Emphysema Discrimination from Raw HRCT Images by Convolutional Neural Networks

Esra Mahsereci Karabulut[1] and Turgay Ibrikci[2]

[1]Gaziantep University, Computer Programming Department, 27310, Sahinbey, Gaziantep, Turkey
mahsereci@gantep.edu.tr
[2]Cukurova University, Electrical-Electronics Engineering Department, 01330, Balcalı, Adana, Turkey
ibrikci@cu.edu.tr

Abstract

Emphysema is a chronic lung disease that causes breathlessness. HRCT is the reliable way of visual demonstration of emphysema in patients. The fact that dangerous and widespread nature of the disease require immediate attention of a doctor with a good degree of specialized anatomical knowledge. This necessitates the development of computer-based automatic identification system. This study aims to investigate the deep learning solution for discriminating emphysema subtypes by using raw pixels of input HRCT images of lung. Convolutional Neural Network (CNN) is used as the deep learning method for experiments carried out in the Caffe deep learning framework. As a result, promising percentage of accuracy is obtained besides low processing time.

1. Introduction

Emphysema is a disease of the lungs described as decrease in the amount of oxygen transferred to the blood and therefore causes shortness of breath. Emphysema is mostly caused by smoking and it is characterized by elasticity loss of the tissue surrounding the alveoli limiting the expanding and shrinking of airspaces. Emphysema with chronic bronchitis is referred to as Chronic Obstructive Lung Disease (COPD) which is the fourth leading cause of death in the United States [1] and affects 5% of the world population [2].

A reliable technique for imaging the pathologies of emphysema is the high resolution computed tomography (HRCT). The possibility of other medical conditions related to lung can be detected and eliminated by HRCT scanning. Two important subtypes of emphysema are centrilobular emphysema (CLE) and paraseptal emphysema (PSE) which can be distinguished via HRCT images. CLE is the most common of the subtypes that arise from long-term cigarette smoking and usually involves the upper half of the lungs. PSE can occur both in smokers and non-smokers, and it is mostly found in younger patients compared to CLE [3]. Fig. 1 represents three example HRCT scans of size 512x512 where normal tissue (NT) describes the non-emphysematous healthy lung.

In the literature, there are several approaches for automated discrimination of emphysema subtypes by using machine learning approaches [4-6]. In most of the studies, as a pre-work, textural features of HRCT images are discovered by using texture analysis methods such as kernel density estimation of local histograms [7], Riesz transfom [8], different intensity measures [6], density mask [9,10], smoothing and thresholding [11]. Sørensen et al. [12] achieved the best in a way that they preprocessed the emphysema data by combining textural features using local binary patterns (LBPs) and classified by k-nearest network (KNN) method.

(a) (b) (c)

Fig. 1. HRCT scans from middle part of the lung (a) NT (b) CLE (c) PSE

Medical image analysis by deep learning is scarce in the relevant literature. Deep learning is a new area in machine learning to solve problems of artificial intelligence with a special emphasis on artificial neural networks. As to the current literature, common fields that make use of deep learning approach are image recognition, natural language processing, speech recognition and other classification tasks which report significant performance improvements in comparison to traditional machine learning methods. The automated medical or biomedical decision making is a potential area that requires use of such new approaches since non-automated decision making is exposed to human-supplied errors and becomes more expensive. As a deep learning approach Convolutional Neural Networks has an architecture design for learning 2D input data. When an image is fed to a CNN, it extracts features from raw pixels at the first layer. As the data is processed through layers, higher-level informative input is generated for the softmax classification layer.

