

Article

# A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images

Vikash Chouhan <sup>1</sup>, Sanjay Kumar Singh <sup>1</sup>, Aditya Khamparia <sup>1</sup>, Deepak Gupta <sup>2</sup>, Prayag Tiwari <sup>3</sup>, Catarina Moreira <sup>4</sup>, Robertas Damaševičius <sup>5,\*</sup> and Victor Hugo C. de Albuquerque <sup>6</sup>

<sup>1</sup> School of Computer Science and Engineering, Lovely Professional University, Punjab 144411, India; vikashcode1997@gmail.com (V.C.); sanjayksingh.012@gmail.com (S.K.S.); aditya.khamparia88@gmail.com (A.K.)

<sup>2</sup> Maharaja Agrasen Institute of Technology, New Delhi 110034, India; deepakgupta@mait.ac.in

<sup>3</sup> Department of Information Engineering, University of Padova, 35131 Padova, Italy; prayag.tiwari@dei.unipd.it

<sup>4</sup> School of Information Systems, Science and Engineering Faculty, Queensland University of Technology, Brisbane City 4000 QLD, Australia; catarina.pintomoreira@qut.edu.au

<sup>5</sup> Faculty of Applied Mathematics, Silesian University of Technology, 44-100 Gliwice, Poland

<sup>6</sup> Graduate Program in Applied Informatics, University of Fortaleza, Fortaleza 60811-905, CE, Brazil; victor.albuquerque@unifor.br

\* Correspondence: robertas.damasevicius@polsl.pl

Received: 17 December 2019; Accepted: 9 January 2020; Published: 12 January 2020



**Abstract:** Pneumonia is among the top diseases which cause most of the deaths all over the world. Virus, bacteria and fungi can all cause pneumonia. However, it is difficult to judge the pneumonia just by looking at chest X-rays. The aim of this study is to simplify the pneumonia detection process for experts as well as for novices. We suggest a novel deep learning framework for the detection of pneumonia using the concept of transfer learning. In this approach, features from images are extracted using different neural network models pretrained on ImageNet, which then are fed into a classifier for prediction. We prepared five different models and analyzed their performance. Thereafter, we proposed an ensemble model that combines outputs from all pretrained models, which outperformed individual models, reaching the state-of-the-art performance in pneumonia recognition. Our ensemble model reached an accuracy of 96.4% with a recall of 99.62% on unseen data from the Guangzhou Women and Children’s Medical Center dataset.

**Keywords:** deep learning; transfer learning; medical image processing; computer-aided diagnosis

## 1. Introduction

Today's deep learning models can reach human-level accuracy in analyzing and segmenting an image [1]. The medical industry is one of the most prominent industries, where deep learning can play a significant role, especially when it comes to imaging. All those advancements in deep learning make it a prominent part of the medical industry. Deep learning can be used in wide variety of areas like the detection of tumors and lesions in medical images [2,3], computer-aided diagnostics [4,5], the analysis of electronic health-related data [6], the planning of treatment and drug intake [7], environment recognition [8] and brain-computer interface [9], aiming to come up with decision support for the evaluation of the person's health. The key element of the success of deep learning is based on the capability of the neural networks to learn high level abstractions from input raw data through a general purpose learning procedure [10].

Although currently, deep learning still cannot replace doctors/clinicians in medical diagnosis, it can provide support for experts in the medical domain in performing time-consuming works, such as examining chest radiographs for the signs of pneumonia.

Pneumonia is an inflammation of the lungs that may be caused by pathogens, like bacteria, viruses and fungi [11]. It can occur to anyone, young or even to healthy people. It becomes life-threatening for infants, people having other diseases, people with an impaired immune system, elderly people, people who are hospitalized and have been placed on a ventilator, people with chronic disease like asthma and people who smoke cigarettes. The cause of pneumonia also determines its severity. Viral pneumonia is milder, and symptoms occur gradually. However, it can become complicated to diagnose if a bacterial infection develops at the same time with viral pneumonia. On the other side, bacterial pneumonia is more severe, and its symptoms can occur gradually or even suddenly, especially among children [12]. This type of pneumonia affects a large part of the lungs, and can lead to affect many lobes of the lung. When multiple lobes of the lungs are affected, a person needs to be hospitalized [13]. Another form of pneumonia is fungal pneumonia, which can occur to persons with weak immune systems. This type of pneumonia could be dangerous also, and requires time for the patient to regain health. Therefore, there is an urgent need to perform research and to develop new methods helping to provide computer-aided diagnosis to reduce pneumonia-related mortality, especially, child mortality, in the developing world [14].

The analysis of chest radiography has a crucial role in medical diagnostics and the treatment of the disease. The Centers for Disease Control and Prevention (CDC, Atlanta, GA, USA) reported that about 1.7 million adults in the US seek care in a hospital due to pneumonia every year, and about 50,000 people died in United States from pneumonia in 2015 [15]. Chronic obstructive pulmonary disease (COPD) is the primary cause of mortality in the United States, and it is projected to increase by 2020 [16]. The World Health Organization (WHO, Geneva, CH) reported that it is one of the leading causes of death all over the world for children under 5 years of age, killing an estimated 1.4 million, which is about 18% of all children deaths under the age of 5 years worldwide [17]. More than 90% of new diagnoses of children with pneumonia happen in the underdeveloped countries with few medical resources available. Therefore, the development of cheap and accurate pneumonia diagnostics is required.

Recently, a number of researchers have proposed different artificial intelligence (AI)-based solutions for different medical problems. Convolutional neural networks (CNNs) have allowed researchers to obtain successful results in wide medical problems like breast cancer detection, brain tumor detection and segmentation, disease classification in X-ray images, etc. [18].

For example, Wang et al. [19] contributed by proposing a new dataset called ChestX-ray8 that contains images of different 32,717 patients, having 108,948 frontal X-ray images. They achieved promising results using a deep CNN. They further stated that this dataset can be extended using more disease labels. Not only in chest X-rays, Ronneburger et al. [20], used the power of data augmentation and CNN, and achieved great results, even training on small sample of images. In another study, Roth et al. [21] showed how deep CNN can be used to detect lymph nodes. They got promising results even in case of an adverse quality of images. Shin et al. [22] worked on different CNN architectures and addressed the problem of lymph and lung disease detection.

Rajpurkar et al. [23] suggested CheXNeXt, a very deep CNN with 121 layers, to detect 14 different pathologies, including pneumonia, in frontal-view chest X-rays. First, neural networks were trained to forecast the probability of 14 abnormalities in the X-ray image. Then, an ensemble of those networks, were used to issue predictions by calculating the mean predictions of individual networks. Woźniak et al. [24] suggested a novel algorithm for training probabilistic neural networks (PNNs) that allows one to construct smaller networks. Gu et al. [25] used a 3D deep CNN and a multiscale forecasting strategy, with cube clustering and multiscale cube prediction, for lung nodule detection. Ho and Gwak [26] used a pretrained DenseNet-121 and four types of local features and deep features

acquired by using the scale-invariant feature transform (SIFT), GIST, Local binary patterns (LBP) and a histogram of oriented gradients (HOGs), and CNN features for classification of 14 thoracic diseases.

Jaiswal et al. [27] used Mask-RCNN, a deep neural network, which utilizes both global and local features for pulmonary image segmentation combined with image augmentation, alongside with dropout and L2 regularization, for pneumonia identification. Jung et al. [28] use a 3D deep CNN (3D DCNN) which has shortcut connections and a 3D DCNN with dense connections. The shortcuts and dense connections solve the gradient vanishing problem. Connections allow deep networks to capture both the general and specific features of lung nodules. Lakhani and Sundaram [29] use AlexNet and GoogLeNet neural networks with data augmentation and without any pre-training to obtain an Area Under the Curve (AUC) of 0.94–0.95. An improved architecture using transfer learning and a two-network ensemble achieved an AUC of 0.99.

