How to Cite:
Natarajan, G., & Dhanalakshmi, P. (2022). Classification of pneumonia using pre-trained convolutional networks on chest X-Ray images. *International Journal of Health Sciences*, *6*(S1), 5378–5390. https://doi.org/10.53730/ijhs.v6nS1.6097

Classification of pneumonia using pre-trained convolutional networks on chest X-Ray images

G. Natarajan
Research Scholar, Department of Computer and Information Science, Annamalai University, Chidambaram
Email natarajang@gmail.com

Dr. P. Dhanalakshmi
Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram
Email: abidhana01@gmail.com

***Abstract*---**Pneumonia is an infection that is caused to the people of all ages with mild to severe inflammation of the lung disease. The most common and best method for the diagnosis of pneumonia is chest radiography. But diagnosing pneumonia from chest radiographs is a difficult task, even for radiologists. To overcome, Pre-Trained Convolutional Neural Networks namely Inceptionv3 and Resnet50 are used as a feature extractor. The exacted features are fed into 1D CNN which is classifies into Normal, Bacterial Pneumonia and Viral Pneumonia. When comparing Inceptionv3 with 1D CNN and resnet50 with 1D CNN, it is analyzed that Inceptionv3 with 1D CNN gives the satisfactory results of 96.04%.

***Keywords*---**Pneumonia, pre-trained convolutional, chest X-Ray images.

Introduction

Pneumonia is an inflammation of the lungs that can be caused by pathogens, such as bacteria, viruses, and fungi [1][2]. It can happen to anyone, even young or even healthy people. It can be fatal in infants, people with other illnesses, people with weakened immune systems, the elderly, people who are hospitalized and on ventilators, people with chronic conditions such as asthma, asthma and smokers. The cause of the pneumonia also determines its severity. Viral pneumonia is milder and symptoms come on gradually. However, the diagnosis can become complicated if a bacterial infection develops along with viral pneumonia. On the other hand, bacterial pneumonia is more severe and its symptoms can appear

Manuscript submitted: 09 Feb 2022, Manuscript revised: 27 March 2022, Accepted for publication: 18 April 2022

gradually or even suddenly, especially in children. Figure 1. Shows the overall framework of the proposed system.

This type of pneumonia affects a large part of the lung and can affect multiple lobes of the lung. When multiple lobes of the lung are affected, a person is hospitalized [3]. Another form of pneumonia is fungal pneumonia, which can occur in people with weakened immune systems. This type of pneumonia can also be dangerous and it takes time for the person to recover. Therefore, the research and development of new methods to provide computer-aided diagnosis to reduce pneumonia-related mortality, especially infant mortality, in developing countries is essential.

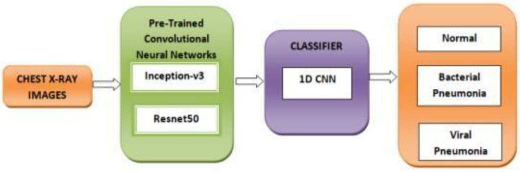


Fig 1. Proposed Method for classification of Pneumonia into Bacterial Pneumonia and Viral Pneumonia

Literature of the work

In the proposed [4] a CheXNet algorithm, a 121-slice CNN, to classify 14 different diseases, including pneumonia seen on a chest x-ray. CheXNet was trained with the Chest Xray14 data set. The performance of CheXNet was compared to that of radiologists and it was found that CheXNet outperforms the average performance of radiologists in the metric F1.

In [5], transfer learning was used to classify pediatric chest x-rays into normal or pneumoniainfected images. In addition, were divided into two categories: bacterial pneumonia and viral pneumonia. In order to classify normal and pneumonia patients using chest x-ray, four common CNN-based deep learning techniques were trained and tested in [6]. These algorithms were DenseNet201, ResNet18, SqueezeNet, and AlexNet. The results showed that DenseNet201 outperforms the other three. The accuracy, precision and recall of the classification of pneumonia images and normal images, images of viral and bacterial pneumonia and normal images only were (98%, 97% and 99%), (95%, 95% and 96th). %) and (93.3%, 93.7% and 93.2%).

