An Efficient Deep Learning Approach for Detecting Lung Disease from Chest X-Ray Images Using Transfer Learning and Ensemble Modeling

Conference Paper · December 2021

DOI: 10.1109/CSDE53843.2021.9718454

6 authors, including:

Mostofa Kamal Sagor
BRAC
3 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Shakib Mahmud Dipto
University of Liberal Arts Bangladesh
30 PUBLICATIONS **56** CITATIONS
SEE PROFILE

Ishrat Jahan
BRAC University
1 PUBLICATION **2** CITATIONS
SEE PROFILE

Md Tanzim Reza
BRAC University
74 PUBLICATIONS **621** CITATIONS
SEE PROFILE

2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) | 978-1-6654-9552-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/CSDE53843.2021.9718454

An Efficient Deep Learning Approach for Detecting Lung Disease from Chest X-Ray Images Using Transfer Learning and Ensemble Modeling

Mostofa Kamal Sagor[1], Shakib Mahmud Dipto[2], Ishrat Jahan[3],
Susmita Chowdhury[4], Md Tanzim Reza[5] and Md Ashraful Alam[6]
[1,2,3,4,5,6]Department of Computer Science and Engineering, BRAC University, 66 Mohakhali, Dhaka 1212, Bangladesh
Email: [1]mostofakamalsagor21@gmail.com, [2]diptomahmud2@gmail.com, [3]ishrat.jahan2@g.bracu.ac.bd,
[4]susmitachowdhury.997@gmail.com [5]rezatanzim@gmail.com, [6]ashraful.alam@bracu.ac.bd

***Abstract*—Among the most convenient bacteriological assessments for the diagnosis and treatment of several health complications is the chest X-Ray. In X-Ray imaging, it is a common technique to standardize the extracted image reconstruction with the usual uniform disciplines taken before the study. Unfortunately, there has been relatively little study on several separate lung disease monitoring, including X-Ray picture analysis and poorly labelled repositories. Our paper suggests an effective automated approach for the detection of lung disease trained on chest X-ray images. Besides, with a weighted binary classifier, a particular technique is also deployed that will optimally leverage the weighted predictions from optimal deep neural networks such as Inception-v3, VGG16 and ResNet-50. In addition to the existing, transfer learning, along with more rigorous academic training and testing sets, is used to fine-tune deep neural networks to achieve higher internal processes. In comparison, 88.14 percent test accuracy was obtained with the final proposed weighted binary classifier, where other models give us about 80.9 percent average accuracy. For a brief recurring diagnosis, the legally prescribed procedure may also be used which may increase the course of the same condition for physicians. For a prompt diagnosis of pneumonia, the suggested approach should be used and can improve the diagnosis process for health practitioners.**

***Index Terms*—lung diseases; chest X-ray images; convolution neural network (CNN); deep learning; transfer learning; diagnostics facilitated by electronic**

I. Introduction

Lung diseases are a pretty common health problem that is usually detected by chest X-Rays. But this is a slow process that requires certain signs to detect. Many patients are also misdiagnosed due to the radiologists' mistakes. Also, coronavirus disease is now the most common and most concerning lung disease that is affecting millions of people around the world. To tackle this problem, we propose a technique that leverages a deep transfer learning algorithm and ensemble approach to make the detection process automated and faster, and also more accurate than the previous models for lung diseases detection from chest X-Ray images.

In this research, we have observed several models individually and we have attempted with Inception-v3 [1], VGG16 [2], VGG19 [2], ResNet-50 [3] and ResNet-101 [4]. We pre-processed the X-Ray image data into a well-defined form of $224 \times 224 \times 3$ in the initial stage. The last layer of the selected models (Inception-v3, VGG16 and ResNet-50) was then selected as a transfer layer and a sequential model was initiated by adding convolutional layers, flatten layer, dropout layer, fully connected layer and a two neuron output layer for classification. Finally, we used an averaging layer to ensemble the outputs of these three classification models.

