Chapter 7

Deep Learning and Transfer Learning based Approaches for Classification of Medical Images.

Highlights of the Chapter

- *Deep Learning based model for classification of images infected by Covid-19.*
- *Network able to distinguish between the nearly similar features of two infections.*
- *A transfer learning based approach is presented for classification of Covid-19 infected images.*

Contribution of the chapter

The COVID-19 pandemic has posed a severe threat to human life. The early detection and treatment of this infection are necessary to save human life [176] [177] [178]. The chapter aims to propose a time efficient and accurate method to classify lung infected images by COVID-19 and viral pneumonia using chest X-ray. The proposed classifier applies end to end training approach to classify the images of the set of normal, viral pneumonia and COVID-19 infected images. The features of the two infected classes were precisely captured by the extractor path, and transferred to the constructor path for precise classification. The classifier accurately reconstructed the classes using the indices and the feature maps. The classifier detected COVID-19 as well as viral pneumonia with the highest evaluation metrics score. For firm confirmation of the classification results, we used the Matthews correlation coefficient (MCC) along with accuracy and F1 scores (1 and 0.5). The classification accuracy of COVID-19 class achieved was about (97 ± 0.03)% with MCC score (0.9151 ± 0.002). The classifier distinguished with great precision between the two nearly correlated infectious classes

(COVID-19 and viral pneumonia). Both the infected classes were classified with the highest accuracy, F score and MCC. This chapter also presents an efficient classification methodology for precise identification of infection caused by covid-19 using CT and X-ray images as both these imaging systems are cheap in cost and commonly available. The proposed classification method applies a transfer learning-based approach for the classification of covid-19 infected CT and X-Ray images. The depth wise separable convolution based model of MobileNet V2 was exploited for feature extraction. The features of infection were supplied to the SVM classifier for training. The SVM classifier learnt the features and produced accurate classification results. For firm evaluation of results, we used twelve classification evaluation metrics. The accuracies obtained for classifying the CT and X-ray images were 0.994 ± 0.2 and 0.986 ± 0.04, respectively. MCC score was used to avoid any mislead caused by accuracy and F1 score as it is a more mathematically balanced metric. The values obtained for MCC were 0.9852 ± 0.03 and 0.9657 ± 0.04 respectively.

7.1 Introduction:

The whole world is in an unprecedented situation because of the outbreak of (SARS-CoV-2) COVID-19. It may attack the respiratory system and can result in various forms of severity ranging from mild to severe. This severity may cause organ failure and can result in the death of the patient. This unusual respiratory disease was initially dominated by pneumonia which is caused by the novel coronavirus. This new virus was initially termed as 2019-nCoV by WHO. It was renamed as COVID-19 disease by WHO on Feb 11, 2020. The incubation period for this disease is normally 4–5 days and develops symptoms within 11-12 days of the infection. The approaches to identifying this disease are gaining considerable attention nowadays. The two important imaging tools for determination of COVID-19 are chest X-ray and computerized tomography (CT) of the chest [179]. The classification can be done using the two approaches. One is handcrafted features based approach, and second is based on the deep learning methods.

In the former approach, the features are extracted using some predefined methods and then supplies to the classifier for learning. The later approach extracts and learns the features on its own by exploiting the morphological properties of images which contains a large number of similar object patterns. The COVID-19 cases are almost comparable to the different types of viral pneumonia infections. Thus the detection of COVID-19 becomes a tedious task as features of COVID-19 overlaps with the other inflammatory diseases of the lung, including different types of pneumonia. The incorrect detection of COVID-19 may lead to the danger of life. Many biomedical problems such as brain tumour, skin cancer, breast cancer etc. apply artificial intelligence and deep learning techniques to diagnose the issues. The image features which are not clearly visible in the original images can be revealed using the deep learning methods. The deep learning-based CNN architecture has shown remarkable improvement in the classification of various biomedical problems. The ability of CNN to learn features automatically from the images of a particular domain is an additional advantage, unlike the machine learning methods where handcrafted features are required to train the classifiers.

Xianghong et al. proposed a tailored VGG-16 model for lung infection identification [180]. The features of COVID-19 class were extracted using covNet developed by Li et al [181]. The accurate identification of thoracic diseases was made by using DenseNet [182]. The transfer learning based approaches are also showing the increased efficiency in the classification of various types of pneumonia and COVID-19 infections [75]. Transfer learning makes use of the acquired knowledge of one domain to perform specific but the related image processing task in other domain. The transfer learning methodologies transfer the knowledge from a source to target to improve the task-specific efficiency of the target. The transfer learning methods can be categorized into three main types, namely 1) Inductive Transfer learning, 2) Transductive Transfer Learning, 3) Unsupervised Transfer Learning. In Inductive Transfer learning the task of source and target are same while in transductive learning, the task of source and target are

different. The unsupervised transfer learning is same as inductive learning, but the target tasks are unsupervised. The machine learning and deep learning approaches which utilized transfer learning showed remarkable improvement in the classification task. A CT analysis method for detection of COVID-19 infection using artificial intelligence was proposed by Gozes et al [183]. It was able to achieve the sensitivity and specificity of about 98% and 92% respectively. A prediction based model which exploits the inception transfer learning achieves 89.5% accuracy was proposed by Wang et al [184]. Narin et al. [185] used InceptionV3, Inception-ResNetV2 and ResNet50 models for effective prediction of COVID-19 using X-ray images. The extracted deep features were transferred to SVM classifier for detection of COVID-19 using X-ray images by Sethy et al [186]. AlexNet [65] and GoogleNet [72] were used by Sundaram et al. [187] for obtaining the area under the curve (AUC) [188] for detection of pneumonia which achieved AUC value of 0.95. Recently DCNN based structure was presented by Linda et al. for detection of COVID-19 from chest X-ray images [189]. A Reset based transfer learning technique was proposed by Ayrton , which achieved validation accuracy of about 96%.