In this paper discrimination between emphysema subtypes and non-emphysematous normal tissue is achieved by deep convolutional neural networks (CNN) which to the best of the authors' knowledge has never been used for classification of CT lung images in the literature. The dataset is provided by Sørensen et al. [12] as grey-scale 16 bit tiff images with a resolution of 61 × 61 pixels. It comprises 168 patches from 115 lung HRCT slices. Each patch is labelled as NT, CLE or PSE. Experimental results are obtained by using Caffe environment. Caffe [13] is a deep learning framework developed by the Berkeley Vision and Learning Center (BVLC) with the support of community contributors. Our CNN model is trained in a supervised fashion in the Caffe deep learning framework. We explored the direct performance of CNN on emphysema medical image data without using any of texture analysis preprocessing. The studies on classification of direct raw images of emphysema are rare in the literature. From this point of view, to our knowledge this study presents the best performance with respect to processing time and accuracy without any preprocessing of the input images.

2. Material and Methods

2.1. Dataset Descriptions

The dataset [12] consists of 168 patches from 115 HRCT slices. Fig. 2. presents samples patches from the dataset. CT scanning was performed using General Electric (GE) equipment (LightSpeed QX/i; GE Medical Systems, Milwaukee, WI, USA) with four detector rows and using the following parameters: in-plane resolution 0.78 x 0.78 mm, slice thickness 1.25 mm, tube voltage 140 kV, and tube current 200 mAs. The slices were reconstructed using a high-spatial-resolution (bone) algorithm [12, 16].

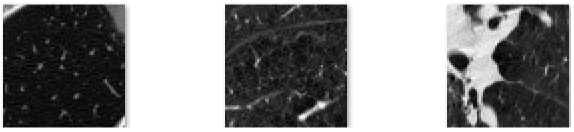

Fig. 2. 61x61 patches from HRCT slices (a) NT (b) CLE (c) PSE

HRCT scans are from 39 people each of which belongs to one of the three categories: a) 9 healthy and non-smokers, b) 10 healthy and smokers (without COPD) and c) 20 smokers and COPD patients. Each patch is manually annotated with one of the three classes; NT, CLE or PSE. Table 1 gives class descriptions and the number of patches in each class.

Table 1. Categories in the dataset

Category	Description	# of samples
1 (NT)	annotated in never smokers	59
2 (CLE)	annotated healthy smokers and smokers with COPD	50
3 (PSE)	annotated healthy smokers and smokers with COPD	59

2.2. Convolutional Neural Networks (CNN)

Lecun et al. [14] introduced CNN which is applied in many areas but showed a superiority when used with two dimensional data such as images and videos [15]. A CNN is a special type of artificial neural network having the advantage of data dimensionality reduction. The “depth” of the model refers to constituting functions into a series of transformations. These transformations are done to convert inputs to intermediate representations and finally to output. Image classification by a traditional neural network requires assigning each input pixel of the image to all of the hidden neuron units. This means a vast number of connections and calculations of weights as data is propagated back and forth through the network. In CNN, each neuron at a hidden layer receives regions of the image with size of the filter. The neurons’ grid-like layout represents the input image of the next hidden layer and this network strategy goes on according to network design requirements. Thereafter the last grid of neurons is input to the hidden layer of an artificial neural network (ANN) in a full-connection form.

The image is fed to the system and convoluted by different filters (or known as convolution kernels) having pre-determined or learnt coefficients. Suppose *nxn* filter **w** is used, convolution result x is calculated by multiplying each image pixel y in layer l-1 by corresponding filter coefficient and the products are all summed up as in equation (1). The image size decreased *n-1* from number both of rows and columns.

$$x_{i,j}^{l} = \sum_{a}^{n}\sum_{b}^{n} w_{ab}\, y_{(i+a)(j+b)}^{l-1} \tag{1}$$

Nonlinear transformation is applied for each value generated by convolution by equation (2)

$$y_{i,j}^{l} = \frac{1}{1+e^{-x_{i,j}^{l}}} \tag{2}$$

By each filter different features or aspects of input image, i.e. feature maps, come out. The output images of convolution is taken by pooling layer and images are subsampled. This operation step is for reducing the dimensionality and preserving critical information. Although several subsampling operations are in use, the widely used ones are 2x2 averaging or max-pooling. The 2x2 averaging means computing the mean of 4 adjacent pixel values and the max-pooling means taking the pixel having maximum value of these 4 adjacent ones.

