Explanation of the Convolutional Neural Network Classifying Chest X-Ray Images Supporting Pneumonia Diagnosis

Hai Thanh Nguyen[1], Hoang Thanh Huynh[1], Toan Bao Tran[2,3], Hiep Xuan Huynh[1,*]

[1] Can Tho University, Can Tho, Vietnam
[2] Center of Software Engineering, Duy Tan University, 550000 Vietnam
[3] Institute of Research and Development, Duy Tan University, 550000 Vietnam

Abstract

Medical images are valuable sources for disease diagnosis. Besides, advancements in deep learning in recent years have been supporting disease diagnosis methods based on images to obtain numerous achievements. However, deep learning algorithms still work as a black-box so it is difficult to interpret output from these algorithms. In this study, we propose a convolutional network architecture to classify Chest X-ray images as well as apply explanation approaches for trained models to support disease diagnosis. The proposed method provides insights in medical images to support Pneumonia diagnosis.

*Corresponding author. Email: hxhiep@ctu.edu.vn

1. Introduction

In recent years, the rapid increase in the development of technologies and a large number of available data in medical contribute to improving treatment outcomes, make it more specific and effective. The medical data are very diverse and abundant, the most popular including images or digital. The medical images, specifically with X-ray imaging, magnetic resonance imaging (MRI), computed tomography (CT) are used widely in clinical analysis, medical intervention, and easy to detect the anomalies by creating visual representations of the inner structures of the body. The outbreaks of SARS-CoV-2 in 2019 have an unintended consequence and the attempts to contain the disease by using medical images are very necessary [1, 2]. X-ray (also commonly referred to as chest radiography or CXR) may be considered as the most common method for the identification of chest and nearby structures abnormalities. Furthermore, the American College of Radiology notes that chest x-ray can be used to minimize the risk of cross-infection [3]; according to [4], chest x-ray produces millions of scans performed globally every year. The improvements of deep learning bring medical technology to light with promising results, for instance, medical events prediction [5, 6], antibiotic discovery [7], or analysis of electronic health records [8]. Moreover, the improvements of image classification, image segmentation [9, 10] offering plenty of encouragement for the developments of medical imaging. Several deep learning approaches in medical imaging have been proposed for disease detection and diagnosis, skin cancer classification [11], or a deep encoder-decoder architecture for 3D image biomedical segmentation [12]. However, the annotation progress in the medical image is based on medical professional knowledge, medical industry standard, and

1

medical system [13]. Thus, it requires a large human and material resources.

2. A literature review on medical images processing

In the field of medical processing, the experimental results of the author in [10] show the improvements of performance on hyperspectral image classification tasks based on a deep feature manifold embedding method. The first extracted features are discovered the discriminant manifold structure by an intrinsic and a penalty graph, then those features are mapped into a low-dimensional embedding space. The results demonstrated that the proposed method is effective. The following approach also works in hyperspectral images with sparse representation and dictionary learning [14]. This approach can learn the sub-dictionary effectively for each class and the multi-scale strategy. The shape of the regions would not be adapted in multi-scale sparse representation based on context structure. There are two process stages, dictionary learning stage, and sparse representation stage. The learning dictionary stage ensures that contain more discriminative information while the sparse representation stage can exploit the spatial context. This combination can explore more information in hyperspectral images and has a great performance. Besides, it can deal with the imbalanced data problem in the hyperspectral image. The authors in [13] proposed semi-supervised learning for CT pathological images analysis of the brain and chest. The generated antagonism network was trained with a small amount of labelled data and combined the extracted features for classifying. This approach reached higher performance in comparison to Convolutional Neural Network (CNN) or the other traditional image classification models. Several different approaches in chest radiography are proposed, the purpose of the study in [15] is to detect the abnormalities and assess the change in findings over serial radiographs from 724 patients. Two thoracic radiologists assessed all the images for abnormalities, four test radiologists assessed the presence of radiographic abnormalities to establish a standard of reference (SOR). The authors used QureAI which is based on CNN to evaluate their study and the results showed that no statistical difference between CNN and SOR but the area under the curve (AUC) is quite a difference. The accuracy of CNN is affected negatively by the presence of the medical devices of the chest. Thus, deep learning with its limitations is unlikely to replace radiologists but it can be helpful for the interpretation of CXR findings. The authors in [16] detected the pneumoconiosis from chest X-ray images in the form of digital radiography by applied CNN. The datasets contain 1881 images and collected from subjects who have worked in a dusty environment. Furthermore, the certified radiologists are involved in this study and compared their performance with CNN. The results demonstrated that CNN had a good performance and can be a solution in screening pneumoconiosis.

