DEEP LEARNING - POWERED DIAGNOSIS OF PULMONARY DISEASES VIA X-RAY IMAGING

Dao Thi Le Thuy

University of Transport and Communications

ARTICLE INFO	ABSTRACT
	Today, machine learning and deep learning have had many positive results in helping to diagnose and treat diseases. Based on data, parameters, and images such as X-ray, ultrasound, and magnetic resonance imaging, machines can help doctors diagnose and treat diseases better. This paper presents initial experiments on using deep learning to identify pulmonary diseases through X-ray image recognition. In experiments, there were three pulmonary diseases: aortic enlargement, lung opacity, and another lesion. There were also cases without disease to identify. The deep learning model with convolution neural network and DenseNet121 were used for our experiments with X-ray image data from Vietnamese samples and provided by VinBigData. The highest average identification accuracy achieved for pleural thickening and pulmonary fibrosis was 91.68% using DenseNet121.

HỌC SÂU – CHẨN ĐOÁN BỆNH PHỔI THÔNG QUA HÌNH ẢNH X-QUANG

Đào Thị Lệ Thủy

Trường Đại học Giao thông Vận tải

THÔNG TIN BÀI BÁO	TÓM TẮT
	Ngày nay, học máy và học sâu đã đạt được nhiều kết quả tích cực trong việc hỗ trợ chẩn đoán và điều trị bệnh. Dựa trên dữ liệu, các thông số và hình ảnh như X-quang, siêu âm và chụp cộng hưởng từ, máy móc có thể giúp các bác sĩ chẩn đoán và điều trị bệnh tốt hơn. Bài báo này trình bày các thử nghiệm ban đầu về việc sử dụng học sâu để xác định bệnh phổi thông qua nhận dạng hình ảnh X-quang. Trong các thử nghiệm, có ba bệnh lý về phổi bao gồm phình động mạch chủ, đục phổi và một tổn thương khác. Ngoài ra, cũng có các trường hợp không mắc bệnh để xác định. Mô hình học sâu với mạng nơ-ron tích chập và DenseNet121 đã được sử dụng trong các thử nghiệm với dữ liệu hình ảnh X-quang từ các mẫu bệnh nhân Việt Nam do VinBigData cung cấp. Độ chính xác trung bình cao nhất đạt được trong việc xác định dày màng phổi và xơ hóa phổi là 91,68% khi sử dụng DenseNet121.

1. Introduction

The use of computers for diagnosing and treating diseases is of utmost importance. Machine learning algorithms and models allow computers to access and process information with very large amounts of data and provide help in a short time. Doing the same thing is extremely difficult for humans. With the advancements in technology achieved today, deep learning has enabled computers to learn and make suggestions, aiding doctors in diagnosing and treating diseases more effectively.

Medical professionals have widely used X-ray images since their discovery by Wilhelm Röntgen in 1895 [1], and X-ray images have greatly improved the accuracy and efficiency of disease diagnosis and treatment in the medical field. The emergence and advancement of computers, artificial intelligence, and deep learning have the potential to aid doctors in diagnosing and treating diseases through the use of X-ray images. This article analyzes recent studies on the automated diagnosis of lung diseases using X-ray images, with a specific focus on those related to the COVID-19 pandemic.

It can be said that in recent times and now, artificial neural networks (ANN) and deep learning have been widely used in disease diagnosis using X-ray images. Many studies have used transfer learning with existing proposed models. Some studies suggest new models. The study referenced in [2] utilized image descriptors based on the spatial distribution of Hue, Saturation, and Brightness values, combined with a neural network and heuristic algorithms (Moth-Flame, Ant Lion), to detect degenerated lung tissues, achieving an average accuracy of 79.06%. In [3], the authors utilized transfer learning and fine-tuning techniques on Xception and Vgg16 models to diagnose pneumonia. Their results showed that Vgg16 outperformed Xception in terms of accuracy, achieving 87% accuracy compared to Xception's 82%. The study in [4] utilized convolutional neural networks (CNN) models (AlexNet, DenseNet121, ResNet18, InceptionV3, GoogLeNet) pre-trained on ImageNet for feature extraction, achieving an accuracy of 96.4% and a recall of 99.62% on data from the Guangzhou Women and Children's Medical Center. In their study [5], the authors proposed a classifier consisting of three binary decision trees to accurately classify chest X-ray images into three categories: normal, tuberculosis, and COVID-19 cases. Their results showed high accuracies of 98% for normal cases, 80% for tuberculosis cases, and an overall average of 95%.

