

Classification Of Lung Diseases Using Deep Learning Models*

Matthew Zak¹ and Adam Krzyżak^{1,2}

¹ Department of Computer Science and Software Engineering
Concordia University
Montreal, Canada H3G 1M8

² Department of Electrical Engineering
Westpomeranian University of Technology
70-313 Szczecin, Poland

zak.matthew@yahoo.com, krzyzak@cs.concordia.ca

Abstract. In this paper we address the problem of medical data scarcity by considering the task of detection of pulmonary diseases from chest X-Ray images using small volume datasets with less than thousand samples. We implemented three deep convolutional neural networks (VGG16, ResNet-50, and InceptionV3) pre-trained on the ImageNet dataset and assessed them in lung disease classification tasks transfer learning approach. We created a pipeline that segmented chest X-Ray (CXR) images prior to classifying them and we compared the performance of our framework with the existing ones. We demonstrated that pre-trained models and simple classifiers such as shallow neural networks can compete with the complex systems. We also validated our framework on the publicly available Shenzhen and Montgomery lung datasets and compared its performance to the currently available solutions. Our method was able to reach the same level of accuracy as the best performing models trained on the Montgomery dataset however, the advantage of our approach is a smaller number of trainable parameters. Furthermore, our InceptionV3 based model almost tied with the best performing solution on the Shenzhen dataset despite being computationally less expensive.

Keywords: Lung disease classification · Transfer learning · Deep learning.

1 Pixel/Voxel-Based Machine Learning

The availability of computationally powerful machines allowed emerging methods like pixel/voxel-based machine learning (PML) in medical image analysis/processing. Instead of calculating features from segmented regions, this technique uses voxel/pixel values in input images directly. Therefore, neither segmentation nor feature extraction is required. The performance of PML's can

* Supported by the Natural Sciences and Engineering Research Council of Canada. Part of this research was carried out by the second author during his visit of the Westpomeranian University of Technology while on sabbatical leave from Concordia University.

possibly get higher than common classifiers [16] as this method is able to avoid errors caused by inaccurate segmentation and feature extraction. The most popular powerful approaches include convolutional neural networks (including shift-invariant neural networks) resulted in false positive (FP) rates reduction in computer-aided design framework (CAD) for detection of masses and microcalcifications [12] in mammography and in lung nodule detection in chest X-ray CXR images [13], neural filters and massive-training artificial neural networks including massive-training artificial neural networks (MTANNs) including a mixture of expert MTANNs, Laplacian eigenfunction LAP-MTANN and massive-training support vector regression (MTSVR) for classification, object detection and image enhancement in malignant lung modules detection in CT, FP reduction in CAD for polyp detection in CT colonography, bone separation from soft tissue in CXR and enhancement of lung nodules in CT [11].

2 Bone Separation from Soft Tissue in Chest Radiographs using MTANNs

Chest X-Ray is one of the most frequently used diagnosis when examining medical images for different lung diseases such as pneumonia or tuberculosis. Roughly 1 million of adults require hospitalization because of pneumonia, and about 50,000 dies from this disease annually in the US only. Examination of lung nodules in CXR can lead to missing of diseases like lung cancer. However, not all of them are visible in retrospect. Studies show that 82-95% of lung cancer cases were missed due to occlusions (at least partial) by ribs or clavicle. To address this problem researchers examined dual-energy imaging, a technique which can produce images of two tissues, which is "soft-tissue" image and "bone" image. This technique has many drawbacks, but undoubtedly one of the most important ones is the exposure to radiation.

The MTANNs models have been developed to address this problem and serve as a technique for ribs/soft-tissue separation. The idea behind training of those algorithms is to provide them with bone and soft-tissue images obtained from a dual-energy radiography system. The MTANN was trained using CXRs as input and corresponding boneless images. The ribs contrast is visibly suppressed in the resulting image, maintaining the soft tissue areas such as lung vessels.

3 Deep Learning Approaches In Chest X-Ray Analysis

Recent developments in Deep Neural Networks [2] lead to major improvements in medical imaging. The efficiency of dimensionality reduction algorithms like lung segmentation was demonstrated in the chest X-Ray image analysis. Recently researchers aimed at improving tuberculosis detection on relatively small data sets of less than 103 images per class by incorporating deep learning segmentation and classification methods from [4]. We will further explore these techniques in this paper.