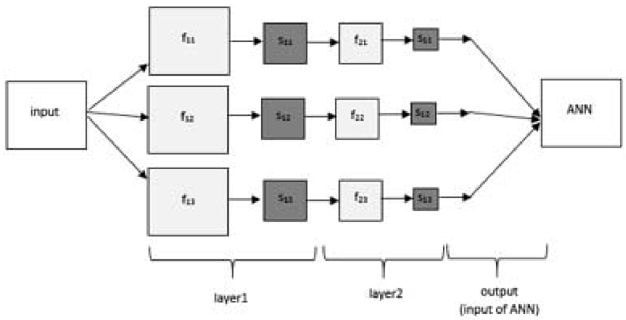

Fig. 3. A two layered example of CNN

The convolution and sampling stages constitute one layer of the deep architecture of a CNN as represented in Fig. 3. There isn’t any greedy layer-wise pre-training like in Deep Belief Networks (DBN), but many layers can be used one after another. After the last sampling process, a traditional ANN continues. The resulting feature maps are categorized by the ANN. The final output of convolutional and subsampling layers is fed to a fully connected layer. More than one fully connected layers can be used and finally a softmax output layer of ANN is utilized for the final decision. The training is achieved in a supervised manner, and the traditional backpropagation algorithm can be used to train the whole model.

2.3. Caffe Deep Learning Framework

Caffe is a fast open-source framework to implement CNNs and other deep learning networks (DNNs) easily for researching and exploring purposes. It is maintained and developed by the Berkeley Vision group. It supports a wide variety of architectures and efficient implementations vital machine learning tasks such as prediction and learning. It provides state-of-the-art deep learning models for research projects and industrial applications in areas of computer vision, speech and multimedia processing. Caffe is licensed under the BSD license. It is designed to be a C++ library but it has bindings for other programming languages such as Python and Matlab.

Computational complexity is high in deep learning models, so Caffe provides means to use both GPU and CPU processing in a parallel fashion. Caffe is claimed to be the fastest CNN implementation [13]. High speed is achieved by Compute Unified Device Architecture (CUDA) GPU computation which is capable of processing over 40 million images in a day with a single K40 or Titan GPU (approximately 2 ms per image) [13]. Such a performance level is of great importance for deep models research.

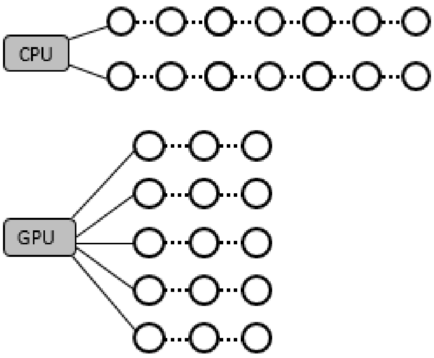

Fig. 2. Parallel computing styles of CPU and GPU

GPU computing puts emphasis on parallel computing rather than higher processing unit performance, whereas CPU has low number of parallel units that are powerful in processing as represented in Fig. 2. However, parallelism outperforms powerful unit performance.

3. Experimental Results

In order to analyze the classification performance of the CNN model for the emphysema disease, several experiments are conducted. In these experiments, a computer with a CPU of 8 cores each at 2.3 GHz is utilized. Also CUDA GPU acceleration is enabled by use of the graphic processor (GeForce 840M with 384 CUDA cores at 1029 MHz clock speed). Ubuntu Linux is the host operating system on which the experiments are conducted.

Since 61x61 pixel images are used, the network has 3721 (=61x61) input neurons. The dataset is split into two parts for training and testing. Training is achieved via 138 out of 168 patches and the remaining 30 patches are for testing purposes. The network parameters used in model are such that; learning rate is 0.0001, momentum 0.9, weight decay 0.0005, and solver mode is a flag that can be switched to CPU or GPU.