In this study, we present a novel method based on the explanation of CNN to support pneumonia diagnosis on chest x-ray images, our contributions include:

- We proposed a method to classify the pneumoconiosis patients and normal from X-ray images with a shallow Convolutional Neural Network.

- We combined the Gradient Class Activation Map (Grad-CAM) [18] with our CNN as a novel method for detecting the abnormalities regions in chest x-ray images and aiding the interpretative decision by visualizing them into images. The method is expected to a good tool for disease diagnosis based medical images.

The remaining of this work is presented as follows. Section 3 introduces the architecture of our Convolutional Neural Networks for supporting pneumonia diagnosis and details of the used dataset. We explain the predictions and our experimental results in Section 4 and 5 respectively. Finally, we discuss and summarize our study in Section 6.

3. Deep learning architecture for predicting Chest X-Ray Images (Pneumonia)

3.1 Dataset

We used a dataset contains 5840 chest X-ray images with 1575 images of normal chest and 4265 images of the pneumoconiosis patients. Chest X-ray images were collected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou [23]. Initially, all chest X-ray images were screened for monitoring the quality by removing all low quality or hard-to-read scans. Then, two expert physicians classified the diagnosis for the images before being cleared. The sample of normal and pneumonia chest X-ray images are visualized in Figure 1. The normal chest (left image) describes a clear lung without any abnormalities regions in the image whereas Pneumonia (right image) exhibits a focal lobar consolidation.

3.2 Learning model

We trained our Convolutional Neural Networks with structures which are illustrated in Figure 2. The CNN contains one Convolutional layer, followed by a Max-Pooling layer and a Fully Connected layer. For more specific, the convolutional layer is related to feature extraction, it generates the feature maps by performing convolution operations between the input and the filter. Our convolutional layer contains 64 filters or kernels, the filter itself is a 3×3 integers matrix. The Max-Pooling layer is used to reduce the dimensions of the generated feature maps and perform the process of extracting a max
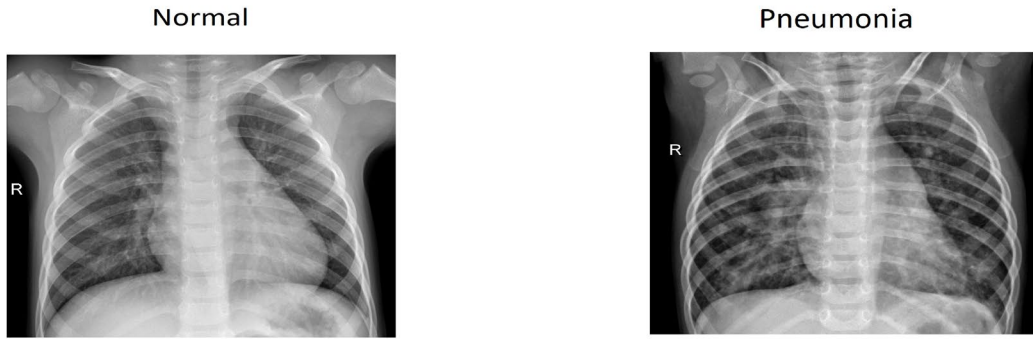
Figure 1. Visualizations of X-ray images of Normal and Pneumonia chest.
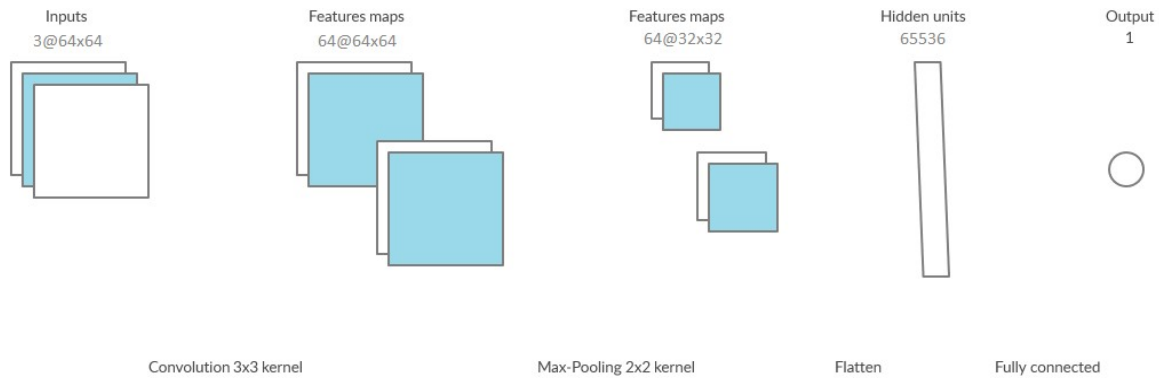
Figure 2. A shallow convolutional Neural Network Architecture.

