International Journal of Electrical, Electronics and Data Communication, ISSN(p): 2320-2084, ISSN(e): 2321-2950
Volume-7, Issue-7, Jul.-2019, http://iraj.in

# THREE-STAGE ENSEMBLE OF IMAGENET PRE-TRAINED NETWORKS FOR PNEUMONIA DETECTION

**[1]ROM UDDAMVATHANAK, [2]SHAILEY, [3]FENG YANG**

[1]School of Engineering & Technology, PSB Academy, Singapore
[2]Department of Computing Science, Institute of High Performance Computing, A*STAR, Singapore, 138632
[3]College of Science and Engineering, James Cook University, Singapore
E-mail: [1]vathanakuddam@gmail.com, [3]yangf@ihpc.a-star.edu.sg

**Abstract -** Focusing on detection of pneumenia disease in the Chest X-Ray images, this paper proposes a three-stage ensemble methodology utilizing multiple pre-trained Convolutional Neural Networks (CNNs). In the first-stage ensemble, k subsets of training data are firstly randomly generated, each of which is then used to retrain a pre-trained CNN to produce k CNN models for the ensemble in the first stage. In the second-stage ensemble, multiple ensemble CNN models based on multiple pre-trained CNNs are integrated to reduce variance and improve the performance of the prediction. The third-stage ensemble is based on image augmentation, i.e., the original set of images are augmented to generate a few sets of additional images, after which each set of images are input to the ensemble models from the first two stages, and the outputs based multiple sets of images are then integrated. In integrating outputs in each stage, four ensemble techniques are introduced including averaging, feed forward neural network-based, decision tree-based, and majority voting. Thorough experiments were conducted on Chest X-Ray images from a Kaggle challenge, and the results showed the effectiveness of the proposed three-stage ensemble method in detecting pneumonia disease in the images.

**Keywords -** Ensemble, Convolutional Neural Network, Chest X-Ray, Pneumonia, Imagenet Pre-Trained Model, Pneumonia Detection.

## I. INTRODUCTION

Computerized clinical guideline and decision support system have been developed to help physicians diagnose and manage diseases [1] [2] [3]. The success of the automated systems highly depends on the availability of the data in the medical information system. Because information from chest X-ray images is significant for detecting pneumonia disease, an automated system is needed to observe and analyze this information to produce the result. For image classification or object detection, Convolutional Neural Networks (CNNs) have achieved state-of-the-art performance in many challenges and domains [4] [5] [6]. Due to the lack of training data, the pre-trained neural network models have been introduced to solve this problem [7] [8].
Ensemble techniques were brought up considering the variance of deep learning models. For example, using an effective recursive based ensemble method [9] had shown improvement in solving multi-class imbalance data classification. Another approach used in [10], where the authors make use of ensemble classifier combination methods for improving word spotting system performance. Other examples include gesture recognition [11], credit scoring [12], and apparel attributes classification [13].

In relation to using an ensemble method for image classification tasks, there have already been several approaches. In one of those approaches, the authors proposed an error-independent method to automatically design a neural network ensemble for high performance image classification systems [14]. Whereas in [15], the authors make use of random subspace ensembles for brain images classification obtained through functional magnetic resonance imaging (fMRI). Furthermore, as proposed in [16], the author ensemble fine-tuned CNN models for medical image classification that achieved higher accuracy than an established CNN model.

Considering the effectiveness of using an ensemble method, this paper proposed a three-stage ensemble method using pre-trained CNNs model for the ensemble. In the first-stage ensemble, K fold cross-validation was used while training individual CNN models to reduce the variance of the model. This is followed by the second-stage ensemble, where multiple pre-trained CNN models are pre-defined for improving the accuracy of the prediction. Lastly, data augmentation was applied to the provided dataset to increase training samples and the variety of the prediction for the third stage ensemble. The outputs from each stage were integrated by using ensemble techniques to produce final predictions. The proposed method was conducted on Kaggle's Chest X-Ray Images for pneumonia detection. The results obtained in this study have proven the effectiveness of using the proposed method for detecting pneumonia in chest X-ray images.

