Int. J. Nonlinear Anal. Appl.
Volume 12, Issue 1, Winter and Spring 2021, 693-700
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2021.4876

Cystoscopic Image Classification by an Ensemble of VGG-Nets

Ehsan Kozegar[a,*]

[a]Faculty of Technology and Engineering (Eastern Guilan), University of Guilan, Guilan, Iran.

(Communicated by Madjid Eshaghi Gordji)

Abstract

Over the last three decades, artificial intelligence has attracted lots of attentions in medical diagnosis tasks. However, few studies have been presented to assist urologists to diagnose bladder cancer in spite of its high prevalence worldwide. In this paper, a new computer aided diagnosis system is proposed to classify four types of cystoscopic images including malignant masses, benign masses, blood in urine, and normal. The proposed classifier is an ensemble of a well-known type of convolutional neural networks (CNNs) called VGG-Net. To combine the VGG-Nets, bootstrap aggregating approach is used. The proposed ensemble classifier was evaluated on a dataset of 720 images. Based on the experiments, the presented method achieved an accuracy of 63% which outperforms base VGG-Nets and other competing methods.

Keywords: Cystoscopy, Classification, Deep Learning, Bootstrap Aggregating
MSC: 68T10.

1. Introduction

According to World Health Organization (WHO) statistics, bladder cancer is the ninth most prevalent cancers worldwide [17]. Similar to other diseases, early diagnosis is a key factor for successful treatment. Early diagnosis of dangerous diseases such as cancer can save thousands of lives annually. In recent years, various types of imaging modalities have been developed to assist experts. Since examination of images by human is potentially error-prone, complex and time consuming in challenging cases, presenting novel intelligent methods to classify medical images can improve the diagnostic accuracy.

*Corresponding author
Email address: Kozegar@guilan.ac.ir (Ehsan Kozegar)

Received: October 2020 *accepted:* February 2021

Urologists encounter with challenges in interpreting bladder cystoscopic images. For example, the diagnosis of benign and malignant tumors from cystoscopic images is very difficult. The doctor's opinions on the type of tumors differ before the pathology results. Moreover, the quality of cystoscopic images is not the same due to different imaging conditions and bladder environment. Also, these types of images are taken at different angles, from different areas of the bladder and in different sizes. In addition, the poor quality of images due to the imaging conditions and the lack of the same format of images have made the task more complicated. Generally, these factors make urologists have poor diagnosis rate on cystoscopic images and this poses many difficulties against its application. Given these issues, implementing a computer aided diagnosis system (CADx) as a second interpreter for classification of cystoscopic images is significant. In this paper, a computer-aided system is proposed for recognizing bladder diseases such as malignant masses, benign masses, and bloody urine. The proposed classifier is based on the convolutional neural networks (CNNs). The main advantage of CNNs is unsupervised learning of discriminative features.

In this paper, a new method is introduced to increase the generalizability of classification. In this regard, an ensemble of VGG16 classifiers is constructed using bootstrap aggregating strategy. To make the ensemble classifier, we focused only on the data level diversity. The rest of this paper is organized as follows: in section 2, we reviewed some state-of-the-art methods based on deep convolutional neural networks for medical image analysis. In section 3, the proposed ensemble classifier is described in details. In section 4, experimental results are presented, and section 5 concludes the paper.

2. Related works

From the 1970s to the 1990s, medical image analysis was performed using low-level pixel processing and mathematical modeling. At the end of the 1980s, supervised learning in which training data were used to train a system were considered in medical image analysis. Supervised techniques included active shape models, the concept of feature extraction, and the use of statistical pattern recognition. The pattern recognition or machine learning approach is still very popular and forms the basis of many commercially successful medical image analysis systems. Hence, there are changes in systems completely designed by humans in such a way that computer algorithms determine the optimal decision boundary in the high dimensional feature space. An important step in designing such systems is to extract distinctive features from the images. This process is still being carried out by human experts. The next step for computers is to extract features from problematic data. This concept is the basis of many deep learning algorithms. Models are made up of many layers that convert input data (e.g. image) into outputs (e.g. existing/absent disease), while learning has higher-level features. Nowadays, the most successful type of models for image analysis is the convolutional neural networks, which include many layers that modify their inputs with small-scale convolutional filters.

