Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network

Asmaa A. Hassan
Mathematics Department of Computer Science
University of Assiut
Assiut, Egypt
asmaa.abbas@science.aun.edu.eg

Mohammed M. Abdelsamea
Mathematics Department of Computer Science, Assiut University, Assiut, Egypt and School of Computer Science, University of Nottingham, Nottingham, UK
m.abdelsamea@aun.edu.eg

***Abstract*—automated classification of tuberculosis in x-ray images is of an increasing interest to all researchers and physicians. Due to the high level of intensity inhomogeneity and variations, statistical machine-learning approaches usually fail to offer a generic solution to image classification. Deep learning approaches, especially convolution neural networks (CNNs), have demonstrated superior effectiveness in computer-aided diagnosis systems. Transfer learning can provide a powerful deep learning solutions to the limited availability of labelled images. In this paper we study the effect of knowledge transferred from a pre-trained ImageNet, in different ways via a pre-trained CNN model, to classify chest x-ray images as having manifestations of tuberculosis or as healthy. We evaluated and compared various models using the learning curve between training and validation set, and receiver operating characteristic (ROC) curve. Our experiments revealed that using fine-tuning technique outperformed both shallow-tuning and deep-tuning techniques and achieved 0.998 for the area under the ROC curve (AUC) with 0.999 for specificity, and 0.997 for sensitivity rate.**

Keywords—computer-aided diagnosis; medical imaging; deep learning; convolution neural network; transfer learning.

I. INTRODUCTION

The historical concept of Computer-aided diagnosis (CAD) for medical imaging has been widely explored through enormous number of approaches ranging from feature engineering [1, 2] to deep learning (DL) [3-6]. Convolutional neural networks (CNNs or ConvNets) are one of the most gained popular algorithms of DL in medical imaging, especially for image segmentation, classification, and recognition [7-9]. The discriminative power of CNN comes from its ability to learn automatically information from training images such that hierarchical level of global and local information can be encoded to make the analysis process much easier and accurate when compared to statistical machine learning frameworks.

Training a deep convolutional neural network can be fulfilled in different strategies. First it can be trained as an end-to-end network [10], where a large amount of labeled training data should be provided, which is impractical in the medical imaging domain. This also could lead to over-fitting due to the high number of parameters that should be learnt. The second strategy is to transfer knowledge that a pre-trained CNN (that has been learned from a huge benchmarked dataset) to a new task to deal with the limited availability of labeled data [8, 11]. Such kind of learning (which is our main focus in this work) can be further accomplished by different learning scenarios: First, shallow-tuning, which deals with a pre-trained model as a fixed feature extractor for the new dataset by only adopting the classification layer. In this case the model can provide an efficient solution (where no learning is required) but the feature maps might not accurately represent the distribution of the different classes. Second, "deep-tuning", which aims to retrain all the convolutional layers in a pertained network using labelled images for the new task, in an end-to-end fashion. Deep-tuning technique usually shares the same drawbacks of deep networks in dealing with the over. Third, "fine-tuning", which works by learning feature maps of some layers by adjusting learning rate accordingly. This is by starting update feature maps of the last layer and repeat this step (by including one more layer each time) until the best performance is achieved. This learning strategy may or may not provide a better performance compare to the other strategies but it can help in better understanding the effect of each layer on the classification performance.

In this paper we aim to study the effect of knowledge transformation from ImageNet using a pre-trained model such as AlexNet [12-16] on chest x-ray (CXR) images for the purpose of identification of manifestation of tuberculosis. Our study aims to provide an answer to an interesting question of what is the best tuning mode for a pre-trained CNN model that can accurately classify CXR images. The overall system can provide a robust and important initial step for more complex task such as object segmentation as a part of a prognosis system.