## II. METHOD

A three-stage ensemble of ImageNet pre-trained network has been proposed for pneumonia detection in chest X-ray images. As can be seen in Fig. 1, the proposed method starts from applying data augmentation and pre-processing technique with the main purpose of increasing the size of training data and removing noise in the training data. Based on the outstanding performance from ImageNet pre-trained models in computer vision domain, multiples deep neural network models were selected for training. In order to adapt these neural networks to learn new task, fully connected layer of the original architecture was replaced by a new fully connected layer, along with a softmax activation function (at the end of the dense layer). In the training process, these models are trained multiple times by sub-sampling of training data, in which the first-stage ensemble was applied (to obtain unique predictions from each model). The second stage ensemble is then carried out to integrate the predictions from multiple deep neural network models. This is followed by the third stage ensemble to integrate the predictions from different models that were trained on different training samples after data augmentation was applied. More details about the proposed method are given in the following subsections.

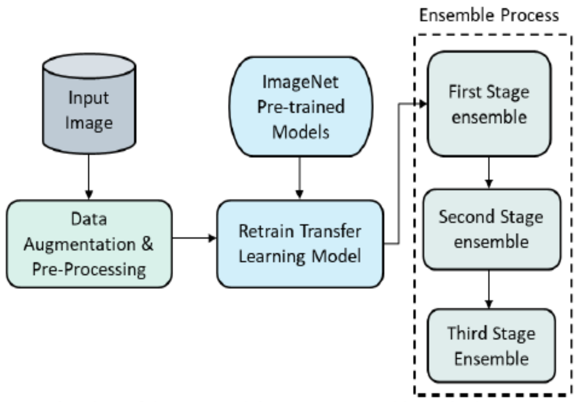



**Fig. 1: Flowchart of the proposed three-stage ensemble method.**

### A. Transfer Learning

The strategy used to train the models begin with initializing the pre-trained models with pre-trained weights, where the knowledge is obtained in a different domain/task. One main drawback in medical image processing is that the size of the provided dataset is small compare to other domains in computer vision. Thus, training a deep convolutional neural network model from scratch is often not achievable. However, this can be solved by using transfer learning approach [17], where the models were used to train on a large dataset to solve related tasks.

In the proposed method, the pre-trained network is used as a feature extractor followed by a fully connected layer (2 dense layers) as well as an output layer to make predictions as shown in Fig. 2. In the training process, the model is re-trained with training samples to adapt to a new domain. All the pre-trained models utilizing the weights that were trained on ImageNet challenge as an initial point.

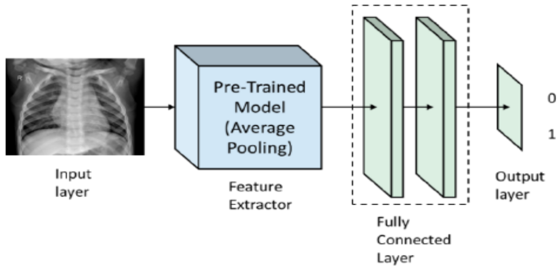



**Fig. 2: a Transfer learning approach**

In referring to the performance of pre-trained CNN models on ImageNet challenge, 7 different architectures of CNN models were selected as given in TABLE II (Xception, Resnet50, Inception, … etc.).

### B. Parameter Fine-tune for Transfer Learning

The default setting of a pre-trained CNN model's parameter may not be appropriate to achieve higher accuracy in this study (as proven in the experiment). For that reason, each pre-trained model is further fine-tuned with K-Fold cross-validation (K = 8). In this model, the global average pooling will be applied to the output of the last convolutional layer. The fully connected layer consists of two dense layers (1024 neurons for the first layer and 512 neurons for the second layer) with the dropout value of 0.25 in each layer. Moreover, RMSprop, with the learning rate of 0.0001, was used as an optimizer in this study.
By using K-Fold cross-validation, K runs of the training process required a lot of time to be trained. In order to reduce the time for training, an early stopping function was implemented which ceases the training when the condition is met. For example, if a loss is not reduce during 5 consecutive epochs and before the pre-defined epoch (30), the training will stop.