3.1 Dataset

In this paper we combine two relatively small datasets containing less than 103 images per class for classification (pneumonia and tuberculosis detection) and segmentation purposes. We selected 306 examples per "disease" class (306 images with tuberculosis and 306 images with pneumonia) and 306 of healthy patients contributing to a set of 918 samples from different patients. Sample images from both datasets are shown in Figure 1.

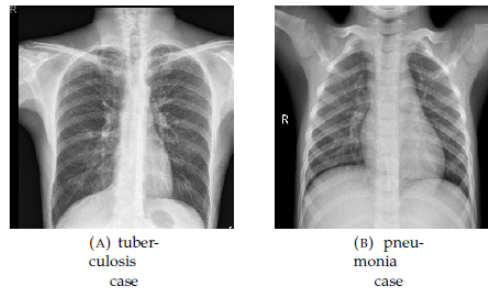


Fig. 1. Sample Chest X-Ray images containing traces of tuberculosis (A) and pneumonia (B) from Shenzhen dataset [2, 6].

The Shenzhen Hospital dataset (SH) [2, 6] containing CXR images was created by the People's Hospital in Shenzhen, China. It includes both abnormal (containing traces of tuberculosis) and standard CXR images. Unfortunately, the dataset is not well-balanced in terms of absence or presence of disease, gender, or age. We extracted only 153 samples of healthy patients (153 from both datasets) and 306 of those labeled with traces of tuberculosis. Selecting information about one class from different resources ensures that the model is not contaminated by the features resulting from the method of taking images, e.g., the lens.

Pneumonia is an inflammatory condition of the lung affecting the little air sacs known as alveoli. Standard symptoms comprise of a blend of a dry hacking cough, inconvenience breathing, chest agony, and fever. The Labeled Optical Tomography and Chest X-Ray Images for Classification dataset [9] includes selected images of patients from the Medical Center in Guangzhou. It consists of data with two classes - normal and those containing marks of pneumonia. All data come from the patient's routine clinical care. The volume of the complete dataset includes thousands of validated optical coherence tomography (OCT) and X-ray images yet for our analysis we wanted to keep the dataset tiny and evenly distributed thus only 153 images were selected (other 153 images come from the tuberculosis dataset) from the resources labeled as healthy and 306 as pneumonia - both chosen randomly. The exact dataset class distribution is presented in external segmentation of left and right lung images (exclusion of

redundant information: bones, internal organs, etc.) was proven to be effective in boosting prediction accuracy. To extract lungs information and exclude outside regions, we used the manually prepared masks included in the extension of the SH dataset, namely, the segmented SH dataset, see Figure 2. Due to nonidentical borders and lung shapes, the segmentation data has high variability although its distribution is quite similar to the regular one when compared to image area distribution.

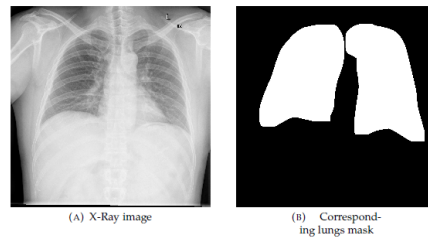


Fig. 2. An X-Ray image and its corresponding lungs mask.

3.2 Image Data Augmentation

Model-based methods greatly improve their predictions when the number of training samples grows. When a limited amount of data is available, some transformations have to be applied to the existing dataset to synthetically augment the training set. Researchers in [10] employed three techniques to augment the training dataset. The first approach was to randomly crop of a 224x224 pixel fixed-size window from a 256x256 pixel image. The second technique was flipping the image horizontally, which allowed capturing information about reflection invariance. Finally, the third method added randomly generated lighting to capture color and illumination variation.