II. Related Work

In recent times, mainly CNN-based algorithms have been used to solve medical image classification related problems. SegNet [5], U-Net [6], AlexNet [7], GoogLeNet [8], VGGNet-16 [2], ChestNet [9], CardiacNet [10] and ResNet-50 [3] are a some of the more well-known models for medical image related classification problems. For determining optimum network hyper-parameters, models like evolutionary-based algorithms [11], BPNN [12] (a multi-layer supervised feed-forward neural network), CpNN [12] (an unsupervised simple neural network with two layers) and reinforcement learning have been developed. For conducting lung nodule detection [13] and pulmonary tuberculosis classification, these algorithms are regularly used.

In addition, the majority vote of a jury of experts acted as a benchmark on the confirmation collection of Chest Radiographs Classification. The reliability of CheXNet on the validation range was contrasted with the level of performance of 9 medical experts using the AU-ROC as the measuring instrument. The average time to go through and classify the around 400 images in the validation set was noticeably longer for the radiologists than for the automated model. The main problem of this study was that both CheXNet and radiologists were not allowed to use patients' previous data and this experiment was limited to a dataset from a single institution [14].

However, in recent studies [15], multi-layer, probabilistic, learning vector quantization, and generalized regression neural networks have been used for the diagnosis of chest and lung diseases. The diagnosis of lung diseases such as TB, pneumonia etc. using chest radiographs in [16] was implemented using a neural network for grouping after pre-processing images using normalization. The research works described in this

paragraph had been used effectively in classifying diseases but their performance was not up to par with the contemporary deep learning models.

III. Proposed Methodology

When we were trying to develop the model, the first focus was how to predict from more than one model and take the average probability to predict the final output label or class. As a consequence, we end up developing an ensemble model of three well-known convolutional neural networks.

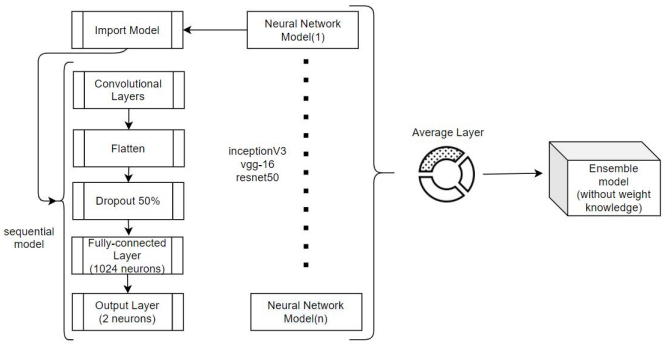

Fig. 1: Ensemble Model using InceptionV3, VGG16 and ResNet50. Using the features extracted from these three models we leveraged fully connected layers and averaging layer for the final prediction

First of all, we have observed several models individually and we have attempted with Inception-v3, VGG16, VGG19, ResNet-50 and ResNet-101. However, we decided to select only three based on their performance over the Pneumonia dataset while training. Because of resource limitations, we could not select more than three models for performing the ensemble operation. Later on, we have added a flatten layer, a dropout layer and 2 dense layers as well after the last convolutional layer of those three models (for InceptionV3, VGG16 and ResNet50, they are "mixed10", "block_pool5", "avg_pool" respectively). Lastly, we have taken the output from these sequential models as the input of an averaging layer and considered the output layer of that averaging layer as our desired classification categories. The detailed overview of our proposed model is given in Fig. 1.

IV. Dataset Details and Processing

We have used the ChestX-Ray8 dataset [10]; which contains 1,12,120 X-Ray images and the Pneumonia dataset found in Kaggle [17]; which contains 5,215 X-Ray images as the training dataset, 626 X-Ray images as the validation dataset and 16 X-Ray images as the testing dataset.

A. Data Sample

The dataset accommodates X-Ray images from patients where some of the patients are diagnosed with pneumonia and some are not. Some sample X-Ray images are shown in Fig. 2.