The authors in [6] developed a convolutional neural network called CheXLocNet for segmenting pneumothorax lesions, achieving an area under the curve (AUC) of 0.87, sensitivity of 0.78, and specificity of 0.78. Study [7] applied transfer learning with CNN models such as DenseNet121, ResNet50, InceptionV3, VGG16, and VGG19. DenseNet121 and InceptionV3 achieved the highest accuracy (100%), while VGG19 had the lowest (78.38%). The study conducted by the authors in [8] utilized YOLOv3 to automatically crop the lung region and evaluated the effectiveness of three different multi-classification methods for this purpose. The model achieved 92.47% accuracy in detecting abnormalities and accuracy rates ranging from 71.94% to 85.71% for specific conditions like bronchiolitis/bronchitis, lobar pneumonia, or normal cases.

Studies employing deep learning models such as SqueezeNet, Inception-v3, DenseNet-161, MobileNet, ResNet, and XCOVNet, combined with techniques like transfer learning, data augmentation, and feature extraction, have achieved high accuracy in classifying chest X-ray and computed tomography (CT) images [9]. A fully automated method using a modified DenseNet-161 to classify chest X-rays into COVID-19, pneumonia, and healthy cases, achieving 100% precision for COVID-19 and pneumonia, and 98% for healthy cases [10]. The study [11] proposed the XCOVNet model for early detection of COVID-19, achieving an accuracy of 98.44%. In [12], a modified MobileNet for X-ray images and a modified ResNet for CT images, achieved high accuracy of 99.6% and 99.3%, respectively. Notable achievements include 99.8% accuracy in classifying COVID-19, viral pneumonia, bacterial pneumonia, and normal cases, as well as 99.9% accuracy in distinguishing between COVID-19 and bacterial pneumonia [13]. Furthermore, COGNEX's VisionPro Deep Learning™ software showed to outperform other models, including COVID-Net [14].

For a long time, Vietnam has not fully collected disease data in a computerized way. On the other hand, this disease data has not been clinical treatment data, but a group of experts diagnosed it after treatment, as seen in the VinDr-CXR dataset. This presents a significant obstacle to automated disease diagnosis with computer support. This paper presents preliminary results on the identification of some lung diseases through the use of deep learning on X-ray images. Using the DenseNet121 and CNN models, the study demonstrated the ability to achieve high accuracy in identifying some lung diseases such as pleural thickening and pulmonary fibrosis, with an accuracy of up to 91.68%. This contributes to developing automatic diagnosis systems. It also expands the scope of artificial intelligence applications to local medical data. This is especially evident with the VinDr-CXR dataset of Vietnam.

2. Materials and Methods

2.1. Data Preprocessing

Data used in this paper include image samples collected from patients in Vietnam and they were taken from the dataset (VinDr-CXR) used in the "VinBigData Chest X-ray Abnormality Detection" competition [15]. This dataset was used for research purposes only. The dataset comprises 18,000 postero-anterior (PA) CXR scans in DICOM format, which were de-identified to protect patient privacy. All images were labeled by a panel of experienced radiologists for the presence of 14 critical radiographic findings as listed below:

0 - Aortic enlargement	8 - Nodule/Mass
1 - Atelectasis	9 - Other lesion
2 - Calcification	10 - Pleural effusion
3 - Cardiomegaly	11 - Pleural thickening
4 - Consolidation	12 - Pneumothorax
5 - ILD	13 - Pulmonary fibrosis
6 - Infiltration	The "No finding" observation (14) was intended
7 - Lung Opacity	to capture the absence of all findings above.