3.3 Transfer Learning In Lung Diseases Classification

Transfer learning is a very popular approach in computer vision related tasks using deep neural networks when data resources are scarce. Therefore, to launch a new task, we incorporate the pre-trained models skilled in solving similar problems. This method is crucial in medical image processing due to the shortage of real samples. In deep neural networks, feature extraction is carried out but passing raw data through models specialized in other tasks. Here, we can refer to deep learning models such as ResNet, where the last layer information serves as input to a new classifier. Transfer learning in deep learning problems can be performed using a common approach called pre-trained models approach. Reuse Model states that pre-trained model can produce a starting point for another

model used in a different task. This involves incorporation of the whole model or its parts. The adopted model may or may not need to be refined on the input-output data for the new task. The third option considers selecting one of the available models. It is very common that research institutions publish their algorithms trained on challenging datasets which may fully or partially cover the problem stated by a new task.

ImageNet [3] is a project that helps computer vision researches in classification and detection tasks by providing them with a large image dataset. This database contains roughly 14 million different images from over 20 thousand classes. ImageNet also provides bounding boxes with annotations for over 1 million images, which are used in object localization problems.

In this work, we will experiment with three models (VGG, ResNet, and Inception) pre-trained on the ImageNet dataset.

3.4 Deep Nets In Lung Diseases Classification

The following deep nets have been considered: VGG16, ResNet-50 and InceptionV3. The VGG convolutional network is a model with 16 layers trained on fixed size images. The input is processed through a set of convolution layers which use small-size kernels with a receptive field 3x3. This is the smallest size allowing us to capture the notion of up, down, right, left, and center. The architecture also incorporates 1x1 kernels which may be interpreted as linear input transformation (followed by nonlinearity). The stride of convolutions (number of pixels that are shifted in every convolution - step size) is fixed and set to 1 pixel; therefore the spatial resolution remains the same after processing an input through a layer, e.g., the padding is fixed to 1 for 3x3 kernels. Spatial downsizing is performed by five consecutive pooling (max-pooling) layers, which are followed by some convolution layers. However, not all of them are followed by max-pooling. The max-pooling operation is carried over a fixed 2x2 pixel window, with a stride of 2 pixels. This pile of convolutional layers ends with three fully-connected (FC) layers where the first two consist of 4096 nodes each and the third one of 1000 as it performs the 1000-way classification using softmax. All hidden layers have the same non-linearity ReLU (rectified linear unit) [10].

The ResNet convolutional neural network is a 50-layer deep model trained on more than a million fixed-size images from the ImageNet dataset. The network classifies an input image into one of 1000 object classes like car, airplane, horse or mouse. The network has learned a large amount of features thanks to training images diversity and achieved 6.71% top- 5 error rate on the ImageNet dataset. The ResNet-50 convolutional neural network is built of 5 stages, each having convolutions and identity blocks. Every convolution block consists of 3 convolutional layers. ResNet-50 is related to ResNet-34, however, the idea behind its sibling model remains the same. The only difference is in residual blocks; unlike those in ResNet-34 ResNet-50 replaces every two layers in a residual block with a three-layer bottleneck block and 1x1 convolutions, which reduce and eventually restore the channel depth. This allows reducing a computational load when a 3x3 convolution is calculated. The model input is first processed through a layer

with 64 filters each 7x7 and stride 2 and downsized by a max-pooling operation, which is carried over a fixed 2x2 pixel window, with a stride of 2 pixels. The second stage consists of three identical blocks, each containing a double convolution with 64 3x3 pixels filters and a skip connection block. The third pile of convolutions starts with a dotted line as there is a change in the dimensionality of an input. This effect is achieved through the change of stride in the first convolution bloc from 1 to 2 pixels. The fourth and fifth groups of convolutions and skip connections follow the pattern presented in the third stage of input processing, yet they change the number of filters (kernels) to 256 and 512, respectively. This model has over 25 million parameters.

The researchers from Google introduced the first Inception (InceptionV1) neural network in 2014 during the ImageNet competition. The model consisted of blocs called "inception cell" that was able to conduct convolutions using different scale filters and afterward aggregate the results as one. Thanks to 1x1 convolution which reduces the input channel depth the model saves computations. Using a set of 1x1, 3x3, and finally, 5x5 size of filters, an inception unit cell learns extracting features of different scale from the input image. Although inception cells use max-pooling operator, the dimension of a processed data is preserved due to "same" padding, and so the output is properly concatenated.

A follow-up paper [17] was released not long after introducing a more efficient solution to the first version of the inception cell. Large filters sized 5x5, and 7x7 are useful in extensive spatial features extraction, yet their disadvantage lies in the number of parameters and therefore computational disproportion.