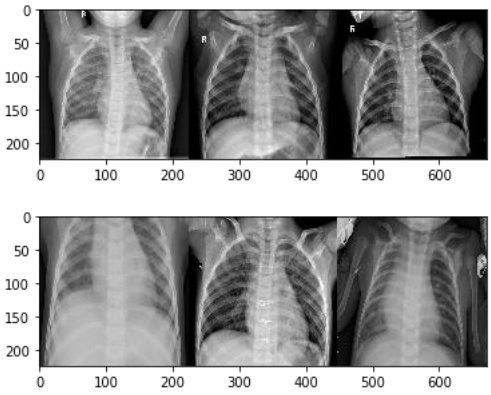

Fig. 2: The first row of images are X-Ray images of patients whose lungs are in normal condition (not pneumonia affected). The second row of images are the X-Ray images of lungs of pneumonia patients

B. Data Pre-processing

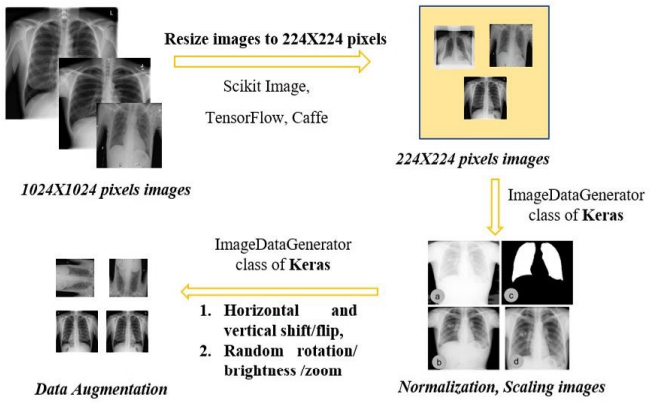

Fig. 3: Overview of the data pre-processing phases- **i)** First the images are re-sized to 224×224 dimension, **ii)** Then, the images are normalized, **iii)** Finally, the images are augmented

1) Resize Images: Since we used pre-trained convolutional neural network models while training the network, each image was re-sized into a fixed size of 224×224. For this purpose, we used scikit-image [18], TensorFlow [19], and Caffe frameworks [20].

2) Scaling Images: We used the ImageDataGenerator class of Keras [21] for scaling our images. We scaled the pixel values (originally between 0 and 255) to the range of 0 to 1.

3) Data Augmentation: We augmented our data by applying a set of random transformations to the images for increasing our model performance. We applied rotation (90°, 180°and 270°) and translation to the images and also horizontally and vertically flipped the images. For this purpose, the ImageDataGenerator class of Keras was used.

V. Implementation

A. Transfer Learning Phase

Transfer learning (TL) is a concept of machine learning (ML), where the model gets trained with an initial weight knowledge and then the learned weights are applied to solve the relatable problem [22].

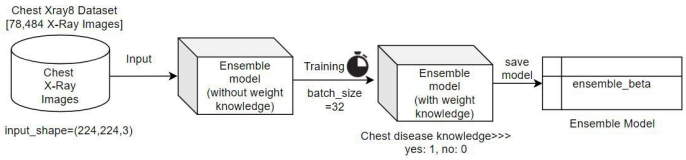

Fig. 4: Transfer learning phase: Training the model that will be used for transfer learning. An intermediate model is trained on ChestX-Ray8 dataset to be used for transfer learning

To apply transfer learning, we first trained our model on the ChestX-Ray8 [10] dataset with a 70%-30% training and validation split. This model was trained for 40 epochs with a batch size of 32. After training, we have saved the model with the gained knowledge of weights and have named it "ENSEMBLE_BETA" which is shown in figure 4. This model with the gained knowledge from the ChestX-Ray8 dataset will then be used to train our main model on the pneumonia dataset which is elaborated in the next subsection.

B. Training and Cross-Validation Phase

We loaded our "ENSEMBLE_BETA" model and re-trained it with the Pneumonia dataset [17], which is known as fine-tuning [22].

