DETECTION AND CLASSIFICATION OF PNEUNOMIA USING TRANSFER LEARNING

G. Natarajan[1] and Dr. P. Dhanalakshmi[2]
[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Chidambaram
[2]Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram
[1]natarajang@gmail.com and [2]abidhana01@gmail.com

INTRODUCTION

detection and classification of Pneumonia using transfer learning is discussed in detail. Section 1.2 explains feature extraction techniques. Section 1.2.1 and Section 1.2.2 explains the Incetion-v3, VGG19 and Mobile net v2 techniques. Section 1.3 explains the modelling of features namely MSVM and 1DCNN. Finally, the experimental outcomes are considered in Section 1.5.

The proposed technique detects and categories the pneumonia images into normal, bacterial pneumonia and viral pneumonia. Inception-v3, VGG19 and Mobile net v2 features are extracted. The performance of the system is studied for all the categories of pneumonia. Most of the samples are correctly detected and it is observed that VGG19 with MSVM model gives a better performance when compared to other techniques.

1.1 INTRODUCTION

Pneumonia is an infection that causes inflammation of the lungs and can be fatal if not detected early. The most common method of detecting pneumonia is a chest x-ray, which requires a careful examination of the chest x-rays by an expert. Pneumonia causes inflammation of the lungs, especially the alveoli, which fill with fluid or pus and can cause coughing and shortness of breath **(Harsh Sharma *et al.,* 2020).** Pneumonia is a life-threatening condition caused by a bacterial or viral infection in the lungs. It can be life-threatening if not acted upon at the right time, so early diagnosis of pneumonia is vital. (Tawsifur Rahman *et al.,* 2020).

Fig .1.1 (a) Normal **(b)** Bacterial Pneumonia **(c)** Viral Pneumonia

There are no explicit instructions for differentiating between bacterial and viral pneumonia in the area of computer-aided detection and diagnosis with imaging methods. In addition, the characteristic of the two types of pneumonia on chest X-ray are similar and confusing. For reasons of diagnostic efficiency, it is very desirable to develop a CAD system that is capable of autonomously diagnosing bacterial and viral pneumonia on chest x-rays (Rachna Jain *et al.,* 2020).

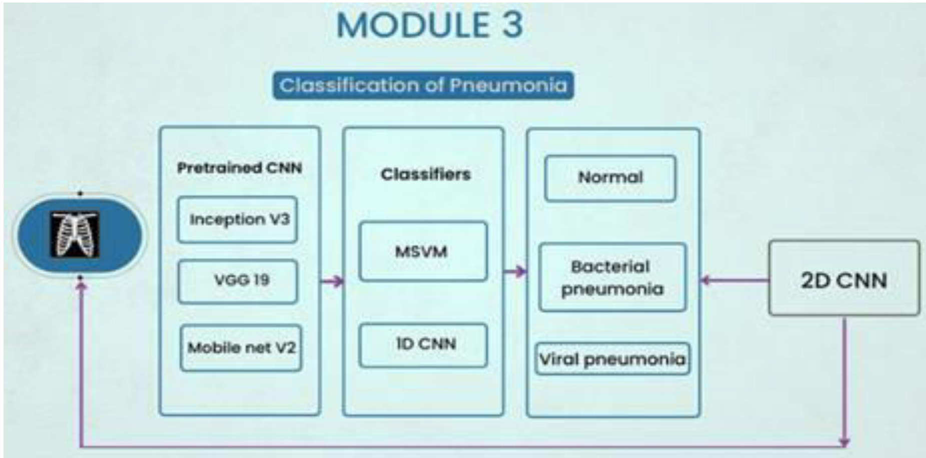

Fig 1.2. Framework of the Proposed work

1.2 Pre-Trained Convolution Neural Networks

The feature extraction performance in chest x-ray images of this chapter 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 VGG19. Inception-v3 and VGG19 use the weights of its network to perform the task of classification of chest x-ray image dataset which is classified into two types namely, Bacterial Pneumonia and Viral Pneumonia.

Transfer Learning

Transfer learning is a technique used to deal with small or large amounts of labelled 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.

