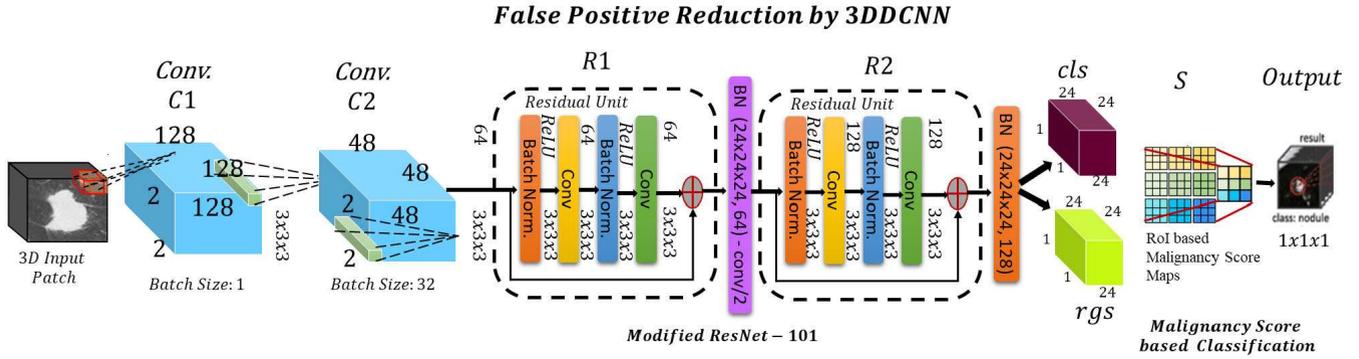
To use the training ability of CNN model for lung cancer, the model should be both end-to-end and trained frequently trainable for ROI detection and classification such as, Faster R-CNN [24]. The challenge for Faster R-CNN [24] in case of pulmonary lesions detection is the diversity in the nodule size and limited labeled dataset. We proposed a novel method, mRPN that has enhanced the feature extraction process (multi-resolution) and uses varied window size for region proposal selection from ROIs. For efficient ROI selection, these ROI extracted from multiple RPNs are merged in an additional layer as shown in Fig. 3. Our proposed novel model mRPN, which is based on based on VGG-16 Net model proposed by [25], takes a CT image (of any size) as input and generates a set of rectangular region proposals, each outputs an objectness score. Our network Multi Region Proposal Network (mRPN) hyper-parameters of all layers from conv1 to conv5 are similar to the VGG16 model. The original VGG-16 model [26] comprises of multiple max pooling layers, which inevitably reduces the image size but simultaneously distort the relatively small sized malignant



**FIGURE 4.** (a) Existing anchor in region proposal network. (b) Proposed anchor in region proposal network.

nodule. We used a small network  $N_s$  to slide through the activation (feature) map output  $M_{out}$  by the final added deconvolutional layer  $deconv_L$ . We proposed a deconvolutional layer  $deconv_L$  [27], 4 kernel size and 4 stride size, to be added after the last feature extracting layer. The deconvolutional layer  $deconv_L$  (or more commonly known as transposed convolutional layer) upsampled the features learned from the input and the feature map  $M_f$ . This  $deconv_L$  upsample the feature maps that are derived from the downsampling stack to generate  $M_{out}$ , while ensuring that both the output  $M_{out}$  and input  $M_{in}$  have the same size. Traditionally, R-CNN depends on the skip connection linked with the deconvolution layer on the upsampling for generating initial results but the deconvolution layer is unable to recover the small-sized objects such as nodules, which are lost after the downsampling. Therefore, they cannot accurately detect small-sized nodules. In our proposed method, we used  $deconv_L$  which ensured the recovery of any loss of small objects such as lung nodules in the downsampling process.

The proposed anchor in RPN [24] is shown in Fig. 4. We explored large number of nodule boundary varying



**FIGURE 5.** Our 3DDCNN model for False Positive Reduction comprising of Convolutional Layer  $convC_1$  and  $convC_2$ , Residual Unit  $R_1 + BN$ ,  $R_2 + BN$ , Classification (cls) and Regression (rgs) layers, Scoring Layer  $S$  and Output Layer generating the malignancy score based classification as nodule (malignant or benign) or non-nodule.

in sizes, and generated seven different sizes of reference bounding boxes which are centered at each sliding spatial window  $w$ , in order to contain nodules of different malignant level, we choose anchor sizes of  $4 \times 4$ ,  $8 \times 8$ ,  $12 \times 12$ ,  $16 \times 16$ ,  $20 \times 20$ ,  $26 \times 26$ , and  $32 \times 32$ . These 7 anchors are divided into RPN levels targeting nodules diameter  $\tau$  ranging from 3mm to 35mm in different aspect ratios and different sizes. These different RPN levels work in a cascade manner and overall increase the efficiency of the proposed model performance as shown in Fig. 4(b). Each of these have a  $1 \times 1$  convolutional layer with about 28 units for the Bounding Box Regression (BBReg) and  $1 \times 1$  convolutional layer having 14 units for the Bounding Box Classification (BBcls). The BBReg with 28 units gives an output of  $(H, W, 28)$  size. This output is used for providing four regression coefficients for each of the seven anchors for each point in the feature map  $(H \times W)$ . These four Regression (Rgs) coefficients are further used to enhance the coordinates of the anchors that is comprised of nodules. On the other hand the BBcls with 14 units provide an output  $(H, W, 14)$  which is used to obtain classification (cls) probabilities for each of the point of feature map  $(H, W)$  whether it contains a nodule within these seven anchors at the given point or not.

Nodules detection carried out using the different levels of RPN results in improvement of the nodule detection since both diameter and volume are considered. Volumetric values (3D input) are in correlation with the diameter values (2D input), therefore the combination of both volume and diameter provides divergence for non-nodules.

## 2) FALSE POSITIVE REDUCTION BY 3DDCNN

False Positive reduction is carried out using by novel 3DDCNN which is inspired by the ResNet-101 network [28]. The 3DDCNN replaces 2D with 3D convolution. This is used to predict the presence of nodules or to classify if a nodule exists on the basis of the malignancy value. VGG16 Net model is the basic layout of the 3DDCNN [26] having 100 convo layers with stride:2 where filters of size  $3 \times 3 \times 3$  used. Furthermore, this architecture network was improved

by adding connection for shortcuts converting it into its comparable counterpart residual network as shown in Fig. 5.

