Hierarchical Deep-Fusion Learning Framework for Lung Nodule Classification

Kazim Sekeroglu[1,a], Omer Muhammet Soysal[1,a], and Xin Li[2]

[1] Southeastern Louisiana University, Hammond LA 70402, USA
[2] Louisiana State University, Baton Rouge, LA 70803, USA
([a] Authors contributed equally)
{kazim.sekeroglu,omer.soysal}@selu.edu
xinli@lsu.edu

Abstract. Lung cancer is the leading cancer type that causes the mortality in both men and women. Computer aided detection (CAD) and diagnosis systems can play a very important role for helping the physicians in cancer treatments. This study proposes a hierarchical fusion based deep learning scheme in a CAD framework for detection of nodules from computed tomography (CT) scans. In the proposed hierarchical approach, a decision is made at each level individually employing the decisions from the previous level. Further, individual decisions are computed for several perspectives of a volume of interest. The proposed model learns the parameters by means of supervised learning. The proposed CAD framework is tested using lung CT scans provided by the LIDC/IDRI database. The experimental results showed that the proposed scheme in the CAD framework significantly improves the performance of lung nodule detection.

Keywords: Computer aided detection, lung cancer, deep learning, hierarchical learning, hierarchical fusion, convolutional neural networks, modular training, and modular learning.

1 Introduction

Although the lung cancer is the second most commonly diagnosed cancer in both men and women, it is the leading cancer type that causes the mortality in both men and women [1]. Lung nodule detection is a very challenging task. The research team in [2] explored the effect of the low-dose CT scans in cancer mortality. Utilizing either low-dose CT or chest radiography, they screened around 53K high lung cancer risk patients three times a year between August 2002 and April 2004. The results of their study show that there is a 20% reduction in mortality of the patients who were screened by low-dose CT scan. Even though the CT scans helps to reduce the mortality rate, the radiologists' decision may differ significantly in identification of the lung nodules from the CT scans. As an example, [3] shared the results of two radiologists examination over 25 CT scans; the results show that one of the radiologist detected 20 nodules, whereas the other radiologist detected 63 nodules from the same CT scans.

A CAD system increases the performance of the nodule detection substantially. The study conducted by [4] showed that the CAD system reduced significantly the number of false positives (FPs). The research in [5] that studied the effect of a CAD system in detection of small nodules shared the results of 6 radiologists examination over 52 CT scans with/without a CAD system. The results show that the CAD system improves a radiologist's performance considerably. In [6], the performance of the commercial CAD software Lung-CAD VB10A and Siemens AG Healthcare were compared with the performance of two independent readers for detecting the pulmonary nodules in NELSON dataset. The study showed that sensitivity of CAD was 96.7% with a 3.7 FPs/scan and sensitivity of double reader was 78.3% with 0.5 FPs/scan. Therefore, CAD system with a higher nodule detection rate can be a good help for radiologist to decrease the number of missed nodules, particularly, the small nodules in their early stages.

In this study, we propose a hierarchical deep-fusion learning method utilizing multiple views of 3D spatial data. Contributions of this study can be summarized as a) utilizing deep learning in a multi-view hierarchical decision-making scheme, b) proposing supervised learning based fusion method to be used in this hierarchical scheme, and c) introducing a modular training approach for the hierarchical scheme.

Rest of the paper is organized as follows: In the following section, the previous work on lung nodule detection is provided. The second chapter introduces the proposed multi-perspective hierarchical deep-fusion learning model. In chapter three, data preparation, experimental results and discussions are provided. Finally, conclusions with the feature directions of the proposed research is covered in chapter four.

1.1 Previous Work

Computer-aided detection and diagnosis (CAD) systems have been studied for decades to get more accurate detection and decrease the work load on the radiologists. A complete computer-aided detection and diagnosis algorithms are usually composed of three main blocks: 1) Detection of the nodule candidates, 2) extraction of the features from the nodule candidates, and 3) false positive reduction and classification. Different approaches are used for the detection of the nodule candidates based on 2D or 3D segmentation. Since the intensity value of the nodule and the other structures in the lung region differ from each other, most of the segmentation methods are based on gray level thresholding. After segmenting out the nodule candidates, the next step is extracting the robust features for classification. The most common features extracted from the nodules are shape and texture based features. Once the features are extracted from the nodule candidates, to reduce the false positives, one of the classification methods such as k-nearest neighbor, support vector machine, linear discriminant, or random forest classifier is used.

