X-Ray Chest Image Classification by A Small-Sized Convolutional Neural Network

Ege Kesim
Department of Computer Engineering,
Koc University,
Istanbul, Turkey.
egekesim@hotmail.com

Zumray Dokur
Department of Electronics and Communication Engineering,
Istanbul Technical University,
Istanbul, Turkey
dokur@itu.edu.tr

Tamer Olmez
Department of Electronics and Communication Engineering,
Istanbul Technical University,
Istanbul, Turkey
olmezt@itu.edu.tr

Abstract— **Convolutional Neural Networks are widely used in image classification problems due to their high performances. Deep learning methods are also used recently in the classification of medical signals or images. It is observed that well-known pre-trained large networks are used in the classification of X-ray chest images. The performances of these networks on the training set are satisfactory, but their practical use includes some difficulties. The usage of the different imaging modalities in the training process decreases the generalization ability of these networks. And also, due to their large sizes, they are not suitable for real-time applications. In this study, new network structures and the size of the input image are investigated for the classification of X-ray chest images. It is observed that chest images are assigned to twelve classes with approximately 86% success rate by using the proposed network, and the training is carried out in a short time due to the small network structure. The proposed network is run as a real time application on an embedded system with a camera and it is observed that the classification result is produced in less than one second.**

Keywords*— X-ray chest image classification, Deep learning, Convolutional neural network, Real-time image processing.*

I. INTRODUCTION

It is observed that Convolutional Neural Networks (CNNs) are widely used in image classification problems. They outperform conventional deep nets with their ability to search for a pattern in a portion of the image rather than looking at the whole picture. In 2012, a deep learning model using CNN which is called AlexNet did a huge difference with other competitors on the ImageNet dataset [1]. AlexNet outperformed the other competitors and showed the power of CNNs in image classification leading to CNNs to become popular. The numbers of layers in the models have been growing, resulting in more complex models that achieve better results than the previous ones. These models are used to solve many problems like image classification, etc. Since CNNs are good at classification they can also be applied to medical classification problems. They are used in problems such as skin disease detection, disease detection from X-ray images, any abnormalities in lungs and etc. [1–7]. In this study, a novel small-sized CNN is proposed to determine the diseases (abnormalities) in the chest region of the human body by using X-ray chest images.

In Liu's study, tuberculosis is detected from chest images using CNN [2]. In this study, pre-trained AlexNet [1] and GoogleNet [8] were used. Normal and tuberculosis images were examined and the images were divided into four classes with 96% success rate. However, the training of these networks took two or three days. In Dong's study [3], lung images were divided into ten different classes using the CNN. The pre-trained Vgg16 and ResNet-101 CNNs were compared. Chest images of 256×256 pixels were fed into these models. Images are classified as normal and abnormal with 82% success rate. The performance was low in the classification of the images into ten different categories. So, a different measure was used to calculate the accuracy. If the classification result for the image was in the top three decisions determined by the network, the classification process was considered successful, and in this way, 90% success was achieved. In Kieu's study [4], chest images were divided into two groups as normal and abnormal. The data set consisted of 300 images, including 150 normal and 150 abnormal. Novel CNN structures were proposed in the study. CNNs consisted of four layers. Therefore, training was completed in a short time of 360 minutes. Chest images of 128×128 pixels were given as input to the CNNs. An average of 92% success was achieved. In Xu's study, four new CNN structures were proposed as the classifier [5]. The first CNN was used to divide the images into two classes as normal and abnormal. Chest images of 512×512 size were given as input to the CNNs. An average of 93.6% success was achieved. In the study of Bar [6], 443 X-ray chest images were divided into four classes. An average of 89% success was achieved. In Abiyev's study, classification performances of three CNN networks were compared [7]. The ChestX-ray 14 dataset was chosen to train the networks [9]. In some parts of the study, there was no clarity: (*i*) Images in the database were labeled by more than one class, and an accuracy measure for the classification performance was not defined. (*ii*) Probability distribution of classes in the database was unbalanced. In addition, whether or not an augmentation operation has been performed on the database was not stated to make the training set balanced. In this case, if 70% of all the images in the database had been used for training, the training of the CNN would not succeed in high performance (*iii*) Images in the dataset were resized from 1024×1024 to 32×32 pixels. It is expected from this over resizing that the classification performance for the images containing nodules should be extremely poor, but an average of 92% success was reported.