The data provided by VinBigData are raw ones, so preprocessing is required. In the raw data, there are images corresponding to certain diseases that are repeated the same many times, so it is necessary to remove those images and keep only one of them. After removing the repeated images, the distribution of the number of images by disease is displayed in Figure 1. Thus, 15 cases numbered from 0 to 14 are considered as 15 classes. For this distribution, it should be noted that the same image may correspond to two diseases or more. The data distribution of image numbers for diseases is shown in Table 1.
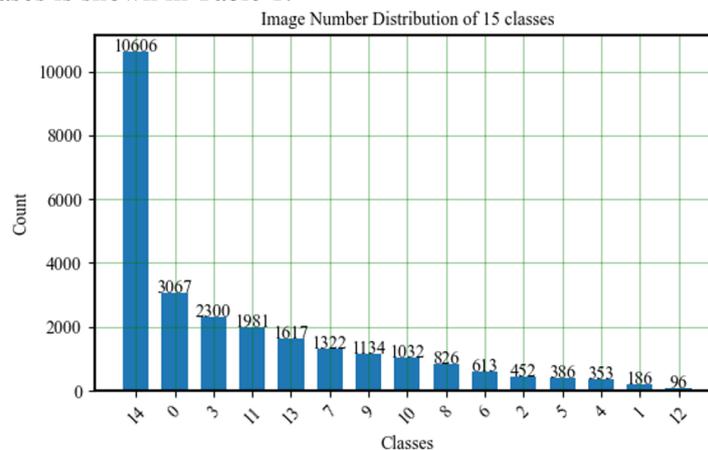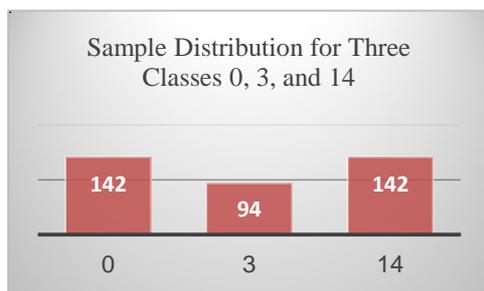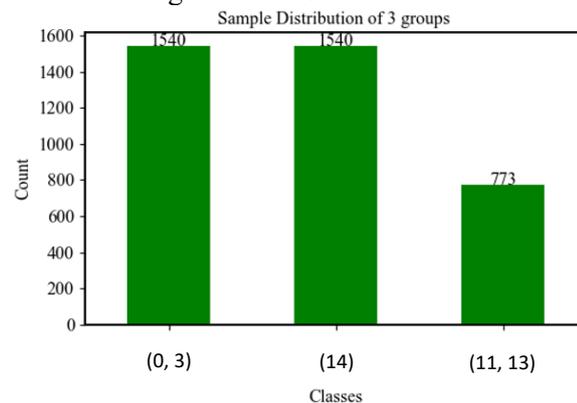

Figure 1. *Image number distribution by classes*

Table 1. *Data distribution of image numbers for diseases*

Class	14	0	3	13	10	8	11	9	2	5	12	7
Sample Number	10606	142	94	38	15	13	12	8	4	1	1	1

The data was filtered and divided into two types for the initial experiment, including (1) The data includes only a single image corresponding to a single disease. The selected data of this case is named DATA1. (2) The data includes images that only correspond to case 14 (No Finding) or to one of the two diseases. The selected data of this case is named DATA2.

For DATA1, there were three classes 0, 3 and 14 with the number of images corresponding to a single disease as shown in Figure 2. The number of images of class 14 was very large, so the number of images of this class was reduced to the same number of images of class 0 to ensure data balance.

Figure 2. *Image number distribution by three classes 0, 3, and 14*

Figure 3. *Image number distribution by three classes (0,3), (14) and (11,13)*

For DATA2 the two pairs with the most samples were selected including (0,3), (11,13), and class (14). Thus, for DATA2, it can be considered as having three groups or three classes including (0,3), (11,13), and (14) respectively. The sample distribution of these three classes is shown in Figure 3. To get this distribution, the images belonging to the group (0,3) but also belonging to groups (11,13) or (14) and vice versa were removed. In other words, the images in each group did not correspond to the disease of the other two groups.

After implementing this filtering process, it was found that the number of images in the (0,3) group was greater than the number in the (11,13) group. Furthermore, the (14) group had the highest number of images. To relatively equalize the number of samples between the three groups, the number of samples of the group (14) was taken equal to the number of samples of the (0,3) group and equal to 1540. The X-ray images were converted from DICOM to PNG format, resizing them to 1024 ×1024 pixels. Before classifying, these images were resized to 224×224 dimensions. Figure 4 showed some examples of X-ray images and corresponding diseases.

Figure 4. *Some examples of X-ray images and corresponding diseases*

2.2. Models used for experiments

Two models were used for the experiments in this paper. The first proposed model was the CNN model, while the second was the DenseNet121 model.

2.2.1. CNN model

The configuration of the CNN model based on traditional CNN is shown in Table 2.