1.2.1 Inception-v3 model

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 Google Net in 2015 and it was named as Inception-v1 **(Szeged *et al*. 2016)**. Next, the inception was refined by batch normalization and then Inception-v2 came into existence. Now, in Inception-v3, 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 proceeded 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 VGG Net and it performs concatenation of many various sized convolutional filters into a new filter **(Xia *et al*. 2017)**. This model reduces the number of parameters which under goes the training and thereby minimizes the computation complexity. Fi-g. 1.3 illustrates the overall architecture of Inception-v3 model. Auxiliary Classifier, Grid Size Reduction, Factorization into Smaller Convolutions, Factorization into Asymmetric Convolution, and Regularization via Label Smoothing make up the architecture.

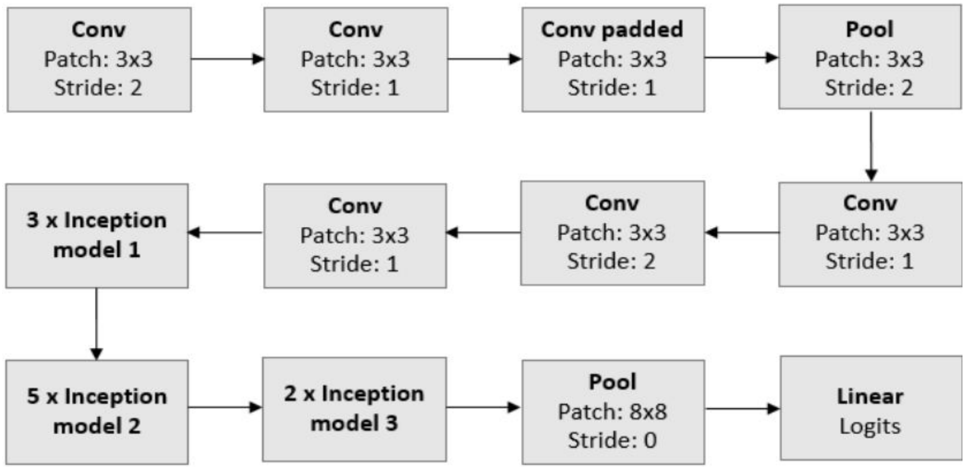

Fig 1.3 Basic architecture of Inception-v3

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. By addressing the disappearing gradient problem in very deep networks, the major goal is to push the crucial gradients to the lower layers during training.

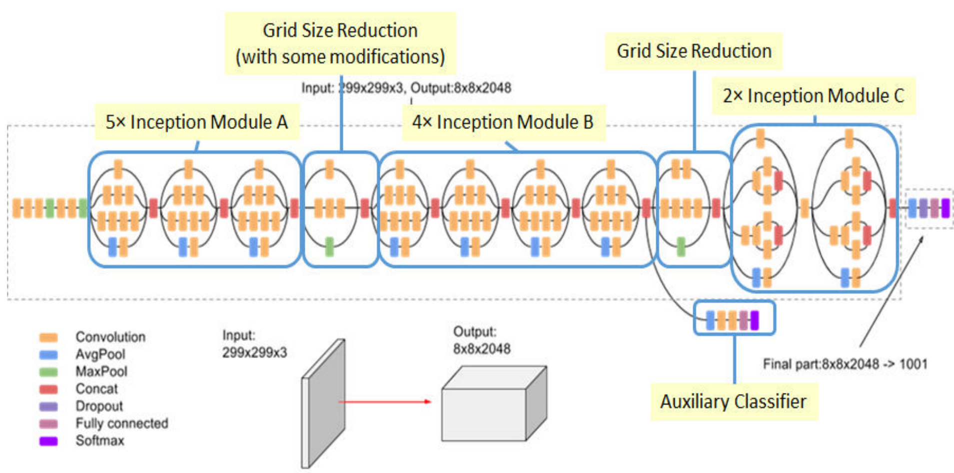

Fig. 1.4 Architecture of Inception-v3 model

We extract a 2048-dimensional feature for one image from the final fully-connected logits layer. The layers of the pre-trained networks before to the fully connected layer 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 MSVM 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. 1.4

1.2.2 VGG19

VGG is a deep Convolutional Neural Network that is used to classify images. The input size of VGG19 is 224 x 224 image which is given as the input to the network. It means the matrix of the shape (224,224, 3). The kernel size is 3 * 3 with the stride size of 1 pixel is covered throughout the image. The spatial padding is used to preserve the spatial resolution of the image **(Dey *et al.,* 2021)**. The max pooling is performed over a 2 * 2 pixel windows with stride 2 followed by Rectified Linear Unit

(ReLu) is introduced for non-linearity to make the model class classify better and to improve computational time Fig .1.5 shows the overall block diagram of VGG19.