In case of 3DDCNN, the feature map is defined as:

$$\omega = \sum_I (x_i, y_i, z_i) \quad (5)$$

The input set was 3D CT image therefore  $\omega$  corresponds to the 3D position (e.g.  $x_i, y_i, z_i$ ).

$$v = \frac{\tau + v + \kappa}{n_{RPN}} \quad (6)$$

where  $\tau$  is the diameter of the detected nodule,  $v$  is the volume,  $\kappa$  is the malignancy score which is obtained by using the algorithm proposed in [11] and  $n_{RPN}$  is the total number of RPN levels used. For our model, value of  $n_{RPN}$  is 7.

$$M_f = (\omega + v + \varphi) \times W_I \times H_I \times \frac{D_I}{\mu} \quad (7)$$

where  $\varphi$  is the confidence score,  $W_I \times H_I \times D_I$  gives the input width, height and depth of  $128 \times 128 \times 128$ , respectively. The confidence score represents the probability value of an anchor to contain a nodule. In our model we used the confidence score as a threshold for determining whether the detected nodule is a FP or a True Positive (TP) result. We set the threshold for confidence score at various levels and obtained 95% confidence interval. Due to memory limitation, the  $\frac{D_I}{\mu}$  is used where the spatial scale factor presented by  $\mu$  is 3. The 3DDCNN is designed to acquire spatial scale  $\mu = 3$  between the input  $M_{in}$  and the output  $M_{out}$  space. The first convolution layer (size of kernel:  $7 \times 7 \times 7$ , stride:2) of 3DDCNN is applied on the input set on bidirection. In 3DDCNN, we have 3D convolution layer (kernel size:  $7 \times 7 \times 7$ , stride:2), followed by BN layer (batch normalization) and ReLu (rectified linear unit) activation. In order to keep low ratio among the feature map output  $M_{out}$  and the feature map input  $M_{in}$ , we proposed to use only two ResNet blocks having a single residual connection (kernel size:  $3 \times 3 \times 3$ ). One of the ResNet block has feature stride:1 while the other has stride:2. In case of similar dimensions for both the output and input, the identity alternatives can be directly used:

$$q = \mathbb{F}(p, \{W_i\}) + p \quad (8)$$

where  $p$  and  $q$  denotes input and output sets respectively, that are fed into and considered by each network layer. Residual mapping learning function is defined as  $\mathbb{F}(p, \{W_i\})$ . The 3DDCNN architecture modification is done by using convolutional layer for generating  $M_f$ , thus, the learnable weights are computed using convolutional layers which are shared on the image level.

According to radiologist annotations, we add a  $k \times k(5+1)$ -channel convolutional layer ( $4 \times 4$  in this paper) as the output layer to generate position sensitive score maps. Note that 5 represents five malignant level of lung nodule, 1 represents non-nodule, and we divided RPN levels proposed ROI into  $4 \times 4$  grid cell. Specifically, for  $32 \times 32$  proposed rectangular, each divided grid has the size of  $8 \times 8$ . Therefore  $4 \times 4 \times 7$  score maps will be generated, and we use the average pooling operation to calculate the relevance score to 7 categories for each split bin:

$$\zeta_c(w, h | \phi) = \sum_{(a,b) \in \text{bin}(w,h)} z_{w,h,c}(a + a_0, b + b_0 | \phi) / n \quad (9)$$

where  $\phi$  denotes parameters of the network,  $\zeta_c(w, h | \phi)$  is the relevance score of  $(w, h)$ th bin to malignant category  $c$ ,  $z_{w,h,c}$  is the score map generated by last convolutional layer,  $(a_0, b_0)$  is the top-left corner of ROI, and  $n$  denotes the total pixel number in the bin. With  $4 \times 4 \times 7$  relevance scores  $\zeta$  being calculated, the scoring layer  $S$  decide the malignancy level of the ROI by simply average voting and also apply cross-entropy evaluation for ranking ROI:

$$\zeta_c(\phi) = \sum_{w,h} \zeta_c(w, h | \phi) \quad (10)$$

Here  $\zeta_c(\phi)$  denotes the relevance score for ROI to class  $c$ , and  $\xi_c(\phi)$  is the softmax response for class  $c$ .

### E. TRAINING PROCESS

In the training process with multiple RPNs providing the region proposals, RPN loss function is applied for each RPN and Fast R-CNN loss function in an iteration. Our loss function can be defined by merging box regression and the cross-entropy loss:

$$\mathbb{L}_{t,\xi} = -\log(\xi_{c^*}) + \frac{1}{N_r} \sum \{L\}_r(t, t^*) \quad (11)$$

$$\mathbb{L}_r(t, t^*) = \begin{cases} 0.5(t - t^*)^2, & \text{if } |t - t^*| < 1 \\ |t - t^*| - 0.5, & \text{\&\{otherwise\}} \end{cases} \quad (12)$$