value from a set of values. We used the Max-Pooling layer with a size of 2×2. The output from the Max-Pooling layer will be flattened into a $1D$ array which forms the input matrix for the Fully Connected layer.

Furthermore, CNN is implemented with Adam optimizer [17] with the default learning rate of 0.001 and a batch of size 16. The proposed CNN architecture has been used in disease prediction tasks [19, 20] and achieved promising results. Furthermore, we used the input of size 64×64, a shallow architecture should be able to work well in this task.

We computed binary cross-entropy loss during training by the formula 1. The goal of cross-entropy loss is to compare the probability distribution output with the true label and evaluate how good are the predicted probabilities. For binary classification tasks, the typical loss function is the binary cross-entropy as we mentioned above.

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (1)$$

Where y_i and $\hat{y}_i$ denote the label and the predicted results respectively.

Alongside with loss function, the selection of optimization algorithms is also important. Adam

optimization is an extension to stochastic gradient descent, showing its advantages in terms of speed of training. Furthermore, Adam combines the characteristic of the AdaGrad [21] and RMSProp [22] that can handle sparse gradients on noisy problems. To avoid overfitting issues, we also used Early Stopping method.

4. Explanations for the predictions

To visualize the areas of abnormal in the images, we combined Grad-CAM with our CNN for detecting characteristic features from X-ray images. In other words, Grad-CAM can detect the areas where CNN focused on and paid more attention to the image. Specifically, let A_{ij}^k represents for the K generated feature maps and refers to the activation at location (i,j), and y^c denotes the score for class c. The heatmap ware generated by computing the importance weights captures the information of the feature map k for class c, by the formula 2:

$$\frac{1}{Z}\sum_{i}\sum_{i}\frac{\partial y^c}{\partial A_{ij}^k} \quad (2)$$

Where Z is the number of pixels in feature map.
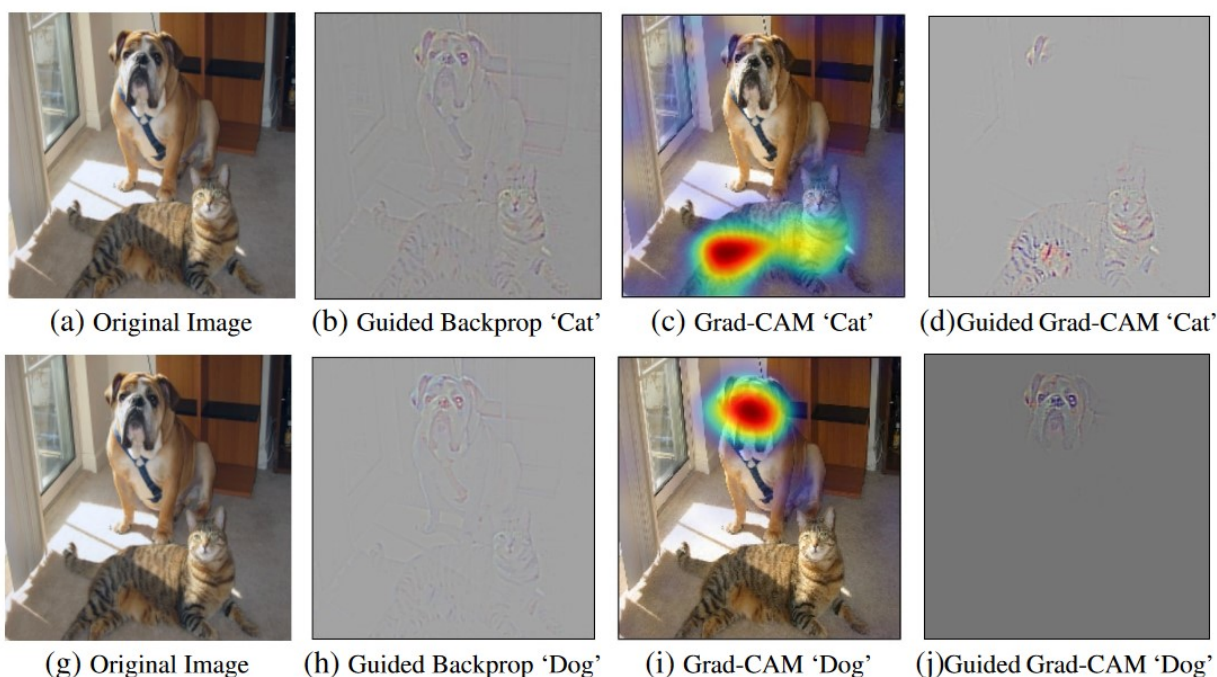
Figure 3. (a) and (g) are original image with cat and dog. (b) and (h) illustrate Guided Backpropagation of cat and dog respectively. The Grad-CAM are visualized in (c) and (i), whereas the combination of Grad-CAM and (b), (c) are presented in (d) and (j) [18].