In the literature it is observed that new CNN structures are created instead of using pre-trained networks for the classification of X-ray chest images [2, 3], and chest images are classified as normal and abnormal since increasing the number of classes reduces the performance [3–5]. Also to obtain high performances from the CNN when the number of

978-1-7281-1013-4/19/$31.00 ©2019 IEEE

classes increases, a different measure is used to calculate the accuracy [3]. The goal of this study is to investigate smaller network structures that provide higher performances for the classification of X-ray chest images. In our work the model is able to detect diseases using only the X-ray image without any additional knowledge about patient's history.

I. Methods

A. General Structure of Convolutional Neural Network

There are basically three processes in CNN's structure; *convolution process*, *nonlinear transformation*, and *max-pooling process*, where a layer is assigned for each process. In convolution layer, a square sub-window (or filters) is moved over the input image. In effect, the pixel values of the sub-window are filter coefficients (weights) which are learned during the training. The gray values of the input image pixels within the sub-window are multiplied with weights and summed to give a scalar value associated with the central pixel of the input image within the sub-window. A new image is formed by sliding this sub-window across the entire input image. The new image is called as feature plane. The size of the sub-window can be chosen freely. This way of using a sub-window is a good process when searching for special patterns in an image. Moreover, since it looks at only a limited number of pixels at a time, it uses less memory than the fully connected network. Usually after a convolution process, a nonlinear activation function such as *sigmoid*, *rectified linear unit function* (ReLU) or *hyperbolic tangent* take place, but generally, the ReLU function is preferred. The output of the ReLU is applied to a max-pooling operation where again there is a sub-window that moves over the image. This sub-window outputs the pixel with the highest value. That single pixel value is the maximum value seen in the sub-window.

B. The Proposed Convolutional Neural Networks

In this study, three CNNs are examined comparatively. The effects of increasing the number of layers in the CNN, on the classification performance of chest images are examined by using the first structure, CNN-1. The size of input signal is 128×128. CNN-1 includes thirteen convolution layers and two fully connected layers. At the output of each convolution layer, there are ReLU layers. But, between some convolution layers, there is no "max-pooling" process. The size of the sub-window (or filters) is 3×3 for all convolution layers. The convolution layers of one to thirteen consist of 32, 32, 32, 32, 64, 64, 64, 128, 128, 128, 256, 256 and 256 feature planes, respectively. The fully connected layers consist of 1024 and 12 nodes (number of classes). Figure 1 shows the structure of CNN-1.

The effects of the size of the input chest images on the classification performance are examined by using the second and the third structures, CNN-2 and CNN-3. CNN-2 has four convolution layers and two fully connected layers. At the output of each convolution layer, there are ReLU layers and max-pooling. The size of input signal is 256×256. The size of the sub-window (or filter) is 3×3 for all the four convolution layers, and the number of feature planes is twelve for all of them. The convolution layers are followed by two fully connected layers. The first and the second fully connected layers contain 1024 and 12 nodes (number of classes), respectively.

CNN-3 also contains four convolution layers and two fully connected layers. At the output of each convolution layer, there are ReLU layers and max-pooling. The size of input signal is 128×128. The size of the sub-window is 3×3 for all the convolution layers, and the numbers of feature planes are 31, 31, 31 and 21 for the first, second, third and fourth layers, respectively. The convolution layers are followed by two fully connected layers. These layers contain 1024 and 12 nodes, respectively. Figure 2 shows the structure of CNN-3.

II. Simulation Results

A. Data and Augmentation

In this study, a subset of the ChestX-ray 14 dataset is chosen to train the networks. ChestX-ray 14 contains patient chest X-ray images with 14 disease types [9]. It also contains healthy/undiagnosed patient images. There are more than 112.000 images, but the data is unbalanced and some patients have multiple diseases. This dataset is used for multi-class classification problem. First the dataset is analyzed to find the distribution of number of patients having single disease in the dataset. After analyzing the dataset, it is realized that some categories contained really small amounts of single diseased patients. In order to have a good balanced dataset, it is decided to eliminate these categories which are *pneumonia* and *hernia*. Also *healthy/undiagnosed* category is eliminated. In order to get a balanced data, the category with the smallest number of images is determined which contained 628 images. Then, from all the images, 628 images are randomly selected from each category which was 7536 images totally. To avoid overfitting, the number of samples in the new dataset was increased by augmenting the images. The images were augmented by randomly zooming and shifting the image. Every image was zoomed and shifted by some random amount. Figure 3 shows an original X-ray chest image and a modified image after the augmentation process. With this method, the number of images has been increased by 6 times, so that after the augmentation, the new dataset has increased to a total of 45216 images. In the training process, binary cross entropy loss function is used since binary labels are used.