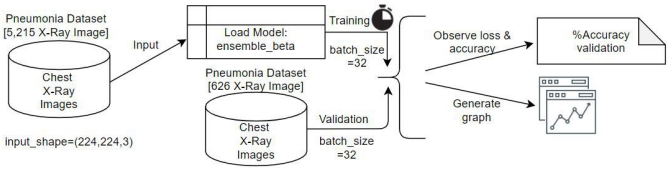

Fig. 5: Training and cross validation phase: Train the main model by loading the pre-trained model "ENSEMBLE_BETA" on the Pneumonia dataset

Along with that, we also cross-checked the validation accuracy as well as validation loss, whether the accuracy is increasing or not and the loss is decreasing or not. The training phase was also carried out for 40 epochs with a batch size of 32. It was already mentioned in IV that this dataset has 5,215 X-Ray images as the training dataset, 626 X-Ray images as the validation dataset and 16 X-Ray images as the testing dataset. This phase is depicted in figure 5.

C. Testing Phase

This is our final step and in this step, we have used 16 X-Ray images of the chest that are unknown to our model and have predicted their label. Table I shows the result of the predictions. Fig. 6 shows the concept behind this phase.

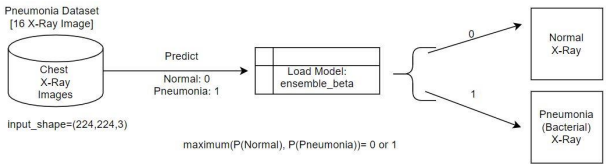

Fig. 6: Testing phase: Predicting on the 16 X-Ray images available as test data

From Table I, we can see that our model has classified the input images into either 0 or 1, naming Normal and Pneumonia respectively.

TABLE I: Predicted output labels of the 16-sample input chest X-Ray images

Actual Image Names	Prediction	Category
NORMAL2-IM-1427-0001	0	Normal
NORMAL2-IM-1430-0001	0	Normal
NORMAL2-IM-1431-0001	0	Normal
NORMAL2-IM-1436-0001	0	Normal
NORMAL2-IM-1437-0001	1	Pneumonia
NORMAL2-IM-1438-0001	0	Normal
NORMAL2-IM-1440-0001	0	Normal
NORMAL2-IM-1442-0001	0	Normal
person1946_bacteria_4874	0	Normal
person1946_bacteria_4875	1	Pneumonia
person1947_bacteria_4876	1	Pneumonia
person1949_bacteria_4880	1	Pneumonia
person1950_bacteria_4881	1	Pneumonia
person1951_bacteria_4882	1	Pneumonia
person1952_bacteria_4883	1	Pneumonia
person1954_bacteria_4886	1	Pneumonia

VI. Results

A. Results on Individual Models

We first trained and tested on the Pneumonia dataset [17] using the well-known models VGG16, VGG19, Inception-v3, ResNet-50 and ResNet-101.

For each model, we have used a different subset of chest X-Ray images as input and have performed particular actions. We have trained these models for 40 epochs with weight parameters settled as 'imagenet'. The results are shown in Table II.

TABLE II: Accuracy and loss in the experimented individual models

Dataset	Keras Models	Accuracy	Loss
Pneumonia	Inception-v3	76.69%	1.9598
	VGG16	82.99%	0.9277
	VGG19	73.74%	1.5533
	ResNet-50	81.03%	0.5689
	ResNet-101	65.62%	1.6825

B. Analysis of ENSEMBLE BETA

We have built an ensemble model and trained it for 40 epochs and we have also settled weight parameters as "imagenet". Secondly, we have re-trained our pre-trained model and

also checked the validation of our model's training accuracy, whether it is under-fitting or over-fitting.

1) Result without Transfer Learning: We trained and tested on the Pneumonia dataset [17] using the Ensemble model without transfer learning. After 40 epochs, we achieved a validation accuracy of 78.45% and a validation loss of 1.5383.