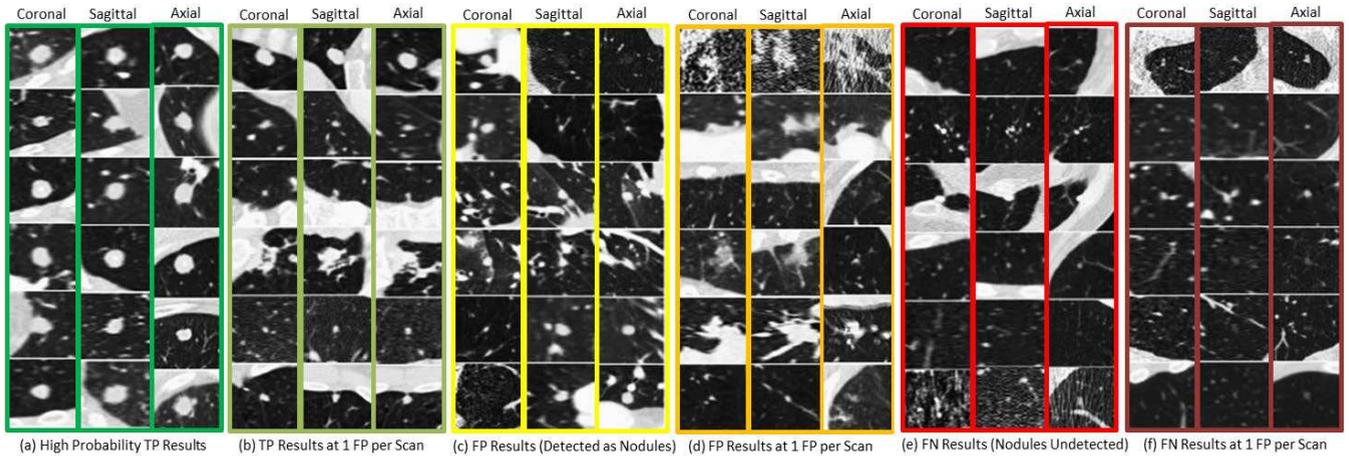
where the left part of the above equation denotes classification cross entropy loss [28],  $N_r$  is the input number of Regression layer,  $\mathbb{L}_r$  is similar to the bounding box regression loss as in [24],  $t^*$  denotes ground truth values while  $t$  denotes predicted values.  $\xi_{c^*}$  represent Intersection-over-Union (IoU) between any two entities that is their overlap volume divided by their union volume. In our model, IoU is used to select the best anchor to acquire nodule feature with least transformation. If an anchor of  $RPN_{L_x}$  has highest IoU, i.e.  $IoU_{max}$  with regards to any of the  $t^*$  or if  $IoU > 0.5$ , then the said anchor

is considered Positive  $An^P$ , whereas those anchors having  $IoU < 0.02$  are considered Negative  $An^N$  and anchors which are neither Positive nor Negative are irrelevant. A hard negative mining method was used to enhance the generalization. The classification imbalance was adjusted by normalizing the weights for classes (non-nodule, benign, malignant). For negative anchors,  $An^N$  weight was the probability of nodule class while on the other hand the positive anchors  $An^P$ , weight was 1. Learning rate was initially set to 0.001 while the decay method for learning rate was done every 30 epochs, the learning rate was halved. The CAD system was trained for 300 epochs using the batch size (size:32). In order to leverage the gradient information from these selected batches, average gradient operation was performed on the  $B$  samples and used as the input gradient estimation for Adam process to iteratively optimize the 3DDCNN. The experiment was conducted on Ubuntu 16.04.3 LTS with 4 processors, Intel(R) Xeon(R) CPU E5-2686 v4@2.3GHz and 64GB total memory space. Our model is trained on Tesla K80 with 12GB Memory. We used Intel Extended Caffe for implementation of 3DDCNN model.

### F. CLOUD-BASED 3DDCNN CAD SYSTEM

In this paper, we have proposed a two stage computer-assisted decision support system for lung cancer detection. To further improve the performance of our proposed method, we integrated cloud computing (Infrastructure as a Service (IaaS) by providing Virtual Machines, and Software as a Service (SaaS) by giving our 3DDCNN model) into our CAD system. The first stage for our proposed CAD system is the training of 3DDCNN model for the nodule candidate screening. The second stage is the reduction of false positive results from the first stage in order to improve the overall diagnosis decision making by our CAD system as shown in Fig. 2. The final decision from the proposed model is provided to the radiologists to assist their diagnosis for lung cancer. The diagnosis decision by the proposed CAD system is sent to the doctors in real-time who determine the cancer stage. These physicians afterwards sent the regular check-up reports and treatment prescription to the patient. Treatment prescription and check-up reports are stored on the cloud storage for further data analysis and improvement of our CAD system. To efficiently identify the effectiveness of each of the prescribed treatment for specific lung cancer stage patient.

The proposed CAD system uses body area network (BAN) comprising of sensors attached to patients body to record physiological information and CT-scan for chest CT which are stored on the cloud storage and undergo pre-processing. Furthermore gateways are used to forward that data to storage cloud for further processing. We deployed 12 VMs, and 24 processing units in our dedicated cloud back-end. For each case the complete processing time is about  $219 \pm 25.47$  seconds. HTCondor tool was used for real-time optimization and monitoring of computing resources, thus the radiologists and the physicians have updated responsive CAD system. 3DDCNN CAD system SaaS model was used to



**FIGURE 6.** True Positive (TP), False Positive (FP), False Negative (FN) results of our proposed CAD system (nodules positioned at the center of  $40 \times 40$ mm Patch).

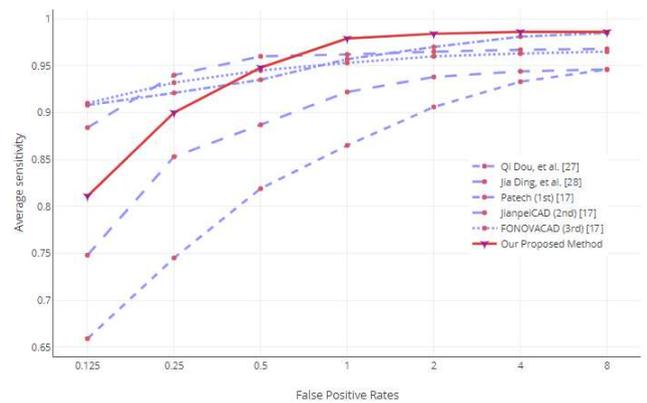
provide supportive decision support system for assistance of the radiologists whereas IaaS provided the GPU-acceleration, fast computation as well as storage resources. This model provides automatic support for on demand scalability of computing and storage resources. Moreover, Cloud-based 3DDCNN CAD system is more efficient and provides cost-effective solution since CAD results can be reviewed in real-time by multiple radiologists while a cloud back-end is taking care of computations. On the other hand, traditional stand-alone CAD systems have low performance and high computational cost with no feedback from multiple radiologists in real-time.