Figure 3 [18] presents the visualizations of cat and dog class. Figure 3b and 3h are not class-discriminative but they are very similar and generated by Guided Backpropagation [24] and Deconvolution [25] respectively. Otherwise, Figure 3c and Figure 3i focused on "cat" and "dog" due to applying Grad-CAM method. The authors in [18] combined the Grad-CAM with Guided Back-propagation and Deconvolution method to create Guided Grad-CAM and visualized the results in Figure 3d and Figure 3j.

The heatmap presents the critical regions on X-ray images and helps the clinicians make faster and more accurate diagnoses. Summarily, the consistent visualizations of heatmap depend on the accuracy of the model due to Grad-CAM conjoined features precisely based on the generated feature map.

Figure 4 illustrates the normal chest and the heatmap one by one. The first row contains the X-ray chest images, the second row presents the visualizations of the abnormalities regions, and so on. The images with tag 79, 81, and 82 are classified to Normal class correctly and its heatmap visualized dark blue images. The others are misclassified due to the imbalanced dataset. The CNN though those images are Pneumonia class and generated the heatmap by Grad-CAM with abnormalities regions which are visualized by light green color.

The similar in Figure 5, the CNN classified to *Pneumonia* class correctly and generated feature maps for Grad-CAM. The highlight regions exhibit the areas where the CNN focused on, also known as the abnormalities regions. These regions are discriminative between *Normal* and *Pneumonia* class.

5. Experimental results

5.1 Metrics for Comparison

We computed the overall Accuracy, Area Under the Receiver Operating Characteristic Curve (ROC-AUC) to evaluate the generalization of the classifiers. The ROC-AUC was designed as a metric for distinguishing noise from not noise, ROC is a probability curve and AUC represents degree or measure of separability. That means, it tells how much model is capable of distinguishing between classes.

5.2 Execution time

The training section is more time consuming, it can take hours or days with complicated models. In this study, we used a shallow CNN, the images with size of 64 × 64, the training section finished after 5 hours.