Li et al. [30] used a CNN-based approach combined with rib suppression and lung field segmentation. For pixel patches acquired in the lung area, three CNNs were trained on different resolution images, and feature fusion was applied to merge all information. Liang et al. [31] designed a novel network with residual structures, which has 49 convolutional layers, only one global average pooling layer and two dense connection layers for pediatric pneumonia diagnosis. Nam et al. [32] used a deep CNN with 25 layers and eight residual connections. The outputs of three networks trained with different hyperparameter values were averaged for the prediction of malignant pulmonary nodules in chest radiographs. Nasrullah et al. [33] used two 3D-customized mixed link network (CMixNet) architectures. Lung nodule recognition was performed with faster R-CNN on features learned from CMixNet and U-Net like encoder-decoder, while a gradient boosting machine (GBM) was used for classification. Pasa et al. [34] proposed a custom neural network consisting of five convolutional blocks (each having two  $3 \times 3$  convolutions with Rectified Linear Units (ReLUs), followed by a max-pooling operation), a global average pooling layer and a fully-connected softmax layer with two outputs.

Pezeshk et al. [35] suggested a 3D fully CNN for the fast screening and generation of candidate suspicious regions. Next, an ensemble of 3-D CNNs was trained using extensive data augmentations obtained from the positive and negative patches. The classifiers were trained on false positive patches using different thresholds and data augmentation types. Finally, the outputs of second stage networks were averaged to produce the final prediction. Sirazitdinov et al. [36] suggested an ensemble of RetinaNet and Mask R-CNN networks for pneumonia localization. First, networks recognized the pneumonia-affected regions, then non-maximum suppression was applied in the predicted lung regions.

Souza et al. [37] used AlexNet-based CNN for lung patch classification. Then, a second CNN model, based on ResNet18, was employed to reconstruct the missing parts of the lung area. The output is obtained by the ensemble combination of initial segmentation and reconstruction. Taylor et al. [38] used standard network architectures (VGG16/19, Xception, Inception and ResNet) and explored the parameter space of pooling and flattening operations. Final results were obtained by fully connected layers followed by a sigmoid function. Wang et al. [39] split X-ray images into six types of patches and used ResNet to classify six patches and recognize lung nodules. They used rotation, translation and scaling techniques for data augmentation. Xiao et al. [40] proposed a multi-scale heterogeneous 3D CNN (MSH-CNN), which uses multiscale 3D nodule blocks with contextual data as input; (2) two 3D CNNs to acquire expression features; (3) and a set of weights defined by backpropagation for feature fusion. Xie et al. [41] suggested a semi-supervised adversarial classification (SSAC) framework that is trained using both labeled and unlabeled data. They used an adversarial autoencoder-based unsupervised network for reconstruction, a supervised network for classification and transition layers that map from the reconstruction network to the classification network.

Xu et al. [42] designed a hierarchical CNN CXNet-m1, which used a novel sin-loss loss function to learn from misclassified and indistinguishable images for anomaly identification on chest X-rays. Yates et al. [43] retrained a final layer of the CNN model, Inception v3, and performed binary normality classification. In addition, da Nóbrega et al. [44] compared the ResNet50 feature extractor combined with the SVM RBF classifier for recognition of the malignancy of a lung nodule.

Ke et al. [45] used the spatial distribution of Hue, Saturation and Brightness in the X-ray images as image descriptors, and an ANN with heuristic algorithms (Moth-Flame and Ant Lion optimization) to detect diseased lung tissues. Finally, Behzadi-khormouji et al. [46] used transfer learning with DCNNs pretrained on the ImageNet dataset and combined with a problem-based ChestNet architecture with unnecessary pooling layers removed, and a three-step pre-processing to support model generalization. Zhang et al. [9] systematically investigated brain signal types and related deep learning concepts for brain signal analysis, and they covered open challenges and future directions for brain signal analysis with help of 230 contributions performed by researchers. Zhou et al. [5] proposed three-stage deep features learning and a fusion framework for identifying Alzheimer disease and its prodromal status. Fan et al. [47] proposed a fully convolutional network deep learning approach for image registration to predict the deformation field which become insensitive to parameter tuning and small hierarchical loss. Liu et al. [1] exploited the usage of a deep convolutional network to explore the uses of local description and feature encoding for image representation and registration.

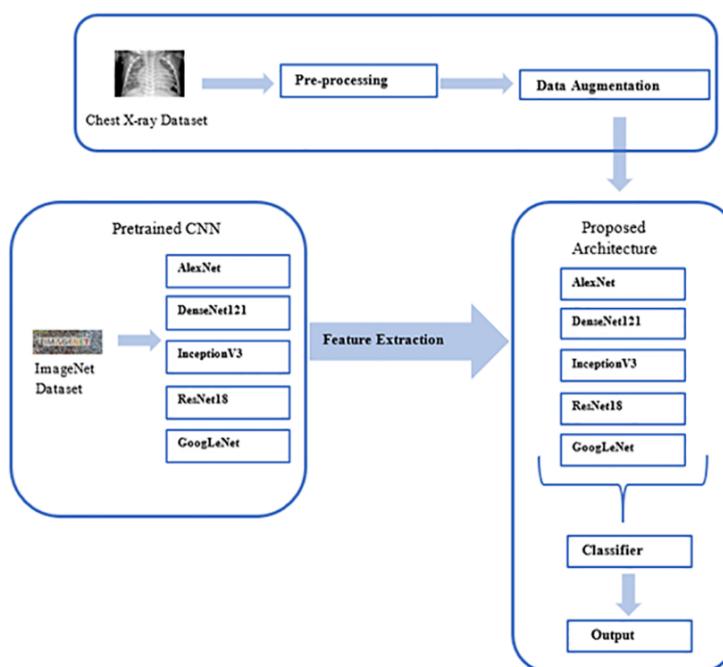
Apart from these significant achievements, CNNs work very well on large datasets. However, most of the time they fail on small datasets if proper care is not taken. To meet the same level of performance even on a small dataset, and to classify pneumonia from normal chest X-rays, we conducted this study to propose a novel approach of transfer learning using pretrained architectures trained on ImageNet. We used five different pretrained models and analyzed their performances. Finally, the combination of all five models was taken, forming a large ensemble architecture and achieving promising results.

Our main contribution in this paper is to provide a novel ensemble approach based on the transfer learning of five different neural network architectures.

## 2. Methods

### 2.1. Outline of Methodology

Our methodology is summarized in Figure 1. It includes the following steps: chest X-ray image preprocessing, data augmentation, transfer learning using AlexNet, DenseNet121, InceptionV3, resNet18 and GoogLeNet neural networks, feature extraction and ensemble classification. Then steps are explained in more detail in the subsequent subsections.



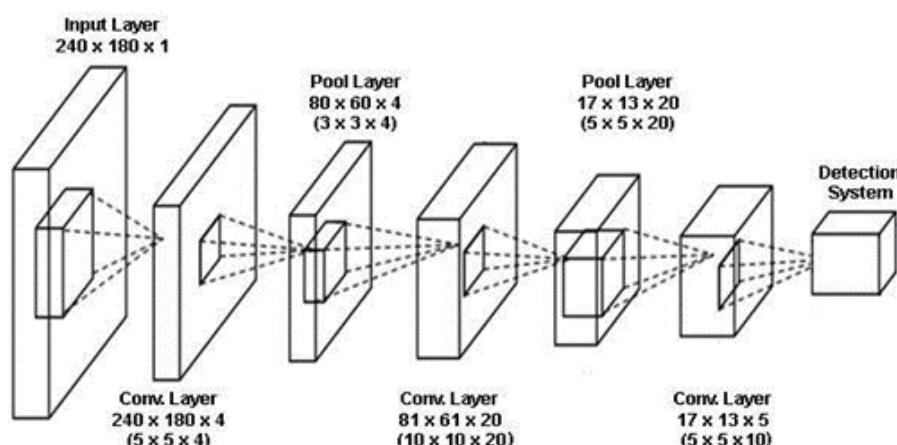
**Figure 1.** Outline of the methodology.

## 2.2. Data Pre-Processing and Augmentation

All of the pre-trained models were quite large to hold this dataset, and each model could be overfitted easily. To prevent this, some noise was added to the dataset; it is well known that by adding some noise to inputs of neural network, in some situations, this leads to significant improvement in generalizing the dataset. Moreover, adding noise acts as some sort of augmentation of the dataset. Furthermore, other augmentation techniques were also used. Since not all augmentation approaches were good for X-ray images, we processed the images in four steps. First, we resized images to  $224 \times 224 \times 3$ , then three augmentation techniques were used, Random Horizontal Flip (to deal with the pneumonia symptoms on either side of the chest), Random Resized Crop (to get deeper relation among pixels), and finally augmenting images with a varying intensity of images.