In the presented work, we classified the chest X-ray images using a modified deep learning architecture for the detection of COVID-19 and viral pneumonia. The values of classification metrics showed good improvement over the other existing methods. The F score, global and classification accuracy of the proposed model were highest among the compared methods. The accuracy and F score can sometimes mislead the classification [190]. Thus we used the Matthews correlation coefficient (MCC) along with these two metrics to confirm the classification effectiveness. MCC produces a high score only if the classifier results in optimal value for all the four cardinalities (TP, TN, FP and FN). It is widely used for both balanced or unbalanced datasets. Moreover, in this work, we got the high score for the FM index also, which is also a good measure for unrelated data. All the evaluation metrics for classification

indicated that the proposed classifier efficiently identified the images infected by COVID-19 and viral pneumonia. Further, it also distinguished both the infected classes more coherent way. The proposed classifier is able to extract feature maps on its own and learns about the indices of these feature maps which is used for the reconstruction of the classes. Thus the proposed classifier efficiently learns the variations of texture for different classes and is clearly able to differentiate the two nearly related classes, i.e. viral pneumonia and COVID-19.

7.2 Methodology

This section describes the detection and classification of Covid-19 infected images from X-ray dataset. We proposed a deep learning-based classifier as shown in figure 7.1 to classify the COVID-19 X-ray images from a set that contain X-ray images of Normal chest, pneumonia infected images and COVID-19 infected images. For classification, a COVID-19 infected images an end to end trainable deep learning classifier is proposed with feature extractor-learner and feature constructer path.

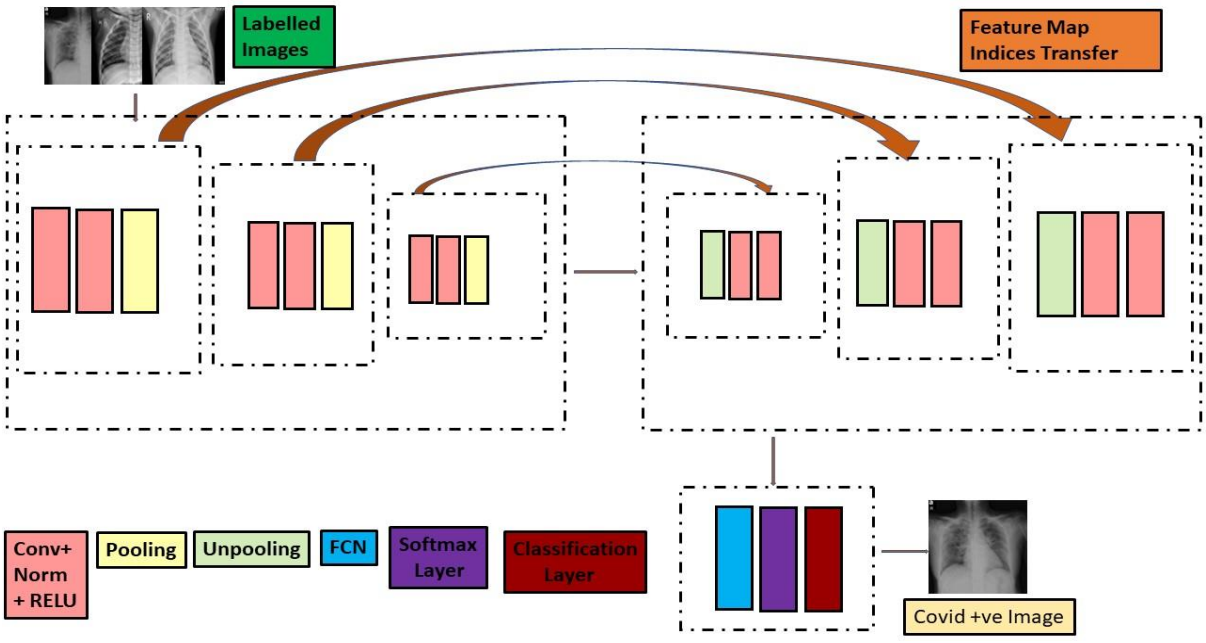

Figure 7.2: Architecture of proposed network.

The proposed classifier extracts the local features of the images by learning the image patterns. The features learnt by the extractor are transferred to the constructer path as feature maps. The indices of the features are stored using the pooling, and these indices of features are transferred to the decomposer path using residual connections. This process enables the classifier to learn the arrangement of features maps in particular locations. The successive convolutional, normalization layers and RELU layer constitutes the main component of the network. The max-pooling layer is used to select the dominant features while suppressing the features that are more correlated [99]. The max-pooling indices are stored and transferred to the constructor path for precise learning of the feature maps. The constructor path constructs the labelled images using the features learnt from the feature maps. The convolution layers produce the feature maps using a series of filters known as Kernels. The filter is a matrix of integers values which are used on the subset of the input pixel values of the same size as the kernel. The output generated by each kernel is a feature map and is given by the following equation:

$$T^m(x,y) = R^n(\mathrm{x},\mathrm{y}) * P^{mn}(\mathrm{x},\mathrm{y}) = \sum_{n=0}^{N-1}\sum_{i=0}^{I-1}\sum_{j=0}^{J-1} R^n(x+i, y+j)\, P^{mn}(i,j) \qquad (7.1)$$