### C. Data Augmentation & Image Pre-processing

The images in the dataset had different levels of brightness, this was corrected using gamma correction. Specifically, it augmented the images into three different categories: darker image, normal image, brighter image. Furthermore, to reduce the noise in the images, zoom-in were carried out, with the value of 10%. Example of data augmentation and pre-processing are shown below in Fig. 3.

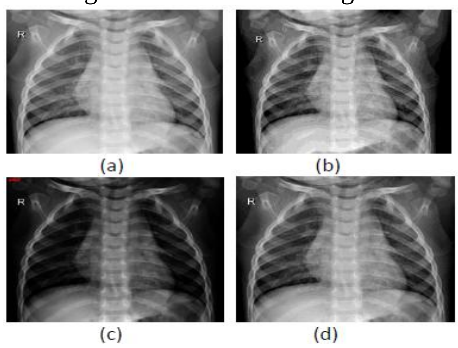



**Fig. 3: Examples of data augmentation and pre-processing. (a): Brighter image; (b) original image; (c) darker image; (d) zoom-in.**

## D. Ensemble Method

There are two approaches used to increase the diverse of the predictions from each type of the pre-trained CNN models before going into the ensemble process. The first approach is known as K-Fold cross-validation where the whole set of training data is first separated into k number of folds, with each fold containing roughly the same number of samples inside. K (K = 8) subsets of samples can be achieved in an iterative training, where each fold of sample is removed at a time.

For each type of CNN model, k different models can be trained (K-fold cross-validation), by applying the CNN model to each of the subsets of training samples. Those k models are then integrated at the first-stage ensemble to produce an individual CNN model. The process is demonstrated in Fig. 4 below (for CNN model i).

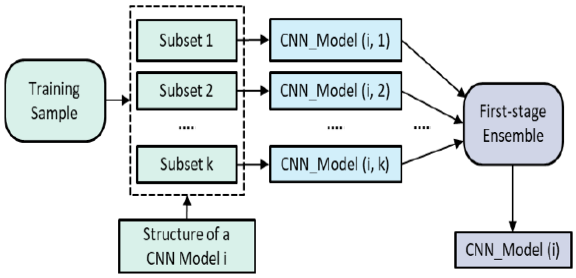



**Fig. 4: Demonstration of the process in the first-stage ensemble.**

As can be seen in TABLE II, 7 pre-trained CNN models were used in this paper. Hence, after the first-stage ensemble, multiple pre-trained CNN models can be achieved. The second stage ensemble is then integrated with the 7 pre-trained CNN models to produce ensemble outputs (Ensemble CNN model) for each training samples (Original image, darker image, and brighter image) as shown in Fig. 5.

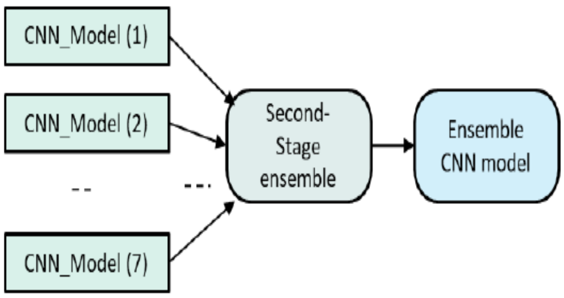



**Fig. 5: Demonstration of the second-stage ensemble**

The two-stage ensemble approach above is applied to the three training samples (original image, darker image, brighter image) that are obtained after data augmentation. This will produce Ensemble CNN model output based on each training samples, which are then integrated in the third stage ensemble to produce a final output using ensemble technique as illustrated in Fig. 6