## 2.3. Convolutional Neural Networks and the Use of Transfer Learning

Currently, the modern deep learning models in computer vision use convolutional neural networks (CNNs). These layers make the explicit assumption that any input to them is an image. Early convolutional layers in a network process an image and come up with detecting low-level features in images like edges. These layers are successful in capturing the spatial and temporal dependencies in an image. This is done with the help of filters. Unlike normal feed-forward layers, these layers have a much lower number of parameters and use a weight-sharing technique, thus reducing computation efforts. The learnable parameters of each layer consist of filters (or kernels), extended through the full depth of the input volume but these have a small receptive field. When an input is subjected to forward pass, each kernel is convolved across the height and width of the input volume, creating a 2-D activation map of that filter. If 'N' filters are used, then stacking those 'N' activation maps along the depth forms the full output of the convolutional layer [48] (see Figure 2).



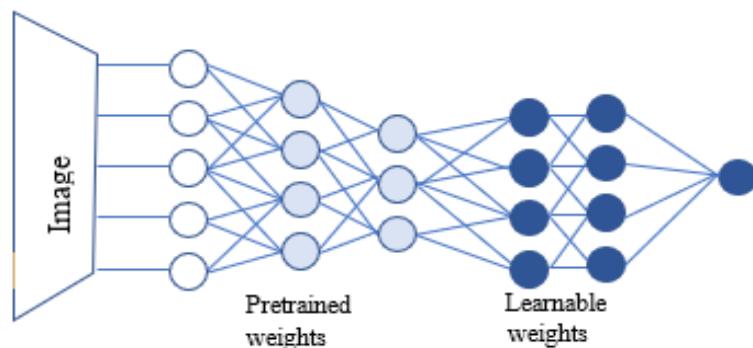
**Figure 2.** Filter operation over convolutional layers.

The activation layer is very useful as it helps to approximate almost any nonlinear function [49]. The feature map from the convolutional layer is taken as input to the activation layer.

Pooling layers are used to reduce the spatial size of representation generated by previous kernels after convolution. This helps in reducing the number of parameters, thus reducing the computation work. These layers are used to extract dominant features that are positional and rotational invariant. It is common practice to include a pooling layer in between two convolutional layers. The most common pooling layer is the max pooling layer; it separates input into squares of a given size, and outputs the maximum value of each square. On the other hand, an average pooling layer finds the average of each square, both the methods reduce the dimensionality and computation efforts [50].

When working on a similar computer vision problem, we can use those pre-trained models, instead of going through the long process of training models from scratch. This method of transferring

learning from one predefined and trained model to some new domain by reusing the network layer weights is called transfer learning (see Figure 3). Transfer learning is a very useful technique and has achieved significant results in computer vision and other areas also [51–54].



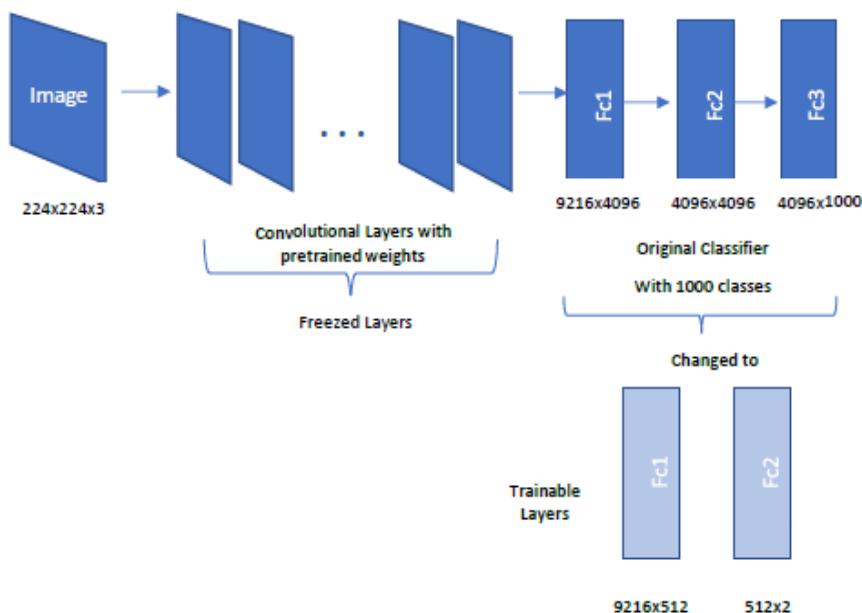
**Figure 3.** Use of general architecture in transfer learning. Layers at last are generally trained on a new domain, and knowledge of the previous domain is used as pretrained weights.

#### 2.4. Pretrained Neural Networks

We employed five different pre-trained models, AlexNet [55], DenseNet121 [56], ResNet18 [57], InceptionV3 [58] and GoogLeNet [59], which were already trained on the ImageNet dataset and then used them on images from the Chest X-ray dataset.

##### 2.4.1. AlexNet Architecture

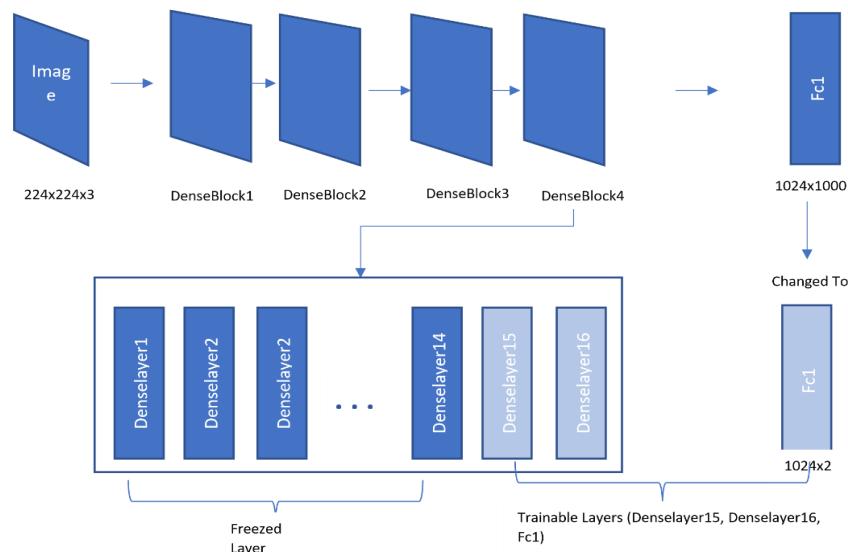
AlexNet is a CNN that is similar to LeNet [60], but deeper. This network replaced the tanh function with a Rectified Linear Unit (ReLU) to add non-linearity. It used dropout layers instead of using regularization to deal with overfitting. Overlapping pooling was also used to reduce the size of the network. We used a pre-trained AlexNet as one of our models and used transfer learning (Figure 4) by freezing the convolutional layers and training the classifier only.



**Figure 4.** AlexNet with trainable and “frozen” layers.

#### 2.4.2. DenseNet121 Architecture

The DenseNet architecture requires fewer parameters than a traditional CNN. DenseNet layers use only 12 filters with a small set of new feature maps (Figure 5). Another problem with DenseNet is the training time, because every layer has its input from previous layers.

