Detection of Lung Pathologies Using Deep Convolutional Networks Trained on Large X-ray Chest Screening Database

Conference Paper · May 2019

CITATION
1

READS
248

2 authors:

Vitali Liauchuk
United Institute of Informatics Problems
40 PUBLICATIONS **178** CITATIONS
SEE PROFILE

Vassili Kovalev
National Academy of Sciences of Belarus
127 PUBLICATIONS **1,183** CITATIONS
SEE PROFILE

Detection of Lung Pathologies Using Deep Convolutional Networks Trained on Large X-ray Chest Screening Database

Vitali Liauchuk, Vassili Kovalev
Department of Biomedical Image Analysis, United Institute of Informatics Problems
Surganova, 6, 220012 Minsk, Belarus. vitali.liauchuk@gmail.com, vassili.kovalev@gmail.com

Abstract: *In this paper we explore the problem of a preliminary computerized diagnosis of lung pathologies under condition of X-ray chest screening.*

Keywords**:** X-ray chest images, Lung diseases, Deep Learning.

1. INTRODUCTION

Screening of the population is an important way of timely discovering of lung diseases and certain abnormalities in chest. The typical feature of such a screening is that the resultant image database contains large fraction of normal cases whereas different kinds of abnormalities represented proportional to their natural incidence rate in the society.

Recently, it was becoming a common case when combination of large database and Deep Learning methods creates a good ground for development of efficient Computer Assisted Diagnosis (CAD) systems [1] which are able to produce reliable "second opinion" on the decision for a detailed examination of suspicions cases.

In this work, the problem of generating of a preliminary diagnosis of lung diseases is considered in the context of availability of large chest x-ray image database resulted from conventional screening system and recent image classification tools based on Deep Learning methods.

2. INPUT IMAGE DATA

Normal cases. All the x-ray image data used with this study are the natively-digital x-ray scans which extracted from a PACS system. The PACS contains information on results of x-ray screening of the chest a two-million city conducted during the years 2001–2014. The version of database we used here contains a total of 1,908,926 records. Each record corresponds to a single digital chest x-ray image. The records include information on patients' age, gender and textual radiological reports. The reports are written in a free-form native, non-English language. The reports were provided collectively by a board of experienced radiologists and contain information on possible lung diseases, diseases of cardiovascular system (heart, blood vessels), and skeleton abnormalities such as scoliosis, deformation of ribs, etc. The lung abnormalities include pneumosclerosis, emphysema, fibrosis, pneumonia, focal shadows, bronchitis, and the lung tuberculosis. Technically, all the x-ray scans were represented by 1-channel 16-bit images which were originally stored in DICOM format. The image resolution varied from rarely occurred small size of 520×576 pixels to relatively large size of 2800×2531 pixels. Since all the images were natively-digital, there were no film scanning artifacts presented in the image data.

An image was categorized into the class of Norm if there were no visible signs of any type of abnormalities in mediastinum, skeleton and the lungs. Normal cases were selected from the database by way of parsing of radiological reports with the help of a set of key phrases which clearly indicate absence of any visible abnormalities. Also, only strict matches of the report texts with one of the key phrases were considered. Following this procedure, a total of 1,215,648 cases were stored to the basic image dataset which was used as a repository of norm in this study.

Pathological cases. Similarly to the norm, labeling of abnormal lung cases was done using a keyword match in the textual descriptions. As a result, 22,355 cases were labeled as belonging to the class of pneumosclerosis, 9,285 recognized as emphysema, 19,844 categorized to fibrosis, 5,718 to pneumonia, 2,897 to focal shadows, 821 to bronchitis, and 793 cases was labeled as tuberculosis. It should be noted that one single case may have several labels of abnormality simultaneously. It should be remembered also that we are dealing with the screening data. This particularly means that the above lung disease labels are preliminary and they are not confirmed clinically. In other words, we should admit presence of certain bias which is caused by (natural for the screening) over-estimation of probability of presence of pathological changes. Finally, as a results of the above selection procedure we end up with a total of 46,882 cases presenting various lung abnormalities. In both occasions, only patients aged from 10 to 69 years were considered.