This article [7] develops four different models by changing the deep learning method used; Two pre-trained models, ResNet152V2 and MobileNetV2, a Convolutional Neural Network (CNN) and a Long and Short Term Memory (LSTM). The proposed models were implemented and evaluated using Python and compared with recent similar research. improves accuracy, precision, F1 value, recall and area under the curve (AUC) by 99.22%, 99.43%, 99.44%, 99.44% and 99.77% respectively As the results clearly show ResNet152V2 model outperforms other recently proposed models Additionally, the other proposed models - MobileNetV2, CNN, and LSTMCNN - scored more than 91% in precision, memory, F1 score, precision, and AUC, outperforming the models recently in the literature were presented.

A. Pre-Trained Convolutional Neural Networks

The feature extraction performance in chest X-ray images presents two re-training techniques known as Transfer-learning and Fine tuning. In this work, Transfer learning has been used on the Inception-v3 and Resnet50. Inception-v3 and Resnet50 use the weights of its network to perform the task of classification of chest x-ray image dataset which is classified into Bacterial Pneumonia and Viral Pneumonia.

B. Transfer Learning

Transfer learning is a technique used to deal with small or large amounts of labeled information from the source domain to a destination domain, in order to build an efficient prediction model. The Inception-v3 CNN model has been pre-trained by utilizing the ImageNet dataset, to freeze its layers rendering them non-trainable. Then the last layer is removed from the fully connected or SoftMax. A new layer is added fully connected, taking the rest of the network for feature extraction and for training the model.

C. Inceptionv3

Inception network is an up-to-date deep learning model. It is mainly used to solving image identification and detecting problems. The Inception deep convolutional architecture was launched by GoogleNet in 2015 and it was named as Inception-v1 [8]. Next, the inception was refined by batch normalization and then Inception-v2 came into existence. Now, in Inceptionv3, more factorization is introduced. Based on the earlier versions, the factorization of 3 × 3 convolutions takes place instead of standard 7 × 7 convolution. A set of 3 standard inception models are incorporated for the network Inception part at 35×35 besides 288 filters each. It is minimized to 17 × 17 grid with 768 filters with grid reduction. It is preceded by 5recurrence of factorized inception modules. The Inception modules consists a set of 8 × 8 level with linked output filter bank size of 2048 for each tile. However, the network quality is relatively stable towards modifications. The Inception-v3 is 42 layers deep, which works more efficiently than VGGNet and it performs concatenation of many various sized convolutional filters into a new filter [9] This model reduces the number of parameters which under goes the training and thereby minimizes the computation complexity. Fig. 2 illustrates the overall architecture of Inception-v3 model. The architecture consists of factorization into smaller Convolutions, Factorization into Asymmetric Convolution, Grid Size Reduction, Auxiliary Classifier and Regularization via Label Smoothing.

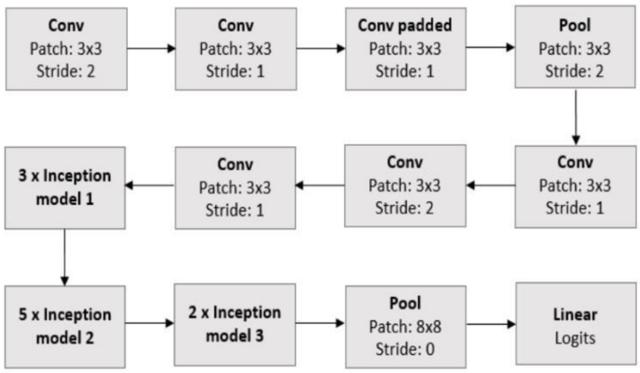


Fig 2 Basic architecture of Inception-v3

Factorization into smaller convolutions

In this stage, the dimensions/parameters are reduced without decreasing the efficiency of the factorizing convolutions.

Grid Size Reduction

At this point, when more convolutional layers are applied, the Grid size decrease is realized by pooling, followed by the convolution operation.