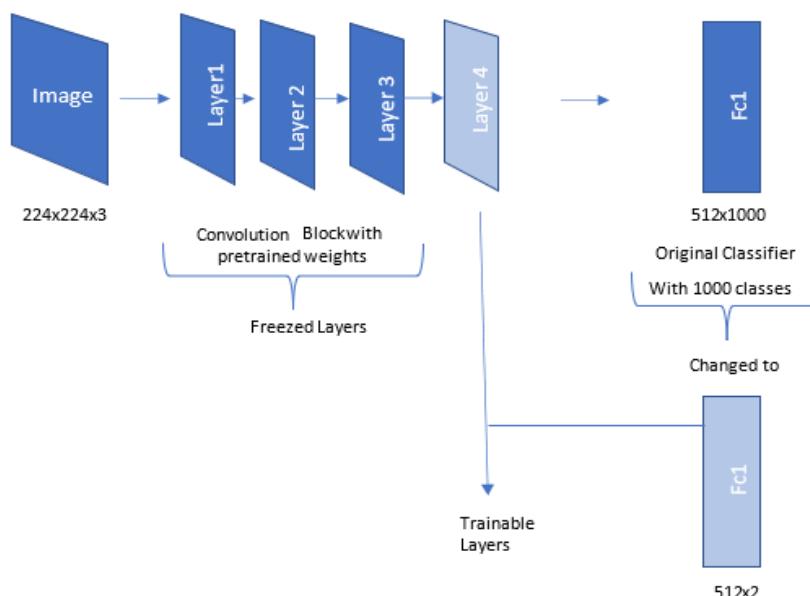


**Figure 5.** DenseNet121 with “frozen” and trainable layers.

However, DenseNet solves this issue by giving access to the gradient values from the loss function and the input image. This significantly reduces the computation cost and makes this model a better choice.

#### 2.4.3. ResNet18 Architecture

The ResNet model comes with a residual learning framework to simplify the training of deeper networks. The architecture is based on the reformulation of network layers as learning residual functions with respect to the layer inputs. The depth of the residual network is eight times deeper than VGG nets [61], but its complexity is lower. Figure 6 shows the architecture used.

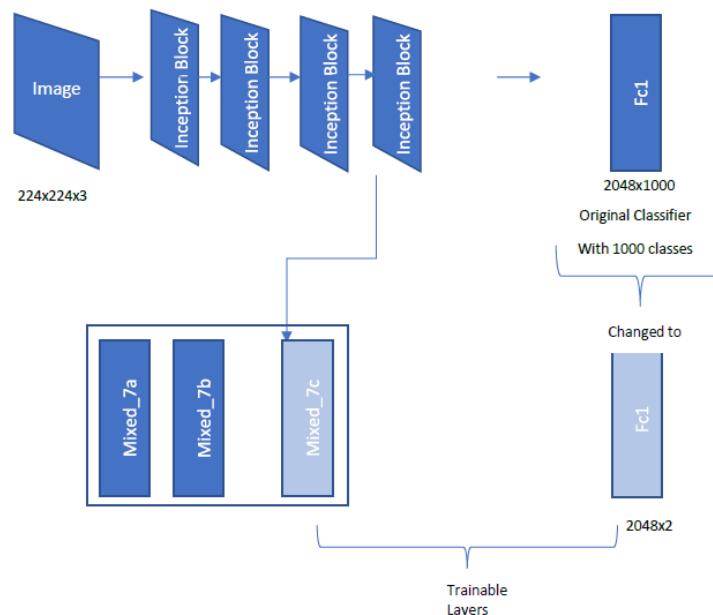


**Figure 6.** ResNet18 architecture with “frozen” and trainable layers.

#### 2.4.4. Inception V3 Architecture

The Inception V3 model allows for increasing the depth and width of the deep learning network, but maintaining the computational cost constant at the same time. This model was trained on the original ImageNet dataset with over 1 million training images. It works as a multi-level feature generator by computing  $1 \times 1$ ,  $3 \times 3$  and  $5 \times 5$  convolutions.

This allows the model to use all kinds of kernels on the image and to get results from all of those. All such outputs are stacked along the channel dimension, and used as input to the next layer. This model achieved top performance for computer vision tasks, by using some advanced techniques. The architecture we use is shown in Figure 7.



**Figure 7.** Inception V3 with “frozen” layers and trainable layers.

#### 2.4.5. GoogLeNet Architecture

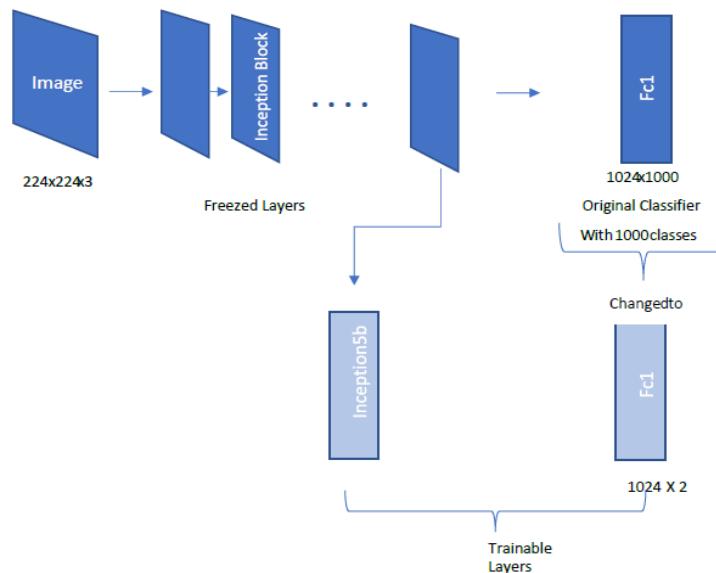
The architecture of GoogLeNet is completely different from AlexNet. This model used global average pooling. It also contains inception modules, which can output convolutions of different types using different kernels on the same input; all outputs are then stacked as the final output of that layer. For instance, the inception module performs  $1 \times 1$ ,  $3 \times 3$  and  $5 \times 5$  convolution and stacks all the outputs together. This model has used  $1 \times 1$  convolution to reduce the dimensionality and computation cost. The architecture used, showing the trainable and “frozen” layers is presented in Figure 8.

#### 2.4.6. Ensemble Classification

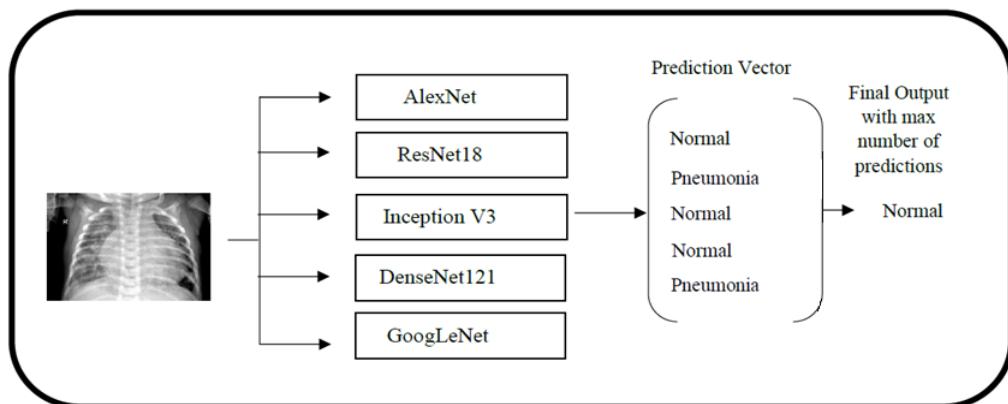
To combine the prediction of five pre-trained neural networks, we use the ensemble classification approach reproduced in Figure 9. The outputs of pre-trained neural networks are combined into a prediction vector, and majority voting is used to come to a final prediction.

#### 2.5. Dataset

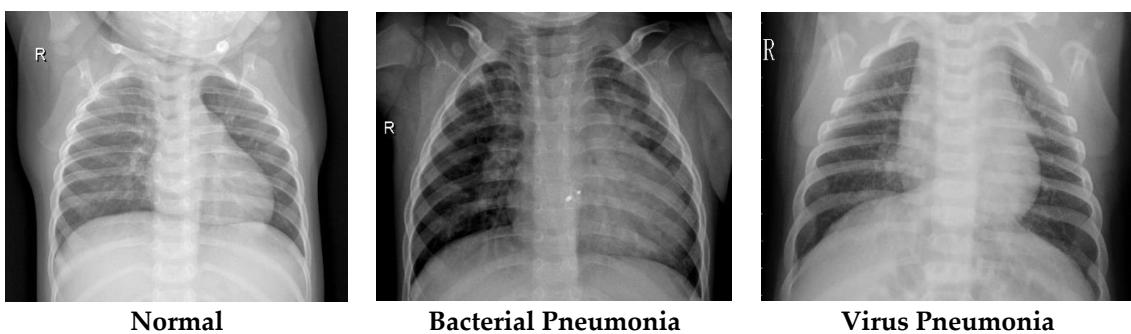
For evaluation we used a dataset from the Guangzhou Women and Children’s Medical Center [62]. The dataset contains a total of 5232 images, where 1346 images belong to the Normal category, 3883 images depicted Pneumonia, out of which 2538 images belongs to Bacterial Pneumonia and 1345 images depicted Virus Pneumonia. The sample images for each class are represented in Figure 10. The partition of this dataset for cross-validation is summarized in Table 1. After the network model is trained using the training test, the test set of images is used to evaluate the accuracy of the model.