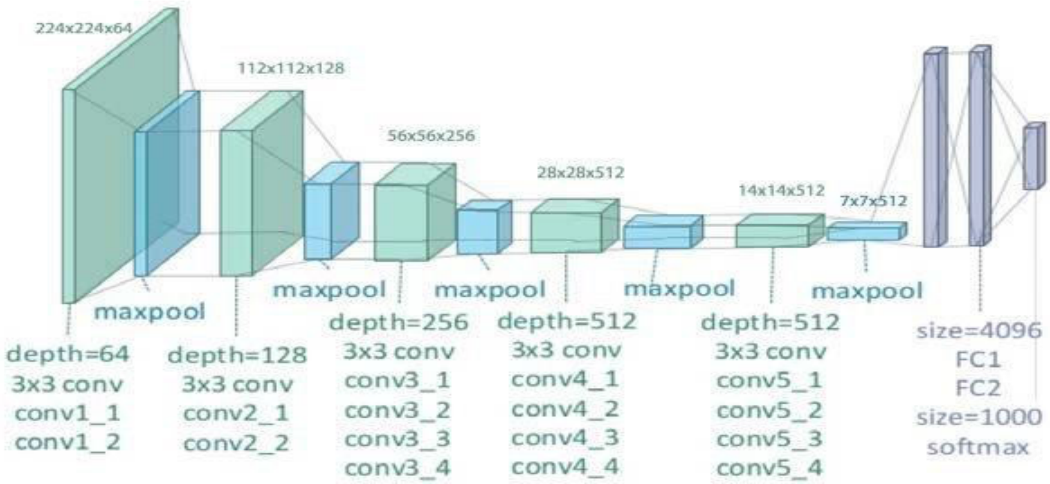

Fig. 1.5 Architecture of VGG19

The three fully connected layers from which first two were of size 4096 and after that a layer with 1000 channels for 1000-way *ILSVRC* classification and the final layer is a softmax function. This model is pre-trained for different ImageNet database classified for 1000 classes, in the proposed work the final network layer is removed and replaced with the softmax classier **(Labhane *et al.,* 2020).**

1.2.3 Mobile net v2

It is originally designed for mobile device by Google. Mobilenetv2 is a fine pre-Trained model that delivers the high output accuracy. Mobile net v2 is built upon the idea of Mobilenetv1. It uses the depth wise separable convolution as efficient building blocks **(Qian Xiang, Xiaodan Wang *et al.,* 2019)**. Mobilenetv2 has two important features namely linear bottleneck between the layers and shortcut connections (residual block in the network) between the bottleneck is shown in the figure 1.6.