Where the two-dimensional input and output are Rn(x,y) and Tm(x,y), respectively. Pmn(x,y) is the convolution kernel, x and y are the dimensions of the map. The number of output and input channels are represented by m and n. Every pixel is multiplied by corresponding kernel values and the result is summed up for a single value. The result is then normalized using normalization layer. Normalization is a process which changes the numeric values in a dataset to a common scale when the data features have a different range of values. The normalize output is presented to the activation layer, i.e. RELU in the proposed network. The normalized input is multiplied by the activation function in this layer before being forwarded to the subsequent layers. RELU gave output as zero when it received negative input. The RELU activation function is given by:

$$\mu(x) = \max(0, x) \tag{7.2}$$

The pooling operation is used to reduce the count of feature maps. These layers minimise the number of learnable parameters and computation complexity. The pooling layer interacts independently with each feature map. The max-pooling is used in the proposed architecture as it selects the feature with the maximum value in a region of the feature map covered by the filter. The location of maximum feature values is stored and passed on to the later layers for learned label classification. The full-connection layers operate next on the inputs. It takes the output from preceding layers and flattens them to turn in to a single vector which can be applied as input to the next stages. Moreover, it performs the feature analysis function that applies weights to predict the correct label. The SoftMax layer interacts with the inputs before they are presented to the final classification layer. The classification layer, followed by a SoftMax layer [100] computes the cross-entropy loss for multi-class classification problems with mutually exclusive classes. The softmax layer provides values to the trained network, and each input is labelled with one of the mutually exclusive class using the cross-entropy loss function [139]. The cross-entropy loss function is given by:

$$loss = -\frac{1}{N}\sum_{i=1}^{M} T_i \log(X_i) \tag{7.3}$$

Where Xi is the response of network, Ti is the target value, M is the total number of responses in the image and N is the total number of response in X. Figure 7.2 shows the whole process of classification using the proposed model. The labelled data is fed to the proposed classifier, which is separated into three groups, namely Covid-19, normal and viral pneumonia. The dataset is further divided in to test and train data. The classification model is then trained using the training dataset. In the final stage, the classification accuracy is evaluated by using the test dataset. The features learnt from the convolution, normalization and RELU are max pooled and they are transferred to the corresponding un-pooling layer of the constructor path. The max-

pooling achieves translational invariance over small spatial shifts. The consecutive convolution and pooling operations may result in loss of spatial resolution. Thus the feature extractor path stores the indices of the features where the regions with the maximum texture variations i.e. deviation from the normal regions are pooled out by pooling layers. These features along with indices are transferred to the feature constructor path. This path uses the indices and analyses the regions with variations considering the rest of the regions as normal. In this process, the time for the analysis of whole features is reduced and the variations are learnt effectively with the help of specific indices. This specific learning capability of the classification network helps it to distinguish between the nearly common features of covid-19 and viral pneumonia.The proposed method is evaluated using three other pre-existing methods. These methods are SVM-classifier [191] trained using features texture features extracted using GLCM [192] and Local Binary Patterns (LBP) [193], RF-classifier [194] trained using features extracted using GLCM and LBP and Alexnet.

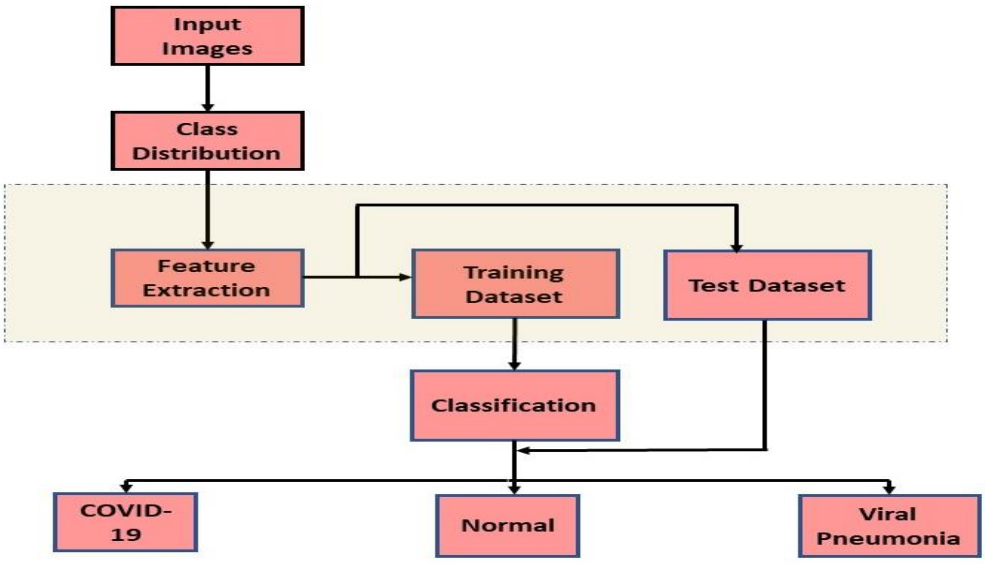

Figure 7.2: Schematic pipeline of proposed network.

The steps to train the SVM and RF classifier are as follows:

Step 1. Label the dataset accurately.

Step 2. Feature set extraction using Local Binary patterns and GLCM [195].

Step 3. Data division into training and test sets.

Step 4. Training of classifier using the texture features obtained from step 2.

Step 5. Finally, evaluation of classification performance parameters.