II. RELATED WORK

Several statistical and deep learning approaches have been extensively proposed for lung diseases classification. For example, Kuruvilla et al. [17] applied a statistical machine learning to classify the lung regions from computed tomography (CT) into normal or cancerous nodule, using six features and thirteen training functions of back-propagation

neural network (BNN). In [18], a support vector machine (SVM) classifier was used with grey-level co-occurrence matrix (GLCM) texture features to the classification of CT lung images into benign or malignant. Qing Li et al. [8] used a shallow CNN approach with a single convolutional layer and three fully connected layers to classify the lung image patches of CT images into five different classes. A patch-based CNN approach [19] has been used as a feature extractor for the identification of abnormalities in lung nodules of CT images, where different scales of the nodule patches have been used for two convolutional layers and two pooling layers to create a multi-scale features. Moreover, a patch-based CNN approach [20] has been used to the identification of interstitial lung disease (ILD) in 2D CT lung images using handcrafted features such as Local Binary Pattern (LBP). Anthimopoulos et al. [21] used a patch-based CNN approach to classify a lung tissue into normal and other six classes. Unlike patch-based CNNs, M. Gao et al. [16] used the whole lung image with different scales to feed a CNN to the classification of a lung image into six different classes to overcome the limitations of traditional patch-based methods in differentiating between diseased and healthy lung.

In our work, we study in details, different scenarios of transferring knowledge from a common image recognition task using a common pre-trained model to classify CXR images into normal and abnormal. We apply different learning schemas to adjust the weights of the classification framework to investigate the effect of different feature maps on the classification of a particular image dataset.

III. METHODS AND MATERIAL

The architecture of our classification model is consisting of 3 stages (as illustrated in Fig. 1). First, a pre-processing and data augmentation techniques have been used to improve the contrast and quality of tissue surface, and to increase the number of training images. Second, knowledge from image recognition task (where AlexNet has already learned their feature maps) is transformed to our binary classification tasks as an initialization step. Third, three different learning techniques (e.g., shallow, deep, and fine tunings) are used to adjust the weights/features to investigate and evaluate the performance of the classification model.

A. Dataset

The dataset contains CXR radiograph, which are acquired from the Japanese Society of Radiological Technology (JSRT) [22, 23]. The dataset has 138 CXR images, divided into 58 cases with manifestation of tuberculosis, and 80 normal cases. All x-ray images have a size of 4020 ×4892 or 4892 × 4020 pixels. The pixel spacing in vertical and horizontal direction is 0.0875 mm, and number of gray levels is 12bits. We divide the dataset into three groups: training, validation and testing sets. We use the training set to train our models. Validation set is used to help in detecting the over-fitting and under-fitting problems, and finding out the optimal parameter settings. Finally, a test set of unseen images are used to evaluate the performance of the different models.

B. Image preprocessing and data augmentation

We applied the histogram equalization technique to normalize/standardize the intensity profiles of the images. Also all images have been down-sampled into 256×256 pixels to accelerate the training process. Moreover, data augmentation based on a different deformation/transformation techniques have been used to expand our dataset. This including flipping, transition, random small noise, rotation with different angles, and color jittering. This phase is resulting in a total of 60.000 lung regions with 256×256 pixels.

C. Model architecture and training

We used AlexNet [12] architecture which consist of five layers of convolutional layers, three different sizes of basis or kernel filters (i.e., 11 × 11, 5 × 5 and 3 × 3) and three fully connected (FC) layers as described in Table I. The convolution layers used to learn the basics patterns in images and to extract different useful features through a hierarchical process. We changed the output classification into 2 classes to be adapted with our problem case (normal and abnormal).