Table 2. *Configuration of CNN model*

Model: "sequential"		
Layer (type)	*Output Shape*	*Param #*
conv2d (Conv2D)	(None, 224, 224, 64)	1792
conv2d_1 (Conv2D)	(None, 224, 224, 64	36928
conv2d_2 (Conv2D)	(None, 224, 224, 64)	36928
max_pooling2d (MaxPooling2D)	(None, 112, 112, 64)	0
dropout (Dropout)	(None, 112, 112, 64)	0
conv2d_3 (Conv2D)	(None, 112, 112, 128)	73856
conv2d_4 (Conv2D)	(None, 112, 112, 128)	147584
conv2d_5 (Conv2D)	(None, 112, 112, 128)	147584
max_pooling2d_1 (MaxPooling2	(None, 56, 56, 128)	0
dropout_1 (Dropout)	(None, 56, 56, 128)	0
conv2d_6 (Conv2D)	(None, 56, 56, 128)	147584
conv2d_7 (Conv2D)	(None, 56, 56, 128)	147584
conv2d_8 (Conv2D)	(None, 56, 56, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 28, 28, 128)	0
dropout_2 (Dropout)	(None, 28, 28, 128)	0
flatten (Flatten)	(None, 100352)	0
dense (Dense)	(None, 256)	25690368
dropout_3 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 128)	32896
dropout_4 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 3)	387
Total params:	26,611,075	
Trainable params:	26,611,075	
Non-trainable params:	0	

Overall, it is evident that this CNN model consists of 9 convolution layers, 3 fully connected layers, and one flattened layer. Additionally, there are layers of MaxPooling and DropOut.
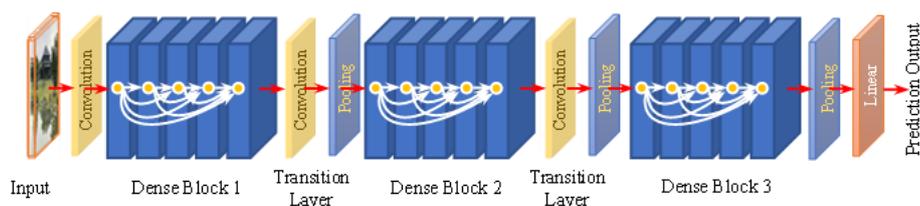
2.2.2. DenseNet121 model

Figure 5. *Illustration of DenseNet Architecture [17]*

DenseNet is considered one of the 7 best models for image classification using Keras [16]. Figure 5 is an illustration of DenseNet Architecture [17]. DenseNet [18] introduces Densely Connected Neural Networks, which aim to provide deeper insights, more efficient training, and accurate outputs. For DenseNet, in addition to the connection between layers like the connection in a CNN network, there is another special type of connection. In the DenseNet architecture, each layer is connected to every other layer. If DenseNet has L layers, there will be L(L+1)/2 direct connections. The input of a layer inside DenseNet is the concatenation of feature maps from previous layers. The architecture of DenseNet contains dense blocks, where the dimensions of the feature maps remain constant within a block, but the number of filters changes between them.

Transition Layers are used to connect dense blocks. As shown in Table 1 from [2], the DenseNet121 has (6, 12, 24, 16) layers in the four dense blocks. The number 121 is deduced as follows: 5 + (6 + 12 + 24 + 16) × 2 = 121, where 5 is (convolution, pooling) + 3 transition layers + classification layer. Multiplying by 2 is because each dense block has 2 layers (1×1 convolution and 3×3 convolution).

3. Results and Discussion

The experiments in this study were performed in a cross-validation way. The data sets DATA1 and DATA2 were both divided into 10 parts. One part, independent of the other nine, was used as test data. Of the remaining 9 parts, one was used for validation data and the other 8 parts were for training data.

The recognition accuracy for each fold, the average accuracy for the folds, and the average accuracy for each class are given in Tables 3, 4, and 5, respectively. Figures 6, 7, and 8 illustrate examples of epoch-based loss and accuracy variation for training and validation data, as well as a confusion matrix.

3.1. DATA1 with DenseNet121

Table 3 presents the results of DATA1 testing with the DenseNet121 model. The results show that class (0) had the highest rate and class (11,13) had the lowest rate. These rates were 81.48% and 44.44% respectively. The average recognition rate of the 3 classes was 65.05%.

The results in Figure 6 showed that overfitting did not occur, as the change in loss over time during training matched the change in loss for validation [19]. The same went for accuracy variation over epochs for training and validation.