**Figure 8.** GoogLeNet with “frozen” layers and trainable layers.



**Figure 9.** Ensemble model using five pretrained different architectures and majority voting.



**Figure 10.** X-ray images for Healthy (Normal), Bacterial Pneumonia and Viral Pneumonia.

**Table 1.** Dataset partition for training and testing.

Category	Training Set (No. of Images)	Test Set (No. of Images)
Normal	1349	234
Pneumonia	3883	390
Total	5232	624

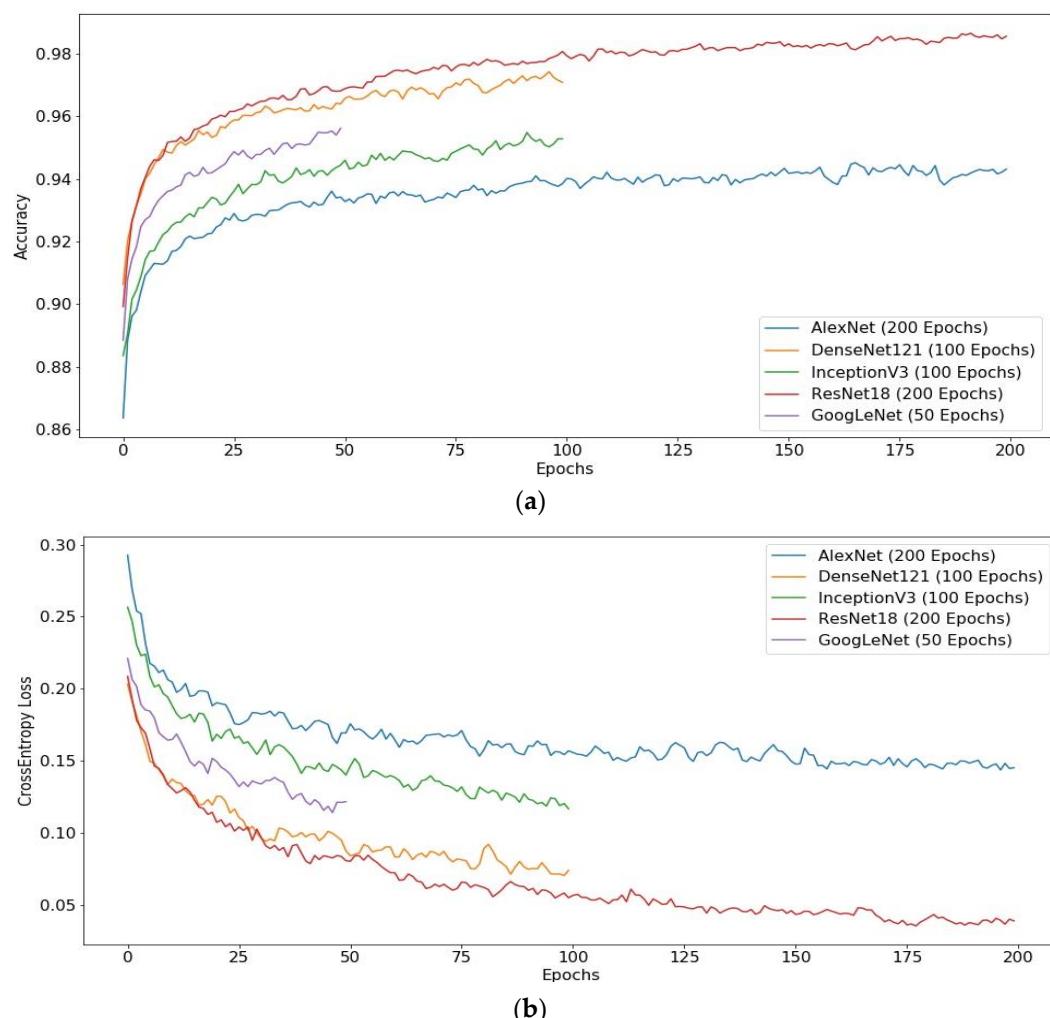
### 3. Results

#### 3.1. Results

The primary goal of our transferring learning approach was to correctly diagnose pneumonia among normal chest X-ray images. For this we prepared all of the models as shown above and trained them separately. We performed training and testing using a computer with Intel(R) Core (TM) i7-6700 (Intel Corporation, Santa Clara, CA, USA) with 3.30 GHz CPU, NVIDIA GeForce GTX 1070 8 GB GPU (NVIDIA Corporation, Santa Clara, CA, USA) and 48 GB of RAM. For training, we used the Adam optimizer [63] and the cross-entropy loss function. The learning rate is started from the value of 0.001 and is reduced by 2 after every three epochs. AlexNet was trained for 200 iterations at learning rate of 0.001 and then trained at very low learning rate of 0.00001; it achieved a test accuracy of 92.86% and a train accuracy of 93.0%. The AUC value was 97.83%. ResNet18 did better than AlexNet and other models; it achieved an area under the ROC curve as 99.36% and test accuracy of 94.23%.

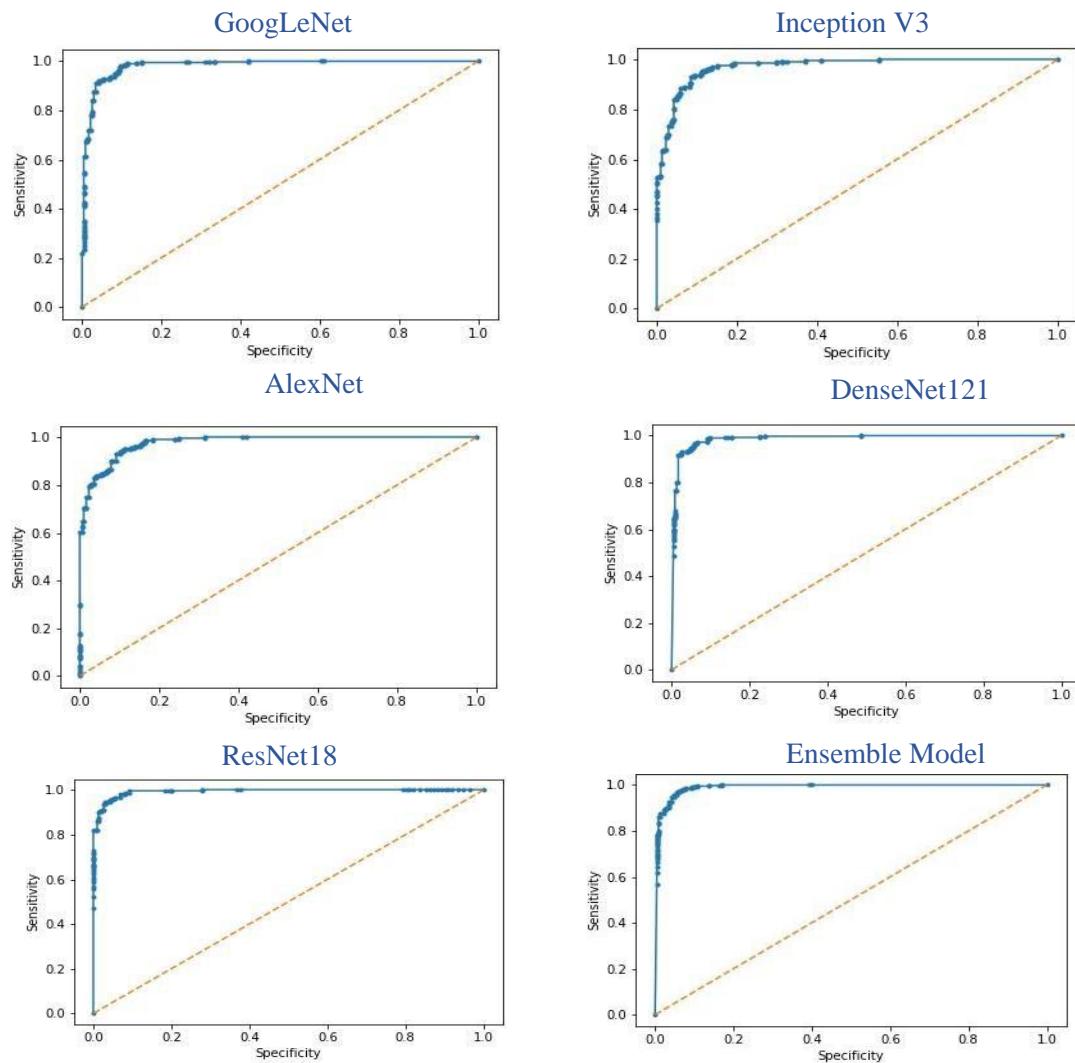
The average computational time for all models on CPU is 0.332 s and for GPU it is 0.043 s, whereas the ensemble model took 0.161 s for computation.