We use the following cross entropy loss function

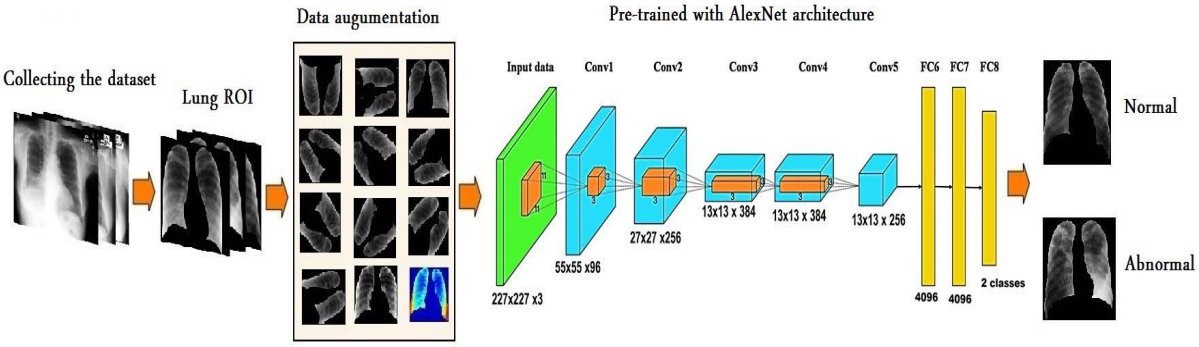


Fig. 1: The architecture of our framework with CNN.

TABLE I: describe the AlexNet architecture used in our experiments

Layer name	Type	Input dimension	Kernel size	Stride	Pad	Output dimension
data	Input data	227×227×3	N/A	N/A	N/A	227×227×3
conv1	Convolution layer	227×227×3	11×11	4	0	55×55×96
Pool 1	Max pooling	55×55×96	3×3	2	0	27×27×96
conv2	Convolution layer	27×27×96	5×5	1	2	27×27×256
Pool 2	Max pooling	27×27×256	3×3	2	0	13×13×256
conv3	Convolution layer	13×13×256	3×3	1	1	13×13×384
conv4	Convolution layer	13×13×384	3×3	1	1	13×13×384
conv5	Convolution layer	13×13×384	3×3	1	1	13×13×256
Pool 5	Max pooling	13×13×256	3×3	2	0	6×6×256
fc6	Fully connected layer	6×6×256	6×6	1	0	4096×1
fc7	Fully connected layer	4096×1	1×1	1	0	4096×1
fc8	Fully connected layer	4096×1	1×1	1	0	2×1

$$\mathrm{F(X,Y)} = \min\left[-\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=0}^{\mathrm{n}} \mathrm{y^i} \ln \mathrm{a(x^i)} + (1-\mathrm{y^i}) \ln\left(1 - \mathrm{a(x^i)}\right)\right] \quad (1)$$

where $\chi = \{x^1, x^2, \dots, x^n\}$ is the set of input images in the training dataset, y = $\{y^1, y^2, \dots, y^n\}$ is the desired output labels for those input samples, and the notation a is the predicted output from a softmax function. We also used *L2* regularization as a part of the total loss function to penalize weight coefficients with large magnitudes and decay every weight toward zero. *L2* regularization can be defined as

$$\mathcal{R} = arg\ min\left(\mathrm{F}(X,Y) + \frac{\mathcal{L}}{2}\|\vec{\omega}\|^2\right) \quad (2)$$

where $L2 : \|\vec{\omega}\|^2 = \sum_{j=0}^{m} \omega_j^2$, $\frac{1}{2}$ factor for back-propagation updates and $\mathcal{L}$ is a parameter to control the regularization strength. To minimize the loss function, we calculated the derivative of Equation 2 as

$$\mathrm{SGD} = \nabla\, \mathrm{F(X,Y)} + \mathcal{L}\omega \quad (3)$$

The weight updating process can be described by the following equation

$$\omega_{t+1} = \omega_t - \eta(\nabla\ \mathrm{F}\left(\omega, X^{(i:i+n)}, Y^{(i:i+n)}\right) + \mathcal{L}\omega) \quad (4)$$

where η is the learning rate.

Finally, the training has been done separately to produce three models: i) shallow-tuned model that relies on features learned by AlexNet directly to our classification task, where no training is accomplished; ii) deep-tuned model that updates all feature maps in an end-to-end way; and iii) fine-tuned model, which allows us to incrementally learn and update feature maps starting from high-level layer and going deeper to the low-level layer (by only considering one layer each time) until the best performance is achieved.