Auxiliary Classifiers

Auxiliary classifiers improve the convergence of very profound networks. The main intention is to push the essential gradients to the lower layers during the training by combating the disappearing gradient issue in very deep networks.

For one image, we extract a 2048-dimensional feature from the last fully-connected logits layer. The layers before the fully connected layer of the pre-trained networks perform the feature extraction for the images and the fully-connected layers are used for classification. In this work, the fully connected layers are replaced by 1D CNN to perform classification. This model is referred as a single transfer learning network. The information related to the features produced from the pre-trained deep CNN is given here. The overall architecture of Inception-v3 is depicted in Fig. 3.

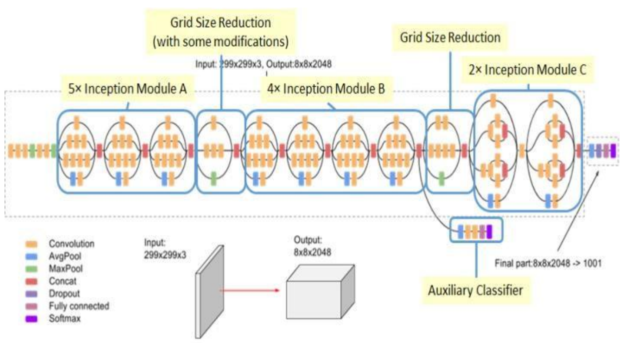


Fig 3. Shows the overall architecture of Inceptionv3

D. Resnet50

Resnet50 is otherwise called the Residual network used to identify mapping by shortcuts. It is the commonly used model in CNN [10] .Residual Networks comprises of various subsequent residual modules, which are the basic foundations block of Resnet. As the network goes deeper and deeper, the training is more difficult. Generally, the input feature map will be followed by the convolutional filter, non-linear activation function and a pooling operation and finally the output is the next layer. Here, back propagation algorithm is implemented. As the network goes deeper and deeper, it is hard to converge.

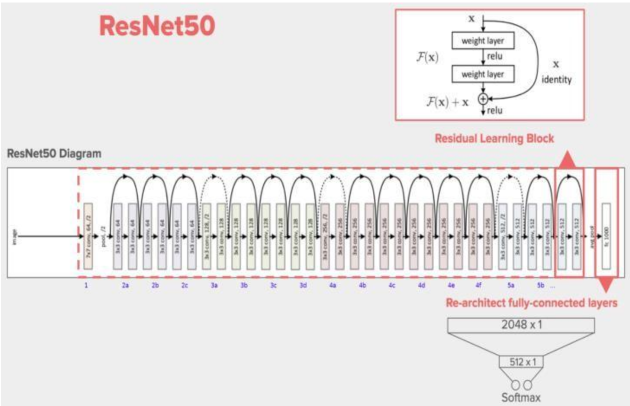


Fig 4. Overall architecture of Resnet50

The architecture of ResNet50 is depicted in Fig. 4.The construction of ResNet50 has 4 stages as shown in Fig 5.4. The input size of the image is 224 x 224 x 3. Every ResNet structure makes the first convolution and max pooling using 7 x 7 and 3 x 3 kernel sizes distinctively. Next, first stage of the network commences and it comprises of 3 Residual blocks containing with 3 layers each. The size of the kernels utilized to perform the convolution operation with all 3 layers of the block of the first stage is 64, 64 and 128 distinctively. The curved arrows refer to the identity connection.

The dashed connected arrow represents that the convolution operation in the Residual block is executed with stride 2, therefore, the size of input will be decreased to half in relation to height and width but the channel width will be doubled. As we move from one stage to the next, the channel width is increased twice as much and the size of the input is decreased to half. For deeper networks like Resnet50, Resnet152, etc, bottleneck design is used. For each residual function F, 3 layers are stacked one over the other. The three layers are 1x1, 3x3, 1x1 convolutions. The 1x1 convolution layers are responsible for decreasing and then replacing the dimensions. The 3x3 layer remains as a bottleneck with less input/output dimensions. Finally, the network has an average pooling layer by a connected layer with 1000 neurons.