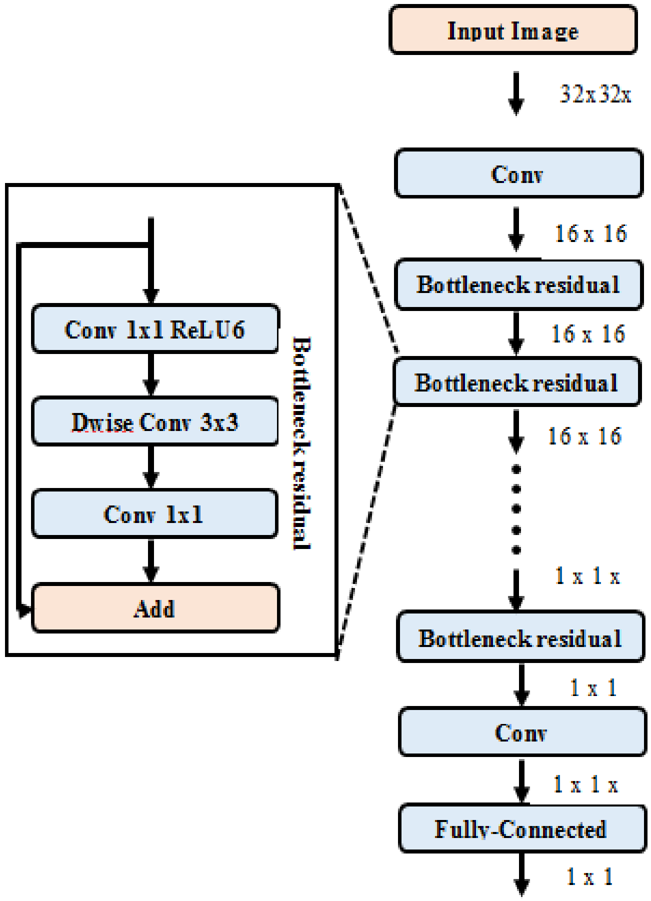

Fig. 1.6 Architecture of Mobile net V2

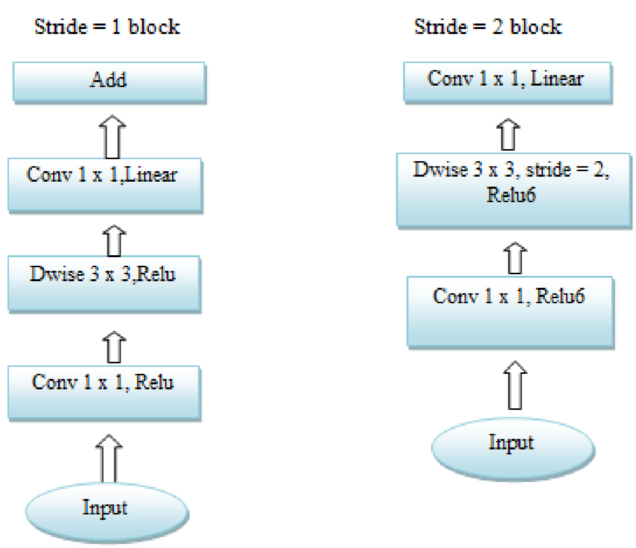

Fig. 1.7 Mobilenetv2 convolutional block

There are two types of blocks in Mobilenetv2. They are residual block with stride of 1 and other block is stride of 2 with downsizing. Fig. 1.7 shows the Mobilenetv2 convolutional blocks. The Table 1.1 shows the bottleneck of Mobilenetv2.

Both type of the block has three layers:

- First layer has 1 x 1 convolution with ReLU as the activation function.
- Second layer has the depth wise convolution.
- The third layer has another 1 x 1 convolution, but without any non-linearity.

Table 1.1 Bottleneck of Mobile net V2

Input	Operator	Output
h X w X k	1 x 1 Conv2d, ReLU6	*h X w X (tk)*
h X w X(tk)	3 x 3 dwise s=s ReLU6	$\frac{h}{s} X \frac{w}{s} X(tk)$
$\frac{h}{s} X \frac{w}{s} X(tk)$	Linear 1 x 1 Conv2d	$\frac{h}{s} X \frac{w}{s} k'$

Bottleneck residual block transforming from *k* to k' channels, with stride s, and expansion factor t, s is the stride, for spatial convolution 3 x 3 kernels are used.

1.3 MODELING THE FEATURES FOR PNEUNOMIA

1.3.1 Multi Support Vector Machine (MSVM)

SVM is a useful statistical machine learning technique that has been successfully applied in the pattern recognition field. If the data cannot be separated linearly but cannot be separated linearly, the support vector classifier is nonlinear will be applied Chang Y.W *et al.,* **(2010)**. The fundamental notion is to transform input vectors into a high-dimensional feature space using a nonlinear transformation **(∅),** and then to do a linear division in feature space as shown in Fig. 1.8.