The SVM classifier finds a hyperplane in an N-dimensional space clearly classifies the provided data. The goal of the SVM is to find the maximum distance between the data for differentiating them into various classes. The SVM is more suitable when there is a distinct margin of separation between the data points to be classified. However, when the data points are more correlated, then the performance of SVM degrades. The RF classifier computes proximities between pairs of cases that can be used in clustering or locating outliers [196]. The prototype which gives information about the relations between classification variables is computed by RF classifier. The local binary patterns are the descriptor of textures. It labels the pixels of an image by thresholding the neighbourhood of each pixel and considers the result as a binary number. The effective differentiating property of LBP makes it a useful operator for the classification task. LBP looks at points surrounding a central point and tests whether the surrounding points are greater than or less than the central point. The LBP is defined by the following equation:

$$LBP\ (x,y) = \sum_{P=0}^{P-1} 2^P\, s(i_p - i_c) \tag{7.4}$$

Where x and y are the coordinates of the central pixel. i_c and i_p are Gray level values of the central pixel and P surrounding pixels in the circular neighbourhood with radius R and the signed function s is defined as:

$$s(x) = \begin{cases} 1 & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases} \tag{7.5}$$

The texture of images is calculated in GLCM using pair of pixels with specific values and spatial relationships. GLCM relate the relative frequencies of the two pixels having intensities (i, j) and separated by a vector distance (d, θ). GLCM is implemented on the basis of joint conditional probability density between the greyscale levels. The functions are as follows:

$$H\left(\frac{i,j}{d,\theta}\right) = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1}(p,q)\,/f(p,q) \qquad (7.6)$$

Where (i, j) is the image pixels separated by distance d and positioned at angle θ which could take values 0^0, 45^0,90^0,135^0. M represents the maximum grey level value. For classification purposes, the eight texture features were calculated: energy, contrast, entropy, mean, homogeneity, correlation, dissimilarity and texture variance.

Image classification task requires the selection of best features is of great importance. This helps in reducing the unnecessary or the most correlated features. In this work, we applied the Pearson correlation coefficient (PCC) approach to select the best parameters . The PCC is given as follows:

$$T = \frac{1}{m-1} * \frac{(A_i-\bar{A})}{V_A} * \frac{(B_i-\bar{B})}{V_B} \qquad (7.7)$$

Where T is PCC values ranged between +1 to -1. The total number of samples are denoted by m. The variables A and B represents the features and classes. V_A and V_B represents the average values while $\bar{A}$ and $\bar{B}$ represents the variance values.

Table 7-12: The PCC values for GLCM

S. No.	GLCM Features	PCC values
1.	Energy	(0.0232, 0.875)
2.	Entropy	(0.086, $2.7097e^{-15}$)
3.	Correlation	(0.5156, $8.0726e^{-22}$)
4.	Mean	(0.4238, $6.0962e^{-08}$)
5.	Dissimilarity	(0.026,$5.428e^{-17}$)
6.	Homogeneity	(0.4359, $6.8124e^{-24}$)
7.	Texture Variance	(0.02,0.8735)
8.	Contrast	(0.4485,$8.7153e^{-14}$)

Table 7-13: The PCC values for LBP

S. No.	LBP Features	PCC values
1.	Feature Vector 1	(0.1453,$1.834e^{-6}$)
2.	Feature Vector 2	(0.1782, $1.653e^{-5}$)
3.	Feature Vector 3	(0.1362, $1.5438e^{-5}$)
4.	Feature Vector 4	(0.1679, $1.9653e^{-7}$)
5.	Feature Vector 5	(0.2876,0.8753)
6.	Feature Vector 6	(-0.15490, 1.8752)
7.	Feature Vector 7	(0.1965, $1.3467e^{-14}$)
8.	Feature Vector 8	(-0.3468,0.6548)
9.	Feature Vector 9	(0.9835,0.7285)
10.	Feature Vector 10	(-0.4983,0.8349)

The useful features were selected using the Table 7-1 and Table 7-2 after the calculation of PCC values. The features with high values are selected, and rest are discarded. From the Table 7-1, it is evident that features such as Correlation, mean, homogeneity and contrast have high positive values. From the Table 7-2, it is clear that feature vector 6,8 and 10 have negative values which indicate the weakness of features. So, based on PCC values feature vectors, 6,8 and 10 are discarded. The Random forest and Support vector classifiers were trained on the image size of 128 × 128. The training of SVM classifier was performed using a custom kernel. The 4-fold cross-validation was used for the training of the classifier. The RF classifier was trained on the default parameters as suggested by the various studies. For training the

classification networks in the study, ADAM optimizer was used with the learning of 0.001. The cross-entropy loss function was used for training on the system having 8 GB GPU and 16 GB RAM.

7.3 Dataset Acquisition and Evaluation Metrics

7.3.1 Dataset Used

The dataset used in this paper is obtained from Covid-19 radiography database of Kaggle [197] which consists of 219 chest X-ray images of COVID-19 positive patients,1341 images of normal chest X-ray and 1345 images of viral pneumonia infection. In addition to this, 301 images were taken from the database of Radiological Society of North America (RSNA) . To increase the data for training, we augmented the data using data augmentation techniques as the deep learning networks need a large amount of dataset to be trained. The data was rotated at the angles of 30^0, 60^0, 90^0,120^0 ,270^0 and 300^0.

7.3.2 Evaluation of dataset:

The evaluation of classification parameters is done by using TP, TN, FP and FN values obtained from the confusion matrix of each class. True positive (TP) values are correctly predicted positive values while true negative (TN) is the correctly predicted negative values. False Positive (FP) are the values predicted by the model as true values, but practically they are false values. False Negative (FN) are the values which are predicted as false values by the classifier, but actually, these values are true values of the class under prediction. The classification achieved by various methods was evaluated using the following parameters:

i) **Accuracy:** Global accuracy is termed as the ratio of correctly classified pixels, irrespective of class, to the total number of pixels in the dataset. It evaluates the correctly identified pixel percentage for every class. Accuracy is given by the following equation:

$$Accuracy = \frac{TP+TN}{N} \quad (7.8)$$

Where TP and TN are true positive and true negative values, respectively. N is the total number of samples.

ii) **F0.5 Score:**. F0.5 score has the effect of raising the importance of precision and lowering the importance of recall. It is given by the following equation:

$$F0.5\ Score = \frac{1.25PR}{0.25P+R} \quad (7.9)$$

iii) **F1 Score** [198] : The F1-score is defined as the harmonic mean of the precision and recall values with a distance error tolerance. It is a measure of a model's accuracy on a dataset. The following equation describes the F1 score:

$$F1\ Score = \frac{2PR}{P+R} \quad (7.10)$$

Where P and R are the Precision and Recall values.