B. Training and Classification Processes

All the analyses were realized by using Python codes running on Ubuntu Linux workstation. The workstation used in this study has 32 core CPUs of 2.7 GHz with GeForce GTX1080 Ti Graphics card. The training process of CNN-1 lasts approximately 14.2 hours and the training processes of CNN-2 and CNN-3 last about 7.3 hours. These values give the elapsed times for the training after determining the optimum network parameters. Determination of the optimum network parameters takes even longer. In particular, it takes six to seven days to find the optimum parameters for CNN-1.

Table 1 shows classification performances of the CNNs. 80% of the 45000 images is used to create the training set and the remaining 20% is used for the test set. The *accuracy* measure is described below:

$$Acc = \frac{1}{N}\sum_{c=1}^{C} TP_c \qquad (1)$$

Table 1 Classification performances of the CNNs

	CNN-1 128×128	CNN-2 256×256	CNN-3 128×128
Acc-Training %	100	100	100
Acc-Test %	67.3	86	84.8

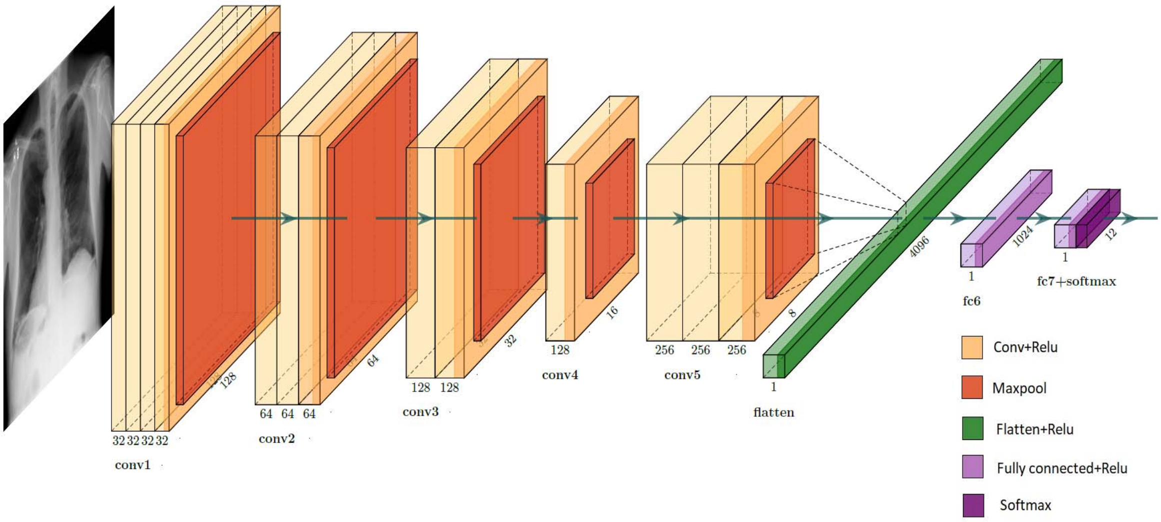


Fig. 1: The structure of CNN-1.