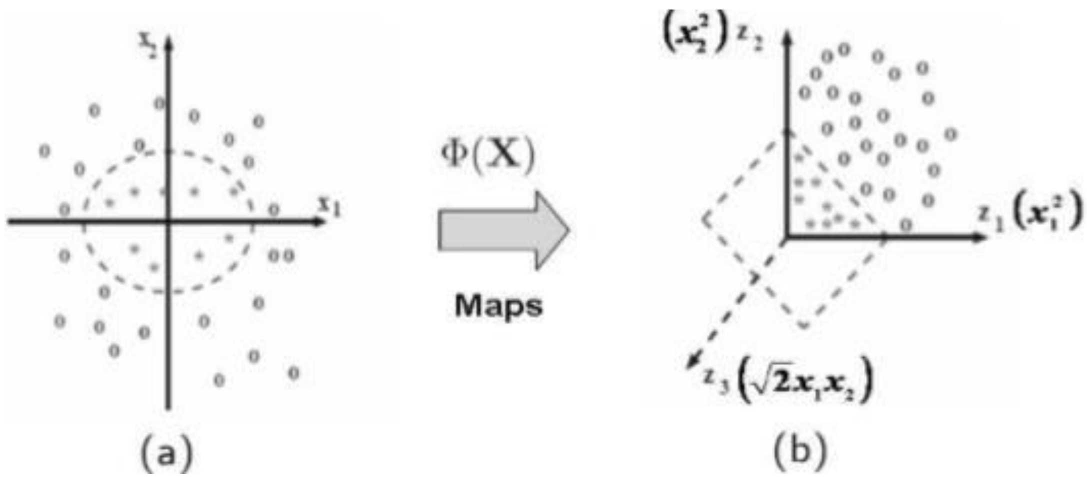

Fig. 1.8 An example for MSVM kernel function∅ (x) maps two-dimensional input space to higher three-dimensional feature space. (a) Nonlinear problem (b) Linear problem

To construct a nonlinear support vector classifier, the inner product (x, y) is replaced by a kernel function K (x, y)

$$f(x) = sgn\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b\right) \tag{1.1}$$

The MSVM has two layers. During the learning process, the initial layer chooses the basis K (x_i, x), i=1, 2, …., N, from the given set of bases defined by the kernel; the second layer constructs a linear function in this space. This is completely equivalent to constructing the optimal hyper plane in the corresponding feature space.

The SVM algorithm can constructs a variety of learning machines by use of different kernel functions. Three kinds of kernel functions are usually used.

They are as follows.

1. Polynomial Kernel of degree d

$$K(X,Y) = ((X,Y)+1)^d \tag{1.2}$$

2. Radial basis function with Gaussian kernel of width C >0

$$K(X,Y) = exp\left(\frac{-|X-Y|^2}{c}\right) \tag{1.3}$$

3. Neural Networks with tan *h* activation function

$$K(X,Y) = tanh(K(X,Y)+\mu) \tag{1.4}$$

where the parameters k and μ are the gain and shift.

Multi-SVM

The aim of multi MSVM is to allocate labels to occurrence by using MSVM in which the labels are taken from a finite set of numerous elements **(Hsu and Lin 2002)**.Each training point belongs to one of N different classes. As the image classification involves multiclass classification and MSVM is generally two-class based pattern classification model, diverse binary MSVMs have to be carried out to produce a MSVM. The MSVM comprises of integrating $(k - 1)$ binary classifiers to a multiclass classifier.

1.3.2 1D CNN

In this work, 1D CNN is developed to model to classify the Parkinson disease. CNN models specifically designed for image classification problems, where the model learning internally represents the two-dimensional input, in a process called feature learning **(Fathurahman *et al.*, 2020)**. The architecture of 1D CNN shown in the Fig. 1.9.

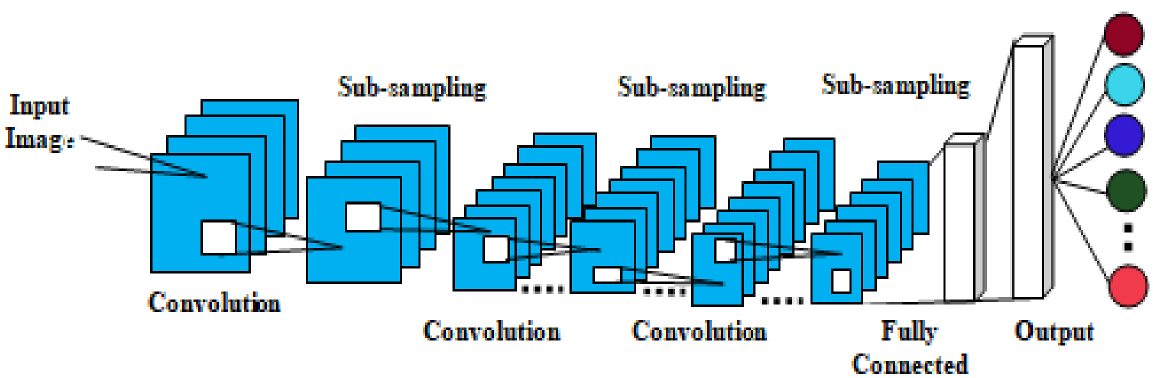

Fig. 1.9 Architecture of 1D CNN

A Convolutional Neural Network has three different sorts of layers:

1. Convolutional Layers.
2. Pooling Layers.
3. Fully-Connected Layers.

Convolutional Layers

Convolution layer consist of Filters and Feature map.