iv) **Matthews correlation coefficient (MCC)** [199]**:** The MCC is also known as the phi coefficient. It is correlation the coefficient between the observed and predicted classifications. MCC is given by the following equation:

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7.11)$$

v) **Fowlkes-Mallows Index** [200]**:** Fowlkes-Mallows Index (FM Index) gives a more accurate representation of unrelated data. The higher value of FM Index gives more significant similarity between classified and ground truth data. It is given by the following equation:

$$FM = \sqrt{\frac{TP}{TP+FP} * \frac{TP}{TP+FN}} \quad (7.12)$$

vi) **Balanced Classification Rate (BCR)** [201]**:** BCR is also known as Balanced Accuracy and is useful for the classification of imbalanced classes. BCR is represented by the following equation:

$$BCR = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \quad (7.13)$$

The F1-score is the harmonic mean of recall and precision while the Fowlkes-Mallows Index is the geometric mean of recall and precision.

7.4 Results:

This section presents the classification analysis of the methods used in this paper. The global accuracies achieved by the methods in this study are listed in Table 7-3. The proposed model outperformed other methods by achieving a global accuracy of about 94%. The increment in global accuracy depicts that the probability of identifying the accurate class is increased. Figure 7.3 shows the comparison of evaluation metrics for the classification of COVID-19 infected images by the different methods. The classification accuracy of COVID-19 infected images achieved by our proposed model is 97% which is relatively higher than other methods. The quantitative assessment of classification accuracy is used to examine the quality of the image classification. The proposed classifier effectively segregate the COVID-19 images from the set of three different types of images as the accuracy of the classification of COVID-19 infected images is much higher in our case. The comparison of evaluation metrics for normal and viral pneumonia infected classes is summarized in figure 7.4 and figure 7.5 The inspection of fig 7.4-7.6 shows the improvement in the evaluation metrics of identifying the viral pneumonia class while the classification ability of a normal class of the proposed classifier is nearly comparable to the other methods in the study.

Table 7-14: Global Accuracies of the classifiers

Network	Global Accuracy
RF	0.8571
SVM	0.8762
Alexnet	0.9095
Proposed	0.9388

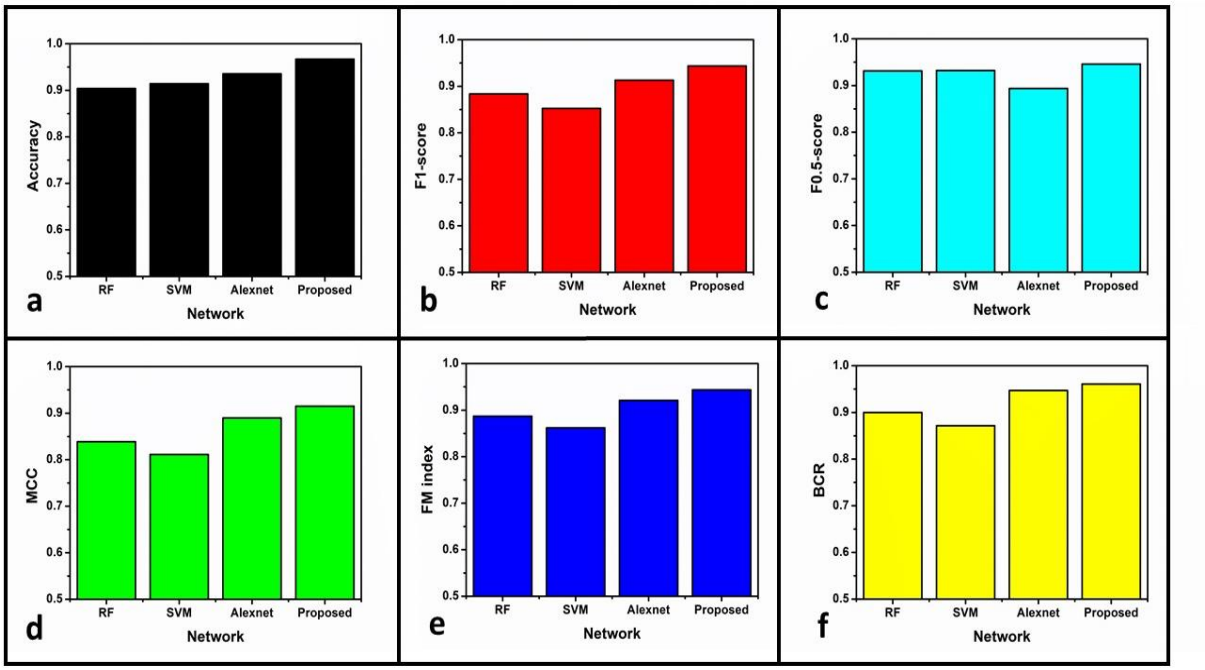

Figure 7.3: Metrics comparison of covid-19 class (a) Accuracy, (b) F1-Score, (c) F0.5-score,(d) MCC,(e) FM Index and (f) BCR.