CNNs contain many layers that change their inputs with small-scale convolutional filters. Many techniques were common for feature learning before introducing AlexNet [2]. These techniques include principal component analysis, image fragment clustering, dictionary approaches and so on. In [2] the introduced CNNs are taught in a part known as Global Training of Deep Models. In [13] the medical image analysis has been briefly performed. In this study, the addressed papers have recently been investigated in a wide range of in-depth training applications in medical image analysis. The purpose of this study is to show that deep learning techniques have penetrated the whole field of medical image analysis [4]. Most deep architectures are based on neural networks [5,15] and can be considered as generalizing a linear regression. For a long time, DNNs were used briefly to effectively train and gain popularity only in [2,6], while it was shown that DNNs train the layer in an unsupervised manner. Then the adjustment of the network microstructure could lead to excellent pattern recognition tools. In [19], a 22-layer network called GoogLeNet was introduced. Their model uses so-called initial blocks. Initial blocks can be interpreted as a network in a network in which the input is divided into several sub- convolutional networks that connect to the end of the block. This structure is the best type of deep structure known as the ImageNet Challenge winner in 2015. The default CNN structure could be the source of multiple information or input profiles in the form of channels that enter the input layer. In principle, different channels can be integrated anywhere in the pipeline. Multi-flow structures are examined under intuition that different tasks require different methods of fusion. These models are sometimes referred to as the dual path structures [9]. Image analysis has been done at multiple scales for decades and is a concept under study. Several medical applications have also successfully utilized multiscale structure [18-20]. Another challenge in the medical field is the new input formats such as 3D data. In early CNN applications of three-dimensional images, the convolutions were fed to a network by different size currents fragments by Volume of Interest (VOI) division. [12] was the first to use this strategy to segment the knee cartilage. Another strategy is to feed this network with multi-angled pieces of 3D space in a multi-current mode practiced by various researchers in the field of medical imaging [16]. In [**?**], fCNN applies exactly the same amount of sampling factor in each direction by changing one pixel at a time. By connecting the result of both, a high-resolution copy is obtained at the final output, except for pixels that are lost due to valid convolutions. In [14] went further than the idea of fCNN and proposed the U-net structure, which included a ”regular” portion of the CNN network, where the convolutions are used to increase image size, contraction, and long paths. A similar approach was used by [3] for 3D data. In [11], a method was proposed for U-Net layout that minimizes the residual blocks and single-layer loss.

To the best of our, few CADx systems have been developed for cystoscopic image classification. Hashemi et al. [8] used a multi-layer perceptron to recognize three abnormalities in a dataset of 540 bladder images including bloody urine, malignant masses, and benign masses. In their proposed method, Genetic Algorithm (GA) is used to initialize the weights of the network. Moreover, the learning rate of the network is adaptively changed during the training phase. Using these contributions, they achieved an accuracy of 41.5%. In their next paper [7], They applied VGG16 and ResNet50 to classify four types of bladder images including bloody urine, malignant masses, benign masses, and normal images in a dataset of 720 cystoscopic images. VGG16 and Resnet50 achieved an accuracy of 58.6 and 54.3, respectively.

3. Material and Methods

In this section, we first describe the dataset used in this study. Afterwards, the proposed ensemble classifier to categorize the bladder cystoscopic images is described in details. The main steps of the proposed method include VGG16 training and bootstrap aggregating.

3.1. Dataset

The cystoscopic images in this study was collected from a medical center in the Netherlands, which contains four classes: normal, blood in urine, benign masses, and malignant masses. All images are biopsy-proven. These images are from different angles and various areas of the bladder taken by the cystoscope. The size and quality of the images are not the same due to different situations. Totally, each class consists of 180 images of size 476x540 pixels. Therefore, robustness to these variations motivated us to construct an ensemble classifier instead of using a single one. Figure 1 illustrates a sample of three abnormalities in the dataset.