IV. Experiment Results

The dataset was randomly divided into three groups as follows: 70% for training, 20% for validation, and 10% as a test set. In this work we used a trial-and-error method for hyper-parameters settings of the three DL approaches using the validation set. As a consequence, all values have been fixed for the hyper-parameters settings in the testing stage, accordingly. In more details, we fix the moderate learning rate for all the CNN layers to 0.001, the last fully connected layer accelerated with 0.01, min batch size was 128, and momentum term was fixed to 0.9. The learning rate has been decreased with value 0.9 every 3 epochs for deep- and fine- tuned models. The mini batch of stochastic gradient descent (SGD) was also used. The value of 0.001 for the weight decay was applied to capture the overfitting and to help in finding the global optima solution. For model performance evaluation, we adopted the sensitivity and specificity metrics

$$\text{sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP+FN}},$$

$$\text{specificity} = \frac{\mathrm{TN}}{\mathrm{TN+FP}},$$

where TP, TN are the correct model predictions for normal and abnormal cases, respectively. FP, FN is incorrect model predictions for normal and abnormal cases, respectively. In Table II, we report the confusion matrix results of training the CNN-based models and the validation error at each layer of fine-tuned model is also reported.

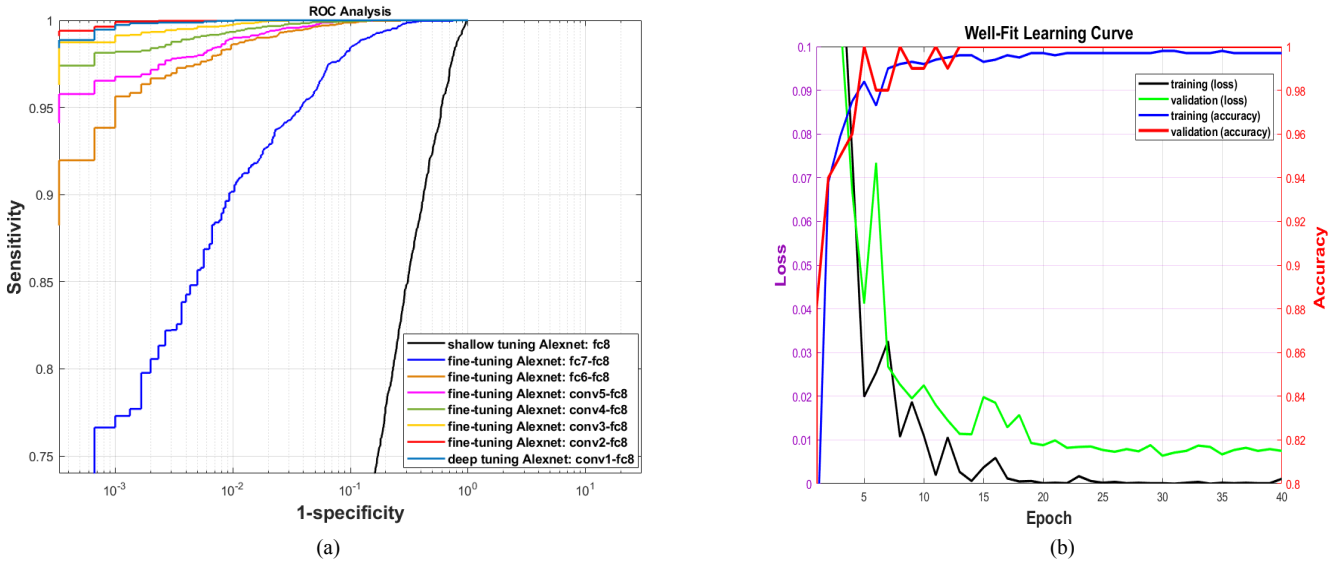
TABLE II: the confusion matrix for normal and abnormal instances in each learning layers, and the hyper parameters we used. η is the learning rate of training the weights for all CNN layers except the last fc(8) has the notation $fc(8)_{\eta}$, γ determines the drop of η over epochs, MB is the size of training samples, $\mathcal{L}$ the regularization strength and μ is the momentum speed .