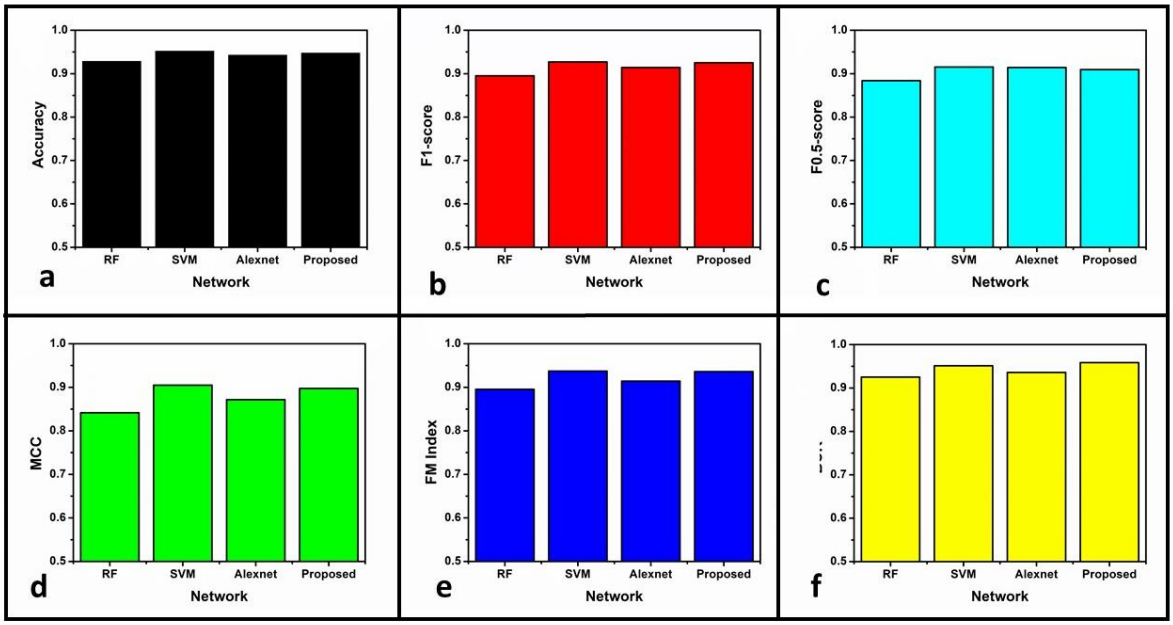

Figure 7.4: Metrics comparison of normal class (a) Accuracy, (b) F1-Score, (c) F0.5-score,(d) MCC,(e) FM Index and (f) BCR.

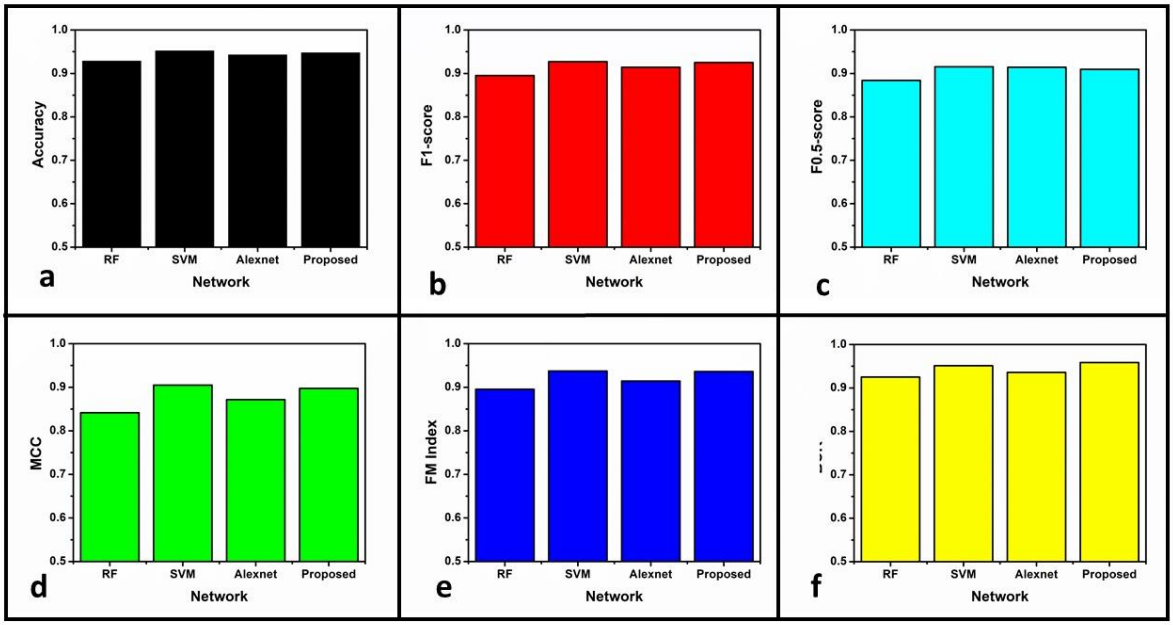

Figure 7.5: Metrics comparison of viral pneumonia class (a) Accuracy, (b) F1-Score, (c) F0.5-score,(d) MCC,(e) FM Index and (f) BCR.

The statistical analysis of the results was performed to validate the results. For this purpose, the Friedman test [165] and Wilcoxon test [166] were used for evaluation. Every method was ranked using the Friedman test, and after that, the p-values were calculated. The null hypothesis assumed that all the methods performed equally. The lower p-value indicated that there is a significant statistical difference between the results obtained by various methods. Thus, the null hypothesis is rejected. The results obtained by the Friedman test applied to MCC values are shown in Table 7-4. Our proposed model got the best rank. The method with the best rank was selected as the controlling method for the Wilcoxon test. The pairwise performance of the controlling method with the other methods was evaluated, and the p-values were calculated, which are presented in Table 7-5. The comparison of the p-values depicted that the proposed model performed statistically well in comparison to other methods in the study as the p-values are less than $\alpha = 0.05$ [145]. This analysis showed that the results produced by our proposed model are statistically more suitable as compared to other methods in the study.

Table 7-15: Mean ranks determined using Friedman Test.

Methods	**RF**	**SVM**	**Alexnet**	**Proposed Method**
Mean of Rank	2.9360	2.0254	1.0738	0.8546

Table 7-16: Results of Wilcoxon signed rank test with $\alpha = 0.05$.