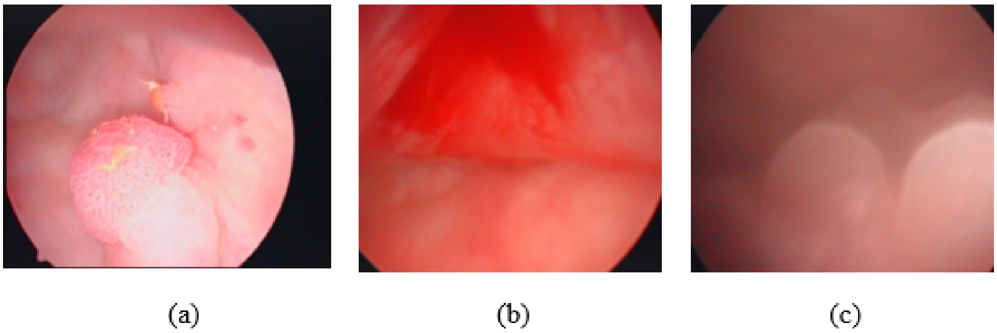
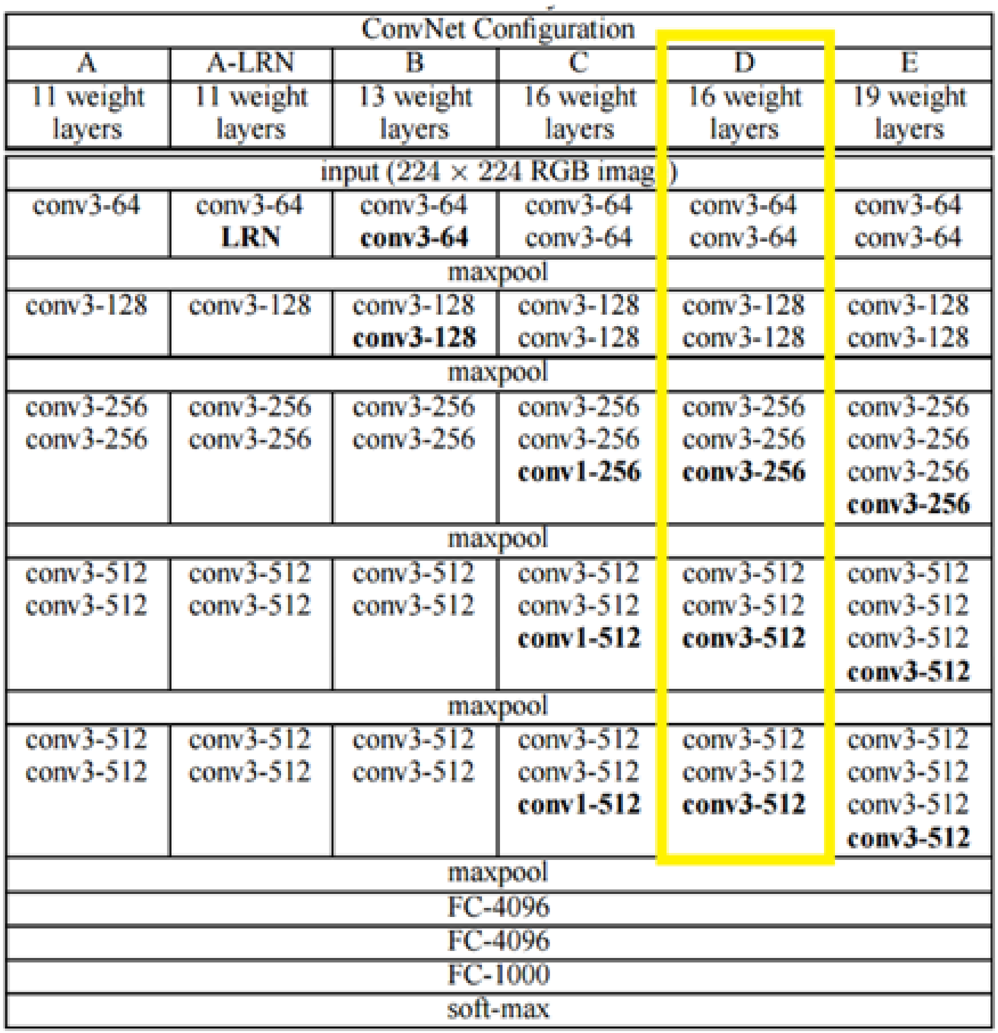


Figure 1: Three abnormalities in cystoscopic images: (a) malignant mass, (b) bloody urine, and (c) benign mass

3.2. Ensemble Classifier

To improves the robustness of classification, it is recommended to fuse simple diverse classifiers instead of a single complex classifier. In this study, diverse VGG-Nets are combined using bootstrap aggregating which is utilized to pump diversity in the ensemble.