Modeling the features for Pneumonia

E. 1D CNN

In this work, 1D CNN is developed to model to classify the Parkinson disease. 1D-CNN models specifically designed for image classification problems, where the model learning internally represents the two-dimensional input, in a process called feature learning The architecture of 1D CNN is shows in Fig 5.

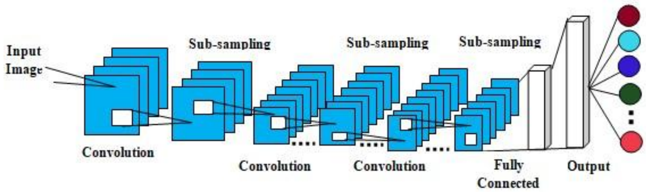


Fig. 5 Architecture of 1D CNN Model

The basic building blocks of 1D-CNN are:

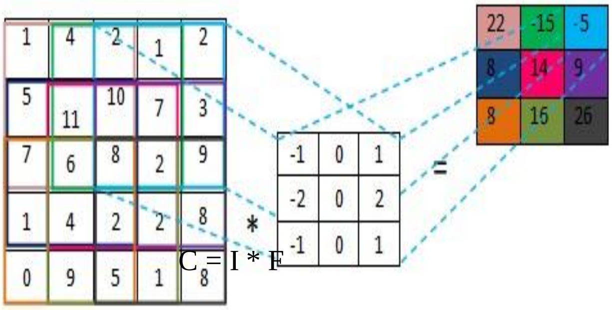
- Convolution layer
- Activation layer
- Pooling layer
- Fully connected layer

F. Convolution Layer

A 'filter' sometimes called a 'kernel' is passed over the image, viewing a few pixels at a time with size of 3x3 or 5x5. The convolution operation is a dot product of the weights defined in the filter with original pixel values as illustrated in Fig 6. The results are summed up into one number that represents all the pixels the filter observed. This setting enables the network to learn different features while keeping the number of parameters tractable. Mathematically, the output feature map $y_{i,j}^{(l)}$ at convolutional layer l is calculated as:

$$y_{i,j}^{(l)} = \sigma^{(l)}\left(\sum_{n=1}^{k}\sum_{m=1}^{k} w_{n,m}^{(l)} . x_{i+n,\, j+m}^{l-1} + b^{(l)}\right) \quad (2)$$

where, the $w_{n,m}^{(l)}$ denoted the convolutional filter with size k x k at layer l, and the $x_{i+n,\ j+m}^{(l)}$ represent the spatial position of the corresponding feature map at the preceding layer l-1. The algorithm passes the convolutional filter throughout the input feature map using the dot product (.) between them with an addition of a bias unit $b^{(l)}$. Moreover, a non-linear activation function $\sigma^{(l)}$ at layer l is taken outside the dot product to strength the nonlinearity.



I - Image Pixel F - Filter 3x3

Fig. 6 Convolution Operation between Image Pixel and Filter

G. Activation Function

Activation functions are really important for learning and making sense of something really complicated and non-linear dynamic functional mapping between inputs and response variable for an Artificial Neural Network. The convolution layer generates a matrix that is much smaller in size than the original image. This matrix is run through an activation function, which introduces non-linearity to allow the network to train itself via back propagation. The most popular types of activation functions are sigmoid, tanh and ReLU.

- Sigmoid

Sigmoid is an activation function of form $f(x) = \frac{1}{1+e^{-x}}$. Its range is between o and 1. It is a S-shaped curve. It is easy to understand and prevent 'jumps' in output. Fig 8 shows the graph of sigmoid activation function.