#### IV. EXPERIMENTAL RESULTS

In this research work, we used two phase validation. The first phase is nodule detection without classification as done by other researchers [7], [29]. The second phase combines the performance of independent detection with the classification results to provide the overall performance evaluation of the CAD system. For the nodule detection, we used Free-Response Receiver Operating Characteristic (FROC) [30] including average sensitivity and the number of FPs per scan (FPs/scan) which is the official evaluation metric for LUNA16, where detection is considered a true positive if the location lies within the radius of a nodule centre. The classification performance is evaluated by the area under the ROC curve (AUROC) which shows the performance of our proposed method on classification of nodules as nodules (malignant or benign) or non-nodules.

##### A. TRAINING

Our 3DDCNN model was trained for 300 epochs during each fold of cross-validation. After approximately training 100 epochs, the loss on the validation set became more stable. The result for each fold was selected to be the one with lowest malignancy prediction loss on the validation dataset. If nodule malignancy is  $M < 3$ , the nodule belongs to benign class, whereas if nodule has  $M \geq 3$  then it is categorized as a



**FIGURE 7.** Performance comparison between our CAD system versus state-of-the-art CAD systems on LUNA16. We compared our results with Top three CAD systems of LUNA16 Challenge namely Patech, JianpeiCAD, FONOVACAD [20] and the two published state-of-the-art CAD systems based on LUNA16 dataset i.e. Dou *et al.* [31] and Ding *et al.* [32].

malignant nodule. Since we obtained the results of detection results on three malignancy levels  $M$ , we used 10-fold cross-validation to merge the detection results of three levels.

##### B. NODULE DETECTION

###### 1) NODULE DETECTION USING LUNA16 DATASET

For evaluation of our proposed method's nodule classification, we compared our results with the two state-of-the-art published methods, i.e. Dou *et al.* [31] and Ding *et al.* [32] along with the best three LUNA16 challenge [20] by calculating the average sensitivity over 7 FPs/scan [0.125, 0.25, 0.5, 1, 2, 4, 8 FPs/scan]. Our method demonstrated best performance for nodule detection sensitivity of 0.812, 0.901, 0.948, 0.978, 0.984, 0.9853, 0.9866 at respective FPs/scan, obtaining an average FROC score of 0.946 shown in Fig. 7.

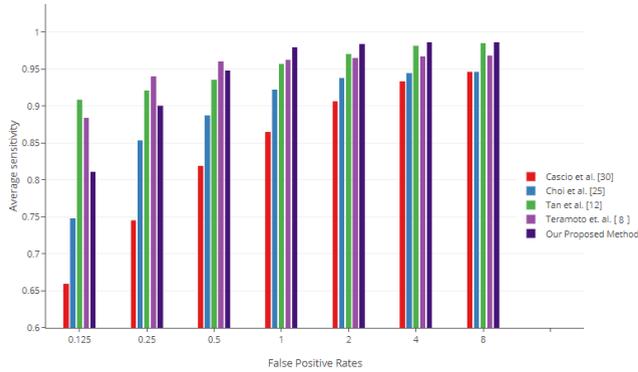
##### C. NODULE DETECTION AND CLASSIFICATION

###### 1) LIDC-IDR AND ANODE09 DATASET

We used a holdout validation set from LIDC dataset [20] and ANODE09 [21] to validate each model from 10-fold cross

**TABLE 1. Comparison of various classifiers’ accuracy (%) on LIDC-IDR and ANODE09 datasets.**

Classifiers	Accuracy(%)
CNN [12]	80.8
MTANNs [33]	88.6
FCN [13]	91.2
Autoencoder [10]	80.29
Massive-feat [9]	83.21
MC-CNN [14]	87.14
Our Proposed Method	98.51



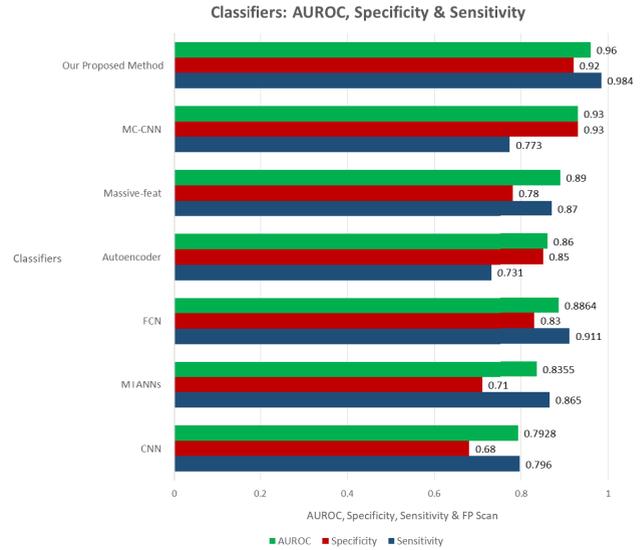
**FIGURE 8. Performance comparison between our CAD system versus state-of-the-art CAD systems on LIDC-IDR dataset.**

validation. For final results on LIDC, we retained 100% recall for validation sets, and reach 94.26% for nodules < 3mm at 2.9 FP/scan. Since it was not easy to process above 1000 test-sets in single step, we acquired result by cascading two FP elimination networks.

The effectiveness of 3DDCNN is verified by comparing with CNN [12], Autoencoder [10], Massive-feat [9], MC-CNN [14], MTANNs [33] and FCN [13]: the results are depicted in Table 1. The performance comparison between our proposed method for candidate nodules versus the state-of-the-art method in terms of sensitivity, specificity, AUROC and FP rate is shown in Fig. 8. Our proposed method improves 3.9% FROC on average over other state-of-the-art systems based on LIDC-IDR. The improvement on holdout test data validates our proposed method as an effective model to exploit potentially large amount of datasets which would not require further costly annotation by expert doctors and can be easily obtained from hospitals.

## 2) LIDC-IDR DATASET

Performance comparison between our proposed CAD system versus state-of-the-art CAD systems on LIDC-IDR dataset is shown in Fig. 9 in terms of the average sensitivity and FPs/scan.  $S$ ,  $SP$ ,  $FP$ ,  $FN$ , and  $TP$  denotes sensitivity, specificity, false positive, false negative and true positive rate, respectively. In our experiment, if a sample with nodule is not predicted as disease in our CAD system, it is  $FN$ . If a sample with nodule is predicted correctly, it represents  $TP$ .