The state of the art computer vision methods for object detection is based on deep learning methods. Therefore, there are existing CAD algorithms for pulmonary nodule detection which are based on the deep learning methods such as convolutional neural networks, deep belief networks, and autoencoders. One of the earliest study that uses deep learning system for lung nodule classification is [7]. In [7], classification of the pulmonary nodules of being malignant or benign by using deep learning methods were

explored. Specifically, the deep belief network (DBN) and convolutional neural network (CNN) models were tested. This is one of the first study that explores the application of deep learning techniques for the classification of pulmonary nodules. LIDC-IDRI dataset that includes 1010 scans and 2545 nodules which are greater than 3mm is used for testing the proposed methods in [7]. For the comparison of deep learning methods and the feature based methods, two of the well performing features SIFT and local binary pattern (LBP) features with k-NN classifier is used. DBN was able to classify pulmonary nodules with 82.2% sensitivity and the SIFT+LBP feature based classifier reached the sensitivity of 66.8%. Another earlier study for classifying the pulmonary nodules as malignant or benign is [8]. The classification is done by using the deep features extracted from 2D images by the autoencoder and classified by the binary decision tree. Publicly available LIDC/IDRI dataset is used to train and test the algorithm. Although there are 1010 CT scans available in LIDC/IDRI dataset, only 157 scans have the proper annotation for the nodules for being benign or malignant. The proposed method achieved the sensitivity of 83.35% with 0.39 FPs/scan over a 10-fold cross validation.

In [9] 3D convolutional neural network based lung nodule classification algorithm is proposed. Authors are claiming that the proposed method can work with weakly labeled 3D data as in the case of only the label of the central voxel and the size of the largest expected nodule are provided. Once they estimate the labels of the 3D training data by using basic thresholding and simple linear iterative clustering (SLIC) [10] super-pixels of the 2D slices, they use these data to train 3D CNN for the nodule classification. The negative samples are extracted from the lung area by randomly sampling the locations based on the threshold. SPIE-AAPM-LUNGx dataset is used to train and test the proposed method. The dataset contains 70 CT scans. 15K positive and 20K negatives samples are labeled by the proposed method. The proposed method achieved 80% sensitivity with 10 FPs/scan.

One of the latest study for lung nodule detection using deep learning methods is done in [11]. Authors are proposing a multi-view CNN for lung nodule detection. In the proposed method, they are extracting the volume of interest as a cube. Then 2-D patches from nine symmetrical perspectives of the extracted volume are fed into a separate CNNs. The outputs of the CNNs are fused based on different architecture. First fusion structure is committee fusion where the fusion is done at decision level. Once the class scores from each CNN is computed, they are fused using a product rule on the output probabilities [12]. Another fusion method is late fusion where the fusion is done at feature level by concatenating the outputs of the first fully connected layers. Lastly, they are using mixed fusion which is the combination of the committee and the later fusion. Although this proposed method is fusing the slices from multi-view, they are using a single slice from each view and the way the fusion is done similar to the previously proposed fusion approaches.

The researcher in [13], [14] and [15] reported that the fusion-based models increase performance of classification. In general, averaging, multiplication, or voting schemes are employed as a fusion method in deep learning [15], [16], [17], and [18]. These strategies utilize a simple approach to reach a final decision out of multiple predictions. Among few studies that explore the deep learning for lung nodule classification, the only method that uses multiple perspectives of a volume is proposed in [14]. However, the method utilizes only a single slice from each perspective and has only one level of

simple fusion. In contrast, this research proposes new kind of ensemble learning strategy "hierarchical fusion-learning" that aims to learn gradually from in-parallel and prior predictions obtained from different views.

2 Multi-Perspective Hierarchical Deep-Fusion Learning Model

In this paper, we propose a hierarchical deep-fusion learning scheme. In the proposed method, there are three levels of hierarchical predictions: 1) Slice level, 2) perspective level, and 3) volume level. We employed three different perspectives of transverse, coronal, and sagittal as shown in Fig. 1. The block diagram of the proposed hierarchical deep-fusion learning scheme is illustrated in Fig. 2. In the proposed method, each module at each level are trained separately in a hierarchical modular fashion; that is, the decision made at each level is predicted based on the decision from the previous layer. Proposed hierarchical learning process is as follows: Once a volume of interest is extracted from an object, a stack of 2D slices for each perspective is generated; the same VOI is represented by three sets of 2D slices, namely V_1, V_2, and V_3. At the first hierarchical level, a decision is produced for each slice of its perspective by a Slice Module (MS). Therefore, there are three Slice Modules MS_i at the first hierarchical level, one per perspective V_i. At the following second level, another learning module, namely Perspective Module (MP), is employed. An MP_i reformats the class scores obtained for each slice of V_i to form its input feature vector. As an example, assuming each stack V_i has 10 slices, the MS_i produces 10 class scores. These scores are formed into an input feature vector of size 10 by 1 for MP_i. At the second level, each MP_i predicts a single class score for each V_i. Similarly, the output scores of MPs are reformatted to form an input feature vector of size 3 by 1 for the last level module, namely Volume Module (MV), of the hierarchical scheme. The MV computes the final decision for the volume of interest (VOI).