Comparative hypothesis	**Proposed Method vs RF**	**Proposed Method vs SVM**	**Proposed Method vs Alexnet**
p-values	0.00463	0.006357	0.003911

7.5 Discussion

The presented work classifies the COVID-19 infected images efficiently, as indicated by the evaluation metrics. The accuracy of classification achieved for COVID-19 class is about 97%

which includes quite a good improvement over the second-best method. Accuracy is a good parameter when only true positive and true negatives are of importance, but when false positive and false negatives are of importance, then the F1 score is a good measure of classification ability. The F1 score, which is a balance between the precision and recall, is also improved by a good margin. The F1 score is a good indicator when the classes are not perfectly balanced. The F1 score is improved by the proposed model, as shown in fig. which ensures a perfect balance among the precision and recall values for COVID-19 class. Accuracy and F1 measure can sometimes mislead the result interpretation because they fail to consider the ratio of positive and negative elements. The MCC gives correct predictions for both majorities of the negative cases, and the majority of the positive cases, independently of their ratios in the overall dataset. It is improved by a margin of 3.5% in the result obtained by our proposed method. Another metrics Fowlkes-Mallows Index (FM), which gives a more accurate description of unrelated data shows a good degree of improvement for classification of COVID-19 class. The F0.5 score, which is more useful for medical applications shows significant betterment when compared with other methods. It minimises the false positive cases, thus reduces the issue of the false detection of disease. The improvement in F0.5 score is a good indicator of enhanced classification accuracy of the classifier.

The normal class is classified with the classification accuracy of about 95% which comparable with the other methods. The other evaluation metrics such as F1sore, MCC, F0.5 score are also showing comparable performance with the other methods described in the study. For normal class, all the classifiers in the study have classified with nearly the same perfection level. The viral pneumonia infected images are classified with the accuracy of about 94%, which is more than 4.5% higher than the second-best method. The F1 and F0.5 score are improved with a margin of about 2%. MCC and FM index is improved with the margin of more than 3.5% and

2% respectively. The evaluation metrics for viral pneumonia shows that the images infected by viral pneumonia are classified with better precision as compared with other methods [202].

The comparison of the class-wise classification performance gives an interesting fact that the classification performance of the proposed classifier is best for the images infected with either COVID-19 or viral pneumonia. The classification of normal class is comparable with other classifiers. Thus the classifiers understudy classifies the normal class with the approximately same perfection. For both the infectious classes (COVID-19 and viral pneumonia), the proposed classifier is best suited as the evaluation metrics suggest. The MCC score, which takes all the four cardinalities of the confusion matrix into account, is improved by the good amount. Moreover, this metric is best suited for the multiclass classification of imbalanced classes as for the imbalanced classes accuracy, and F1 score sometimes mislead the classification evaluation. The Balanced classification accuracy, which is a measure of the quality of predicting positive values by the classifier shows a considerable improvement in the results obtained by the proposed classifier. The proposed classifier is comparable in performance in classifying the normal class because the normal images are highly correlated and lack abrupt texture variations. The infected images have a variation in texture because of the disease. These random variations of texture are correctly learnt by the proposed classifier in the learning phase, which results in a more precise classification of the infected images. The feature maps produced by the convolution layers are captured by the classifier along with their indices, and these indices are passed on to the decoder stage to correctly identify the variations along with the locations. Sometimes the location of infection is also of great importance while detecting the disease. These variations are effectively captured by the classifier for the infected classes, and the difference between the two types of infections is also learnt quite effectively. In the identification of nearly correlated diseases, the texture variations become an important parameter. The proposed classifier correctly learns the abrupt variations and this type of

learning is very helpful for classification purpose. Moreover, it was able to perfectly classify the infected images of both types, i.e. COVID-19 or viral pneumonia. The normal chest X-ray images are highly correlated and possess approximately the same texture variations which are easily learnt by all the classifiers in the study [203] [204]. So for the normal class, all the classifiers are classifying the images with comparable effectiveness. According to the pathological studies of the lung on the autoptic tissues for covid-19 and viral pneumonia, the widespread thrombosis with microangiopathy and vascular angiogenesis are the main discriminators for texture variations. These texture variations are efficiently distinguished by the proposed classifier for COVID-19 and viral pneumonia. The results of classification by the proposed classifier are more clinically pertinent in terms of accurate segregation of COVID-19 infected images. The Friedman test and Wilcoxon signed-rank test indicated the statistical superiority of the results obtained by the proposed model [205]. The Wilcoxon signed-rank test gives the p-values less than 0.05 ($\alpha = 0.05$) [144] (Table -5), which indicates the better performance of the proposed method.

7.6 Transfer Learning Based Approach for Classification of Covid-19 infected images:

7.6.1 Methodology:

The presented work utilizes a transfer learning-based approach for classifying the covid-19 infected CT and X-Ray images as both types of image modalities are used for detection of covid-19 symptoms. The SVM classifier [191] is used to in this work for the classification purpose. MobileNet V2 architecture was used for extracting the features of the covid-19 infected images. The extracted features were transferred to the SVM classifier, which effectively classified the infected images. For comparison, various other networks were also used for feature extraction. The features extracted by various network architectures for comparison were also transferred to the SVM classifier. The use of transfer learning facilitated

the fast classification of the images which were pertinent from the medical point of view. The figure 7.6 shows the pipeline of the proposed work.