Filters

The "neurons" of the layer are the filters. They take weights as input and provide a value. A patch or receptive field is a solid square that serves as the input size. The input patch is made up of pixel values when the convolutional layer serves as an input layer (Rekeczkyet al., 1998). The convolutional layer takes data from a feature map of the preceding layer if the network architecture is deeper.

Feature Map

The feature map is the output of a filter that is applied to the previous layer. A certain filter is drawn across on the entire layer above, it is moved pixel by pixel. The neuron is activated at each place, and the output is gathered in the feature map.

gathering layer

The preceding layers' feature map is down-sampled by the pooling layers. Following one or more convolutional layers, pooling layers are intended to combine the features that were learned and expressed in the feature map from the preceding layers (Wu and Gu, 2015). As a result, pooling may be taken into consideration as a method to generalise or compress feature representations and overall lessen the model's overfitting of the training data.

Fully connected layer

Fully connected layers are the ordinary flat feed-forward neural network layer (**Basha *et al.,* 2020**). These layers can also additionaly have a non-linear activation function or a softmax activation to generate output probabilities of class predictions.

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, mobilenetv2 and vgg16. 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 into Bacterial Pneumonia and Viral Pneumonia.

1.4 Performance Measures

To evaluate the performance of MSVM and 1D CNN classifiers utilizing Inception-v3 and VGG19 feature extraction techniques, a set of evaluation parameters namely Accuracy, Precision, recall and F-score are utilized. The definition and the formulas used for these measures are given in Section 1.3.6.

1.5 Experimental Results

In this chapter, the architecture of the Inception-v3 and VGG19 is provided in the keras package. By analyzing the effectiveness of the network and by utilizing the Pre-Trained network as a feature Extractor, the weights are transferred to the proposed system. The optimal setup values vary while fine tuning the deep networks. In the Inception-v3 and VGG19 the fully connected layer followed by an output layer was selected and was replaced by the MSVM and 1D CNN classifier. Implementing the chest x-ray images.

1.5.1 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

1.5.2 Feature Extraction using Inception -v3

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.

1.5.2.1Evaluation of Inception-v3 using MSVM

The training process analyses pneumonia training data to find an optimal way to classify pneumonia affected images in their distinctive classes. A nonlinear support vector classier is used to discriminate the various stages. The N class classification problem can be solved using N SVMs. Every individual SVM differentiates a particular class from the remaining classes (one-vs-rest approach). Support vector machine is trained to identify pneumonia features into its types. Three SVMs are created for each type. For training 4192 feature vectors, each of 2048-dimension are extracted from the images. The training process analyses the pneumonia training data to find an optimal way to classify OSCC affected images into its respective types namely, normal(Category 0), bacterial pneumonia(Category 1) and viral pneumonia (Category 2). The derived support vector is utilized to categorize the images. For testing the 400 feature vectors, each of 2048 dimensions are given as input to the MSVM model and the distance between each of the feature vector and the MSVM hyper plane is calculated. The average distance is calculated for each model. The average distance gives a better performance than using distance for each feature vector. The types of pneumonia are decided based on the maximum distance. The performance of pneumonia classification for Polynomial, Gaussian and Sigmoidal kernels is studied. From the analysis, Gaussian kernel gives the best performance using MSVM classification.

The classification results are analyzed based on four measures namely, Precision, Recall, F-score and Accuracy as presented in Section 3.6. Table 1.2 shows the classification performance of pneumonia images using Inception v3.

Table. 1.2 Performance of MSVM Kernel for Chest X- images using Mobilenetv2

Pneumonia Types	Gaussian Kernel (in %)	Polynomial Kernel (in %)	Sigmoid Kernel (in %)
Normal	80.00	79.12	77.54
Bacterial Pneumonia	80.71	78.21	75.42
Viral Pneumonia	88.00	86.11	83.40

From the above table, it is clear that Gaussian kernel achieves better when compared to polynomial and Sigmoidal kernel.

1.5.2.2Evaluation of Inception-v3 using 1DCNN

The training process analyses the chest x-ray images of Pneumonia into its types as normal, Bacterial Pneumonia and Viral Pneumonia training data to find an optimal way to classify into their respective classes. Here the 1D CNN is used as a classifier. The extracted features are reshaped into three-dimensional input which is suitable input format used in 1D Convolutional Neural Network.