Table 3. *The average accuracy for the folds and the average accuracy for each class*

Folds	CRV1	CRV2	CRV3	CRV4	CRV5	CRV6	CRV7	CRV8	CRV9	**Aver. Acc.**
Acc. (%)	57.81	67.7	55.96	64.37	61.87	73.52	68.48	64.72	61.07	**65.05**
					Classes	*0*	*3*	*14*		
				Aver. Acc. (%) for each class	**81.48**	**69.23**	**44.44**		

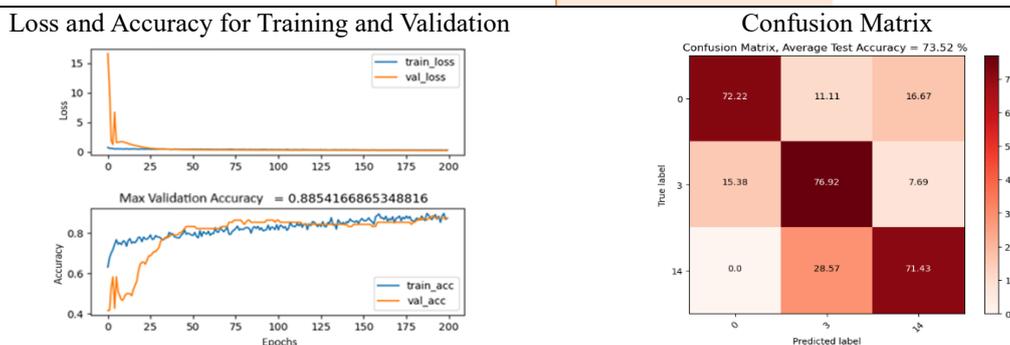Loss and Accuracy for Training and Validation   Confusion Matrix

Figure 6. *Examples of loss and accuracy variation by epoch for training and validation data (DATA1 with DenseNet121) and confusion matrix*

3.2. DATA2 with CNN and DenseNet121

This section presents the experimental results of DATA2 data using the CNN and DenseNet121 models. Each image corresponds to either one disease (14) or two diseases (0,3) or (11,13). The sample distribution for the three classes (0,3), (14), and (11,13) is shown in Figure 3.

3.2.1. Results with CNN

Table 4 displays the results of the DATA2 test using CNN, which achieved an average success rate of 73.42% across the three classes. The highest success rate was achieved by class (11,13) at

83.23%, followed by class (0,3) at 78.99%, and the remaining class (14) at 58.02%.

Figure 7 was an example of the variations of loss, training accuracy, and validation accuracy according to epochs for the CNN model with DATA2. The results in Figure 7 indicated that overfitting did not occur.

Table 4. *The average accuracy for the folds and the average accuracy for each class (CNN)*

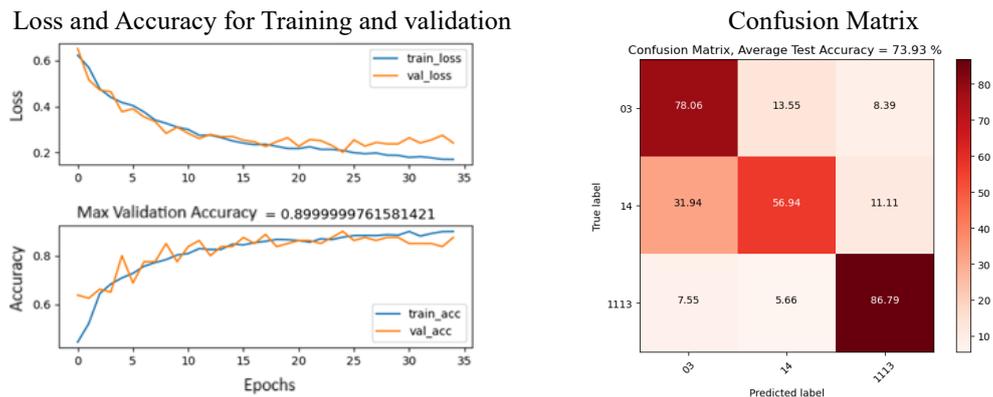| Folds	CRV1	CRV2	CRV3	CRV4	CRV5	CRV6	CRV7	CRV8	CRV9	**Aver. Acc.**
Acc. (%)	73.52	74.58	71.77	75.85	73.07	73.93	73.12	73.31	71.59	**73.42**
						Classes	*(0,3)*	*(14)*	*(11,13)*	
					Aver. Acc. (%) for each class		**78.99**	**58.02**	**83.23**	