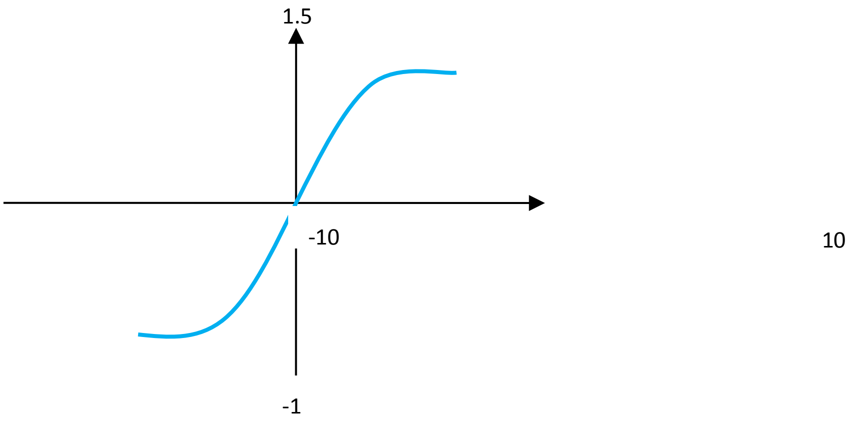


Fig. 7 Sigmoid Activation Function

- tanh

tanh is another activation function of form $f(x)=\frac{2}{1+e^{-2x}}-1$. It is a scaled sigmoid function. This has characteristics similar to sigmoid. Its rage is between -1 to 1. Fig 9 shows the graph of tanh activation function.

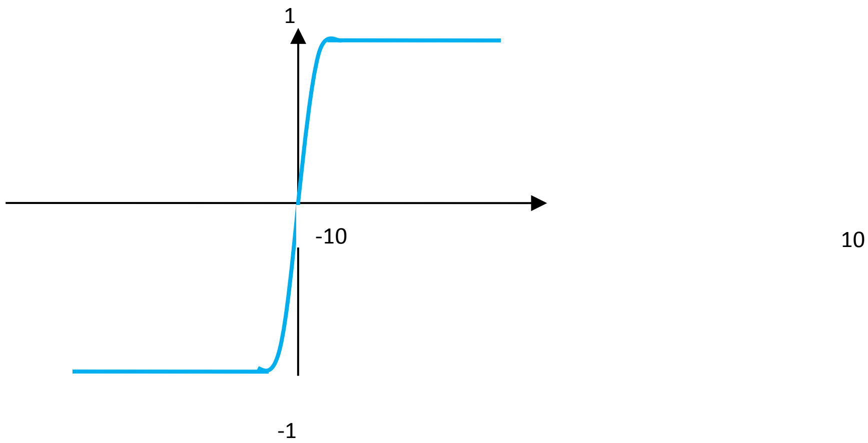


Fig. 8Tanh Activation Function

- ReLU

ReLU activation function given an output x if x is positive and 0 otherwise. ReLU function is the most widely used activation function in neural networks. One of the greatest advantages ReLU has over other activation functions is that it does not activate all neurons at the same time. In practice, ReLU converges six times faster than tanh and sigmoid activation functions. Fig.10 illustrates the graph of Rectified Linear Unit (ReLU).

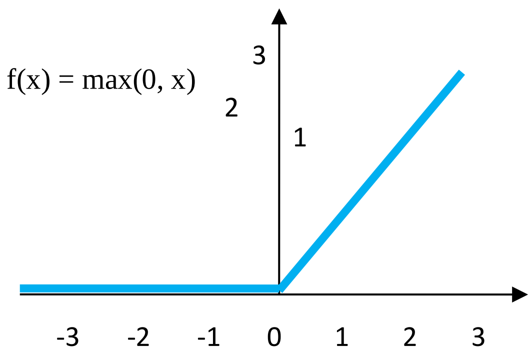


Fig. 9 Rectified Linear Unit (ReLU)

Pooling layer

The pooling layer can generalize the convolved features through down-sampling and thereby reduce the computational complexity during the training process. A filter is passed over the results of the previous layer and selects one number out of each group of values (typically the maximum, this is called max pooling). This allows the network to train much faster, focusing on the most important information in each feature of the image. Given a pooling/subsampling layer q, the feature output F^q can be derived from the preceding layer $f^{(q-1)}$ through

Fi,qj = max($f1+q-p1(i-1),1+p(j-1)$,..., $f\ piq-,l1+p(i-1)$,..., $f1+q-p1(i-1),\ pj$,..., $f\ piq-,pj1$ (3)

where p x p is the size of the local spatial region, and 1□ i, j □ $(m - n + 1)/p$, here m refers to the size of input feature map, while n corresponds to the size of the filter. Then it simply summarizes the input features within local spatial region using the maximum value. Max pooling is a pooling operation that calculates the maximum, or largest, value in each patch of each feature map.