Hyper parameters	**CNN Training Techniques**	**Predicted Class**		**True Class**		**Validation Error**
		Normal TP	***Abnormal TN***	***Abnormal FP***	***Normal FN***	
$\eta = 0.001$, $\gamma = 0.9$ $\mathcal{L} = 0.001$, MB = 128 $\mu = 0.9$, $fc(8)_{\eta} = 0.01$	(shallow)Fine-tuned AlexNet: only fc8	2636	2071	364	929	0.515
	Fine-tuned AlexNet:fc7-fc8	2959	2738	41	262	0.165
	Fine-tuned AlexNet:fc6-fc8	2976	2942	24	58	0.043
	Fine-tuned AlexNet:conv5-fc8	2976	2954	24	46	0.038
	Fine-tuned AlexNet:conv4-fc8	2988	2967	12	33	0.026
	Fine-tuned AlexNet:conv3-fc8	2990	2983	10	17	0.017
	Fine-tuned AlexNet:conv2-fc8	**2997**	**2992**	**3**	**8**	**0.011**
	(deep) Fine-tuned AlexNet:conv1-fc8	2996	2992	4	8	0.007

Table II demonstrates the pre-trained CNN when we only fine-tune the last layer (fc8), which is achieved the lowest sensitivity of 0.74, specificity of 0.85 and accuracy of 0.785. On the other hand, fine-tuned the layers (conv2: fc8) of the pre-trained CNN gives the highest accuracy of 0.998 with sensitivity of 0.997, and specificity of 0.999, which means that most of the testing images were correctly classified, and the well-fit learning error and accuracy curves between training and validation set shows high robustness without under-fitting or overfitting problems, as shown in Fig. 2(b).

We also plot the area under the ROC as illustrated in Fig. 2(a). Note how ideal the ROC curve is. We also observed that fine-tuned layers from conv2 until the last fully connected layer (fc8) behaves similar to deep-tuned (conv1: fc (8)). This observation comes from the fact that knowledge has been successfully transferred to describe low-level features such as edges, curves and blobs. The visualization of the first (conv1) layer and deep features of (conv5) layers is also provided in Fig.3.

We also investigated the effect of various training sizes on the classification performance by decreasing the volume of training data gradually. The learning curves of Fig. 4(a) and 4(b) confirm the improvements in the performance by increasing the size of training set.



(a) (b)

Fig. 2: ROC analysis for lung classification. (a) Comparison between CNN techniques by incremental fine-tuning, shallow tuning, and deep tuning approach. (b) The optimal learning curve error and accuracy obtained.

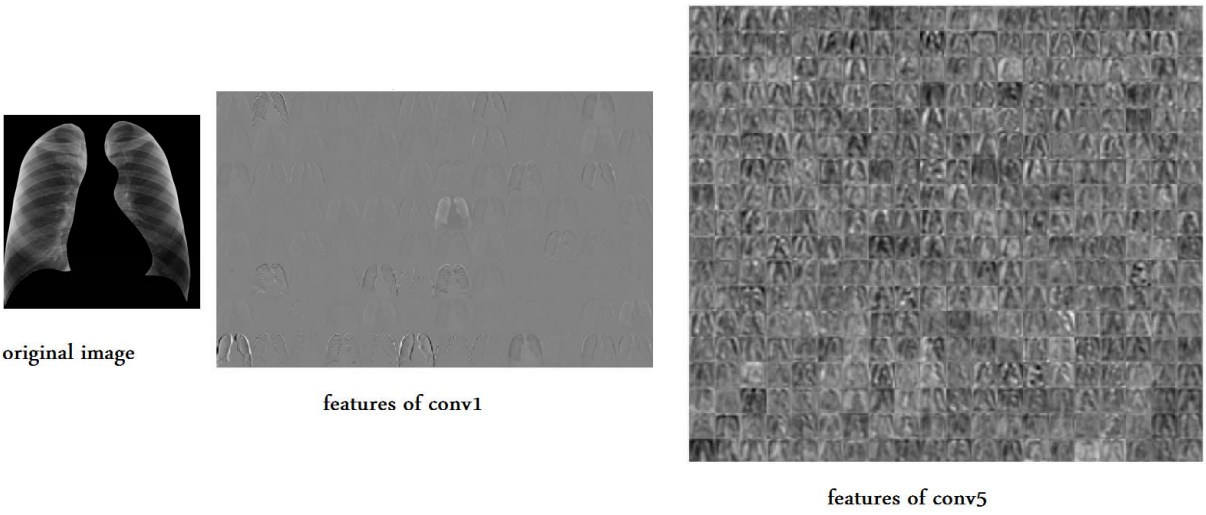


Fig. 3: Feature visualization of the first and the fifth convolution layer of an associated image.