Figure 7. *Examples of loss and accuracy variation by epoch for training and validation data (DATA2 with CNN) and confusion matrix*

3.2.2 Results with DenseNet121

The results of the DATA2 experiment using DenseNet121 are presented in Table 5. The average recognition rate for the 3 classes achieved a 5.01% improvement compared to CNN. Specifically, classes (11,13) and (0,3) showed significantly higher recognition rates of 91.68% and 88.96%, respectively, compared to CNN. However, the recognition rate for class (14) was lower.

Table 5. *The average accuracy for the folds and the average accuracy for each class (DenseNet121)*

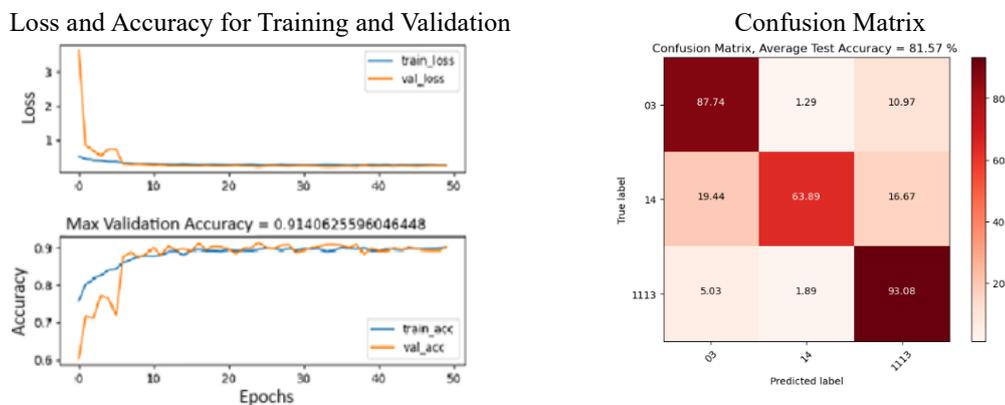| Folds	CRV1	CRV2	CRV3	CRV4	CRV5	CRV6	CRV7	CRV8	CRV9	**Aver. Acc.**
Acc. (%)	76.4	81.11	76.8	77.17	80.27	74.2	81.57	78.72	79.57	**78.43**
						Classes	*(0,3)*	*(14)*	*(11,13)*	
					Aver. Acc. (%) for each class		**88.96**	**54.63**	**91.68**	

Figure 8. *Examples of loss and accuracy variation by epoch for training and validation (DATA2 with DenseNet121) and confusion matrix*

Figure 8 was an example of the variations of loss, training accuracy, and validation accuracy according to epochs for the DenseNet model with DATA2. Similar to the results illustrated in Figures 6 and 7, these variations showed no overfitting.

The experiments were conducted on computers with NVIDIA's 2080 GPU. The computation time for the CNN model is much shorter than that of the DenseNet121 model because CNN has a simpler structure than DenseNet121. With the DATA2 set, the average accuracy of DenseNet121 was about 5% higher than that of CNN. DenseNet121 and CNN's highest average accuracy for class (11,13) was 91.68% and 83.23% respectively. With a recognition rate of 91.68%, this result is much higher than the results of many studies presented in Section 1. A team of experts has manually labeled VinDr-CXR. With a given X-ray image, the corresponding number of diseases can vary by experts. This is also often the case with emotional speech corpus. The same sentence may evoke different emotions in different listeners. In cases involving large amounts of data, it is crucial to utilize computers to aid in decision-making.

4. Conclusion

This paper presents the initial findings on the classification of lung diseases using VinDr-CXR image data. The DenseNet121 model has a much more complex architecture than the CNN model, so it has shown an overall advantage in recognition accuracy compared to the CNN model. The results of identifying some diseases are also positive, but the disease bounding box data on the provided X-ray images has not yet been exploited. Therefore, the next line of research will focus on leveraging this parameter to achieve more effective results.

REFERENCES

[1] Wikipedia, "Wilhelm Röntgen," December 17, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Wilhelm_R%C3%B6ntgen. [Accessed December 26, 2024].