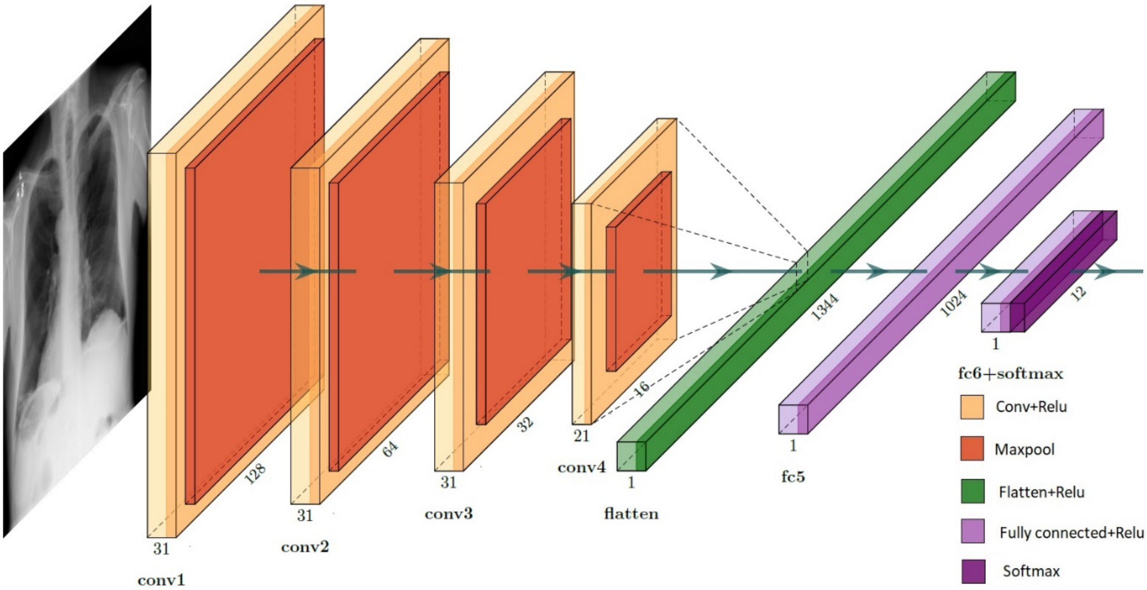
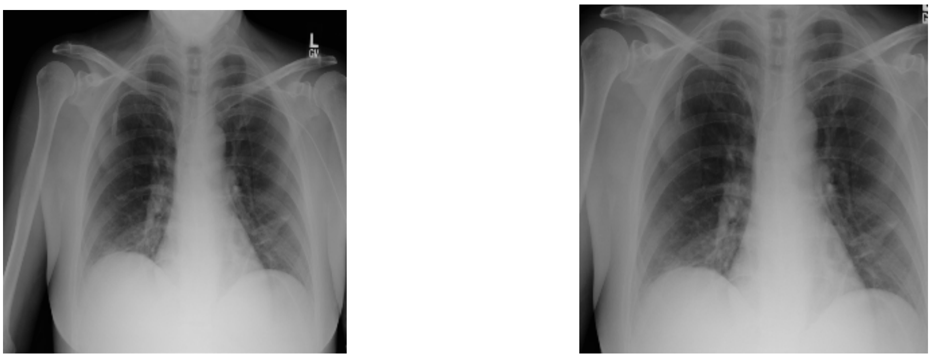