We implemented this system in Tensorflow, trained and tested on a 64-bit Windows system equipped with an NVIDIA GeForce GTX 1070 and 8 GB of memory.
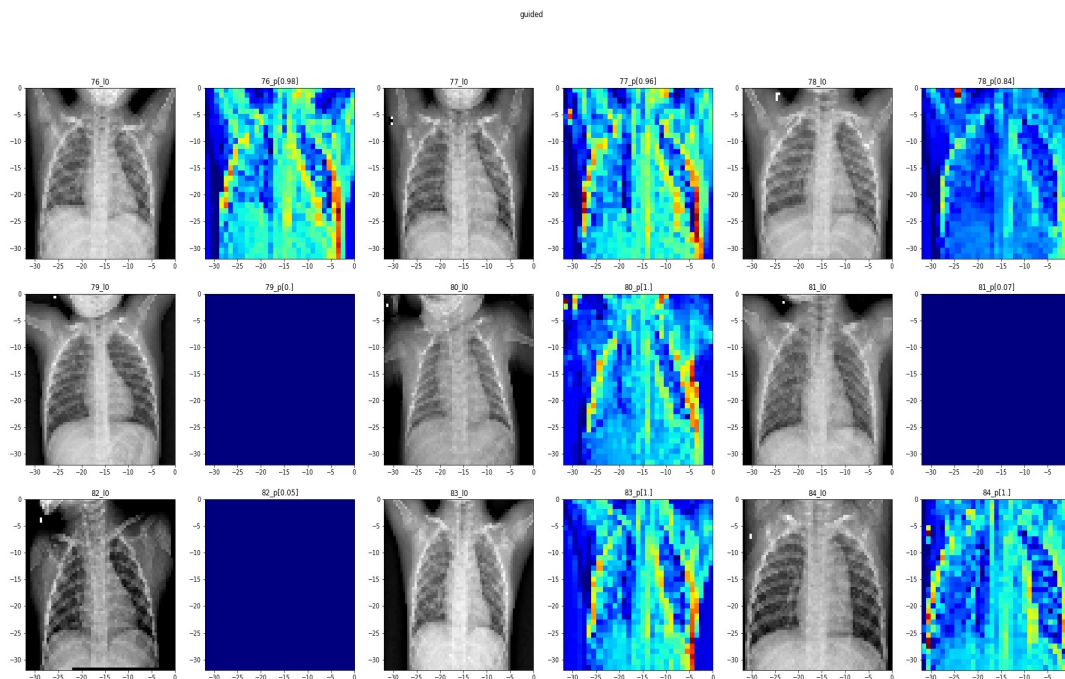
Figure 4. Visualizations of X-ray images and the generated abnormalities regions images on normal person respectively.
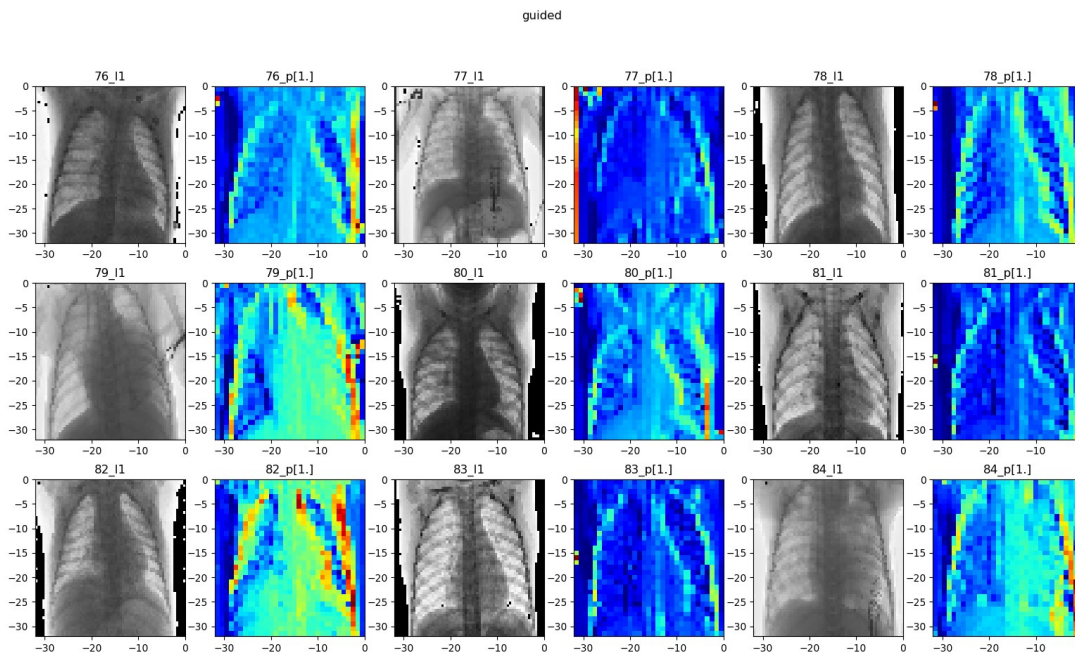

Figure 5. Visualizations of X-ray images and the generated abnormalities regions images on pneumoconiosis patients respectively.

5.3 Results

We trained our CNN model with 10-fold cross-validation, reached the overall 99.3% of training and 84.8% validation accuracy. The *Pneumonia* cases were classified almost correctly in each validation fold but the *Normal* class was misclassified. The training and validation accuracy during training are visualized in Figure 6. In general, the difference between training and validation accuracy is not significant, the classifiers focused on *Pneumonia* features for generating the feature map which will be used by Grad-CAM.