The InceptionV3 model contains over 23 million parameters. The architecture can be divided into 5 modules. The first processing block consists of 3 inception modules. Then, information is passed through the effective grid size reduction and processed through four consecutive inception cells with asymmetric convolutions. Moving forward, information flows to the 17x17 pixels convolution layer connected to an auxiliary classifier and another effective grid size-reduction block. Finally, data progresses through a series of two blocs with wider filter banks and consequently gets to a fully-connected layer ended with a Softmax classifier. Visualization of the network architecture can be found in Figure 3.

4 Experiments

4.1 Image Segmentation Using Deep Neural Networks

Many vision-related tasks, especially those from the field of medical image processing expect to have a class assigned to every pixel, i.e., every pixel is associated with a corresponding class. To conduct this process, we propose so-called U-net neural network architecture described in [17]. This model works well with very few training image examples yielding precise segmentation. The motivation behind this network is to utilize progressive layers instead of a building system, where upsampling layers are utilized instead of pooling operators, consequently increasing the output resolution. High-resolution features are combined with the upsampled output to perform localization. The deconvolution layers consist of a high number of kernels, which better propagate information and result in

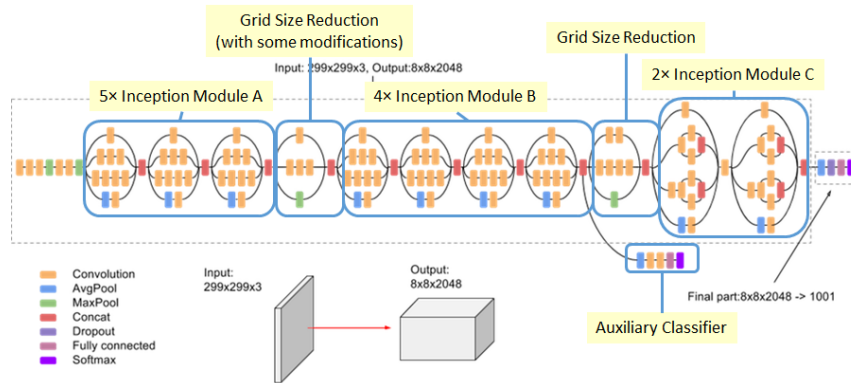


Fig. 3. InceptionV3 architecture. Batch normalization and ReLU units are used after every convolution layer.

outputs with higher resolution. Owing to the described procedures, the deconvolution path is approximately symmetric to the contracting one and so the architecture resembles the U shape. There are no fully connected layers, therefore, making it possible to conduct the seamless segmentation of relatively large images extrapolating the missing context by mirroring the processed input.

4.2 U-net Architecture

The network showed in Figure 4 consists of an expansive path (right) and a contracting one(left). The first part (contracting) resembles a typical convolutional neural network; the repeated 3x3 convolutions followed by a non-linearity (here ReLU), and 2x2 pooling with stride 2. Each downsampling operation doubles the number of resulting feature maps. All expansive path operations are made of upsampling of the feature channels followed by a 2x2 deconvolution (or "up-convolution") which reduces the number of feature maps twice. The result is then concatenated with the corresponding feature layer from the contracting path and convolved with 3x3 kernels, and each passed through a ReLU. The final layers apply a 1x1 convolution to map each feature vector to the desired class.

4.3 Lung Segmentation

Following the approaches presented in the literature we wanted to use deep convolutional neural networks to segment lungs [8] before processing it through the classification models mentioned in Section 2.4. Researchers in [8] indicate that U-Net architecture and its modifications outperform the majority of CNN-based models and achieve excellent results by easily capturing spacial information about the lungs. Thus, we propose a pipeline that consists of two stages; first segmentation and then classification.