Figure 11 shows plots of accuracy and loss against epochs. The best results were obtained by the ResNet18 network both in terms of accuracy and loss values.



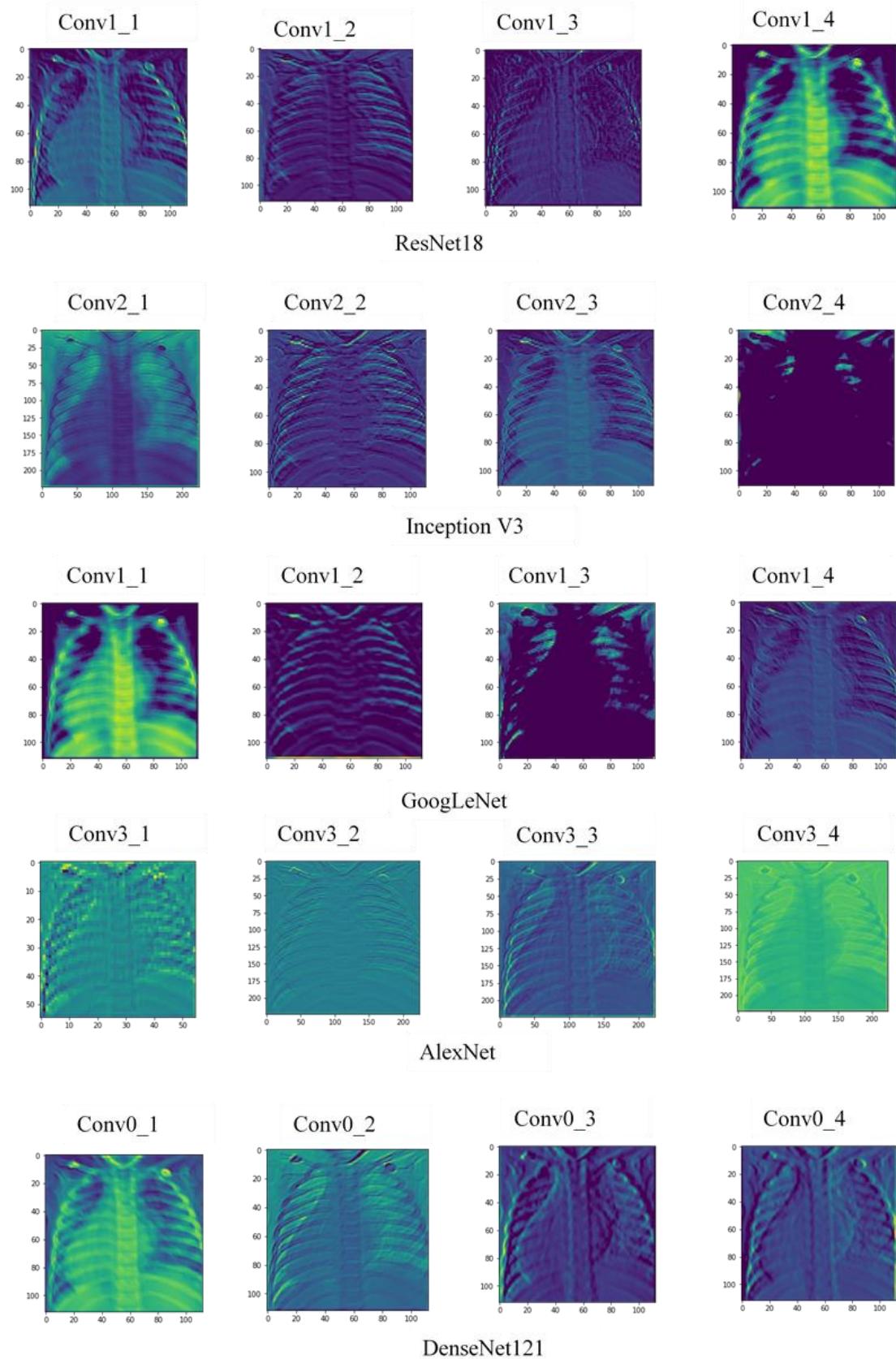
**Figure 11.** Accuracy against epoch (a) and cross-entropy loss against epoch (b) for all trained models. DenseNet121 and InceptionV3 were trained for 100 epochs, and GoogLeNet was trained for 50 epochs. This was done to prevent overfitting and get a better generalization. Plots are drawn for the training set.

Figure 12 shows the plot of sensitivity against specificity for each model and the ensemble model on the testing set. Note that the ensemble model has performed better.



**Figure 12.** Sensitivity and specificity for each model and the ensemble model on the testing set.

Figure 13 shows the activation map in early convolutional layers for each of the models. Note that the networks can capture and learn similar activation maps, albeit at different layers. For example, the Conv1\_4 layer activation map of ResNet18 is very similar to the Conv1\_1 layer activation map of Inception V3. Note that activation maps show that the network in most cases correctly differentiates between bacterial pneumonia, which is characterized by activations in the regions of lobar consolidations, and viral pneumonia, which typically has diffuse interstitial patterns across the lungs.



**Figure 13.** Plots of activation maps for each model in early convolutional layers. Activation maps are calculated for the positive class and are shown for the last layer of each convolutional block before pooling.

Table 2 shows the results for each neural network model. After analyzing the obtaining of the predictions of all models, we combined the results of 5 models: AlexNet, DenseNet121, InceptionV3, GoogLeNet, and ResNet18, and predicted the class which was most frequently predicted by all models. This combination gave the remarkably best performance and achieved a test accuracy of 96.39%, with an area under the ROC curve of 99.34%, and the sensitivity as 99.62%.

**Table 2.** Comparative results for each model on the test set.

Model	Epoch	Recall (%)	Precision (%)	AUC (%)	Test Accuracy (%)
AlexNet	200	98.97	90.21	97.83	92.86
DenseNet121	100	99.23	91.18	98.78	92.62
InceptionV3	100	98.46	90.30	97.33	92.01
GoogLeNet	50	99.48	90.44	98.29	93.12
ResNet18	200	99.48	91.58	99.36	94.23
Ensemble model	—	99.62	93.28	99.34	96.39

### 3.2. Comparison

We compare our results with the results of other authors using the same dataset. The comparison is given in Table 3. In the comparison of pneumonia versus normal, Kermany et al. [62] used Tensorflow and adapted an Inception V3 architecture pretrained on the ImageNet dataset. Retraining consisted of initializing the convolutional layers with loaded pretrained weights. They obtained a sensitivity of 93.2% and a specificity of 90.1%, with an accuracy of 92.8%. Cohen et al. [64] used an implementation of the CheXnet DenseNet-121 model [23], which is based on the DenseNet-121 [50]. It is trained using Adam optimization with default parameters values ( $b1 = 0.9$  and  $b2 = 0.999$ ), a learning rate of 0.001, and a learning rate decay of 0.1, and achieved an AUC of 98.4% (unfortunately, no other metric value was provided in their paper for this dataset).