5
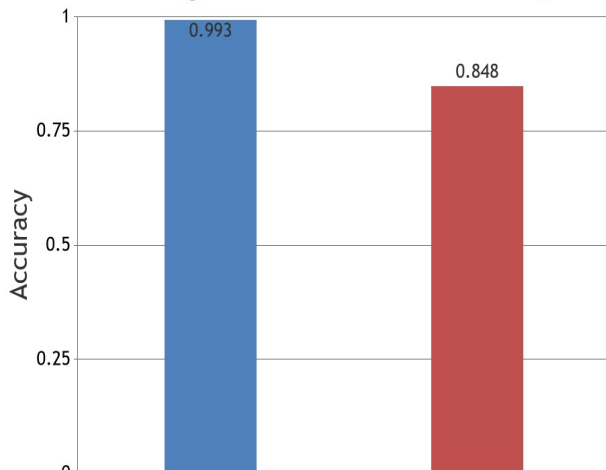
Figure 6. The visualization of accuracy during training.

In the case of imbalanced datasets, the accuracy is considered to be a misleading metric and the ROCAUC seems to be more effective metrics to evaluate the performance. We achieved 0.830 and 0.802 for training - validation AUC respectively, it is quite good for classification tasks. The AUC is illustrated in Figure 7. These results stated the performance of CNN is acceptable for classification tasks. As we mentioned above, the good model is the well-generated visualizations for abnormalities regions in X-ray chest images. The abnormalities regions in Figure 5 was built based on the findings of CNN. But in Figure 4 it still visualizes the same things on several images due to the poor performance of CNN on *Normal* class.

Figure 4 and Figure 5 show explanations generated by Grad-Cam using visualizations with heatmap color to point out signals which can aim to distinguish classes of a normal person and pneumonia patients. The output from the explanation method can provide important areas (in reds) in the images so that doctors can identify regions of injury in patients' lungs.

6. Conclusion

We presented an approach to use Convolutional Neural Network and explanation from the Grad-Cam algorithm to support Pneumonia diagnosis. The experiments provide interesting results in image analysis for predicting this disease.

There are significant differences in medical advancements among countries from various regions in the world where doctors' abilities are also different. Research on disease diagnosis based on computers can promote to reduce Adverse events in image-based medical diagnosis.
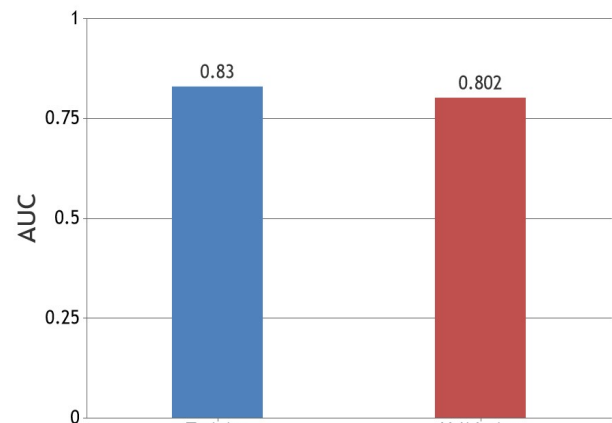
Figure 7. The visualization of AUC during training.

As an attempt for using an explanation for supporting disease diagnosis and due to some limitations of computation resources, we only deployed a shallow CNN architecture for image classification tasks on small-size medical images of Chest X-ray. In the future, deeper architectures should be investigated on larger images.

Although the detailed medical discussion of the areas visualized in the X-ray images and the quality of the obtained signatures is out of the scope of this study and is to be done by doctors doing pre-clinical research, we expect that our framework can help to better determine signal of pneumonia and to develop methods of image-based diagnosis.

References

[1] Xu X, Yu C, Qu J, et al. Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2. Eur J Nucl Med Mol Imaging. 2020;47(5):1275-1280. doi:10.1007/s00259-020-04735-9

[2] Fang Y, Zhang H, Xie J, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR [published online ahead of print, 2020 Feb 19]. Radiology. 2020;200432. doi:10.1148/radiol.2020200432

[3] American College of Radiology. ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. March 11, 2020.