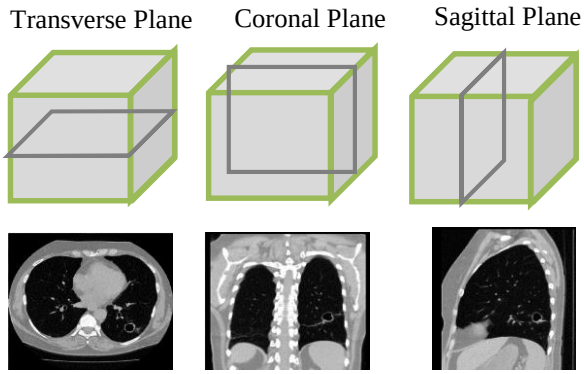


Fig. 1. Slices from three different perspectives

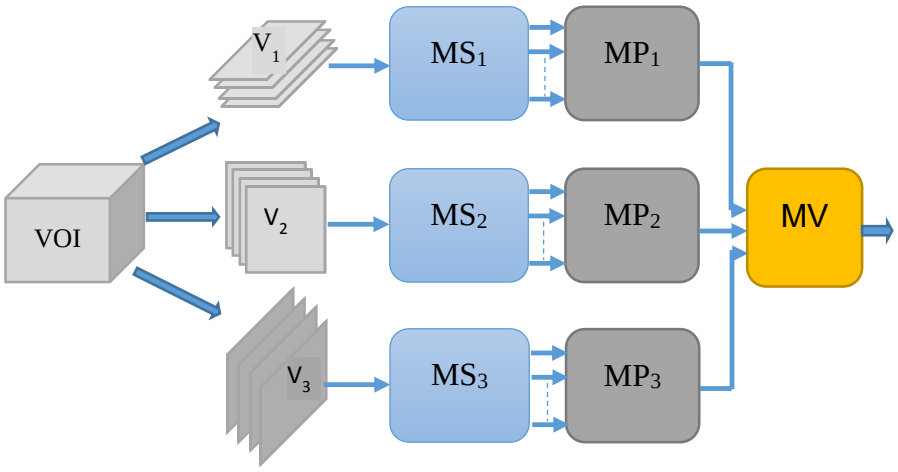


Fig. 2. Block diagram of the proposed model.

In the proposed hierarchical deep learning scheme, a Slice Module MS_i is of the type Deep Convolutional Neural Network (DCNN) as illustrated in Fig. 3. Although the structure of each MS_i is the same, they are trained separately. Its DCNN structure is consists of four convolutional and four pooling layers following with the regular one-hidden layer feed-forward neural network. In this structure, the input size of each 2D slice is 56x56 pixels. At the first convolution layer, there are 8 filters in the size of 3x3. The number of filters at the second, third, and the fourth convolutional layers are the double of the number of filters at their previous convolutional layer. Hence, at the last convolutional layer, there are 64 filters. After the last pooling layer, there is a fully connected layer comprised of 32 number of elements. At the convolutional and fully connected layers rectified linear function defined by (1) and at the output layer softmax function defined by (2) are used. The filters at each convolution layer are adjusted by back-propagating the error obtained at the output based on the cross-entropy loss function defined by (3).

$$f(x) = \max(0, x) \tag{1}$$

$$\sigma(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}} \tag{2}$$

where K is the total number of neurons in the layer, and j is the index of the neuron at the output layer.

$$L = -\sum_j t_j \log(p_j) \tag{3}$$

where t is the target and p is the predicted values at the output layer, and j is the index of the neuron at the output layer.