The study groups. In order to compose a balanced study group, X-ray examinations which labeled as Pathology were mapped to the Norm cases using subjects' age and gender. In total, 42 754 healthy subjects with similar age-gender status were selected. Thus, the study group consisted of 89 636 X-ray images. Finally, the study group was split into training (75%) and validation (25%) datasets.

3. IMAGE DATA PREPROCESSING

Preprocessing of the original X-ray images included automated segmentation lungs and intensity normalization. Segmentation was performed using the UNet [2] network model trained on publicly available JSRT [3] dataset. The trained model is available at [4]. Before the inputting into the lung segmentation network the original X-ray images were downsized to 256×256 pixels and normalized using commonly known histogram equalization algorithm. The binary masks of lungs resulted from segmentation were resized in the same way separately. Next, the mean and standard deviation (STD) of pixel intensity values inside of the lung regions were calculated. Intensity normalization was performed based on these statistical moments so that that the lung pixel intensities became zero mean and STD=0.143. Finally, regions of interest (ROIs) were calculated for each image as bounding boxes of the lung masks with certain margins. Subsequently, before inputting into the

classification networks, these ROIs are resized to fit the architecture of each specific network.

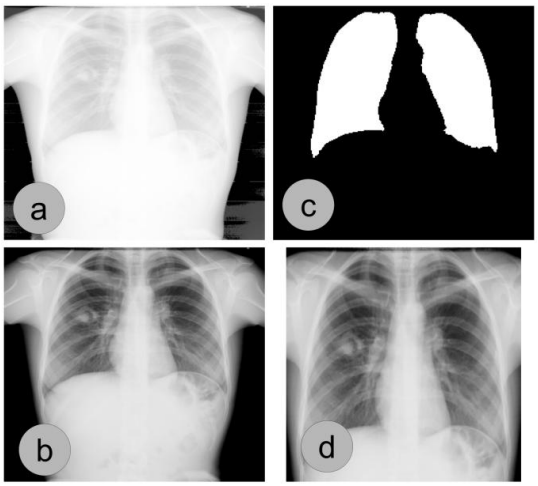

Fig. 1 – Stages of X-ray image preprocessing: a) original X-ray image (full intensity range); b) extracted lung regions; c) normalized by lung pixels intensities; d) cropped ROI

4. METHODS

In this study, the following three popular deep convolutional network architectures were examined: VGG16, VGG19 and InceptionV3. Although each network type can operate with different image sizes, the input image size was set correspondingly to the default input size for each network: 224×224 for VGG16/19, and 299×299 for InceptionV3. The deep networks were evaluated in two different setups described below.

Setup-1. The networks were trained for solving a binary classification problem to predict mutually exclusive image classes: normal and abnormal lungs. For prediction of the presence of abnormalities, the trained network was used as a classifier. In order to predict presence of specific lung diseases (e.g. pneumonia, tuberculosis), the network was used as a feature extractor. Here, the image features were extracted from the network layer prior to the output layer. For VGG16/19 network architecture this was the second fully-connected layer consisting of 4096 elements, and for InceptionV3 this was the global average pooling layer containing 2048 elements. Finally, a Logistic Regression classifier was trained on the feature vectors extracted from the training images and then evaluated on the feature vectors of the validation images. The Logistic Regression was used because it is conceptually agreed with the transform typically used in network architectures.

Setup-2. The networks were trained for a multi-class and multi-label classification problem to predict all the available image class labels including the general Norm-Pathology status and seven abovementioned particular pathologies. This case the network outputs were used to predict the presence of specific lung pathologies directly.