[2] Q. Ke, J. Zhang, W. Wei, D. Połap, M. Woźniak, L. Kośmider, and R. Damaševĭcius, "A neuro-heuristic approach for recognition of lung diseases from X-ray images," *Expert Systems with Applications,* vol. 126, pp. 218-232, 2019.

[3] E. Ayan and H. M. Ünver, "Diagnosis of pneumonia from chest X-ray images using deep learning," in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, IEEE, 2019, pp. 1-5.

[4] V. Chouhan, S. K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaševičius, and V. H. C. D. Albuquerque, "A novel transfer learning-based approach for pneumonia detection in chest X-ray images," *Applied Sciences,* vol. 10, no. 2, 2020, Art. no. 559.

[5] S. H. Yoo, H. Geng, T. L. Chiu, S. K. Yu, D. C. Cho, J. Heo, M. S. Choi *et al.,* "Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging," *Frontiers in medicine,* vol. 7, 2020, Art. no. 427.

[6] H. Wang, H. Gu, P. Qin, and J. Wang, "CheXLocNet: Automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks," *PLoS One,* vol. 15, no. 11, 2020, Art. no. e0242013.

[7] J. Velasco, J. R. Ang, J. Caibigan, F. M. Naval, N. Arago, and B. Fortaleza, "Identification of Normal and Diseased Lungs using X-ray Images through Transfer Learning," *Int. J. Adv. Trends Comput. Sci. Eng.,* vol. 9, no. 4, pp. 6227-6231, 2020.

[8] K.-C. Chen, H.-R. Yu, W.-S. Chen, W.-C. Lin, Y.-C. Lee, H.-H. Chen, J.-H. Jiang, *et al.,* "Diagnosis of common pulmonary diseases in children by X-ray images and deep learning," *Scientific reports,* vol. 10, no. 1, pp. 1-9, 2020.

[9] K. T. Islam, S. N. R. Wijewickrema, A. Collins, and S. J. O'Leary, "A Deep Transfer Learning Framework for Pneumonia Detection from Chest X-ray Images," in *VISIGRAPP (5: VISAPP),* pp. 286-293, 2020.

[10] J. de Moura, J. Novo, and M. Ortega, "Fully automatic deep convolutional approaches for the analysis of Covid-19 using chest X-ray images," *Applied Soft Computing*, vol. 115, 2022, Art. no. 108190.

[11] V. Madaan, A. Roy, C. Gupta, P. Agrawal, A. Sharma, C. Bologa, and R. Prodan, "XCOVNet: Chest X-ray Image Classification for COVID-19 Early Detection Using Convolutional Neural Networks," *New Generation Computing,* vol. 39, pp. 583–597, 2021.

[12] G. Jia, H.-K. Lam, and Y. Xu, "Classification of COVID-19 chest X-ray and CT images using a type of dynamic CNN modification method," *Computers in biology and medicine,* vol. 134, 2021, Art. no. 104425.

[13] P. Vieira, O. Sousa, D. Magalhes, R. Rablo, and R. Silva, "Detecting pulmonary diseases using deep features in X-ray images," *Pattern Recognition,* vol. 119, 2021, Art. no. 108081.

[14] A. Sarkar, J. Vandenhirtz, J. Nagy, D. Bacsa, and M. Riley, "Identification of Images of COVID-19 from Chest X-rays Using Deep Learning: Comparing COGNEX VisionPro Deep Learning 1.0™ Software with Open-Source Convolutional Neural Networks," *SN Computer Science,* vol. 2, no. 3, pp. 1-16, 2021.

[15] H. Q. Nguyen, K. Lam, L. T. Le, H. H. Pham, D. Q. Tran, D. B. Nguyen, D. D. Le, *et al.,* "VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations," *arXiv preprint arXiv:2012.15029*, 2020.

[16] D. Subhash, "7 Best Models for Image Classification using Keras," November 17, 2018. [Online]. Available: https://www.it4nextgen.com/keras-image-classification-models/. [Accessed December 26, 2024].

[17] T. L. T. Dao, V. L. Trinh, B. T. Chu, and H. C. Nguyen, "Music genre classification using densenet and data augmentation," *Computer Systems Science and Engineering*, vol. 47, no. 1, pp. 657–674, 2023.

[18] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700-4708.

[19] J. Admin, "How to Treat Overfitting in Convolutional Neural Networks," September 7, 2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/09/overfitting-in-cnn-show-to-treat-overfitting -in-convolutional-neural-networks. [Accessed December 21, 2024].