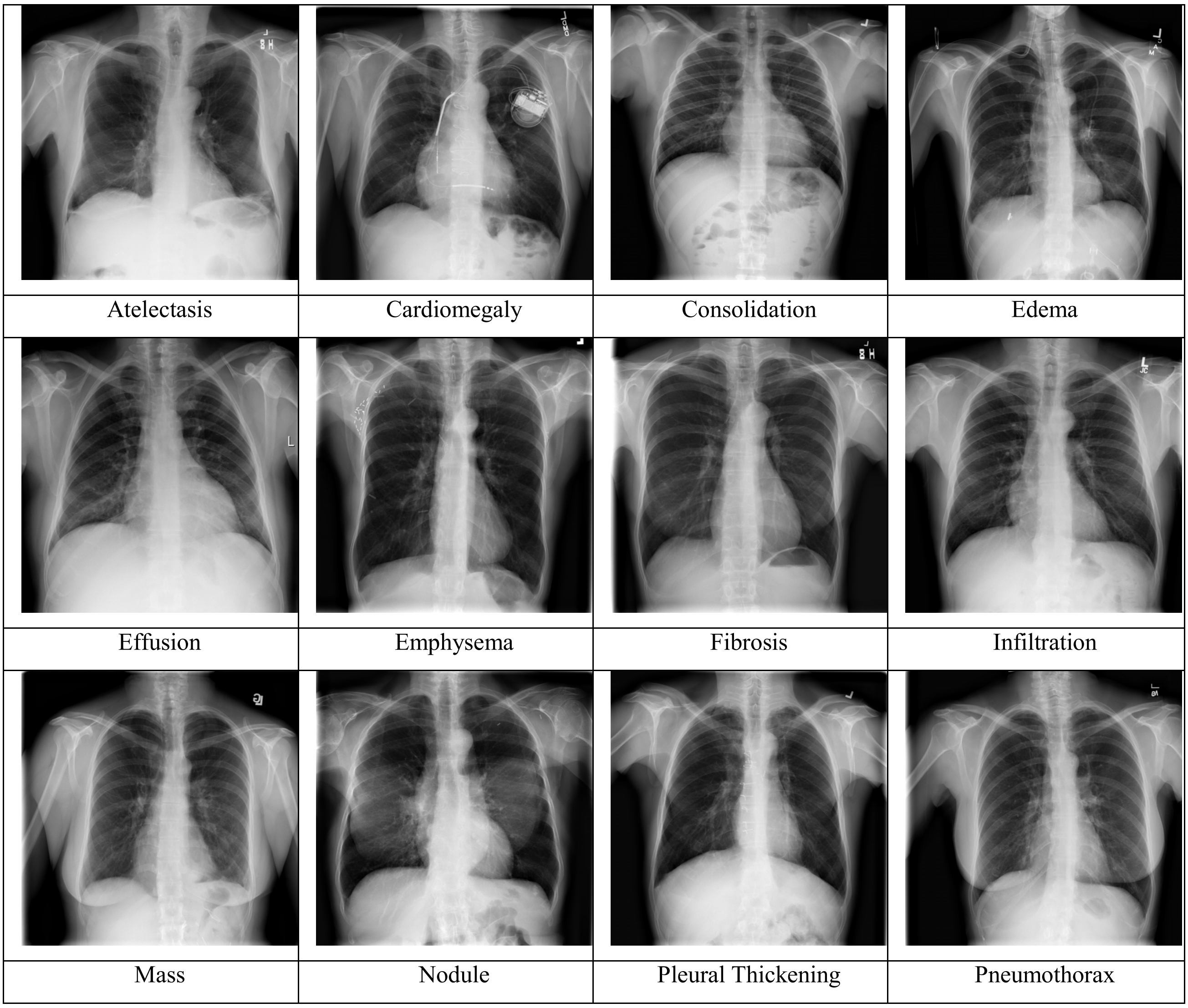
Fig. 2: The structure of CNN-3.



(a)

(b)

Fig. 3 (a) Original X-ray chest image, and (b) a modified image after the augmentation process



Atelectasis | Cardiomegaly | Consolidation | Edema

Effusion | Emphysema | Fibrosis | Infiltration

Mass | Nodule | Pleural Thickening | Pneumothorax

Fig. 4 Original X-ray chest images of twelve different classes

where C represents the number of classes, TP_c represents the number of *true-positives* of the class c.

If the input image is 128×128 or 256×256 the classification performances of CNNs do not change. However, for images smaller than 128×128, performances greatly decrease. In addition, it is observed that increasing the number of CNN layers did not improve the generalization of the network.

C. Memory Needs for Real-Time Applications

In this subsection, memory needs of CNNs will be examined.

CNN-1 has thirteen convolution layers and two fully connected layers. The convolution layers of one to thirteen consist of 32, 32, 32, 32, 64, 64, 64, 128, 128, 128, 256, 256 and 256 feature planes, respectively. The fully connected layers have 1024 and 12 nodes. The structure of the CNN-1 comprises of 1×3×3×32, 32×3×3×32, 32×3×3×32, 32×3×3×32, 32×3×3×64, 64×3×3×64, 64×3×3×64, 64×3×3×128, 128×3×3×128, 128×3×3×128, 128×3×3×256, 256×3×3×256, 256×3×3×256, (4×4×256)×1024 and 1024×12 parameters.

CNN-2 has four convolution layers and two fully connected layers. Each of the four convolution layers consists of 12 feature planes, and the fully connected layers have 1024 and 12 nodes. The structure of the CNN-1 comprises of 1×3×3×12, 12×3×3×12, 12×3×3×12, 12×3×3×12, (16×16×12)×1024 and 1024×12 parameters.

CNN-3 has four convolution layers and two fully connected layers. The convolution layers of one to four consist of 31, 31, 31 and 21 feature planes, respectively. The fully connected layers have 1024 and 12 nodes. The structure of the CNN-1 comprises of 1×3×3×31, 31×3×3×31, 31×3×3×31, 31×3×3×21, (8×8×21)×1024 and 1024×12 parameters.