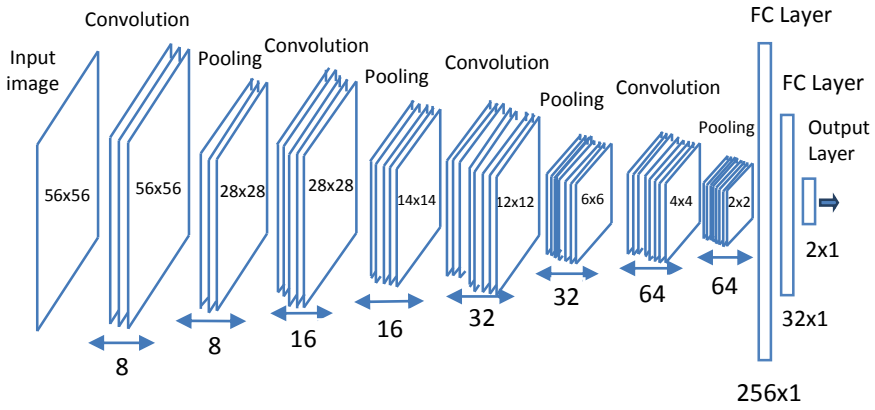


Fig. 3. DCNN structure.

In the perspective module and the volume module a supervised classifier such as Support Vector Machine, ANN, Bayesian Network, or a multi-dimensional regression model can be used. In this study, a regular feedforward ANN is used for the perspective and the volume level predictions.

3 Experiments and Results

3.1 Data Preparation

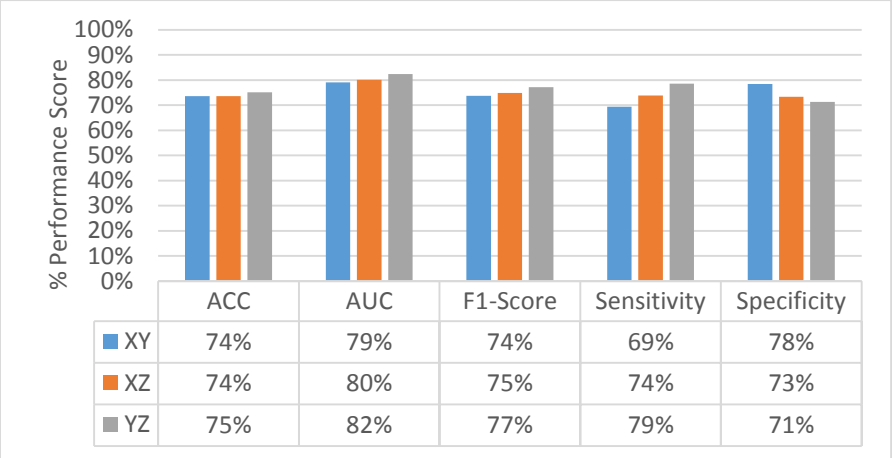
Publicly available lung CT scan database created by the Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) [19] is used to test the proposed CAD framework. LIDC/IDRI database contains 1010 CT scans which has the annotations for the nodules and the non-nodules whose diameter ≥ 3mm. Annotations made by the radiologists belong to one of these three groups; nodule ≥ 3mm, nodule ≤ 3mm, or non-nodule ≥ 3mm. The annotations of the CT scans are accomplished by 4 expert radiologists in 2 phases, blinded-read phase and unblinded-read phase. In the initial blinded-read phase, each of the radiologists examined the scans independently without knowing the opinion of the others. In the second unblinded-read phase, they examined the CT scans while knowing the annotations made by 3 other radiologists. While the surrounding boundary for the nodules ≥ 3mm are annotated, the nodules ≤ 3mm or non-nodules ≥ 3mm are annotated by only their volume center.

During the anotation process, each radiologist individually marks the surrounding boundary of the nodules ≥ 3mm so that the volume center of the same annotated nodule most probably differs from one radiologist to another. Therefore, as an initial step of extraction of VOI, the average volume center of each annotated nodule is computed based on the provided annotations by each radiologist. If the center coordinates of the annotated nodules are closer than the threshold, they are assumed to be the same nodule. Hence, at the next step, the average volume center for each nodule with the number of

radiologists' approval is found. Similar approach is used for detecting the average volume center and the number of radiologists' approval for the non-nodules ≥ 3mm. There is a possibility that some of the objects might be annotated as nodule by one radiologist and non-nodule by the other(s) or vice versa. To overcome this problem, once the average volume centers are computed for nodules and non-nodules, if the volume centers of nodules and the non-nodules are closer than the threshold, they are eliminated from the dataset. Once the volume centers of the objects are determined, 30x30x30 mm^3 region around the volume center is extracted as the volume of interest. The reason for using 30x30x30 mm^3 bounding cube is due to the longest axis of the annotated largest nodule in the dataset can be 30mm as provided in [19]. In LIDC-IDRI dataset, CT scans are collected from different CT scanners. Although all slices from all scans are of 512x512 pixels, the physical size of a single pixel is not the same for all scans. Thus, 30x30x30 mm^3 bounding cube corresponds to different size of pixel resolution. However, the input data for training and testing the proposed model should be the same size. Therefore, all extracted 30x30x30 mm^3 are normalized to the maximum resolution of 56x56x56 pixels. 100 CT scans from LIDC/IDRI database is used to create a dataset to test the proposed method. The dataset is balanced and it is composed of the 604 nodule and non-nodule objects that are approved by at least one radiologist. Dataset is split into 2 parts, 70% for the training and 30% for the testing. Therefore, training data has total of 422 nodules and non-nodules, and the testing data has total of 182 nodules and non-nodules.