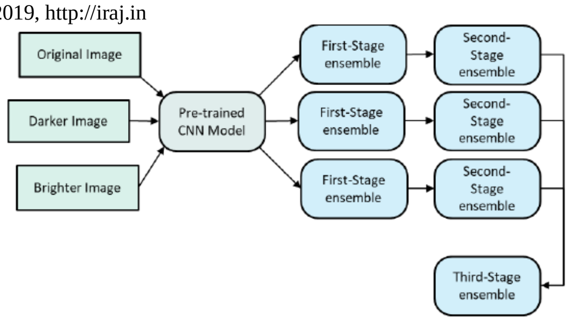



**Fig. 6: Demonstration of the third-stage ensemble.**

## E. Ensemble Techniques

The predictions from the CNN model are the probability value between 0 and 1 (Softmax activation) to be a class of an image. Several ensemble techniques were introduced to combine multiple outputs from the proposed methods. Here is the detailed description of ensemble techniques used in this study:

### 1) Averaging

Averaging (AVG) is one of the most common ensemble technique used to combine multiple values in machine learning. Generally, the outputs in the first-stage and second-stage ensemble is integrated by averaging multiple outputs to produce an ensemble output. For instance, if the predictions value is 0.9, 0.8, and 0.7, the ensemble output will be equal to 0.8.

### 2) Feedforward Neural Network

Feedforward Neural Network (FNN) is a simple neural network commonly used in deep learning domain. The main idea of how FNN is used as an ensemble technique is that the outputs from multiple predictions act as vectors to fit into the FNN (see Fig. 7). The hidden layer contains three dense layers (1024, 512, and 256 neurons respectively) with the dropout value of 0.25. This will produce the same ensemble outputs as other ensemble techniques.

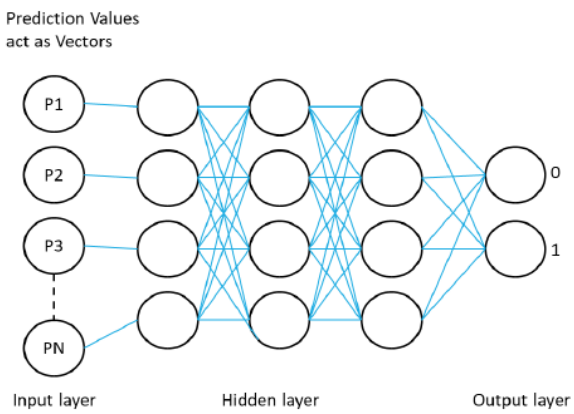



**Fig. 7: Architecture of Feedforward Neural Network.**

### 3) Decision Tree

Decision Tree Classifier (DTC) was also used as an ensemble technique to integrate the outputs from multiple predictions (The same procedure as FNN's). Some of the default parameter settings in DTC were implemented with the max depth of 4 and criterion of gini. DTC model was provided by Sci-kit learn framework in this experiment.

#### 4) Majority Voting

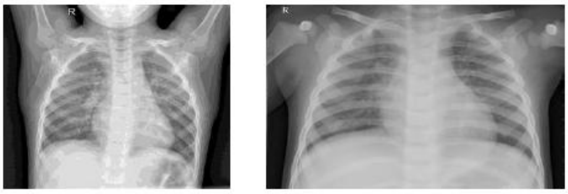
Unlike previous ensemble technique, Majority Voting algorithm [18] was used only in the Thirst-Stage ensemble of the proposed method. This technique was used to select only the highly confident predictions from the second-stage ensemble. This can be done by evaluating the prediction from different models and then selecting the prediction with the highest voted prediction. For instance, if four models are confident that the output should be true and only one model confident that the output should be false, the final output will be equal to true based on Majority Voting. Since Majority Voting needs pre-defined value, the max and min value in this paper is 0.7 and 0.2.

## III. EXPERIMENTS AND RESULTS

This section presents the experiments as well as the results obtained by using the proposed ensemble method for pneumonia detection. More detail is given in the following subsections.

### A. Dataset

This image data was published by Kaggle's "Chest X-Ray Images (Pneumonia)" with a total of 5840 images. In available data, each image is either a healthy (normal) lung or a pneumonia lung. There are 5216 images in the training sample with 3876 images being pneumonia lungs and the rest being normal lungs. For testing samples, there are 624 images with 390 being pneumonia lungs. The examples of image in both classes are shown in Fig. 8.