For training 4192 feature vectors, each of 2048-dimension are extracted from the images. The training process analyses the MRI images to categorize into normal/abnormal image. 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

Table. 1.3 Performance of 1D CNN for Chest X- images using Inception-v3

	Precision (in %)	Recall (in %)	F-Score (in %)	Accuracy (in %)
Normal	92.00	82.00	86.71	90.00
Bacterial Pneumonia	60.10	83.00	69.61	82.00
Viral Pneumonia	92.00	79.00	83.51	89.00

1.5.3 Feature Extraction using VGG 19

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 1 x 1 x 512 which is referred as the feature extraction of the model, while the remaining of the network is regarded as the classification of the model. In this work, VGG19 is used to extract the features hence we arrive at 512 feature vectors for an individual image.

1.5.3.1 Evaluation of VGG 19 using MSVM

The training process analyses pneumonia training data to find an optimal way to classify pneumonia affected images into their distinctive classes. A nonlinear support vector classier is utilized to differentiate the various stages. The N class classification problem can be solved using N MSVMs. Each MSVM separates a single class from all the classes (one-vs-rest approach). Support vector machine is trained to discriminate pneumonia features of a stage from all the rest. Three MSVMs are created for each stage. For training 4192 feature vectors each of 512-dimension are extracted from the images. The training process analyses the pneumonia training data to find an optimal way to categorize pneumonia affected images into its distinctive types, namely, normal (Category 0), bacterial pneumonia (Category 1) and viral pneumonia (category 2). The derived support vector used to classify the images. For testing 400 features, vectors each of 512 dimensions are given as input to the MSVM model and the distance between each of the feature vectors and the MSVM hyperplane is procured. The average distance is calculated for each model. The average distance gives better a performance than using distance for each feature vector. The stages of pneumonia are decided based on the maximum distance. The performance of pneumonia classification for Polynomial, Gaussian and Sigmoidal kernels is studied. From the analysis Gaussian kernel gives the best performance using MSVM classification. The pneumonia belonging to three categories namely normal, bacterial pneumonia and viral pneumonia.

The classification results are analyzed based on four measures namely Precision, Recall, F-score and Accuracy as shown in Section 3.6. Table 1.3. Show the classification performance of chest x-ray images using Resnet50.

Table 1.3 Performance of MSVM Kernel for Chest x-ray images using VGG19

Pneumonia Types	**Gaussian Kernel (in %)**	**Polynomial Kernel (in %)**	**Sigmoidal Kernel (in %)**
Normal	80.00	79.12	77.54
Bacterial Pneumonia	80.71	78.21	75.42
Viral Pneumonia	88.00	86.11	83.40

1.5.3.2 Evaluation of VGG 19 using 1DCNN

The training process analyses the chest x-ray images of Pneumonia into its types as normal, Bacterial Pneumonia and Viral Pneumonia training data to find an optimal way to classify into their respective classes. Here the 1D CNN is used as a classifier. The extracted features are reshaped into three-dimensional input which is suitable input format used in 1D Convolutional Neural Network.

For training 4192 feature vectors, each of 512-dimension are extracted from the images. The training process analyses the MRI images to categorize into normal/abnormal image. For testing 400 feature vectors each of 512 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.

Table 1.4 Performance of 1D CNN for Chest X-Ray Images using VGG19

	Precision (in %)	Recall (in %)	F-Score (in %)	Accuracy (in %)
Normal	88.00	95.61	91.60	94.62
Bacterial Pneumonia	92.15	88.42	90.00	93. 31
Viral Pneumonia	96.10	92.31	96.12	96. 00

1.5.4. Feature Extraction using Mobile net v2

The input layer accepts 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 7x7x 1280 which is regarded as the feature extraction of the model, while the rest of the network is presented as the classification of the model. In this work, Mobilenetv2 is applied to extract the features and we arrive at 1280 feature vector for a single image.

1.5.4.1. Evaluation of Mobile net V2 using MSVM

The training process analyses the chest x-ray images of Pneumonia into its types as normal, Bacterial Pneumonia and Viral Pneumonia training data to find an optimal way to classify into their respective classes. Here the MSVM is used as a classifier. The extracted features are fed in to the MSVM classifier.

The performance of pneumonia classification for Polynomial, Gaussian and Sigmoidal kernels is studied. From the analysis Gaussian kernel gives the best performance using MSVM classification. The pneumonia belonging to three categories namely normal, bacterial pneumonia and viral pneumonia.