Firstly, the effect of the amount of training over classification performance is investigated. To this end, training is carried out over 100000 iterations and test procedure is repeated at each 200 iterations. When Fig. 3 is examined, it is clear that up to a certain point (approximately around 20000 iterations) there is a positive relationship between the amount of training and the accuracy. After that point, the accuracy stays almost intact. A maximum accuracy of %84.25 is achieved through these tests. With sufficient number of training iterations, the network proves itself capable of classifying given instances with an accuracy of %80±4.25 which is a mostly adequate performance for automated classification of emphysema disease for decision support systems.

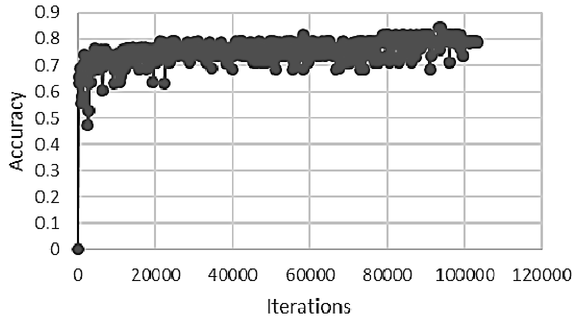

Fig. 3. Accuracy values by iterations

Secondly, the computation load of the classifier system is examined. Owing to the complex nature of neural networks, a deep CNN is expected to operate relatively slower than shallow counterparts. Caffe framework presents opportunities to deal with this disadvantage such as the utilization of GPU as a parallel processor. As can be seen in Table 2, 1000 iterations of the training phase is completed within approximately 142 seconds whereas the GPU powered version of the network completes the same task in almost 29 seconds.

Table 2. CPU and GPU processing time while emphysema training of CNN

	Computing time [s]	
#Iterations	CPU	GPU
0	4.091787	0.380072
200	31.61855	6.136533
400	59.09462	11.99544
600	86.54946	17.80014
800	114.0059	23.58323
1000	142.2619	29.4446

Fig. 4 reveals that the required computing time related to each iterations is reduced if the GPU processor is utilized.

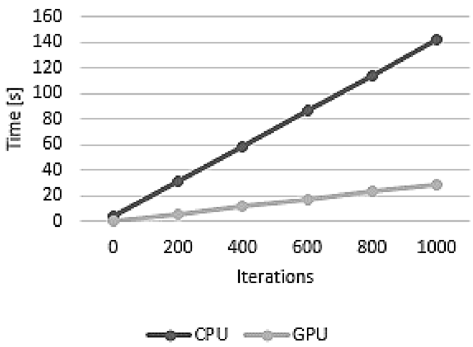

Fig. 4. CPU vs. GPU computing time

The accuracy of %84.25 achieved in this study is very reasonable for an automated decision support system. However, some literature studies such as that of Sørensen et al. [12] reached a better accuracy of %95. The main reason behind the fact is that this study uses the raw image pixels and relies on the intensity values of each pixel for classification purposes whereas

Sørensen et al. [12] utilized more advanced texture techniques such as combining two different texture groups (i.e., local binary patterns and a filter bank based on Gaussian derivatives). In another relevant study, Azim and Niranjan [5] presented classification results for the same dataset using k-NN, condensed nearest neighbor and Restricted Boltzman Machine (RBM) classifiers that were %46.04, %46.06 and %47.85, respectively. It is obvious that CNN is capable of outperforming some other classifiers when sole classifier is used with raw images without any preprocessing. However, the same study [5] also reported that 3-way factored RBM with the input images cropped into the size of 31x31 pixels performed better and produced an accuracy of %86.97.

As a result, CNN as a classifier proves itself successful in classifying raw emphysema images. Without any preprocessing, the achieved accuracy of %84.25 proves this fact. However, since pathology can be localized in small areas of the lung and the region of interest (ROI) is generally smaller than the whole image, advanced texture techniques and preprocessing are required to improve the performance of CNN in this task.