2) Result of with Transfer Learning: We trained and tested on the Pneumonia dataset [17] using the Ensemble model with transfer learning. The result is shown in Table III. In

TABLE III: Validation accuracy and loss after transfer learning into the ensemble model after 40 epochs

Dataset	Phase of the model	Accuracy	Loss
ChestX-Ray8	Ensemble model without weight knowledge	61.05%	1.234
Pneumonia	Ensemble model with weight knowledge	88.14%	0.5033

machine learning, learning curves are most often used and are frequently a plot that exhibits iterations or time or history on the x-axis and the consistency of learning or classification on the y-axis. It helps to test as well as evaluate the model at the beginning of training. We included a tqdm callback to see the percentage of training and minimize the progress callback to decrease the number of learning when another ratio halted progressing with a patience value of 3. For 40 periods of history or oscillations, we have trained and validated our model and established the following prediction performance and validation loss graph.

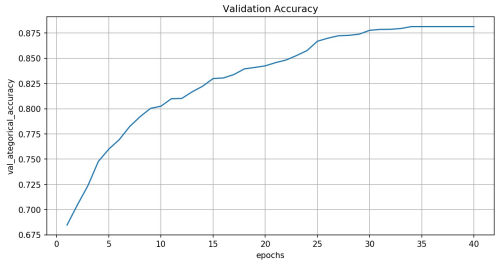

Fig. 7: Validation accuracy while training. Number of epochs is shown along the X-axis and the accuracy is displayed along the Y-axis

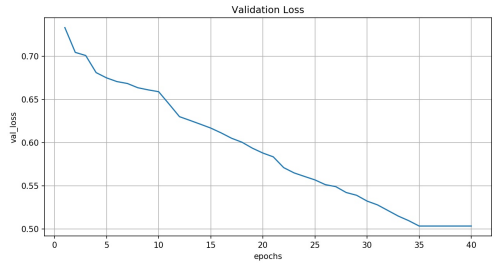

Fig. 8: Validation loss while training. Number of epochs is shown along the X-axis and the loss is displayed along the Y-axis

From Figure 7, we can observe that the validation accuracy has an increasing slope and it stops increasing and becomes constant at the very end of training and gets fixed with an accuracy of 88.14%. On the other hand, from the Figure 8, we can observe that the validation loss is decreasing having a negative slope and it illustrates the efficiency of our model ENSEMBLE BETA.

C. Comparison with Other Models

We have compared our model with the other existing models that we have covered during over literature reviewing and have listed the accuracy in Table IV. A bar chart of the test accuracy of the models are also shown in Figure 9.

TABLE IV: Comparison of other models versus our model

Model	Accuracy
CheXNet	76.80%
CNN with Lightened Image on Increased Contrast	75.65%
CNN with Lightened Image on Increased Contrast with ResNet	78.73%
DenseNet-161 [23]	84.50%
Our model (Ensemble_Beta)	88.14%

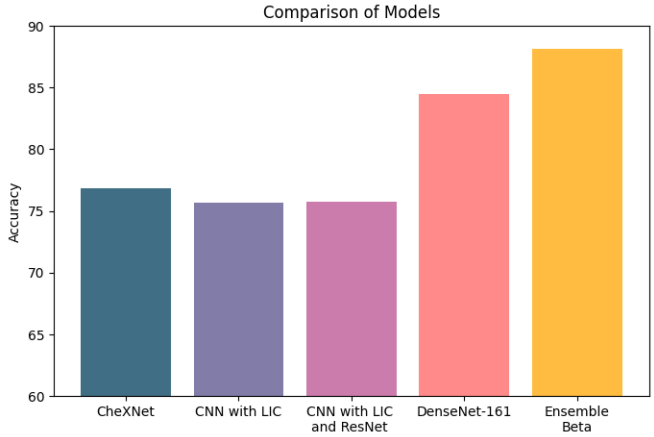

Fig. 9: Test accuracy of previous models versus our model using a bar chart

Since it has two positive aspects, our ensemble model with transfer learning is efficient. First of all, the ensemble model gives higher accuracy as it reduces the over-fitting and during our model testing, we have seen an increasing sloped curve for validation accuracy and decreasing sloped curve for validation loss. Secondly, the ensemble technique helps us to reduce the bias and variance error by maintaining a trade-off between these two parameters and this helps to learn less noisy data while training. As a result higher accuracy was obtained compared to other existing models.