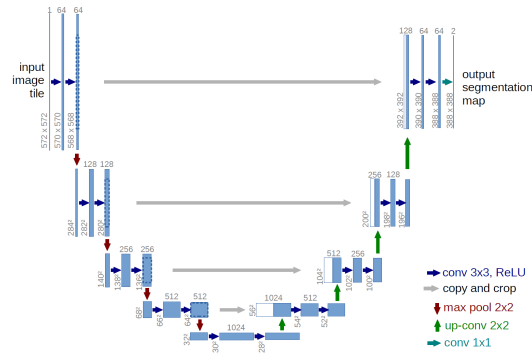


Fig. 4. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multichannel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

4.4 Dataset

The phase of extracting valuable information (lungs) is conducted with a model presented in Section 3.2. Our algorithms trained for 500 epochs on an extension of the SH dataset. The input to our U-shaped deep neural network is a regular chest X-Ray image, whereas the output is a manually prepared binary mask of lung shape, matching the input.

4.5 Software and Hardware

The code for the transfer-learning models is publicly available through a python API, Keras. Our algorithms were trained on servers equipped with GPU provided by Helios Calcul Québec, which consists of fifteen computing nodes each having eight Nvidia K20 GPUs and additionally six computing nodes with eight nVidia K80 boards each. Every K80 board includes two GPU's and so the total of 216 GPU's in the cluster.

4.6 Training

As mentioned before, our model was trained for 500 epochs using a dataset partitioned into 80%, 10%, and 10% bins, for training, validation and test parts, respectively using the models introduced in Section 2.4 using the batch size of 8 samples, augmentation techniques briefed in Section 2.2, Adam optimizer and categorical cross-entropy as a loss function for pixel-wise binary classification. The training results are shown in Figure 5. As we can easily notice, the validation error is slowly falling throughout the whole training, whereas there is no major change after the 100th epoch. The final error on the validation set is right below 0.05 and slightly above 0.06 on the test set.

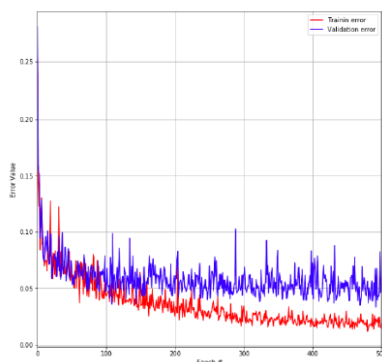


Fig. 5. U-Net training and validation losses change during training.

4.7 Segmentation Results

Our algorithm learns shape-related features typical for lungs and can generalize well further over unseen data. Figure 6 shows the results of our U-Net trained models. It is clear that the network was able to learn chest shape features and exclude regions containing internal organs such as the heart. The incredibly promising results allowed us to process the whole dataset presented earlier and continue our analysis on the newly processed images.

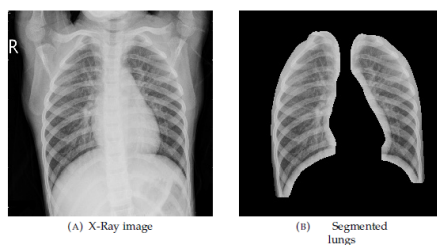


Fig. 6. (A) training X-ray lung image and (B) segmented lung image.

5 Training Deep Learning Models On Segmented Images

We propose a two-stage pipeline for classifying lung pathologies into pneumonia and tuberculosis consisting of two stages: first for chest X-ray image segmentation and second for lung disease classification. The first stage (segmentation) is trained during experiments described in the previous section. The second stage utilizes deep models described in Section 2, whereas we investigate potential

improvements in performance depending on the type of model used. Our classification models were trained using the same setup as described in Section 2. Here, we conduct our experiments using the data described in Section 2.1. The difference is in prior segmentation, which extracts valuable information for the task, namely lungs. Figure 6 shows the training samples; the left and right column correspond to input and output, respectively.

5.1 Inception Results

We tried all models with three deep net classifiers (VGG16, ResNet-50, InceptionV3) in the task of classification of lung images into two classes: pneumonia and tuberculosis. We observed that InceptionV3 based model performed best and thus due to lack of space we display only its performance results. The confusion matrix in Figure 8 (A) shows that the new model improved the number of true positives (TP) in all classes in comparison with the VGG16 and ResNet-50 based models. Image (B) shows that the AUC score for healthy, tuberculosis and pneumonia cases were 90%, 93%, and 99%, respectively.