3.2 Results and Discussion

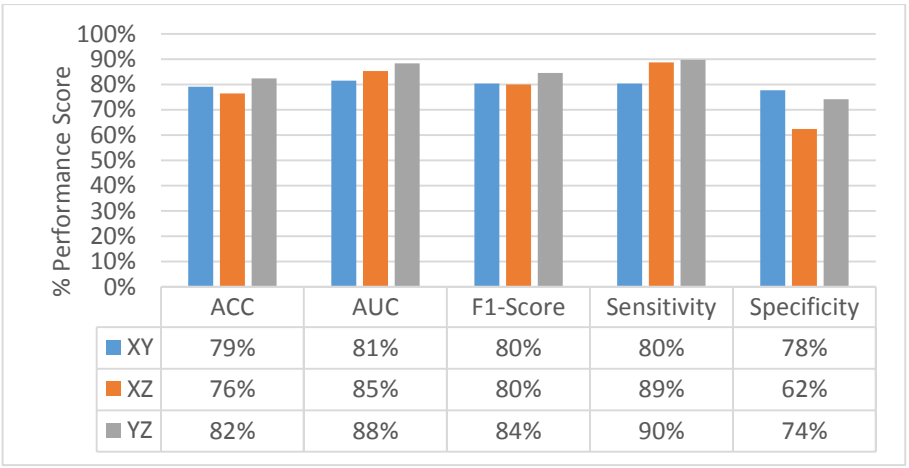
Classification performace of each module in the proposed hierarchical learning scheme as well as the classification performance across different perspectives are experimented. In the experimental results section, XY-perspective refers to transverse plane, XZ-perspective refers to coronal plane, and YZ-perspective refers to sagittal plane. Fig. 4 shows the change of the slice level classification performances across different perspectives. Although the slices from YZ-perspective gives the highest ACC, AUC, F1-score and sensitivity, specificity of the model created using slices from YZ-perspective is the smallest. On the other hand, ACC, AUC, F1-score and sensitivity of the model created using XY-slices is smallest among all 3 models. However, the specificity of the model created using XY-slices is highest among all 3 models. These results also can be interpreted as the model uses the slices from YZ-perspective has higher tendency towards type-I error and has higher FP. Alternatively, the model uses the slices from XY-perspective has higher tendency towards type-II error and has higher FN.



	ACC	AUC	F1-Score	Sensitivity	Specificity
XY	74%	79%	74%	69%	78%
XZ	74%	80%	75%	74%	73%
YZ	75%	82%	77%	79%	71%

Fig. 4. Change of slice level classification performances across different perspectives.

The classification performance of the second level, perspective level, of the proposed hierarchical learning model is provided in Fig. 5. At the perspective level classification, while the model uses slices from XY-perspective still has the lowest type-I error and the highest type-II error, the tendency toward type-I error of the model that uses YZ-perspectives is decreasing. At the perspective level classification, still the model which uses the slices from YZ-perspective has the highest performance score of ACC, AUC, F1-Score and sensitivity.



	ACC	AUC	F1-Score	Sensitivity	Specificity
XY	79%	81%	80%	80%	78%
XZ	76%	85%	80%	89%	62%
YZ	82%	88%	84%	90%	74%

Fig. 5. Change of perspective level classification performances across different perspectives