VII. Experimental Setup

We ran all our experiments using Google Colab (colab.research.google.com/) and a regular PC with an Nvidia GeForce GTX TITAN Xp 12 GB graphics card, cuDNN v7.0 library and 9.0 CUDA Toolkit.

VIII. Discussion and Conclusion

Our ensemble model is more superior to the previous models in two aspects. First of all, our model achieved higher accuracy than the previous models. Secondly, the ensemble technique helped reduce the bias and variance error and this helped us to learn less noisy data while training. As a result higher accuracy was obtained compared to the other existing models.

We analyzed the existing CNN lung disease classification techniques, their correlations and also stated in-depth our planned "ENSEMBLE_BETA" structure in terms of its proposed architecture, transfer learning stage, classification model, implementation stage, predictive validity and effectiveness in this research paper. Furthermore, this research has future demands, as now we are going through a Coronavirus pandemic and now the lung diseases are getting prioritized to be detected in the early stage. However, we encountered some challenges during the implementation phase when training the dataset with a sufficiently smaller like 16 or greater like $64, 128, \ldots, k$ as batch size during the implementation phase.

IX. Future Work

We might use GoogleNet, AlexNet and other powerful machine learning models to the ensemble and check whether the experiment gives better performance or not. We might also use a more powerful machine which will help us experiment with a more complex model. In conclusion, this research project can be further progressed by classifying different lung diseases in the future, rather than just predicting pneumonia from standard X-Ray images.

Acknowledgment

We are indebted to the CVIS (Computer Vision and Intelligent Systems) Research Lab of the CSE Department, BRAC University for providing us with the necessary equipment for running our experiments.

References

[1] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] Z. Wu, C. Shen, and A. Van Den Hengel, "Wider or deeper: Revisiting the resnet model for visual recognition," *Pattern Recognition*, vol. 90, pp. 119–133, 2019.

[5] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," *arXiv preprint arXiv:1710.10501*, 2017.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[9] H. Wang and Y. Xia, "Chestnet: A deep neural network for classification of thoracic diseases on chest radiography," *arXiv preprint arXiv:1807.03058*, 2018.

[10] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[11] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, "Interleaved text/image deep mining on a very large-scale radiology database," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1090–1099.

[12] R. H. Abiyev and M. K. S. Ma'aitah, "Deep convolutional neural networks for chest diseases detection," *Journal of healthcare engineering*, vol. 2018, 2018.

[13] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, "Interleaved text/image deep mining on a large-scale radiology database for automated image interpretation," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 3729–3759, 2016.

[14] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists," *PLoS medicine*, vol. 15, no. 11, p. e1002686, 2018.

[15] O. Er, N. Yumusak, and F. Temurtas, "Chest diseases diagnosis using artificial neural networks," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7648–7655, 2010.

[16] S. Khobragade, A. Tiwari, C. Patil, and V. Narke, "Automatic detection of major lung diseases using chest radiographs and classification by feed-forward artificial neural network," in *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*. IEEE, 2016, pp. 1–5.

[17] P. Mooney, "Chest x-ray images (pneumonia)," https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia/.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[19] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.

[20] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[21] F. Chollet *et al.* (2015) Keras. [Online]. Available: https://github.com/fchollet/keras

[22] M. Kurtz, "How transfer learning can make machine learning more efficient," https://thenewstack.io/how-transfer-learning-can-make-machine-learning-more-efficient/.

[23] S. B. AlSumairi and M. M. B. Ismail, "X-ray image based pneumonia classification using convolutional neural networks."