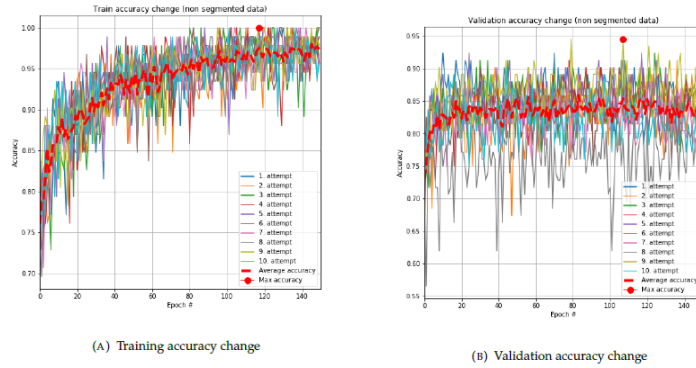


Fig. 7. InceptionV3 based model training and validation accuracy versus number of epochs.

5.2 Comparison Of Results On Non-Segmented And Segmented Images

After comparing the results obtained by models without transfer learning we observe that transfer-learning models perform well in lung diseases classification using segmented images tasks even when the data resources are scarce. In this section, we compare the performance of our models to the results achieved in the literature over different datasets.

The algorithm that scored the best in the majority results was InceptionV3 trained on the segmented images. What is more, it produced incredibly high

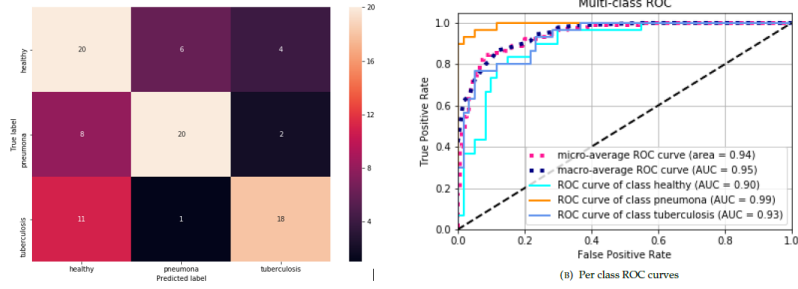


Fig. 8. Classification results obtained by InceptionV3 model on segmented lung images.

| | Non segmented CXRs | | | Segmented CXRs | | |
|----------------------------|--------------------|-----------|-------------|----------------|-------------|-------------|
| | VGG16 | ResNet-50 | InceptionV3 | VGG16 | ResNet-50 | InceptionV3 |
| Accuracy | 0.64 | 0.72 | 0.81 | 0.70 | 0.75 | 0.82 |
| AUC (healthy) | 0.68 | 0.84 | 0.89 | 0.75 | 0.77 | 0.90 |
| AUC (pneumonia) | 0.80 | 0.84 | 0.92 | 0.81 | 0.91 | 0.99 |
| AUC (tuberculosis) | 0.82 | 0.76 | 0.87 | 0.90 | 0.82 | 0.93 |
| F1 score (healthy) | 0.55 | 0.66 | 0.76 | 0.55 | 0.62 | 0.76 |
| F1 score (pneumonia) | 0.64 | 0.82 | 0.90 | 0.84 | 0.89 | 0.93 |
| F1 score (tuberculosis) | 0.72 | 0.66 | 0.75 | 0.79 | 0.73 | 0.78 |
| precision (healthy) | 0.49 | 0.77 | 0.80 | 0.67 | 0.66 | 0.75 |
| precision (pneumonia) | 0.87 | 0.76 | 0.89 | 0.92 | 0.93 | 0.90 |
| precision (tuberculosis) | 0.71 | 0.67 | 0.74 | 0.68 | 0.71 | 0.81 |
| sensitivity (healthy) | 0.62 | 0.60 | 0.73 | 0.48 | 0.62 | 0.77 |
| sensitivity (pneumonia) | 0.53 | 0.88 | 0.92 | 0.78 | 0.85 | 0.95 |
| sensitivity (tuberculosis) | 0.77 | 0.68 | 0.77 | 0.95 | 0.78 | 0.75 |

Fig. 9. Comparison of all results obtained for segmented and non-segmented data.

scores for the "disease" classes showing that a random instance containing marks of tuberculosis or pneumonia has over 90% probability to be classified to the correct class. Although the scores of the healthy class are worse than the diseased ones, its real cost is indeed lower as it is always worse to classify a sick patient as healthy. The InceptionV3 based model scored the best, reaching better accuracy than VGG16 algorithms by over 12 percentage. Although the interpretability of our methods is not guaranteed, we can clearly state that using transfer-learning based algorithms on small datasets allows achieving competitive classification scores on the unseen data. Furthermore, we compared the class activation maps shown in Figure 10 in order to investigate the reasoning behind decision making. The remaining features, here lungs, force the network to explore it and thus make decisions based on observed changes. That behavior was expected and additionally improved the interpretability of our models as the marked regions might bring attention of the doctor in case of sick patients.