**FIGURE 9. Performance comparison between our method and other existing classifiers for nodule detection on LIDC-IDR Dataset.**

We calculate FP rate between the number of non-nodule samples falsely predict as nodule with a certain level of malignancy and the total amount of non-nodule samples. In the following we give the definition of FP rate:

$$\frac{FP}{N} = \frac{FP}{FP + TN} \tag{13}$$

Sensitivity and specificity are calculated as:

$$S = \frac{TP}{TP + FN} \tag{14}$$

$$SP = \frac{TN}{TN + FP} \tag{15}$$

To convert the malignancy probability output by the classifier to a binary response, we used threshold (e.g.  $\tau = 0.5$ ). However, decreasing or increasing  $\tau$  will cause the classifier to produce more positive or negative predictions. We can observe in Table 1, that our proposed CAD system (3DDCNN) has attained the best performance with accuracy of 98.51%. From Table 2, we can deduce that 3DDCNN’s performance is highest in terms of sensitivity, which was found for 98.4% with a lowest FP rate of 2.1 per CT Scan among these CAD system. Our 3DDCNN performs better with mRPN than the original 3DDCNN proposed model. Fig. 8 shows the accuracy of our proposed method in comparison to the published classifiers for the nodule candidate detection. The comparison between our proposed system and the previously published CAD systems to investigate the perspectives of our 3DDCNN system was done using the average FPs/Scan as parameter shown in Table 3.

Fig. 6 shows true nodules (marked in green) that were missed in the traditional CNN method [12], but were detected by our proposed method, when the false positives per scan lies within the range of 1 to 4 with overall sensitivity of 0.9743. The false negatives are marked in red and are shown in the

**TABLE 2. Performance comparison of various classifiers' sensitivity, specificity, AUROC, FPs/Scan, and classification Time.**

Classifiers	Total Cases	Total Nodules	Sensitivity	Specificity	AUROC	FPs/Scan	Time (sec)	Classes
CNN [12]	800	1738	0.796	0.68	0.7928	3.63	179.5	Malignant
MTANNs [33]	120	493	0.865	0.71	0.8355	2.62	128.6	Malignant & Benign
FCN [13]	300	893	0.911	0.83	0.8864	3.16	69.75	Malignant & Benign
Autoencoder [10]	495	1010	0.731	0.85	0.86	3.11	0.01	Malignant & Benign
Massive-Feat [9]	880	1243	0.870	0.78	0.89	3.43	32.76	Malignant & Benign
MC-CNN [14]	340	825	0.773	0.93	0.93	2.97	0.23	Malignant & Benign
Proposed 3DDCNN	1190	2361	0.984	0.92	0.96	3.39	0.011	Malignant & Benign
Proposed 3DDCNN (with mRPN)	1190	2361	0.991	0.94	0.9743	2.10	0.025	Malignant & Benign

**TABLE 3. Performance comparison of our proposed CAD system with state-of-the-art CAD systems to detect and classify lung cancer on LIDC Dataset.**

CAD Systems	Nodule Size (3mm)	Sensitivity(%)	FPs/Scan
Cascio <i>et al.</i> [34]	$\geq 3$	97.0%	6.1
Teramoto and Fujita [8]	5-20	80.0%	4.2
Choi and Choi [29]	3-30	95.3%	2.3
CNN [12]	$\geq 3$	81.7%	3.7
Our Proposed Method	$\geq 3$	98.4%	2.1

**TABLE 4. Quantitative results associated with training, testing errors, of 3DDCNN and Cloud-Based 3DDCNN CAD system on different datasets.**

Dataset	3DDCNN CAD Mode	Training Error	Testing Error	Sensitivity	FP/scan
LIDC-IDRI [22]	Stand-Alone	0.00448	0.00661	0.952	2.4
	Cloud-Based	0.00391	0.00398	0.974	2.1
ANODE09 [21]	Stand-Alone	0.00542	0.00478	0.950	2.9
	Cloud-Based	0.00419	0.00354	0.976	2.3
LUNA16 [20]	Stand-Alone	0.00391	0.00328	0.961	2.2
	Cloud-Based	0.00368	0.00274	0.988	1.97
SPH6 Dataset	Stand-Alone	0.00364	0.00288	0.964	2.1
	Cloud-Based	0.00294	0.00224	0.987	1.97

"SPH6" refers to Clinical Dataset from Shanghai Sixth People's Hospital

last two rows; these have similar appearance to nodules but our proposed system detected them as non-nodules using the characteristics of lung nodules obtained by our 3DDFCN, such as the example in Fig. 6 third and fourth row marked in yellow.

### 3) CLINICAL DATASET

We investigated the performance of our 3DDCNN system in comparison to Cloud-Based 3DDCNN system on 120 cases from Shanghai Sixth People's Hospital using sensitivity and average FPs/Scan as parameter shown in Table 4. Training and Testing error were obtained for both Stand-alone and Cloud-based CAD mode. With the integration of cloud computing, Cloud-Based 3DDCNN CAD system allowed the cloud server to efficiently perform the nodule detection, and enables the cloud server to reduce overall storage costs by more than 59%, while ensuring improved lung cancer detection.

### D. QUANTITATIVE EVALUATION

As mentioned before the performance evaluation metric include Sensitivity, Specificity, AUROC, and false positives

**TABLE 5. Quantitative results for 3DDCNN in terms of Mean IoU and Average Precision against 3DDCNN network layers. Three AP are considered ( $AP_{50}$ ,  $AP_{75}$  and  $AP_m$  at different IoU thresholds) were selected showing mean IoU values of 3DDCNN layers  $i$  to  $n$ .**