(a) (b)

**Fig. 8: Example of a normal lung image (a) and pneumonia lung (b).**

### B. Evaluation Metric

The accuracy score is used to evaluate the performance of the classification model. The metric for the accuracy score is express as below:

$$Accuracy = \frac{y^p}{y}$$

Where $y_p$ is the number of the correct predictions made by the model and y is the total number of predictions.

### C. Result

TABLE I presents the performance of the proposed methods after the third-stage ensemble, whereas TABLE II present the performance of individual models as well as the second-stage ensemble model. In TABLE I, different combinations of the ensemble techniques were conducted in the experiment. ("3rd_Ensemble (All)") referred to the ensembling outputs from all the models that were integrated using the three ensemble techniques (FNN, DTC, and AVG) mentioned in Section E for the second-stage ensemble. It can be seen that the performance of each model varies, with the best performance being 0.884 and 0.8661 for second-stage and Third-Stage ensemble models respectively. On the other hand, the best performance with accuracy of 0.842 for an individual model was achieved from the ensemble Xception model (using AVG) on the brighter images.

| Model | Accuracy |
|---|---|
| 3rd_Ensemble (All) | 0.818 |
| 3rd_Ensemble (FNN + DTC) | 0.852 |
| 3rd_Ensemble (FNN) | 0.8661 |
| 3rd_Ensemble (DTC) | 0.7999 |
| 3rd_Ensemble (AVG) | 0.7884 |

**TABLE I: Performance of the proposed method.**

### D. Discussion

The results obtained in this study using the provided image dataset has proved the effectiveness of using the proposed method for pneumonia detection on chest x-ray images. The proposed three-stage ensemble method using FNN outperform most of the models in this study except the second-stage ensemble with FNN as an ensemble technique on darker images. To reduce computational burden, Majority Voting was applied to the third stage ensemble. This study has shown that it is possible to do both two stages and three stages ensemble to get a desired output or accuracy score. Although the third-stage ensemble does not achieve the best performance in this study, it shows that the proposed method is very generic and powerful and also provides more options to other researchers. Hence, it shall be easily adjusted to other applications or domains in the A.I. field.

## IV. CONCLUSION

This paper proposed a method with the three-stage ensemble of pre-trained networks for pneumonia detection. K fold cross-validation was used while training CNN models to reduce the variance of the model in the first-stage ensemble. Subsequently, in the second-stage ensemble, numbers of pre-trained models were pre-defined to improve the performance of the predictions. Finally, the provided images were augmented in the third-stage ensemble for improving the variety of predictions. The outputs in each ensemble stage were integrated using various ensemble techniques, which yield two best possible results of 0.884 and 0.8661. The experimental results have shown the effectiveness of using the proposed method on the given dataset. This ensemble process is comprehensive and adaptable that make it easier to apply to other applications/domains in the machine learning field.

## REFERENCES


[1] R. a. G. P. a. P. J. a. W. L. Marée, "Biomedical image classification with random subwindows and decision trees," in International Workshop on Computer Vision for Biomedical Image Applications, 2005.

[2] M. M. a. A. S. K. a. T. G. R. Rahman, "A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback," IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 4, pp. 640-646, 2011.

[3] M. P. a. S. R. a. Z. Y. a. Z. J. M. a. E. A. S. a. o. Wachowiak, "An approach to multimodal biomedical image registration utilizing particle swarm optimization," IEEE Trans. Evolutionary Computation, vol. 8, no. 3, pp. 289-301, 2004.

[4] a. S. I. a. H. G. E. Krizhevsky, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012.

[5] a. R. M. a. K. T. Wachinger, "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," NeuroImage, vol. 170, pp. 434-445, 2018.

[6] Q. a. H. Y. a. Z. Z. a. Z. Z. a. Z. L. a. Y. Y. Guan, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," arXiv preprint arXiv:1801.09927, 2018.