VGG architecture was developed at ILSVRC 2014 competition and reached the error rate of 7.3%, with the view that deeper architecture improves the accuracy. As shown in figure 2, there are different configuration for VGG. We used configuration D in the proposed ensemble. The VGG16 network contains 16 convolution layers or 16 parametric layers. Compared to AlexNet, VGG uses much more convolutional layers with smaller filter size within each layer. Moreover, it consists of consequent convolutional layers which increase the field of view of filters. For example, convolving two 3×3 filters results in a 5×5 filter having wider field of view.



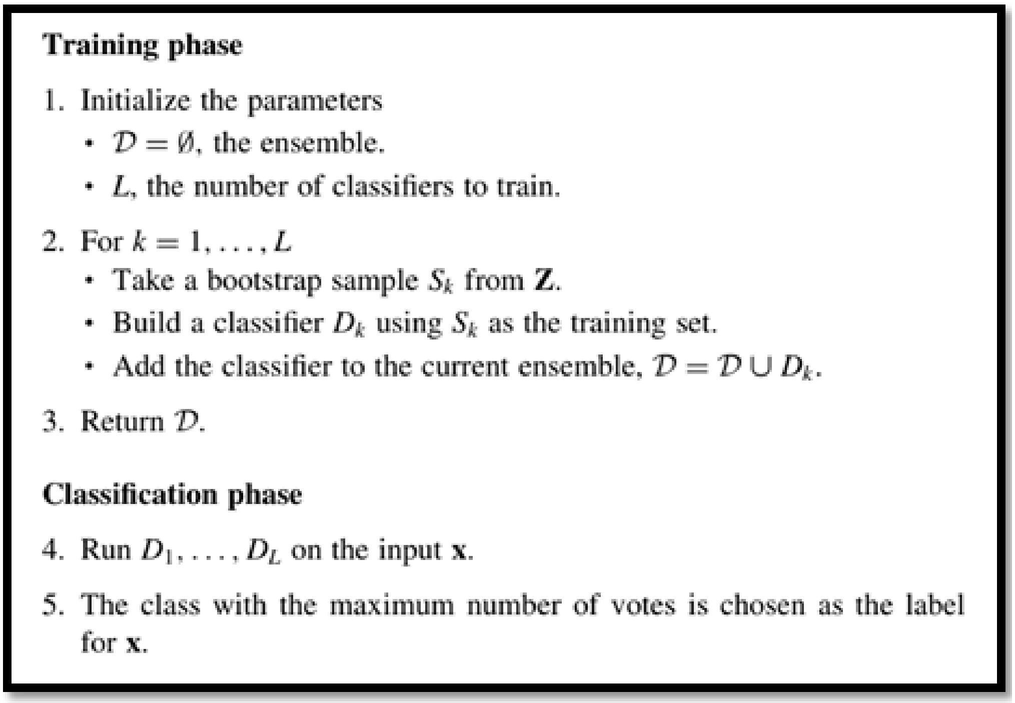
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 2: Six different configurations of VGG

VGG has too many parameters to be trained that probably encounter the network with overfitting problem, especially for small data sets such as the dataset used in this paper. To avoid the overfitting, transfer learning approach is considered in the network training. In this approach, a pre-trained VGG which is already trained on ImageNet dataset is modified. To this end, all convolutional layers except the last one are frozen. Then, the last convolutional layer is vectorized using a global average pooling. Afterwards, the obtained vector is fed to two dense layers simulating a MLP classifier. This is the final architecture of the modified VGG-Nets which are applied in the ensemble classifier.