Method	Layer	Testing Dataset			
		mean IoU mIoU	Average Precision ( $AP_m$ )	$AP_{50}$	$AP_{75}$
3DDCNN	$Layer_i$	37.7	49.4	39.5	34.8
	$Layer_j$	<b>48.7</b>	47.4	35.2	36.2
	$Layer_k$	33.4	48.9	<b>51.1</b>	32.3
	$Layer_n$	42.5	<b>50.8</b>	50.7	<b>37.5</b>
3DDCNN Using mRPN	$Layer_i$	40.3	36.5	<b>54.2</b>	33.9
	$Layer_j$	37.1	52.4	48.2	35.7
	$Layer_k$	42.0	42.6	48.9	<b>47.2</b>
	$Layer_n$	<b>45.8</b>	<b>53.5</b>	51.6	44.2

per scan (FPs/Scan). For further quantitative evaluation of our proposed work with the existing RCNN based methods, we used the statistical performance evaluation based on error calculation metrics which were used to determine the error rate on the testing dataset i.e. Clinical Dataset (120 cases from Shanghai Sixth People's Hospital). To calculate the average difference between the radiologists nodule detection and the proposed method nodule detection we proposed the Detection Error Rate ( $e_D$ ). The Detection Error Rate ( $e_D$ ) was calculated as:

$$e_D(I_\phi) = \sum_{i=0,k} e_I(G_k(x_i), h | I_\phi) \quad (16)$$

where  $e_D(I_\phi)$  is metric that was used to calculate the error rate of detection phase,  $k$  represents the number of nodules,  $\phi$  represents nodule in Image  $I$ ,  $h$  is the score derived for each layer of proposed model,  $x_i$  is the real value of the resultant metric for each layer, and  $G$  is the mean value for outputs delivered from previous layer. Another metric was proposed to calculate the average error between the nodule classification by doctors from Shanghai Sixth People's Hospital and the proposed methods classification. The metric, Classification Error Rate ( $e_C$ ) can be defined as:

$$e_C = \frac{1}{2} \log \frac{1 - e_D(I_\phi)}{e_D(I_\phi)} \quad (17)$$

Since, we obtained the detection results on three malignancy levels  $m$ , we used 10-fold cross-validation to merge

**TABLE 6.** Comparison of proposed 3DDCNN with state-of-the-art on clinical dataset using different statistical metrics, namely, mean detection error rate (mean  $e_D$ ), mean classification error rate (mean  $e_C$ ), variance detection error rate (Var  $e_D$ ), variance classification error rate (Var  $e_C$ ), standard deviation detection error rate (Std  $e_D$ ), standard deviation classification error rate (Std  $e_C$ ), mean average precision (m-AP) and processing-time (Time).

Methods	Score	Mean $e_D$	Mean $e_C$	Var $e_D$	Var $e_C$	Std $e_D$	Std $e_C$	m-AP (%)	Time (min)
Mask R-CNN [35]	83.63± 1.34	2.20±1.2	2.38±4.5	2.3±0.5	1.23±0.27	4.18 ±3.30	4.3±1.8	54.3±2.5	11
RetinaNet [36]	83.68± 4.10	3.94±0.71	1.93±2.8	4.6±1.2	1.62±0.30	3.40 ±4.31	3.9±3.0	56.4±2.8	7-8
Retina U-Net [37]	89.4± .91	2.64±2.3	1.84±2.3	1.1±0.7	1.79±0.95	2.82±2.32	4.2±1.7	62.9±3.4	5-6
Fast R-CNN [38]	79.5± 3.70	5.69±3.94	-0.04±4.4	5.8±1.3	1.97±0.90	8.58 ±8.02	9.9±2.2	44.9 ±3.0	21.7
Faster R-CNN [24]	76.3± 2.31	6.07±2.92	1.04±6.8	1.7± 0.6	1.61±0.99	7.41±4.31	9.31±5.1	52.8 ±10.7	17.2
Proposed 3DDCNN	84.2± 1.20	3.32±2.8	2.96±1.3	1.2± 2.35	1.24±0.51	3.69±3.87	4.01±5.2	55.8 ±4.1	4-6
Proposed 3DDCNN (Using mRPN)	87.8± 1.35	3.30±0.58	1.28±0.19	1.4±2.48	1.10±0.26	2.88±3.61	4.5±3.8	59.7±3.2	2-4

the detection results of three levels. The performance of nodule classification was validated using the LIDC dataset and the LUNA16's dataset distribution criteria of 10-fold cross-validation of patient-data. Owing to the minor differences among the malignant and the benign nodules, 900 epochs were used for the various learning rates [0.001,0.001,0.0001]. Detailed step-by-step performance of our proposed method is provided in the Table 5. We used mean Average Precision also refer to as  $AP$  for the detection phase evaluation. Table 5 represents the  $AP$  metric which averages  $AP$ s across IoU thresholds from 0.5 to 0.95 with an interval of 0.05. For our proposed model  $AP$ , we took three different IoU thresholds referring to three  $AP$  i.e.  $AP_{50}$ ,  $AP_{75}$  and  $AP_m$ .