[7] D. a. D. M. a. E. T. a. S. U. Marmanis, "Deep learning earth observation classification using ImageNet pretrained networks," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1, pp. 105-109, 2016.

[8] L.-C. a. P. G. a. K. I. a. M. K. a. Y. A. L. Chen, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834-848, 2018.

[9] T. a. A. C. F. a. Z. S. A. a. K. M. A. H. a. I. M. T. Alam, "An Effective Ensemble Method for Multi-class Classification and Regression for Imbalanced Data," in Industrial Conference on Data Mining, 2018.

[10] M. a. S. C. Y. Khayyat, "Improving Word Spotting System Performance using Ensemble Classifier Combination Methods," in 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2018.

[11] H. a. W. L. Liu, "Gesture recognition for human-robot collaboration: A review," International Journal of Industrial Ergonomics, vol. 68, pp. 355-367, 2018.

[12] X. a. X. Z. a. Z. B. a. D. Y. a. Q. J. Feng, "Dynamic weighted ensemble classification for credit scoring using Markov Chain," Applied Intelligence, pp. 1-14, 2018.

[13] T. a. F. Z. a. L. H. Han, "Two-Layer Mixture Network Ensemble for Apparel Attributes Classification," International Conference on Artificial Intelligence on Textile and Apparel, pp. 153-159, 2018.

[14] G. a. R. F. Giacinto, "Design of effective neural network ensembles for image classification purposes," Image and Vision Computing, vol. 19, no. 9-10, pp. 699-707, 2001.

[15] L. I. a. R. J. J. a. P. C. O. a. L. D. E. a. J. S. J. Kuncheva, "Random subspace ensembles for fMRI classification," IEEE transactions on medical imaging, vol. 29, no. 2, pp. 531-542, 2010.

[16] a. K. J. a. L. D. a. F. M. a. F. D. Kumar, "An ensemble of fine-tuned convolutional neural networks for medical image classification," IEEE journal of biomedical and health informatics, vol. 21, no. 1, pp. 31-40, 2017.

[17] S. J. a. Y. Q. a. o. Pan, "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345-1359, 2010.

[18] H.-C. a. P. S. a. J. H.-M. a. K. D. a. B. S. Y. Kim, "Pattern classification using support vector machine ensemble," in Pattern Recognition, 2002. Proceedings. 16th International Conference on, 2002.


| Pre-trained Model | Normal Image | | | Brighter Image | | | Darker Image | | |
|---|---|---|---|---|---|---|---|---|---|
| | Averaging | Neural Network | Decision Tree | Averaging | Neural Network | Decision Tree | Averaging | Neural Network | Decision Tree |
| Xception | 0.8221 | 0.8205 | 0.822 | 0.842 | 0.8222 | 0.8333 | 0.841 | 0.825 | 0.807 |
| Resnet50 | 0.786 | 0.772 | 0.767 | 0.759 | 0.767 | 0.778 | 0.769 | 0.777 | 0.78 |
| InceptionV3 | 0.762 | 0.774 | 0.782 | 0.794 | 0.804 | 0.8044 | 0.809 | 0.782 | 0.774 |
| InceptionResNetV2 | 0.793 | 0.741 | 0.785 | 0.783 | 0.791 | 0.7788 | 0.79 | 0.802 | 0.791 |
| DenseNet201 | 0.791 | 0.772 | 0.7804 | 0.788 | 0.783 | 0.785 | 0.801 | 0.782 | 0.798 |
| DenseNet169 | 0.785 | 0.774 | 0.783 | 0.775 | 0.756 | 0.764 | 0.814 | 0.788 | 0.78 |
| VGG19 | 0.758 | 0.75 | 0.758 | 0.77 | 0.7996 | 0.7644 | 0.7644 | 0.753 | 0.764 |
| Second-Stage Ensemble Model | 0.786 | 0.817 | 0.823 | 0.804 | 0.806 | 0.794 | 0.785 | 0.884 | 0.769 |

**TABLE II: Performance of different models**

★★★