By combining classifiers we are aiming at a more accurate classification decision at the expense of increased complexity. Instead of looking for the best set of features and the best classifier, now we look for the best set of classifiers and then the best combination method. In classifier combination, diversity is vital for the success of the ensemble. There are four levels to make the base classifiers of the ensemble diverse: (1) data level diversity, (2) feature level diversity, (3) classifier level diversity, and (4) combiner level diversity [10]. In This work, to make the base classifiers (i.e. the VGG-Nets) more diverse, we focused on data level diversity. Based on this level, the base classifiers must use different data subsets for training. Bootstrap aggregating is adopted to produce different training sets for the base VGG-Nets.

The idea of bootstrap aggregating is simple and appealing: the ensemble is made of classifiers built on bootstrap replicates of the training set. The classifier outputs are combined by the plurality vote. The diversity necessary to make the ensemble work is created by using different training sets. Ideally, the training sets should be generated randomly from the distribution of the problem. In practice, we can only afford one labelled training set, $Z = \{z_1, ..., z_N\}$, and have to imitate the process or random generation of L training sets. We sample with replacement from the original training set to create a new training set of length N. To make use of the variations of the training set, the base classifier should be unstable, that is, small changes in the training set should lead to large changes in the classifier output. Otherwise, the resultant ensemble will be a collection of almost identical classifiers, therefore unlikely to improve on a single classifier's performance. An example of unstable classifier is neural networks which is the base classifier of the proposed ensemble. Figure 3 shows the training and operation of bootstrap aggregating. Bootstrap aggregating is a parallel algorithm in both its training and operational phases. The L ensemble members can be trained on different processors if needed.



```
Training phase
1. Initialize the parameters
    • $\mathcal{D} = \emptyset$, the ensemble.
    • $L$, the number of classifiers to train.
2. For $k = 1, \dots, L$
    • Take a bootstrap sample $S_k$ from $\mathbf{Z}$.
    • Build a classifier $D_k$ using $S_k$ as the training set.
    • Add the classifier to the current ensemble, $\mathcal{D} = \mathcal{D} \cup D_k$.
3. Return $\mathcal{D}$.

Classification phase
4. Run $D_1, \dots, D_L$ on the input $\mathbf{x}$.
5. The class with the maximum number of votes is chosen as the label
   for $\mathbf{x}$.
```

Figure 3: Bootstrap aggregating algorithm

After training each VGGNet on its own bootstrap, the training phase of the ensemble is accomplished. Now, the ensemble classifier is ready for testing (i.e. the classification) phase. In the testing phase, a simple majority vote is used to make the final decision based on the votes of the base VGGNets.

4. Results

In this section, evaluation criteria and analysis of the experiments are explained. All experiments are implemented by Python language using Keras framework on a computer with 8 GB of RAM and NVIDIA GeForce GTX 950M graphics card.

In this presented study, standard 10-fold cross validation is used to evaluate the perforamance of the proposed CADx system because just one run of training and testing phases is not reliable. In this strategy, the cystoscopic image dataset is randomly divided into 10 non-overlapping sections of equal size, $X_i, i = 1, 2, \ldots, 10$. To produce each pair of train and test data, one of the 10 sections is used for testing and the other 9 sections for training. This operation is repeated 10 times. In this way, 10 pairs are obtained as follows:

$$\begin{aligned} &V_1 = X_1 \quad , \quad T_1 = X_2 \cup X_3 \cup \ldots X_{10} \\ &V_2 = X_2 \quad , \quad T_2 = X_1 \cup X_3 \cup \ldots X_{10} \\ &: \qquad\qquad : \\ &: \qquad\qquad : \\ &V_{10} = X_{10} \quad , \quad T_{10} = X_1 \cup X_2 \cup \ldots X_9, \end{aligned} \tag{4.1}$$

At the end of the training procedure, there are 10 copies of each predictor, each version being trained on 10 datasets. The total error is summation of errors on these 10 datasets. The classification performance criteria is accuracy which is formulated as:

$$Accuracy = \frac{TN + TBU + TMM + TBM}{N} \tag{4.2}$$

where TN, TBU, TMM, and TBM are number of true normal, true bloody urine, true malignant masses, and true benign masses, respectively. The denominator (N) is the total number of tested images. Based on the experiments, the proposed ensemble classifier achieved an accuracy of 63% on 720 images. Table 1 compares the result of the proposed method with two studies which are tested on the same dataset considering a similar experimental setup. As shown in the table, while a single VGG16 achieved an accuracy of 58.6%, combination of VGG-Nets using bootstrap aggregating significantly improves the accuracy of classification.