The classification results are analyzed based on four measures namely Precision, Recall, F-score and Accuracy as shown in Section 3.6. Table 1.5. Shows the classification performance of MSVM kernel for chest x-ray images using Mobile net v2.

Table 1.5 Performance of MSVM Kernel for Chest x-ray images using Mobile net V2

Pneumonia Types	Gaussian Kernel (in %)	Polynomial Kernel (in %)	Sigmoidal Kernel (in %)
Normal	80.00	79.12	77.54
Bacterial Pneumonia	80.71	78.21	75.42
Viral Pneumonia	88.00	86.11	83.40

1.5.4.2. Evaluation of Mobile net V2 using 1DCNN

The training process analyses the chest x-ray images of Pneumonia into its types as normal, Bacterial Pneumonia and Viral Pneumonia training data to find an optimal way to classify into their respective classes. Here the 1D CNN is used as a classifier. The extracted features are reshaped into three-dimensional input which is suitable input format used in 1D Convolutional Neural Network.

The classification results are analyzed based on four measures namely Precision, Recall, F-score and Accuracy as shown in Section 3.6. Table 1.4. Shows the classification performance 1DCNN of chest x-ray images using Mobile net v2.

Table 1.4 Performance of 1D CNN for Chest X-Ray Images using Mobile net V2

Pneumonia Types	Precision (in %)	Recall (in %)	F-Score (in %)	Accuracy (in %)
Normal	80.00	92.61	91.60	87.62
Bacterial Pneumonia	89.15	87.42	89.00	87. 31
Viral Pneumonia	90.10	92.31	88.12	86. 00

1.5.3 Comparative Results analysis

This section made a comparative study of three different feature extractions namely, Inception-v3, VGG19 and Mobile net v2 on the MSVM and 1D CNN. The results are provided in Table 1.5 and also in Fig.1.10. From the Fig. 1.10, the results show that VGG19 with MSVM achieved comparably better results when compared with other techniques. It is noted that the highest accuracy of 98.90 % is obtained with MSVM using VGG19 features.

Table 1.5. Overall Performance of Pre-Trained Models with MSVM and 1D CNN

Pre-Trained Models	Classifiers	Accuracy (in %)
Inception v3	MSVM	82.5
Inception v3	1DCNN	81.6
VGG19	**MSVM**	**98.90**
VGG19	1D CNN	90.98
Mobilenetv2	MSVM	82.90
Mobilenetv2	1D CNN	81.37

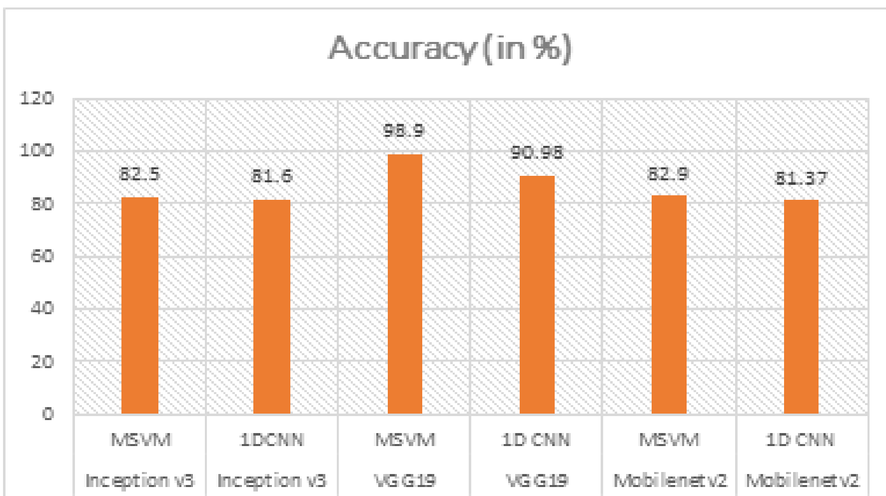

Fig 1.10 Overall Comparison

SUMMARY

In this chapter the proposed technique detects and categories the pneumonia images into normal, bacterial pneumonia and viral pneumonia. Inception-v3, VGG19 and Mobile net v2 features are extracted. The performance of the system is studied for all the categories of pneumonia. Most of the samples are correctly detected and it is observed that VGG19 with MSVM model gives a better performance when compared to other techniques.