[4] NHS, NHS England: Diagnostic imaging dataset statistical release. February2019,https://www.england.nhs.uk/, (accessed 30 July 2019)

[5] Kuhlmann L, Lehnertz K, Richardson MP, Schelter B, Zaveri HP. Seizure prediction - ready for a new era. Nat Rev Neurol. 2018;14(10):618-630. doi:10.1038/s41582018-0055-2

[6] Kwon JM, Lee Y, Lee Y, Lee S, Park J. An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest. J Am Heart Assoc. 2018;7(13):e008678. Published 2018 Jun 26. doi:10.1161/JAHA.118.008678

[7] Stokes JM, Yang K, Swanson K, et al. A Deep Learning Approach to Antibiotic Discovery [published correction

appears in Cell. 2020 Apr 16;181(2):475-483]. Cell. 2020;180(4):688-702.e13. doi:10.1016/j.cell.2020.01.021

[8] B. Shickel, P. J. Tighe, A. Bihorac and P. Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589-1604, Sept. 2018, doi:10.1109/JBHI.2017.2767063.

[9] C. Ma, S. Xu, X. Yi, L. Li and C. Yu, "Research on Image Classification Method Based on DCNN," 2020 International Conference on Computer Engineering and Application (ICCEA), Guangzhou, China, 2020, pp. 873876, doi: 10.1109/ICCEA50009.2020.00192.

[10] Jiamin Liu, Song Yang, Hong Huang, Zhengying Li & Guangyao Shi (2020) A deep feature manifold embedding method for hyperspectral image classification, Remote Sensing Letters, 11:7, 620-629, DOI: 10.1080/2150704X.2020.1746855

[11] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017). https://doi.org/10.1038/nature21056

[12] K. Tseng, Y. Lin, W. Hsu and C. Huang, "Joint Sequence Learning and Cross-Modality Convolution for 3D Biomedical Segmentation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3739-3746, doi: 10.1109/CVPR.2017.398.

[13] H. Tang and Z. Hu, "Research on Medical Image Classification Based on Machine Learning," in IEEE Access, vol. 8, pp. 93145-93154,2020, doi: 10.1109/ACCESS.2020.2993887.

[14] Xiao Tu, Xiaobo Shen, Peng Fu, Tao Wang, Quansen Sun, Zexuan Ji, Discriminant SubDictionary Learning with Adaptive Multiscale Superpixel Representation for Hyperspectral Image Classification, Neurocomputing, 2020, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.05.082.

[15] Singh R, Kalra MK, Nitiwarangkul C, et al. Deep learning in chest radiography: Detection of findings and presence of change. PLoS One. 2018;13(10):e0204155. Published 2018 Oct 4. doi:10.1371/journal.pone.0204155

[16] Wang X, Yu J, Zhu Q, et al. Potential of deep learning in assessing pneumoconiosis depicted on digital chest radiography [published online ahead of print, 2020 May 29]. Occup Environ Med. 2020;oemed-2019-106386. doi:10.1136/oemed-2019-106386

[17] Diederik P. Kingma and Jimmy Lei Ba. Adam : A method for stochastic optimization. 2014. arXiv:1412.6980v9

[18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[19] Nguyen T.H. (2019) Metagenome-Based Disease Classification with Deep Learning and Visualizations Based on Self-organizing Maps. In: Dang T., Küng J., Takizawa M., Bui S. (eds) Future Data and Security Engineering. FDSE 2019. Lecture Notes in Computer Science, vol 11814. Springer, Cham

[20] T. Hai Nguyen, E. Prifti, N. Sokolovska and J. Zucker, "Disease Prediction Using Synthetic Image Representations of Metagenomic Data and Convolutional Neural Networks," 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF), Danang, Vietnam, 2019, pp. 1-6, doi: 10.1109/RIVF.2019.8713670.

[21] Duchi, John & Hazan, Elad & Singer, Yoram. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research. 12. 2121-2159.

[22] Tieleman, T. and Hinton, G. (2012) Lecture 6.5-rmsprop: Divide the Gradient by a Running Average of its Recent Magnitude. COURSERA: Neural Networks for Machine Learning, 4, 26-30.

[23] Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", Mendeley Data, v2 http://dx.doi.org/10.17632/rscbjbr9sj.2

[24] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Ried-miller. Striving for Simplicity: The All Convolutional Net.CoRR,abs/1412.6806, 2014

[25] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. InECCV, 2014

7