Table 1: Comparison of cystoscopic images classification mehtods

Method	**Number of images**	**Number of classes**	**Accuracy (%)**
MLP+GA [19]	540	3	49.3
VGG16+Transfer Learning [20]	720	4	58.6
Proposed Method	720	4	63

Despite training an ensemble of deep convolutional neural network is time consuming, the testing phase takes a few seconds which is ignorable in CADx systems like the system developed in this study. It is notable that training phase is offline but testing phase is online. Hence, for medical fields in which accuracy plays more important role than speed, combining deep convolutional neural networks would be a trend.

5. Conclusions

In this paper, an ensemble of deep convolutional neural network to recognize abnormalities in cystoscopic images was introduced. The base classifiers of the proposed ensemble were VGGNets which were combined using bootstrap aggregating method. The proposed method was evaluated on 720 images via 10-fold cross validation and achieved an accuracy of 63%. In this study, data level diversity was considered in combining base classifiers. Considering other aspects of diversity such as classifier level and feature level are foreseen.

References

[1] Y. Bengio, et al., Greedy layer-wise training of deep networks. in Advances in neural information processing systems. (2007).

[2] Y. Bengio, A. Courville and P. Vincent, Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35(8): (2016) 1798-1828.

[3] Ö. Çiçek, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. in International conference on medical image computing and computer-assisted intervention. (2016) Springer.

[4] H. Greenspan, B. Van Ginneken and R.M. Summers, Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging,35(5), (2016) 1153-1159.

[5] J. Gu, et al. ,Recent advances in convolutional neural networks. Pattern Recognition, 77 (2018) 354-377.

[6] G.E.Hinton, S. Osindero and Y.W. Teh, A fast learning algorithm for deep belief nets. Neural computation, 18(7), (2006) 1527-1554.

[7] S. Hashemi, H.Hassanpour, E.Kozegar, and Tao Tan. Cystoscopy Image Classification Using Deep Convolutional Neural Networks. International Journal of Nonlinear Analysis and Applications, 10(1) , (2019) 193-215.

[8] S. Hashemi, H.Hassanpour, E. Kozegar and Tao Tan. Cystoscopic Image Classification Based on Combining MLP and GA. International Journal of Nonlinear Analysis and Applications, 11(1) (2020) 93-105.

[9] K. Kamnitsas, et al., Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis, 36 (2017) 61-78.

[10] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, ISBN 0-471-21078-1 Copyright 2004 John Wiley & Sons, Inc.

[11] F. Milletari, N. Navab, and S.A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. in Fourth International Conference on 3D Vision (3DV). (2016) IEEE.

[12] A. Prasoon, et al., Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. in International conference on medical image computing and computer-assisted intervention. (2013) Springer.

[13] D. Ravì, Deep learning for health informatics. IEEE journal of biomedical and health informatics,21(1) (2016) 4-21.

[14] O. Ronneberger, P. Fischer and T. Brox. U-net: Convolutional networks for biomedical image segmentation. in International Conference on Medical image computing and computer-assisted intervention.(2015) Springer.

[15] J. Schmidhuber, Deep learning in neural networks: An overview. Neural networks, 61 (2015) 85-117.

[16] A. Setio, et al., Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE transactions on medical imaging, 35(5), (2016) 1160-1169.

[17] N. Simforoosh, Iranian textbook of urology, Tehran, Shahid Beheshti University of Medical Sciences, (2007).

[18] Y. Song, et al., Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. IEEE Transactions on Biomedical Engineering, 62(10) , (2015) 2421-2433.

[19] C. Szegedy, et al., Going deeper with convolutions. in Proceedings of the IEEE conference on computer vision and pattern recognition. (2015).

[20] W. Yang, et al. Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. Medical image analysis, 35: (2017) 421-433.