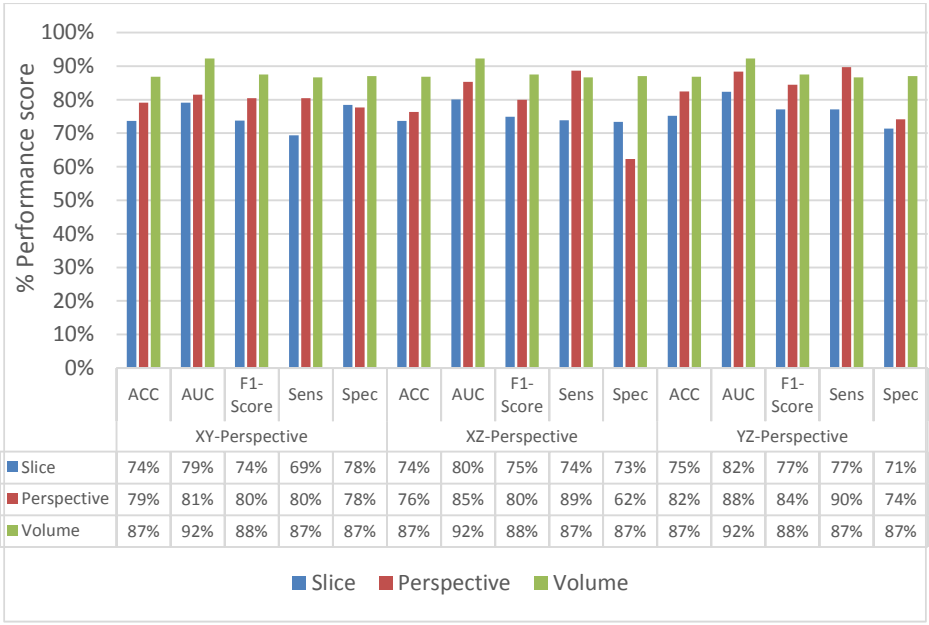
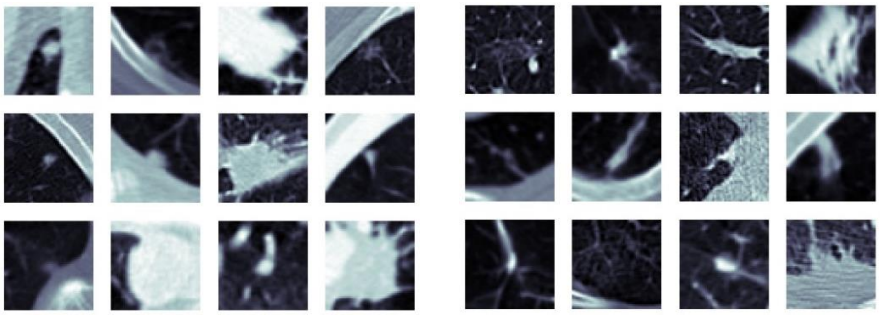


Fig. 6. Change of classification performance for slice, perspective and volume level classifications for each perspective.



(a) (b)

Fig. 7. (a) Missed nodules (FN) by the proposed method (b) Missed non-nodules (FP) by the proposed method.

As seen in Fig. 6, change of the classification performance for slice, perspective and volume level classifications, the classification performance increases at each of the proposed hierarchical learning model. Slices level classification gives the highest ACC as 75% using the slices from YZ-perspective. When the class scores from multiple slices are fused at the perspective level, the highest classification ACC is increasing from 75% to 82%. Finally adding another hierarchical fusion level, which fuses the class scores from all three perspectives, increases the highest classification ACC from

82% to 87%. Similarly, AUC, F1-score, sensitivity and specificity scores are also increasing from slice level classification to perspective level classification and volume level classification. At the volume level classification both the tendency toward type-I error and type-II error are the same while having the 87% sensitivity and specificity. Furthermore, the increase in the classification performance uisng hierarchical fusion can be seen in Fig. 8, Fig. 9, and Fig. 10 where the comparison of ROC curves for slice, perspective and volume level classifications for the slices from XY-perspective, XZ-perspective and YZ-perspectives are provided. In addition, the missed nodules (FN) and missed non-nodules (FP) by the proposed method are provided in Fig. 7.

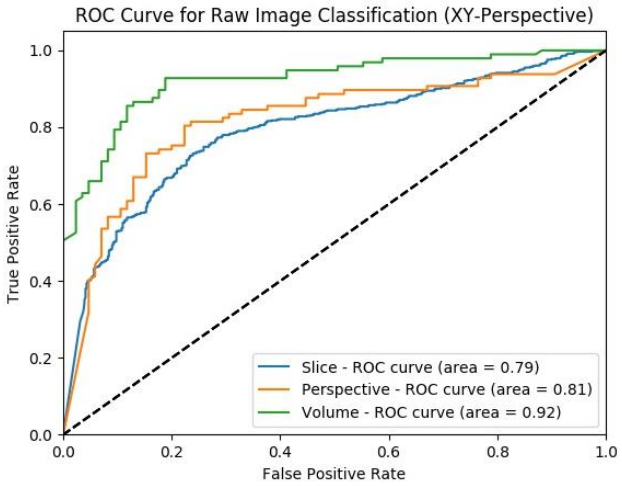


Fig. 8. Slice, perspective and volume level classification performance comparison on the ROC curve for the model created by using the slices from XY-perspective