Generation of heatmaps for highlighting suspicious pathology regions. In order to build heatmaps indicating the locations of possible abnormal lung regions we used activations presented in the last convolutional layer of each network. The activations were averaged over the spatial dimensions. The Pearson's correlation coefficients between the "abnormal lung class" binary variable and the average activation values for each channel were calculated. Finally, to build the heatmap for a test image, the activations in each channel were multiplied by the corresponding correlation coefficients and summed up.

5. RESULTS

Training of the deep neural network models was performed on a personal computer equipped with a GPU of Nvidia TITAN X type with 3072 CUDA Cores and 12 GB of GDDR5 onboard memory. Keras framework with Tensorflow backend was used for running training and prediction procedures. The training lasted for 30 epochs with Adam optimizer and Learning Rate set to 10^{-5}. In Setup-1 we used the softmax activation and categorical cross-entropy metrics. A sigmoid activation and binary cross-entropy metrics were used in Setup-2 because they better suited for multi-class and multi-label classification task. On-the-fly image data augmentation was utilized which included appropriate random image rotations and rescaling.

Since the frequency of different labels in the dataset varies significantly, the classification performance was evaluated using the Area Under ROC-Curve metrics (AUC) which is well suited for the unbalanced image datasets.

Setup-1. ROC-curves for Norm vs. Pathology classification calculated over the validation dataset with the use of outputs from three different networks are shown in Fig. 2. Here the networks were trained for binary classification tasks accordingly. The AUC values obtained are 0.851 for VGG16, 0.847 for VGG19 and 0.841 for InceptionV3. Table 1 shows the AUC values for prediction of presence of different lung pathologies independently on the presence of others.

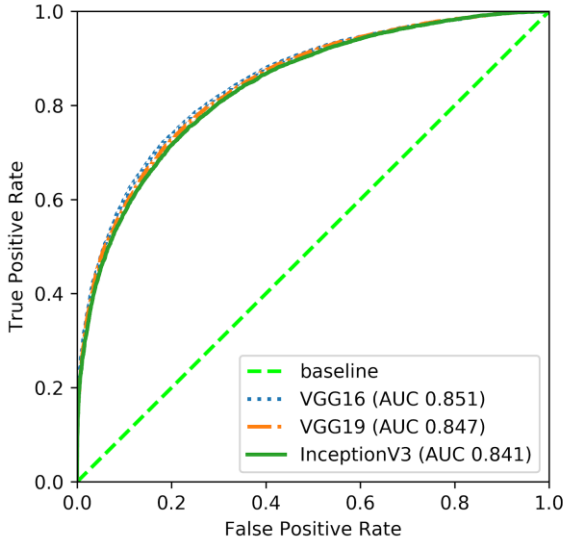

Fig. 2 – ROC-curves for Norm vs. Pathology classification and networks trained for binary classification.

Setup-2. ROC-curves for Norm vs. Pathology classification calculated over the validation dataset are shown in Fig. 3. Here the networks were trained for multi-class multi-label classification tasks and the network outputs from the Pathology class node were used. The AUC values obtained are 0.859 for VGG16, 0.851

for VGG19 and 0.851 for InceptionV3. Table 2 presents the AUC values for prediction of presence of different lung pathologies independently under condition of presence of some others.

Table 1. AUC values for lung disease prediction, obtained using network-driven image descriptors and Logistic Regression classifier.

Pathology	AUC		
	VGG16	VGG19	InceptionV3
bronchitis	0.730	0.701	0.798
emphysema	0.826	0.81	0.841
fibrosis	0.756	0.755	0.757
focal shadows	0.874	0.862	0.880
pneumonia	0.877	0.859	0.884
pneumo-sclerosis	0.838	0.827	0.844
tuberculosis	0.844	0.833	0.806

Table 2. AUC values for predicting lung abnormalities obtained using networks trained for a multi-class and multi-label task.