- *Min pooling* is a pooling operation that calculates the minimum or smallest value in each patch of each feature map.
- *Average pooling* involves calculating the average for each patch of the feature map. This means that each 2×2 square of the feature map is down sampled to the average value in the square. Once the higher level features are extracted, the output feature maps are flattened into a one-dimensional vector, followed by a fully connected output layer.

Fully Connected layer

The objective of the fully connected layer is to take the results of the convolution/pooling process and use them to classify the images into a label. It is a traditional multilayer perceptron structure. Its input is a one-dimensional vector representing the output of the previous layers. The 1D CNN cannot classify the image directly. It is mainly used for sequence classification. In this work, the 1D CNN used as classifier the features are extracted from the pre-trained model Inceptionv3 and Resnet50. The extracted features are transformed as three-dimensional input (*samples, time steps, features*) and then fed into the 1D CNN model. The 1D CNN model classifies the Pneumonia diseases normal, Viral Pneumonia and Bacterial Pneumonia.

Performance Measures

The performance of a classification model on a set of test data for which the true values are known. The basis of confusion matrix shown in Table 1.

Table 1
Basis of Confusion Matrix

Actual class		Predicted class	
		P	N
	P	True positive (TP)	False Negative (FN)
	N	False positive (FP)	True Negative (TN)

The performance measures used for evaluation are Accuracy (A), Precision (P), Recall (R) and F-measure (F) which are calculated as given below.

$$Accuracy(A) = \frac{No.of TP + No.of TN}{No.of TP + No.of TN + No.of FP + No.of FN}$$

$$Precision\ (P) = \frac{No.of TP}{No.of TP + No.of FP}$$

$$Recall(R) = \frac{No.of TP}{No.of TP + No.of FN}$$

$$F - Measure\ (F) = \frac{2 * P * R}{P + R}$$

Experimental Results

Datasets

The datasets were collected from kaggle Datasets. A total 4592 chest x-ray images were collected from different patients. A total 4592 x-ray images were collected from different patients. Here 4192 images were utilized for training, 400 images were used for testing. In this 1082 represents Normal images, 1957 represents Bacterial Pneumonia and 1153 represents Viral Pneumonia Pre-Trained Convolutional Neural Network used for Feature Extraction.

Pre-Trained Inceptionv3 as a Feature Extractor

The input layer takes an image in the size of 299 x 299 x 3 and the output layer is the softmax prediction on 1000 classes. From the input layer to the last is the max pooling layer by 8x8x2048 which is referred as the feature extraction of the model, while the rest of the network is regarded as the classification of the model. In this work, inception-v3 is used to extract the features hence we arrive at 2048 feature vector for an individual image.

Classification using 1D CNN

1D CNN analyses the chest x-ray images as normal and Pneumonia affected images and classifies into Bacterial Pneumonia and viral pneumonia. Here 1D CNN is used as a classifier.

The extracted features are reshaped into three-dimensional input format and fed as input to 1D CNN.