V. CONCLUSIONS

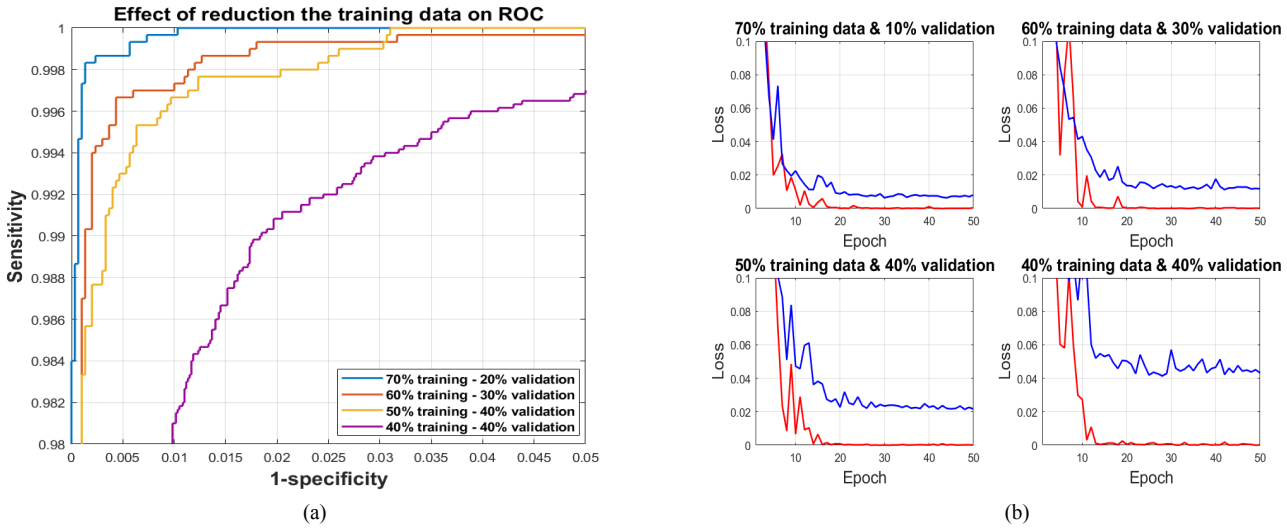
In this paper we studied the effect of learning different level of information based on knowledge transferred by AlexNet (when learned on ImageNet) to classify the lung region within CXR into two classes as normal or abnormal. We achieved state-of-the-art results in diagnosing lung region into normal and abnormal about AUC of 0.998with 0.999 for specificity, and 0.997 for sensitivity rate.

We conclude that a) shallow-tuned CNN does not encode accurate information and result in poor classification performance; and b) fine-tuned network has a slightly better performance than deep-tuned network. Moreover, fine-tuned CNN outperformed other classification models of [17, 18, 20] as shown in TABLE III.

This study can provide an important initial step for the development of an accurate multi-task learning framework for prognosis and diagnosis system.

ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro P5000GPU used for this research. We would like also to thank Dr. Sameer Antani for providing us with the images that have been used in this research.



(a) (b)

Fig. 4: (a) Comparison of receiver operating characteristic curves for different training data size. (b) Explained how the gap between training and validation set was increasing with the decreasing of training data size.