**Table 3.** Comparative results for each model on test set.

Model	Specificity	Recall/Sensitivity (%)	Precision (%)	Area Under Curve (%)	Test Accuracy (%)
Chester [64]	—	—	—	98.4	—
Kermany et al. [62]	90.1	93.2	—	—	92.8
This paper	—	99.62	93.28	99.34	96.39

Despite other authors not providing full data for evaluation, our model has outperformed other solutions in recall (sensitivity), accuracy and Area Under the Curve (AUC).

## 4. Discussion

Correct diagnostics requires deeper understanding of the radiological features visible in chest X-rays. Unfortunately, deep neural networks are known for providing no explanation as to how the final decision is made. To make the decision support process more useful, the deep network should provide also explanations beyond plain decisions [65]. Such explanations can take the form of semantic segmentation with explanations in natural language assigned to each identified segment of a chest X-ray photo. Correct implementation of such tasks may require the use of additional eHealth data from the patients, and high-quality annotated datasets.

Another limitation is introduced by the scarcity of image data representing all types of pneumonia pathologies, which prevents for achieving a higher accuracy or using a deeper network with more parameters. Successful deep learning models such as AlexNet, GoogLeNet and ResNet, have been trained on more than a million images, which are hardly available in the medical domain. Training deep neural networks with limited data available also may lead to over-fitting and prevent from good generalization.

The accuracy results reached in this paper could be still improved by adding more sophisticated deep networks to the ensemble and training the networks with a larger dataset.

Future research directions will include the exploration of image data augmentation techniques [66] to improve accuracy even more, while avoiding overfitting.

## 5. Conclusions

In this article, our goal is to propose a deep learning-based approach to classify pneumonia from chest X-ray images using transfer learning. In this framework, we adopted the transfer learning approach and used the pretrained architectures, AlexNet, DenseNet121, Inception V3, GoogLeNet and ResNet18 trained on the ImageNet dataset, to extract features. These features were passed to the classifiers of respective models, and the output was collected from individual architectures. Finally, we employed an ensemble model that used all five pretrained models and outperformed all other models. We observed that performance could be improved further, by increasing dataset size, using a data augmentation approach, and by using hand-crafted features, in future.

Our findings support the notion that deep learning methods can be used to simplify the diagnostic process and improve disease management. While pneumonia diagnoses are commonly confirmed by a single doctor, allowing for the possibility of error, deep learning methods can be regarded as a two-way confirmation system. In this case, the decision support system provides a diagnosis based on chest X-ray images, which can then be confirmed by the attending physician, drastically minimizing both human and computer error. Our results suggest that deep learning methods can be used to improve diagnosis relative to traditional methods, which may improve the quality of treatment. When compared with the previous state-of-the-art methods, our approach can effectively detect the inflammatory region in chest X-ray images of children.

**Author Contributions:** Conceptualization, P.T., C.M., and V.H.C.d.A.; methodology, C.M., and V.H.C.d.A.; software, V.C., S.K.S., A.K., D.G., P.T., and C.M.; validation, V.C., S.K.S., A.K., D.G., P.T., C.M., and V.H.C.d.A.; formal analysis, C.M., R.D. and V.H.C.d.A.; investigation, V.C., S.K.S., A.K., D.G., P.T., and C.M.; resources, V.C., S.K.S., A.K., and D.G.; data curation, V.C., S.K.S., and A.K.; writing—original draft preparation, V.C., S.K.S., A.K., D.G., P.T., and C.M.; writing—review and editing, R.D. and V.H.C.d.A.; visualization, V.C., S.K.S., A.K., D.G., and P.T.; supervision, C.M. and V.H.C.d.A. All authors have read and agree to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Liu, N.; Wan, L.; Zhang, Y.; Zhou, T.; Huo, H.; Fang, T. Exploiting Convolutional Neural Networks With Deeply Local Description for Remote Sensing Image Classification. *IEEE Access* **2018**, *6*, 11215–11228. [[CrossRef](#)]
2. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
3. Brunetti, A.; Carnimeo, L.; Trotta, G.F.; Bevilacqua, V. Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: A survey based on medical images. *Neurocomputing* **2019**, *335*, 274–298. [[CrossRef](#)]
4. Asiri, N.; Hussain, M.; Al Adel, F.; Alzaidi, N. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. *Artif. Intell. Med.* **2019**, *99*. [[CrossRef](#)] [[PubMed](#)]
5. Zhou, T.; Thung, K.; Zhu, X.; Shen, D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Hum. Brain Mapp.* **2018**, *40*, 1001–1016. [[CrossRef](#)]
6. Shickel, B.; Tighe, P.J.; Bihorac, A.; Rashidi, P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1589–1604. [[CrossRef](#)]
7. Meyer, P.; Noblet, V.; Mazzara, C.; Lallement, A. Survey on deep learning for radiotherapy. *Comput. Biol. Med.* **2018**, *98*, 126–146. [[CrossRef](#)]

8. Malūkas, U.; Maskeliūnas, R.; Damaševičius, R.; Woźniak, M. Real time path finding for assisted living using deep learning. *J. Univers. Comput. Sci.* **2018**, *24*, 475–487.
9. Zhang, X.; Yao, L.; Wang, X.; Monaghan, J.; McAlpine, D. A Survey on Deep Learning based Brain Computer Interface: Recent Advances and New Frontiers. *arXiv* **2019**, arXiv:1905.04149.
10. Bakator, M.; Radosav, D. Deep Learning and Medical Diagnosis: A Review of Literature. *Multimodal Technol. Interact.* **2018**, *2*, 47. [CrossRef]
11. Gilani, Z.; Kwong, Y.D.; Levine, O.S.; Deloria-Knoll, M.; Scott, J.A.G.; O'Brien, K.L.; Feikin, D.R. A literature review and survey of childhood pneumonia etiology studies: 2000–2010. *Clin. Infect. Dis.* **2012**, *54* (Suppl. 2), S102–S108. [CrossRef] [PubMed]
12. Bouch, C.; Williams, G. Recently published papers: Pneumonia, hypothermia and the elderly. *Crit. Care* **2006**, *10*, 167. [CrossRef] [PubMed]
13. Scott, J.A.; Brooks, W.A.; Peiris, J.S.; Holtzman, D.; Mulholland, E.K. Pneumonia research to reduce childhood mortality in the developing world. *J. Clin. Investig.* **2008**, *118*, 1291–1300. [CrossRef] [PubMed]
14. Wunderink, R.G.; Waterer, G. Advances in the causes and management of community acquired pneumonia in adults. *BMJ* **2017**, *358*, j2471. [CrossRef] [PubMed]
15. National Center for Health Statistics (NCHS); Centers for Disease Control and Prevention (CDC) FastStats: Pneumonia. Last Updated February 2017. Available online: <http://www.cdc.gov/nchs/fastats/pneumonia.htm> (accessed on 21 November 2019).
16. Heron, M. Deaths: Leading causes for 2010. *Natl. Vital. Stat. Rep.* **2013**, *62*, 1–96.
17. World Health Organization. *The Top 10 Causes of Death*; World Health Organization: Geneva, Switzerland, 2017; Available online: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (accessed on 10 November 2019).
18. Kallianos, K.; Mongan, J.; Antani, S.; Henry, T.; Taylor, A.; Abuya, J.; Kohli, M. How far have we come? Artificial intelligence for chest radiograph interpretation. *Clin. Radiol.* **2019**, *74*, 338–345. [CrossRef]
19. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. 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, Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA; pp. 2097–2106.
20. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer International Publishing: New York, NY, USA; pp. 234–241.
21. Roth, H.R.; Lu, L.; Seff, A.; Cherry, K.M.; Hoffman, J.; Wang, S.; Liu, J.; Turkbey, E.; Summers, R.M. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In Proceedings of the 17th International Conference on Medical Image Computing and Computer-Assisted Intervention, Boston, MA, USA, 14–18 September 2014; Springer International Publishing: New York, NY, USA; pp. 520–527.
22. Shin, H.-C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [CrossRef]
23. Rajpurkar, P.; Irvin, J.; Ball, R.L.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.P.; et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med.* **2018**, *15*, e1002686. [CrossRef]
24. Woźniak, M.; Połap, D.; Capizzi, G.; Sciuto, G.L.; Kośmider, L.; Frankiewicz, K. Small lung nodules detection based on local variance analysis and probabilistic neural network. *Comput. Methods Programs Biomed.* **2018**, *161*, 173–180. [CrossRef]
25. Gu, Y.; Lu, X.; Yang, L.; Zhang, B.; Yu, D.; Zhao, Y.; Gao, L.; Wu, L.; Zhou, T. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. *Comput. Biol. Med.* **2018**, *103*, 220–231. [CrossRef]
26. Ho, T.K.K.; Gwak, J. Multiple feature integration for classification of thoracic disease in chest radiography. *Appl. Sci.* **2019**, *9*, 4130. [CrossRef]
27. Jaiswal, A.K.; Tiwari, P.; Kumar, S.; Gupta, D.; Khanna, A.; Rodrigues, J.J.P.C. Identifying pneumonia in chest X-rays: A deep learning approach. *Meas. J. Int. Meas. Confed.* **2019**, *145*, 511–518. [CrossRef]