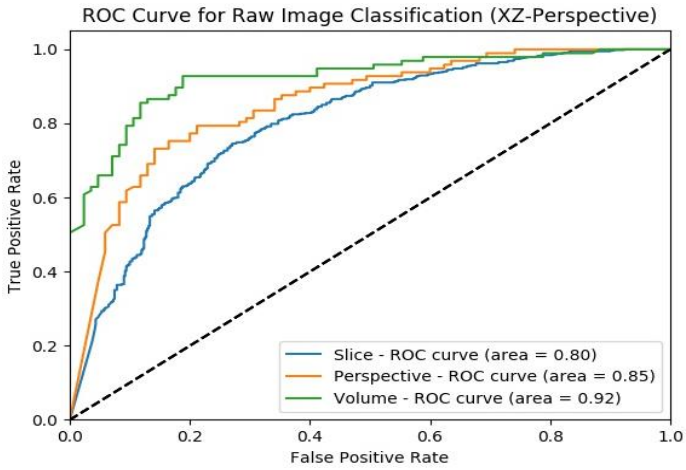


Fig. 9. Slice, perspective and volume level classification performance comparison on the ROC curve for the model created by using the slices from XZ-perspective

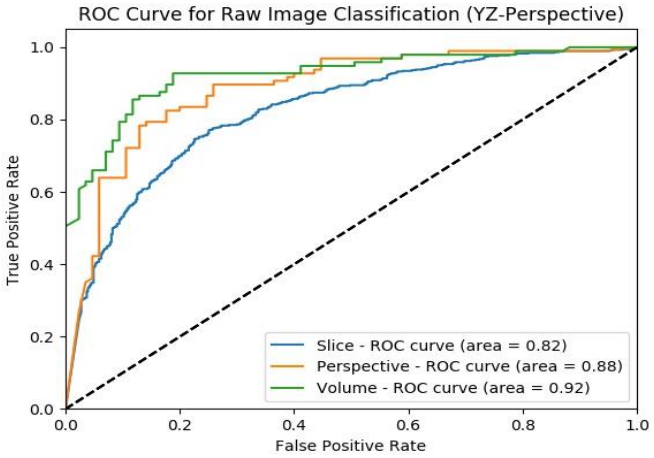


Fig. 10. Slice, perspective and volume level classification performance comparison on the ROC curve for the model created by using the slices from YZ-perspective

4 Conclusions

Lung cancer is the leading cancer type in terms of causing the mortality in both men and women. As reported in previous studies, screening the lung cancer using CT scans is very common and effective method. However, detecting pulmonary nodules in CT scans is a very challenging problem, particularly for the nodules in their early stages. CAD systems can be used by the radiologists during the examination of CT scans to increase the nodule detection rate as well as to decrease the false positives.

In this research, a multi-perspective hierarchical fusion based deep learning scheme is proposed for lung nodule detection from CT scans. The proposed method employs three level of multi-perspective hierarchical fusion of deep learning models. Our proposed fusion method is based on supervised learning, and the models are trained in a modular fashion such that each model is trained individually and the output of one model is utilized as an input for the next model.

To test the classification performance of the proposed method, total of 604 nodule and non-nodule objects are extracted from 100 CT scans from LIDC/IDRI database, and 70% of the data is used for training and 30% of the data is used for testing the proposed model. Experimental results show that the proposed hierarchical fusion based deep learning model achieved the ACC of 74% and AUC of 81% at the first, slice level, classification. At the second level, perspective level, classification, ACC increased to 79% and AUC increased to 85%. Finally, at the last level, volume level, classification, ACC of 87% and AUC of 92% is achieved. As seen from the results, proposed multi-perspective hierarchical fusion approach increases the classification performance significantly from slice level to volume level.

As a future work, single-feature image and multi-feature image based hieararchical deep-fusion models will be explored. In contrast to the proposed model, feature image-based representation can eliminate redundancy which is naturally embedded in raw

images, where the variance among the object classes is expected to be higher. A feture image will be generated by a specific filter that emphasizes certain characteristics of an image content.

Acknowledgement. National Natural Science Foundation of China under Grant No. 61728206.