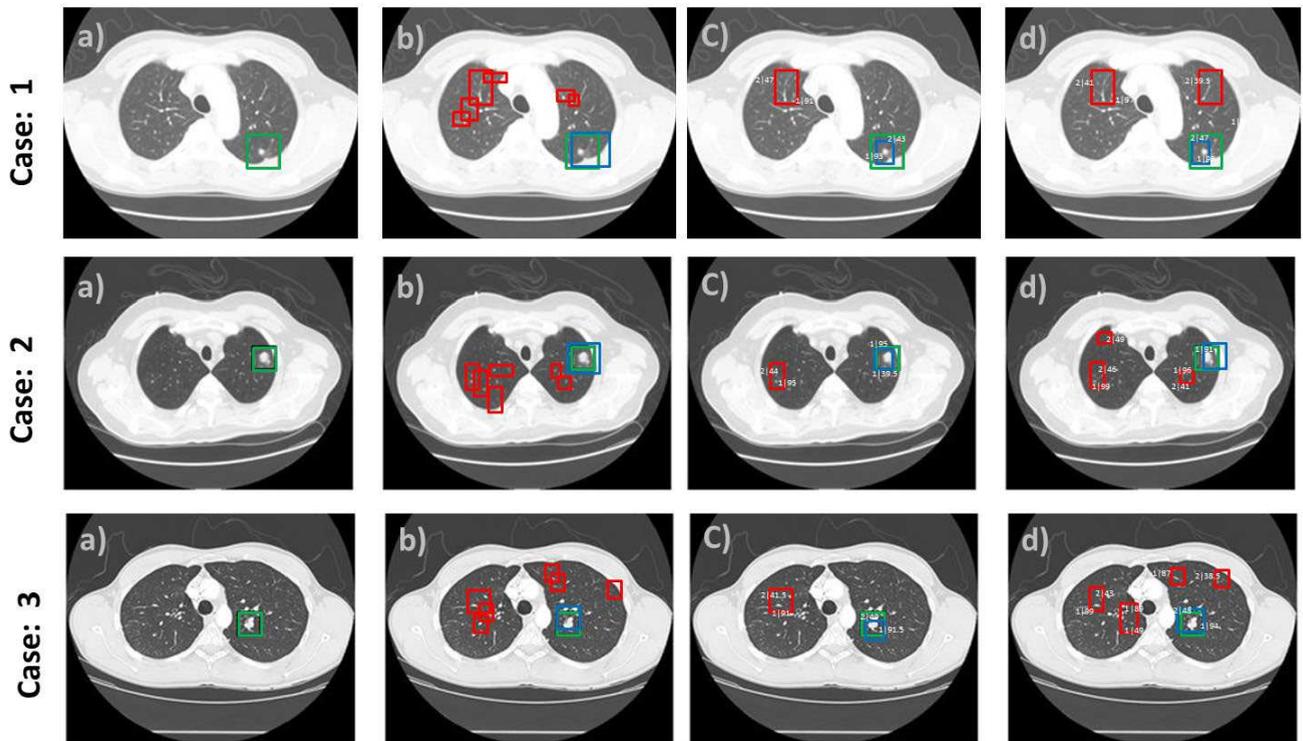
To quantitatively evaluate the results for our proposed method, we have measured Mean Detection Error Rate (Mean  $e_D$ ), Mean Classification Error Rate (Mean  $e_C$ ), Variance Detection Error Rate (Var  $e_D$ ), Variance Classification Error Rate (Var  $e_C$ ), Standard Deviation Detection Error Rate (Std  $e_D$ ), Standard Deviation Classification Error Rate (Std  $e_C$ ), Mean Average Precision (m-AP) and Processing-time (Time) of the CT images in the testing set of our clinical dataset with Mask R-CNN [35], RetinaNet [36], Retina U-Net [37], Fast R-CNN [38], Faster R-CNN [24] techniques. It can be seen in Table 6 that, our proposed method achieved comparatively good results for Detection and Classification than the state-of-the-art.

Table 6 presents the scoring criteria: statistical parameters such as mean, variance, standard deviation which are used for both types of error rates i.e. Detection error rate and Classification error rate. Mean Average Precision m-AP as 100% indicate perfect detection, the metric  $e_D(I_\phi)$  represents the error rate between the radiologists from Shanghai Sixth People's Hospital nodule detection and the detection done by our proposed method and the term  $e_C$  represents the nodule classification by doctors from Shanghai Sixth People's Hospital and our proposed methods classification. Computed results are presented in Table 6. Usually, it is difficult to compare the proposed study with state-of-the-art due to diverse datasets and different evaluation metrics. However, in our case, we compared our proposed work with other deep learning based methods which were designed

for object detection but performed well for nodule detection. The evaluation results of Fast R-CNN [38] and Faster R-CNN [24] are costly due to the rigorous object detection process, they are not specifically designed for the lung nodules, therefore, their detection error rate is comparatively higher than the rest of the methods. On the other hand, RetinaNet [36], Mask R-CNN [35] and Retina U-Net [37] performed better than the Fast R-CNN and Faster R-CNN. Although the assessment is conducted on the clinical dataset, Table 6 still emphasizes 3DDCNN advantages in terms of computational duration (of around  $219 \pm 25.47$  seconds) over other methods.

### E. QUALITATIVE EVALUATION

Qualitative results from Fig. 10 show that our proposed model 3DDCNN performed well in most of the clinical dataset cases (total 120 cases) from Shanghai Sixth People's Hospital while the state-of-the-art Retina U-Net [37] method outperformed our proposed method in some of the CT-scan from clinical dataset. We randomly selected 3 cases for our qualitative analysis. The step-by-step evaluation of the proposed work with Retina U-Net in the set of four visualizations of the central CT-scan slices for ground-truth and the classification accuracy is shown in the form of set of images (a), (b), (c), and (d). Images marked as (a) in each case (left-most images in each row) represents the ground-truth of given CT scan. Images (b) in each case depicts the box-prediction of the Retina U-Net (red boxes) while the prediction of our proposed method is represented with blue boxes. There are minor deviations of the markings of our proposed model (shown in blue) from the ground truth (shown in green) as shown in images (c). The last images (d) in each row shows the final results after the FP reduction, resulting in the correct prediction with approximately 87% confidence score. The proposed 3DDCNN performs well for nodule detection whereas the Retina U-Net performs better in case of nodule classification by achieving higher classification accuracy. Qualitative validation of our proposed 3DDCNN model against the annotations demonstrated that the detection accuracy of our proposed 3DDCNN performed equally well



**FIGURE 10.** Qualitative results from the performance comparison between our CAD system versus state-of-the-art CAD system on Clinical Dataset having a confidence threshold of 95%. Green represents the Ground truth box. Red as predicted by Retina U-Net [37] and blue represents our proposed model predictions. Three random cases out of 120 cases from Shanghai Sixth People’s Hospital are selected. Left-most images (a) show the ground truth of lung CT scan, images (b) in all cases show the box-predictions of Retina U-Net (2 foreground classes) as well as the prediction of our proposed 3DDCNN model, images (c) represent the remaining predictions after prediction of nodule along with some False Positive predictions, images (d) After the False Positives reduction step, leaving the correct prediction with confidence scores as high as 87%.

as the radiologists while in some cases even better than the radiologists annotations.