TABLE III: comparison between our result and different approaches for classification lung image

Method	AUC	specificity	sensitivity
BNN-based method [17]	93.3%	100%	91.4%
SVM-based method [18]	94 %	93 %	100%
Patch-based CNN [20]	93.33%	91.6%	94.4%
Fine-tuned based CNN	0.998	0.999	0.997

REFERENCES

1. M.N. Wernick, Y. Yang, J.G. Brankov, G. Yourganov, and S.C. Strother, "Machine learning in medical imaging." IEEE signal processing magazine, 2010. VoL:27: p. 25-38.
2. J. Shiraishi, Q. Li, D. Appelbaum, and K. Doi. "Computer-aided diagnosis and artificial intelligence in clinical imaging," in Seminars in nuclear medicine.Vol.41, p. 449-462, 2011.
3. L. Lu, Y. Zheng, G. Carneiro, and L. Yang, "Deep Learning and Convolutional Neural Networks for Medical Image Computing." 2017, Springer.
4. K. Suzuki, "Overview of deep learning in medical imaging." Radiological physics and technology, 2017. VoL:10: p. 257-273.
5. R.M. Summers, "Deep Learning and Computer-Aided Diagnosis for Medical Image Processing: A Personal Perspective," in Deep Learning and Convolutional Neural Networks for Medical Image Computing, 2017, Springer, p. 3-10.
6. D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, et al., "Deep learning for health informatics." IEEE journal of biomedical and health informatics, 2017. VoL:21: p. 4-21.
7. G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, et al., "A survey on deep learning in medical image analysis." arXiv preprint arXiv:1702.05747, 2017. VoL.
8. Q. Li, W. Cai, X. Wang, Y. Zhou, D.D. Feng, et al. "Medical image classification with convolutional neural network,"IEEE. p. 844-848, 2014,in Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on. 2014.
9. J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis." IEEE Access, 2018. VoL:6: p. 9375-9389.
10. Z. Wang, W. Yan, and T. Oates. "Time series classification from scratch with deep neural networks: A strong baseline,"IEEE. p. 1578-1585, 2017,in Neural Networks (IJCNN), 2017 International Joint Conference on. 2017.
11. X. Yin, W. Chen, X. Wu, and H. Yue. "Fine-tuning and visualization of Convolutional Neural Networks,"IEEE. p. 1310-1315, 2017,in Industrial Electronics and Applications (ICIEA), 2017 12th IEEE Conference on. 2017.
12. A. Krizhevsky, I. Sutskever, and G.E. Hinton. "Imagenet classification with deep convolutional neural networks." p. 1097-1105, 2012,in Advances in neural information processing systems. 2012.
13. N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, et al., "On the necessity of fine-tuned convolutional neural networks for medical imaging," in Deep Learning and Convolutional Neural Networks for Medical Image Computing, 2017, Springer, p. 181-193.
14. Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, et al. "Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally." p. 7340-7349, 2017,in IEEE conference on computer vision and pattern recognition, Hawaii. 2017.
15. N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?." IEEE transactions on medical imaging, 2016. VoL:35: p. 1299-1312.
16. M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, et al., "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks." Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2018. VoL:6: p. 1-6.
17. J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images." Computer methods and programs in biomedicine, 2014. VoL:113: p. 202-209.
18. T. Manikandan and N. Bharathi, "Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier." Journal of medical systems, 2016. VoL:40: p. 181.
19. W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian. "Multi-scale convolutional neural networks for lung nodule classification,"Springer. p. 588-599, 2015,in International Conference on Information Processing in Medical Imaging. 2015.
20. P. Hattikatti. "Texture based interstitial lung disease detection using convolutional neural network,"IEEE. p. 18-22, 2017,in Big Data, IoT and Data Science, 2017 International Conference on. 2017.
21. M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network." IEEE transactions on medical imaging, 2016. VoL:35: p. 1207-1216.
22. S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, et al., "Automatic tuberculosis screening using chest radiographs." IEEE transactions on medical imaging, 2014. VoL:33: p. 233-245.
23. S. Candemir, S. Jaeger, K. Palaniappan, J.P. Musco, R.K. Singh, et al., "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration." IEEE transactions on medical imaging, 2014. VoL:33: p. 577-590.