28. Jung, H.; Kim, B.; Lee, I.; Lee, J.; Kang, J. Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method. *BMC Med. Imaging* **2018**, *18*, 48. [[CrossRef](#)] [[PubMed](#)]
29. Lakhani, P.; Sundaram, B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology* **2017**, *284*, 574–582. [[CrossRef](#)] [[PubMed](#)]
30. Li, X.; Shen, L.; Xie, X.; Huang, S.; Xie, Z.; Hong, X.; Yu, J. Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection. *Artif. Intell. Med.* **2019**, *101744*. [[CrossRef](#)] [[PubMed](#)]
31. Liang, G.; Zheng, L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput. Methods Programs Biomed.* **2019**, *104964*. [[CrossRef](#)] [[PubMed](#)]
32. Nam, J.G.; Park, S.; Hwang, E.J.; Lee, J.H.; Jin, K.; Lim, K.Y.; Park, C.M. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* **2019**, *290*, 218–228. [[CrossRef](#)]
33. Nasrullah, N.; Sang, J.; Alam, M.S.; Mateen, M.; Cai, B.; Hu, H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors* **2019**, *19*, 3722. [[CrossRef](#)]
34. Pasa, F.; Golkov, V.; Pfeiffer, F.; Cremers, D.; Pfeiffer, D. Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. *Sci. Rep.* **2019**, *9*, 6268. [[CrossRef](#)]
35. Pezeshk, A.; Hamidian, S.; Petrick, N.; Sahiner, B. 3-D convolutional neural networks for automatic detection of pulmonary nodules in chest CT. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 2080–2090. [[CrossRef](#)]
36. Sirazitdinov, I.; Kholiavchenko, M.; Mustafaev, T.; Yixuan, Y.; Kuleev, R.; Ibragimov, B. Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Comput. Electr. Eng.* **2019**, *78*, 388–399. [[CrossRef](#)]
37. Souza, J.C.; Bandeira Diniz, J.O.; Ferreira, J.L.; França da Silva, G.L.; Corrêa Silva, A.; de Paiva, A.C. An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Comput. Methods Programs Biomed.* **2019**, *177*, 285–296. [[CrossRef](#)] [[PubMed](#)]
38. Taylor, A.G.; Mielke, C.; Mongan, J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. *PLoS Med.* **2018**, *15*, e1002697. [[CrossRef](#)] [[PubMed](#)]
39. Wang, Q.; Shen, F.; Shen, L.; Huang, J.; Sheng, W. Lung nodule detection in CT images using a raw patch-based convolutional neural network. *J. Digit. Imaging* **2019**, *32*, 971–979. [[CrossRef](#)] [[PubMed](#)]
40. Xiao, Z.; Du, N.; Geng, L.; Zhang, F.; Wu, J.; Liu, Y. Multi-scale heterogeneous 3D CNN for false-positive reduction in pulmonary nodule detection, based on chest CT images. *Appl. Sci.* **2019**, *9*, 3261. [[CrossRef](#)]
41. Xie, Y.; Zhang, J.; Xia, Y. Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT. *Med. Image Anal.* **2019**, *57*, 237–248. [[CrossRef](#)]
42. Xu, S.; Wu, H.; Bie, R. CXNet-m1: Anomaly detection on chest X-rays with image-based deep learning. *IEEE Access* **2019**, *7*, 4466–4477. [[CrossRef](#)]
43. Yates, E.J.; Yates, L.C.; Harvey, H. Machine learning “red dot”: Open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification. *Clin. Radiol.* **2018**, *73*, 827–831. [[CrossRef](#)]
44. da Nóbrega, R.V.M.; Rebouças Filho, P.P.; Rodrigues, M.B.; da Silva, S.P.P.; Dourado Júnior, C.M.J.M.; de Albuquerque, V.H.C. Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks. *Neural Comput. Appl.* **2018**, *1*–18. [[CrossRef](#)]
45. Ke, Q.; Zhang, J.; Wei, W.; Połap, D.; Woźniak, M.; Kośmider, L.; Damaševičius, R. A neuro-heuristic approach for recognition of lung diseases from X-ray images. *Expert Syst. Appl.* **2019**, *126*, 218–232. [[CrossRef](#)]
46. Behzadi-khormouji, H.; Rostami, H.; Salehi, S.; Derakhshande-Rishehri, T.; Masoumi, M.; Salemi, S.; Keshavarz, A.; Gholamrezanezhad, A.; Assadi, M.; Batouli, A. Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images. *Comput. Methods Programs Biomed.* **2020**, *185*, 105162. [[CrossRef](#)] [[PubMed](#)]
47. Fan, J.; Cao, X.; Yap, P.T.; Shen, D. BIRNet: Brain image registration using dual-supervised fully convolutional networks. *Med. Image Anal.* **2019**, *54*, 193–206. [[CrossRef](#)] [[PubMed](#)]
48. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6.
49. Goyal, M.; Goyal, R.; Lall, B. Learning Activation Functions: A new paradigm of understanding Neural Networks. *arXiv* **2019**, arXiv:1906.09529.

50. Bailer, C.; Habtegebrial, T.; Varanasi, K.; Stricker, D. Fast Feature Extraction with CNNs with Pooling Layers. *arXiv* **2018**, arXiv:1805.03096.
51. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H.; Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N.D.; Weinberger, K.Q. (Eds.) How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems 27, Proceedings of the Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13 December 2014*; Neural Information Processing Systems Foundation, Inc. (NIPS): Montreal, QC, Canada, 2014; pp. 3320–3328.
52. Dai, W.; Chen, Y.; Xue, G.-r.; Yang, Q.; Yu, Y.; Koller, D.; Schuurmans, D.; Bengio, Y.; Bottou, L. (Eds.) Translated Learning: Transfer Learning across Different Feature Spaces. In *Advances in Neural Information Processing Systems 21, Proceedings of the Neural Information Processing Systems 2008, Vancouver, BC, Canada, 8–10 December 2008*; Neural Information Processing Systems Foundation, Inc. (NIPS): Vancouver, BC, Canada, 2008; pp. 353–360.
53. Raghu, M.; Zhang, C.; Kleinberg, J.M.; Bengio, S. Transfusion: Understanding Transfer Learning with Applications to Medical Imaging. *arXiv* **2019**, arXiv:1902.07208.
54. Ravishankar, H.; Sudhakar, P.; Venkataramani, R.; Thiruvenkadam, S.; Annangi, P.; Babu, N.; Vaidya, V. Understanding the Mechanisms of Deep Transfer Learning for Medical Images. In *Deep Learning and Data Labeling for Medical Applications; DLMIA 2016, LABELS 2016; Carneiro, G., Ed.*; Springer: Cham, Switzerland, 2016; Volume 10008.
55. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
56. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2016**, arXiv:1608.06993.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
58. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv* **2015**, arXiv:1512.00567.
59. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. *arXiv* **2014**, arXiv:1409.4842.
60. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
61. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015*.
62. Kerney, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)] [[PubMed](#)]
63. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference for Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015*.
64. Cohen, J.P.; Bertin, P.; Frappier, V. Chester: A Web Delivered Locally Computed Chest X-Ray Disease Prediction System. *arXiv* **2019**, arXiv:1901.11210.
65. Holzinger, A.; Langs, G.; Denk, H.; Zatloukal, K.; Müller, H. Causability and explainability of artificial intelligence in medicine. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1312. [[CrossRef](#)]
66. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).