References

1. A. C. Society, "Cancer Facts & Figures 2016," American Cancer Society, Inc, Atlanta, 2016.
2. N. L. S. T. R. Team, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *N Engl J Med*, vol. 2011, no. 365, pp. 395--409, 2011.
3. Armato, S. G., Roberts, R. Y. a. Kocherginsky, M. a. Aberle, D. R. a. Kazerooni, E. A. a. MacMahon, H. a. v. Beek, E. J. a. Yankelevitz, D. a. McLennan, G. a. McNitt-Gray and M. F. a. others, "Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of "truth"," *Academic radiology*, vol. 16, no. 1, pp. 28-38, 2009.
4. Rubin, G. D. Lyo, J. K. Paik, D. S. Sherbondy, A. J. Chow, L. C. Leung, A. N. Mindelzun, R. S. Desmond, P. K. Zinck, S. E. Naidich and D. P. others, "Pulmonary nodules on multi detector row ct scans: Performance comparison of radiologists and computer-aided detection 1," *Radiology*, vol. 234, no. 1, pp. 274-283, 2005.
5. Sahiner, B. a. Chan, H.-P. a. Hadjiiski, L. M. a. Cascade, P. N. a. Kazerooni, E. A. a. Chughtai, A. R. a. Poopat, C. a. Song, T. a. Frank, L. a. Stojanovska and J. a. others, "Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: analysis of an observer performance study by nodule size," *Academic radiology*, vol. 16, no. 12, pp. 1518-1530, 2009.
6. Zhao, Y. a. d. Bock, G. H. a. Vliegenthart, R. a. v. Klaveren, R. J. a. Wang, Y. a. Bogoni, L. a. d. Jong, P. A. a. Mali, W. P. a. v. Ooijen, P. M. a. Oudkerk and Matthijs, "Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume," *European radiology*, vol. 22, no. 10, pp. 2076-2084, 2012.
7. Hua, K. L. a. Hsu, C. H. a. Hidayati, S. C. a. Cheng, W. H. a. Chen and Y. Jen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets and therapy*, vol. 8, 2015.
8. Kumar, D. a. Wong, A. a. Clausi and D. A, "Lung nodule classification using deep features in ct images," in *Computer and Robot Vision (CRV), 2015 12th Conference on*, Halifax, NS, 2015.

9. Anirudh, R. a. Thiagarajan, J. J. a. Bremer, T. a. Kim and Hyojin, "Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data," in *SPIE Medical Imaging*, 2016.
10. R. e. a. Achanta, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274-2282., 2012.
11. Setio, A. A. A. a. Ciompi, F. a. Litjens, G. a. Gerke, P. a. Jacobs, C. a. v. Riel, S. J. a. Wille, M. M. W. a. Naqibullah, M. a. Sanchez, C. I. a. v. Ginneken and Bram, "Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1160--1169, 2016.
12. v. Ginneken, B. a. Setio, A. A. a. Jacobs, C. a. Ciompi and Francesco, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *IEEE*, 2015.
13. O. M. Soysal, J. Chen and H. Schneider, "Efficient photometric feature extraction in a hierarchical learning scheme for nodule detection," *International Journal of Granular Computing, Rough Sets and Intelligent Systems*, vol. 2, no. 4, pp. 314--326, 2012.
14. Setio, A. A. A. a. Ciompi, F. a. Litjens, G. a. Gerke, P. a. Jacobs, C. a. v. Riel, S. J. a. Wille, M. M. W. a. Naqibullah, M. a. Sanchez, C. I. a. v. Ginneken and Bram, "Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1160--1169, 2016.
15. Karpathy, A. a. Toderici, G. a. Shetty, S. a. Leung, T. a. Sukthankar, R. a. Fei-Fei and Li, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014.
16. B. K. Kim, J. Roh, S. Y. Dong and S. Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," *Journal on Multimodal User Interfaces*, vol. 10, no. 2, pp. 173-189, 2016.
17. K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in Neural Information Processing Systems*, 2014.
18. E. Park, X. Han and T. L. B. a. A. C. Berg, "Combining multiple sources of knowledge in deep CNNs for action recognition," in *EEE Winter Conference on Applications of Computer Vision* , Lake Placid, NY, 2016.
19. A. III, S. G. a. McLennan, G. a. Bidaut, L. a. M. Gray, M. F. a. Meyer, C. R. a. Reeves, A. P. a. Zhao, B. a. Aberle, D. R. a. Henschke, C. I. a. Hoffman and E. A. a. others, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical physics*, vol. 38, no. 2, pp. 915--931, 2011.