4. Conclusions

Emphysema recognition from HRCT scans requires high level anatomical knowledge of a physician that must be supported by automated computer recognition. In literature emphysema is researched in detail and many computer based solutions are presented. But the result of direct classification results don't take place in most of them, a model is used together with texture analysis preprocessing. Differently we directly processed raw HRCT images in order to emphasize the classification performance of CNN. To our knowledge we gained the best comparing to this kind of studies. The results lead to expectation of more success by incorporating such texture analysis preprocessing methods as future work.

5. Acknowledgements

This study was financially supported by the Cukurova University Research Foundation (Project No: FDK-2015-4395).

6. References

[1] US Department of Health and Human Services. Chronic obstructive pulmonary disease. Washington, DC: USDHHS; 2003. Data fact sheet.

[2] T. Vos et al., "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010". *Lancet* 380 (9859): 2163–96, 2012.

[3] K. Satoh, T. Kobayashi, T. Misao, Y. Hitani, Y. Yamamoto, Y. Nishiyama, M. Ohkawa, "CT assessment of subtypes of pulmonary emphysema in smokers", *Chest*, 120:725-729, 2001.

[4] M. J. Gangeh, L. Sørensen, S. B. Shaker, M. S. Kamel, M. De Bruijne, M. Loog, "A texton-based approach for the classification of lung parenchyma in CT images", in *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, Springer Berlin Heidelberg, 2010, pp. 595-602.

[5] T. Azim, M. Niranjan, "Texture Classification with Fisher Kernel Extracted from the Continuous Models of RBM", in *International Joint Conference on Computer Vision Theory and Applications (VISAPP)*, 2014.

[6] M. Ibrahim, R. Mukundan, "Multi-fractal Techniques for Emphysema Classification in Lung Tissue Images", in *International Conference on Environment, Chemistry and Biology*, Singapore, 2014.

[7] C. Mendoza, G. Washko, J. Ross, A. Diaz, D. Lynch, J. Crapo, E. Silverman, B. Acha, C. Serrano, R. Estepar, "Emphysema quantification in a multi-scanner HRCT cohort using local intensity distributions" in *9th IEEE International Symposium.on Biomedical Imaging (ISBI)*, 2012, pp. 474–477.

[8] A. Depeursinge, A. Foncubierta–Rodriguez, D. Van de Ville, H. Müller, "Multiscale lung texture signature learning using the Riesz transform", In *Medical Image Computing and Computer-Assisted Intervention–MICCAI,* 2012, pp. 517-524.

[9] N. L. Müller, C. A. Staples, R. R Miller, R. T. Abboud, "Density Mask-An Objective Method to Quantitate Emphysema Using Computed Tomography" *Chest*, 94(4), 1988, pp. 782-787.

[10] M. Kinsella, N. L. Müller, R. T. Abboud, N. J. Morrison, A. DyBuncio, "Quantification of emphysema by computed tomography using a density mask program and correlation with pulmonary function tests", *Chest*, 97:315–321, 1990.

[11] O. Friman, M. Borga, M. Lundberg, U. Tylén, and H. Knutsson, "Recognizing emphysema a neural network approach", in *International Conference on Pattern Recognition*, 2002, pages 512–515.

[12] L. Sørensen, S. B. Shaker, M. de Bruijne, "Quantitative Analysis of Pulmonary Emphysema using Local Binary Patterns", *IEEE Transactions on Medical Imaging*, 29(2): 559-569, 2010.

[13] Y. Jia, "Caffe: An open source convolutional architecture for fast feature embedding". http://caffe. berkeleyvision.org/, 2013.

[14] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, 86(11):2278-2324, 1998.

[15] I. Arel, D. C. Rose, T. P. Karnowski, "Deep machine learning-a new frontier in artificial intelligence research [research frontier]", *Computational Intelligence Magazine, IEEE,* 5(4):13-18, 2010.

[16] S. B. Shaker et al., "Identification of patients with chronic obstructive pulmonary disease (COPD) by measurement of plasma biomarkers", *The Clinical Respiratory Journal* 2(1): 17–25, 2008.