5.3 Comparison With Other Works

In this section, we compare performance of our models with the results in the literature using over different datasets. In order to do so we trained our algorithms on the Shenzhen and Montgomery datasets [6] ten times, generated the

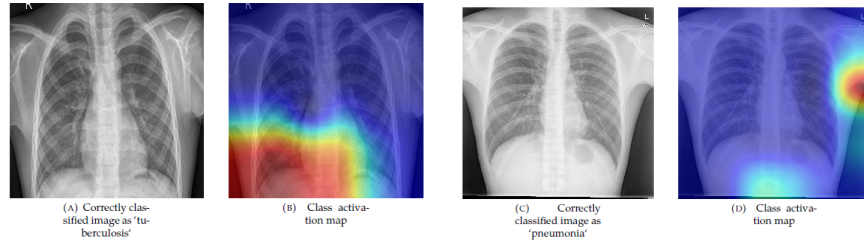


Fig. 10. Two pairs of correctly labelled images containing traces of tuberculosis and pneumonia with their class activation maps.

results for all the models and averaged their scores: accuracy, precision, sensitivity, specificity, F1 score and AUC. Table 1 presents comparison of different deep learning models trained on the Shenzhen dataset [6]. Although our approach does not guarantee the best performance, it is always close to the highest even though it is typically less complex. Researchers in [5] used various pre-trained models in the pulmonary disease detection task, and the ensemble of them yields the highest accuracy and sensitivity. To compare, our InceptionV3-based model achieves accuracy smaller by only one percent and has identical AUC, which means that our method gives an equal probability of assigning a positive case of tuberculosis to its corresponding class over a negative sample. Although we could not outperform the best methods, our framework is less complicated. Furthermore, in Table 2 we compared the performance of our framework trained on the Montgomery dataset [6] to the literature. Our InceptionV3-based model tied with [14] in terms of accuracy, yet showed higher value of AUC. ResNet-50 and VGG16 based models performed worse, however not by much as they reached accuracies of 76% and 73% respectively, which is roughly 3 and 6 percent less than the highest score achieved.

| Model | Accuracy | Precision | Sensitivity | Specificity | F1 score | AUC |
|------------------|----------|-----------|-------------|-------------|----------|------|
| [1] | 0.82 | - | - | - | - | - |
| [14] | 0.84 | - | - | - | - | 0.90 |
| VGG16 [5] | 0.84 | - | 0.96 | 0.72 | - | 0.88 |
| ResNet-50 [5] | 0.86 | - | 0.84 | 0.88 | - | 0.90 |
| ResNet-152 [5] | 0.88 | - | 0.80 | 0.92 | - | 0.91 |
| Ensemble [5] | 0.90 | - | 0.88 | 0.92 | - | 0.94 |
| VGG16 [5] | 0.84 | 0.88 | 0.80 | 0.89 | 0.83 | 0.86 |
| ResNet-50 [5] | 0.85 | 0.97 | 0.73 | 0.98 | 0.83 | 0.92 |
| InceptionV3 [17] | 0.89 | 0.96 | 0.80 | 0.97 | 0.88 | 0.94 |

Table 1. Comparison of different deep learning based solutions trained on the Shenzhen dataset [6]. Although our result is not the best, it performs better than any single model (excluding Ensemble). Horizontal line means that corresponding results were not provided in literature.

| Model | Accuracy | Precision | Sensitivity | Specificity | F1 score | AUC |
|------------------|----------|-----------|-------------|-------------|----------|-------|
| [14] | 0.790 | - | - | - | - | 0.811 |
| [15] | 0.674 | - | - | - | - | 0.884 |
| [7] | 0.783 | - | - | - | - | 0.869 |
| VGG16 [5] | 0.727 | 0.842 | 0.581 | 0.872 | 0.669 | 0.931 |
| ResNet-50 [5] | 0.764 | 0.814 | 0.691 | 0.836 | 0.744 | 0.891 |
| InceptionV3 [17] | 0.790 | 0.822 | 0.745 | 0.836 | 0.779 | 0.884 |

Table 2. Comparison of different deep learning based solutions trained on the Montgomery dataset [6]. Our average performance is almost identical to [14].