There is also a need for memory space in order to save intermediate values calculated between layers. To be able to represent weights by floating point, 8 bytes will be allocated for each weight. Therefore, the above results will be multiplied by 8. The usage of large-sized networks such as

CNN-1 network is difficult on portable systems (as FPGA, embedded Linux cards) in real time. Therefore, new network structures that have smaller sizes and produce high success rates are investigated in this study.

After the training process, the weights of all the layers of the networks are stored on the disk. These weights are then loaded to create the network in the target embedded system. In the study, OrangePi zero plus 2 with a camera was chosen as the target embedded system. All the layers (convolution, max-pooling and ReLU) of the network were modeled by using the OpenCV library and C language separately. Both models are called from the Python program. The configurations and dependencies are related with the embedded system used (Raspberry Pi, NonoPi, OrangePi, BananaPi, imx platform etc.). Therefore, neural network structures are modeled in C language to avoid this variability. Neural networks modeled in C, can also be used on mobile phones. Because the hardware of embedded systems is less powerful than the normal computers, it is not always possible to call the neural network models from OpenCV library in embedded systems. These libraries are difficult to install on embedded systems, and a high number of dependent libraries must be installed. Configuration and memory problems occur in this context. In C patform, 'Omap' library was used in order to increase the speed. By selecting this library of the C language, four CPUs of OrangePi were used in parallel. In both models, an X-ray chest image displayed on a different computer was classified in less than one second using OrangePi's camera and the Python program developed in this study.

III. Conclusions and discussion

In the literature, the generalization problem arising from the use of pre-trained networks is mentioned [5]. In order to solve this problem, the network can be trained by using biomedical images acquired with the same imaging modality. As a future study, the network used in the training of chest images can be retrained to recognize the abnormalities in MR images.

The size of the original X-ray chest images is 1024×1024. If this size of images were fed into the networks, the classification performances would be higher. However, applying such large-size input data would take a very long time of learning. The risk of being caught by local optima during the training is naturally high. Because of such problems, the size of the input images is usually reduced in the literature [4–7]. However, this causes a decrease in the classification performance. It is observed that if the size of the input image is 128×128, classification performances for the X-ray chest images do not decrease dramatically.

The aim of our project (İTÜ-BAP: MYL-2018-41621) is to classify chest images captured by the camera of a portable embedded Linux system by using a Python program developed. In the first stage of the project, modeling of the convolutional neural network by using the C language and speed of this model were tested. Images of twelve different classes captured by the embedded system's own camera were then classified quickly and accurately by using the Python program running on the embedded system. In the succeeding stages of the project, the network model and the main program as Android mobile phone application will be coded in C language using QT ide. Problems encountered in the classification of images captured by the camera will be examined. The training set will also be enriched with chest images captured by the system's camera. It is expected that the classification performance will be improved at this stage.

Ackowledgment

This study is supported by the Istanbul Technical University Scientific Research Project Unit (ITU-BAP project number MYL-2018-41621).

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[2] C. Liu et al., "TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network", 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 2314–2318. doi: 10.1109/ICIP.2017.8296695

[3] Y. Dong, Y. Pan, J. Zhang and W. Xu, "Learning to Read Chest X-Ray Images from 16000+ Examples Using CNN", 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Philadelphia, 2017, pp. 51–57. doi: 10.1109/CHASE.2017.59

[4] P.N. Kieu , H.S. Tran, T.H. Le , T. Le and T.T. Nguyen, "Applying Multi-CNNs model for detecting abnormal problem on chest x-ray images", 10th International Conference on Knowledge and Systems Engineering (KSE), 2018.

[5] S. Xu , H. Wu and R. Bie , "CXNet-m1: Anomaly Detection on Chest X-Rays With Image-Based Deep Learning", IEEE Access, 2018.

[6] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen and H. Greenspan, "Chest pathology detection using deep learning with non-medical training", 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, 2015, pp. 294–297.
doi: 10.1109/ISBI.2015.7163871.

[7] R.H. Abiyev and M.K.S. Ma'aitah, "Deep Convolutional Neural Networks for Chest Diseases Detection", J Healthc Eng. 2018; doi: 10.1155/2018/4168538.

[8] C. Szegedy, W. Liu, Y. Jia, Sermanet et al., "Going deeper with convolutions", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[9] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, "Chest X-ray 8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases", Proceedings of IEEE CVPR 2017, 2017.