For training 4192 feature vectors, each of 2048-dimension are extracted from the images. The training process analyses the X-Ray images to categorize into normal, Bacterial Pneumonia and Viral Pneumonia affected images. For testing 400 feature vectors each of 2048 dimensions are given as input to the 1D CNN model. The convolution and max pooling layers are processing with features and the softmax classifier used to predicting the output by using dense layer. The performance of Inceptionv3 with 1D CNN shown in the Table 2

Table 2
Performance of 1D CNN for Chest X- images using Inceptionv3

	Precision (in %)	Recall (in %)	F-Score (in %)	Accuracy (in %)
Normal	88.00	95.61	91.60	94.62
Bacterial Pneumonia	92.1 5	88.42	90.00	93. 31
Viral Pneumonia	96.1 0	92.31	96.12	96. 04

Pre-Trained Resnet50 as a Feature Extractor

The input layer adopts an image in the size of 224 x 224 x 3 and the output layer is the softmax prediction on 1000 classes. From the input layer to the end is the max pooling layer by 7x7x2048 which is regarded as the feature extraction of the model, while the remnant of the network is presented as the classification of the model. In this work, inception-v3 is utilized to extract the features and we conclude at 2048 feature vector for a single image.

Classification using 1D CNN

1D CNN analyses the chest x-ray images as normal and Pneumonia affected images and classifies into Bacterial Pneumonia and viral pneumonia. Here 1D CNN is used as a classifier. The extracted features are reshaped into three-dimensional input format and fed as input to 1D CNN. For training 4192 feature vectors, each of 2048-dimension are extracted from the images. The training process analyses the X-Ray images to categorize into normal, Bacterial Pneumonia and Viral Pneumonia affected images. For testing 400 feature vectors each of 2048 dimensions are given as input to the 1D CNN model. The convolution and max pooling layers are processing with features and the softmax classifier used to predicting the output by using dense layer. The performance of Inceptionv3 with 1D CNN shown in the Table 3.

Table 3
Performance of 1D CNN for Chest X- images using Resnet50

	Precision (in %)	Recall (in %)	F-Score (in %)	Accuracy (in %)
Normal	92.00	82.00	86.71	90.11
Bacterial Pneumonia	60.10	83.00	69.61	82.20
Viral Pneumonia	92.41	79.00	83.51	89.14

Conclusion

This section made a comparative study of two different feature extractions namely, Inceptionv3 as well as Resnet50 is fed on 1D CNN classifier. From the above experiments, it is analyzed that Inceptionv3 with 1D CNN gives the best results of the highest accuracy of 96.04% when compared with other techniques.

Table 4
Overall Performance of Pre-Trained Model Features with 1D CNN

Pre-Trained Models	Classifier	Accuracy (in %)
Inceptionv3	1D CNN	96.04
Resnet50		90.11

References

1. Vikash Chouhan, Sanjay Kumar Singh, Aditya Khamparia, Deepak Gupta, Prayag Tiwari,
   Catarina Moreira, RobertasDamaševičius, Victor Hugo C. de Albuquerque, "A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images", Applied Sciences, Vol 10, Issue 2, 2020.
2. 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", Clinical. Infect. Dis, Vol54, Issue 2, pp S102–S108, 2012.
3. Scott, J.A.; Brooks, W.A.; Peiris, J.S.; Holtzman, D.; Mulholland, E.K," Pneumonia research to reduce childhood mortality in the developing world", Journal of Clinical. Investiation, Vol 118, pp 1291–1300, 2008.
4. P. Rajpurkar et al. "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017.
5. D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, F.; et al., "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," Cell, 172, pp. 1122–1131, 2018.
6. Mrs.G.Revathy and Dr.K.Selvakumar, "Channel assignment using tabu search in wireless mesh networks",Wireless personal communication ISSN NO 09296212.

7. for Pneumonia Detection using Chest X-ray” Appl. Science, Vol 10, Issue 3233, 2020.
8. Nada M. Elshennawy, and Dina M. Ibrahim, “Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images”, Diagnostics, Vol 10, Issue 649, 2020.
9. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., &Wojna, Z,“Rethinking the inception architecture for computer vision”, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826), 2016.
10. Jie, Z., Xia, H., Zhong, S. L., Feng, Q., Li, S., Liang, S., & Kristiansen, K,” The gut microbiome in atherosclerotic cardiovascular disease”, Nature communications, Vol8, Issue 1, pp 1-12, 2017.
11. He, K., Zhang, X., Ren, S., & Sun, J, “Identity mappings in deep residual networks. In European conference on computer vision”, pp. 630-645, Springer, 2016