6 Conclusions

We created lung diseases classification pipeline based on transfer learning that was applied to small datasets of lung images. We evaluated its performance in classification of non-segmented and segmented chest X-Ray images. In our framework we used U-net segmentation network and InceptionV3 deep model classifier. Our framework was compared with the existing models. We demonstrated that models pre-trained by transfer learning approach and simple classifiers such as shallow neural networks can successfully compete with the complex systems.

References

1. R. H. Anuj Rohilla and A. Mittal, Tb detection in chest radiograph using deep learning architecture, *International Journal of Advance Research in Science and Engineering*, vol. 6, pp. 1073–1084, August 2017.
2. S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration, *IEEE Transactions on Medical Imaging*, vol. 33, pp. 577–590, February 2014.
3. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, Imagenet: A large-scale hierarchical image database, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR’2009)*, pp. 248–255, 2009.
4. Y. Gordienko, P. Gang, J. Hui, W. Zeng, Y. Kochura, O. Alienin, O. Rokovyi, and S. Stirenko, Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer, *Computing Research Repository*, vol. abs/1712.07632, 2017.
5. M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, Abnormality detection and localization in chest x-rays using deep convolutional neural networks, *ArXiv*, vol. abs/1705.09850, 2017.
6. S. Jaeger, S. Candemir, S. Antani, Y.-X. Wang, P.-X. Lu, and G. Thoma, Two public chest x-ray datasets for computer-aided screening of pulmonary diseases, *Quantitative Imaging in Medicine and Surgery*, vol. 4, pp. 475–7, December 2014.
7. S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y. Wang, P. Lu, and C. J. McDonald, Automatic tuberculosis screening using chest radiographs, *IEEE Transactions on Medical Imaging*, vol. 33, pp. 233–245, February 2014.

8. Q. Kang, Q. Lao, and T. Fevens, Nuclei segmentation in histopathological images using two-stage learning, *Medical Image Computing and Computer Assisted Intervention (MICCAI'2019)*, pp. 703-711, October 2019.
9. K. Z. D. Kermany and M. Goldbaum, Large dataset of labeled optical coherence tomography (oct) and chest X-ray images, *Cell*, pp. 1122– 1131, 2018.
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems (NIPS'2012)*, vol. 25, pp. 1097–1105, 2012.
11. F. Li, H. Arimura, K. Suzuki, J. Shiraishi, Q. H. Li, H. Abe, R. M. Engelmann, S. Sone, H. MacMahon, and K. Doi, Computer-aided detection of peripheral lung cancers missed at ct: ROC analyses without and with localization., *Radiology*, vol. 237, no. 2, pp. 684–90, 2005.
12. S.-C. B. Lo, H. Li, Y. J. Wang, L. Kinnard, and M. T. Freedman, A multiple circular path convolution neural network system for detection of mammographic masses, *IEEE Transactions on Medical Imaging*, vol. 21, pp. 150–158, 2002.
13. S.-C. B. Lo, H.-P. Chan, J.-S. Lin, H. Li, M. T. Freedman, and S. K. Mun, Artificial convolution neural network for medical image pattern recognition, *Neural Networks*, vol. 8, pp. 1201–1214, December 1995.
14. F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization, *Scientific Reports*, vol. 9, pp. 1–9, December 2019.
15. K. Sangheum Hwang, Hyo-Eun Kim, A novel approach for tuberculosis screening based on deep convolutional neural networks, *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785, pp. 1–23, 2016.
16. K. Suzuki, Pixel-based machine learning in medical imaging, *International Journal of Biomedical Imaging*, vol. 2012, pp. 1–1, February 2012.
17. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the inception architecture for computer vision, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2016)*, pp. 2818–2826, June 2016.