Pathology	AUC		
	VGG16	VGG19	InceptionV3
bronchitis	0.817	0.795	0.830
emphysema	0.859	0.850	0.861
fibrosis	0.782	0.785	0.771
focal shadows	0.892	0.900	0.900
pneumonia	0.910	0.919	0.912
pneumo-sclerosis	0.863	0.859	0.863
tuberculosis	0.800	0.820	0.829

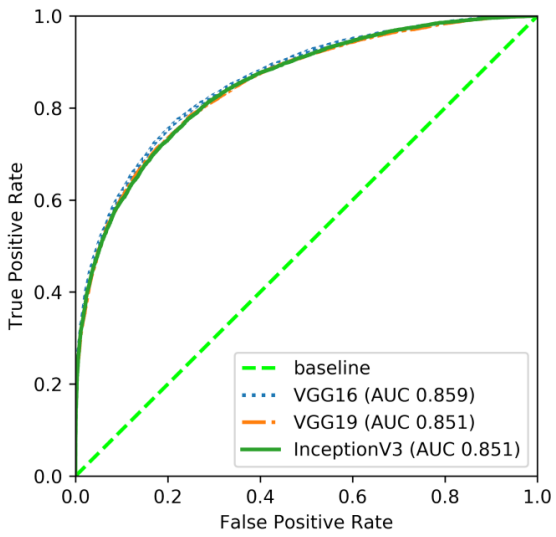

Fig. 3 – ROC-curves for Norm vs. Pathology classification and networks trained for multi-class and multi-label classification.

The heatmaps for localization of abnormalities were built using the networks trained for binary classification task. Fig. 4. illustrates examples of the resultant heatmaps built for three X-ray images from the validation set. The images were labeled as having fibrosis only, pneumonia only and tuberculosis only visible signs.

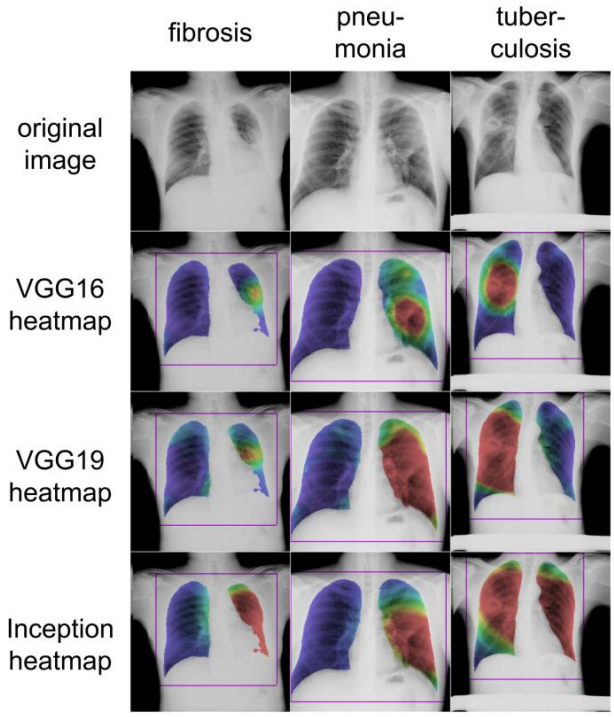

Fig. 4 – Heatmaps built over the validation X-ray images

6. CONCLUSIONS

Results obtained in this study allow drawing the following conclusions.

1. Development of X-ray lung CAD systems providing practically acceptable accuracy of multi-label classification is a challenging problem.

2. Among the three network architectures being examined including VGG16, VGG19, and InceptionV3 the VGG16 has demonstrated better results in generating heatmaps.

3. An additional research work is necessary for tuning global parameters and testing alternative network architectures.

7. REFERENCES

[1] S. Zaidi. Evaluation of the diagnostic accuracy of Computer-Aided Detection of tuberculosis on Chest radiography among private sector patients in Pakistan, *Sci. Reports* 8(12339) (2018). p. 1-11

[2] O. Ronneberger. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany 5-9 October, pp. 234-241.

[3] URL: http://db.jsrt.or.jp/eng.php

[4] URL: https://github.com/imlab-uiip/lung-segmentation-2d