## V. CONCLUSION

For the detection of lung cancer, CAD systems are developed to assist the radiologist in the process of nodule detection by providing a reference opinion. From the performance comparisons it is evident that our proposed model 3DDCNN attained the highest results against other state-of-the-art systems for sensitivity and FPs per scan. Although the current tested performance metric of 3DDCNN is relatively high, it could be further improved. The performance was relatively less accurate in detecting micro nodules, therefore future work will investigate the detection of micro nodules whose diameter is less than 3 mm. To ensure our solution is scalable, future work will consider extending the training stage to include data from hospitals worldwide. Integrating more data-augmentation methods to increase the training sample in order to achieve more robustness and reduce the overfitting problem brought by local optimal. Another future direction for lung cancer CAD system is to propose CAD system that performs well on all nodule types maintaining good performance in terms of sensitivity and FPs/Scan, even if the dataset contains relatively less amount of such nodule types samples.

## REFERENCES

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2018,” *CA, Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, Jan. 2018.
- [2] S. Park, S. J. Lee, E. Weiss, and Y. Motai, “Intra- and inter-fractional variation prediction of lung tumors using fuzzy deep learning,” *IEEE J. Transl. Eng. Health Med.*, vol. 4, 2016, Art. no. 4300112.
- [3] T. Tan *et al.*, “Optimize transfer learning for lung diseases in bronchoscopy using a new concept: Sequential fine-tuning,” *IEEE J. Transl. Eng. Health Med.*, vol. 6, 2018, Art. no. 1800808.
- [4] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, “An automatic detection system of lung nodule based on multigroup patch-based deep learning network,” *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1227–1237, Jul. 2018.
- [5] X. Du, S. Yin, R. Tang, Y. Zhang, and S. Li, “Cardiac-DeepIED: Automatic pixel-level deep segmentation for cardiac bi-ventricle using improved end-to-end encoder-decoder network,” *IEEE J. Transl. Eng. Health Med.*, vol. 7, 2019, Art. no. 1900110.
- [6] H. Chung, H. Ko, S. J. Jeon, K.-H. Yoon, and J. Lee, “Automatic lung segmentation with juxta-pleural nodule identification using active contour model and Bayesian approach,” *IEEE J. Transl. Eng. Health Med.*, vol. 6, 2018, Art. no. 1800513.
- [7] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, “An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification,” *Expert Syst. Appl.*, vol. 128, pp. 84–95, Aug. 2018.
- [8] A. Teramoto and H. Fujita, “Fast lung nodule detection in chest CT images using cylindrical nodule-enhancement filter,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 8, no. 2, pp. 193–205, 2013.
- [9] M. S. Brown, M. F. McNitt-Gray, J. G. Goldin, R. D. Suh, J. W. Sayre, and D. R. Aberle, “Patient-specific models for lung nodule detection and surveillance in CT images,” *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1242–1250, Dec. 2001.

- [10] L. Lu, Y. Tan, L. H. Schwartz, and B. Zhao, "Hybrid detection of lung nodules on CT scan images," *Med. Phys.*, vol. 42, no. 9, pp. 5042–5054, 2015.
- [11] A. Masood *et al.*, "Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images," *J. Biomed. Inf.*, vol. 79, pp. 117–128, Mar. 2018.
- [12] M. Tan, R. Deklerck, B. Jansen, M. Bister, and J. Cornelis, "A novel computer-aided lung nodule detection system for CT images," *Med. Phys.*, vol. 38, no. 10, pp. 5630–5645, Oct. 2011.
- [13] B. van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2015, pp. 286–289.
- [14] W. Shen *et al.*, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognit.*, vol. 61, pp. 663–673, Jan. 2017.
- [15] A. A. A. Setio *et al.*, "Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1160–1169, May 2016.
- [16] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, "3D convolutional neural network for automatic detection of lung nodules in chest CT," in *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 1013409.
- [17] C. Zhao, J. Han, Y. Jia, and F. Gou, "Lung nodule detection via 3D U-Net and contextual convolutional neural network," in *Proc. IEEE Int. Conf. Netw. Netw. Appl. (NaNA)*, Oct. 2018, pp. 356–361.
- [18] J. Zhang, Z. Huang, T. Huang, Y. Xia, and Y. Zhang, "Lung nodule detection using combined traditional and deep models and chest CT," in *Proc. Int. Conf. Intell. Sci. Big Data Eng.* Cham, Switzerland: Springer, 2018, pp. 655–662.
- [19] B. Wang, G. Qi, S. Tang, L. Zhang, L. Deng, and Y. Zhang, "Automated pulmonary nodule detection: High sensitivity with few candidates," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.* Cham, Switzerland: Springer, 2018, pp. 759–767.
- [20] A. A. A. Setio *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, Dec. 2017.
- [21] B. van Ginneken *et al.*, "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study," *Med. Image Anal.*, vol. 14, no. 6, pp. 707–722, 2010.
- [22] S. G. Armato, III, *et al.*, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, 2011.
- [23] S. Wang *et al.*, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Med. Image Anal.*, vol. 40, pp. 172–183, Aug. 2017.
- [24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.
- [26] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Neural Inf. Process. Syst.*, 2016, pp. 379–387.
- [27] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [29] W.-J. Choi and T.-S. Choi, "Automated pulmonary nodule detection system in computed tomography images: A hierarchical block classification approach," *Entropy*, vol. 15, no. 2, pp. 507–523, 2013.
- [30] Y. Tsuchiya, Y. Kodera, R. Tanaka, and S. Sanada, "Quantitative kinetic analysis of lung nodules using the temporal subtraction technique in dynamic chest radiographies performed with a flat panel detector," *J. Digit. Imag.*, vol. 22, no. 2, pp. 126–135, 2009.
- [31] Q. Dou, H. Chen, Y. Jin, H. Lin, J. Qin, and P.-A. Heng, "Automated pulmonary nodule detection via 3D convnets with online sample filtering and hybrid-loss residual learning," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2017, pp. 630–638.
- [32] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2017, pp. 559–567.
- [33] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *Proc. Conf. Comput. Robot. Vis.*, 2015, pp. 133–138.
- [34] D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, "Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models," *Comput. Biol. Med.*, vol. 42, no. 11, pp. 1098–1109, 2012.
- [35] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.
- [36] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.
- [37] P. F. Jaeger *et al.*, "Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," 2018, *arXiv:1811.08661*. [Online]. Available: <https://arxiv.org/abs/1811.08661>
- [38] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.