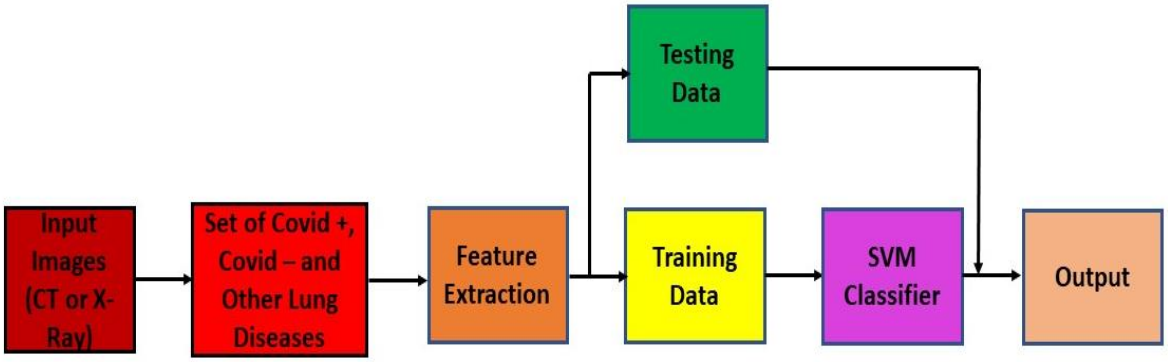

Figure 7.6: Pipeline of the proposed work.

7.6.2 Transfer Learning:

The transfer learning eliminates the need for training the classifier from scratch as it utilizes the knowledge which the gained by a network on with much available training labelled dataset [206]. This approach is useful, where the less amount of data is available. Transfer learning is a sort of design methodology in the field of machine and deep learning. For computer vision-related application, the deep learning models generally extracts the features of edges in the initial layers. The middle layers extract the specific features related to the particular task for which the network is intended. So the features learnt by these layers are transferred to the other network for performing the same or different applications. In the transfer learning approach, generally, the last layers are trained. The advantage of using this approach is that it saves time for training the large networks, and it eliminates the need for large datasets for training. We used the knowledge gained by MobileNet model to train SVM classifier for classification of CT and X-ray images infected by SARS-CoV-2. Figure 7.7 illustrates the whole process of transfer learning.

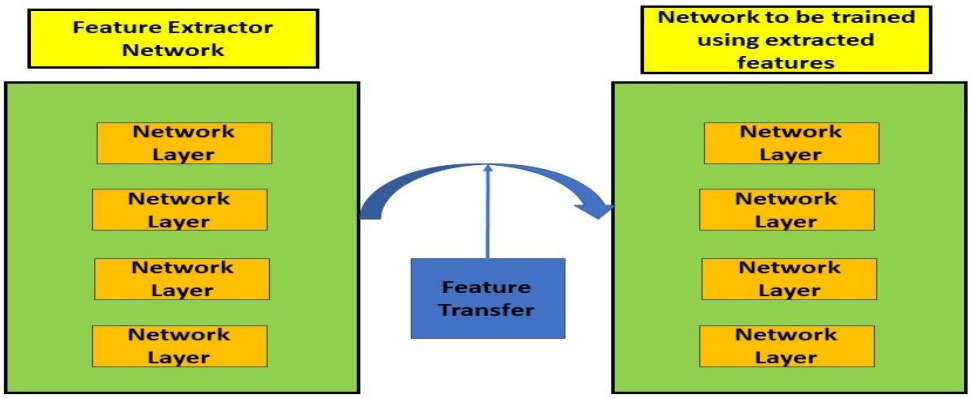

Figure 7.7: Feature transfer using transfer learning.

7.6.3 MobileNet V2 architecture:

The MobileNet V2 [73] architecture was developed by Google Brain, which is a next-generation architecture for general-purpose computer vision applications. Its immediate predecessor was MobileNet V1 which enabled the deep learning networks to be operated on the personal mobile devices. MobileNet V2 has significantly enhanced architecture than MobileNet V1 as it enables the mobile visual object detection, classification and other computer vision-related tasks. MobileNet [207] family is low power and low latency models used to accomplish various computer vision use cases. Version 2 of MobileNet is incorporated with the linear bottlenecks between the layers and the shortcut connections between the bottleneck blocks [208]. The bottlenecks encode the in-between inputs and outputs. The transformation from lower-level units such as pixels to higher-level illustration such as image is performed by the middle layers. The residual bottleneck block for MobileNet V2 is shown in the figure 7.8. The depthwise convolutional Layer and Projection Layer are standard in both the versions of MobileNets. The depthwise convolution is more efficient than the standard convolution. A single filter is applied to each input channel by the incorporation of dept wise convolution. For combining the outputs of depthwise convolution, a pointwise 1×1 convolution is applied. This combined process of filtering and combining is termed as depthwise separable

convolution. The depthwise separable convolution splits this into two layers, a different layer for filtering and a different layer for combining as shown in figure 7.8.

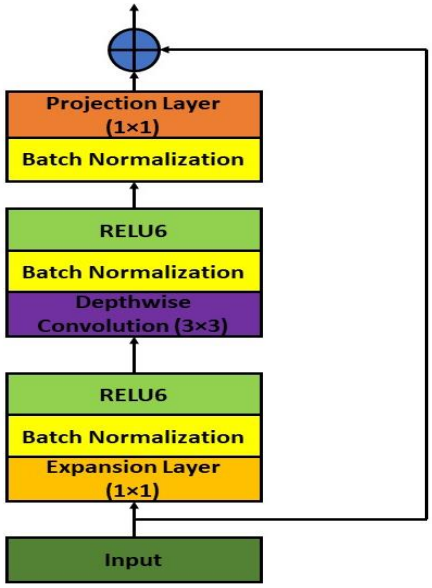

Figure 7.8: Depth wise Separable convolution block.

The initial layer in the figure is the Expansion Layer. This layer is also a 1×1 convolutional layer. It expands the channel count in the data before it is passed to the depthwise convolution layer. Expansion Layer is just opposite to the projection layer as it has a higher number of output channels than the input channels as shown in figure 7.9. The incorporation of residual connections in MobileNet V2 architecture assists the flow of gradient through the network. In each layer, Batch Normalization and activation function Relu6 is incorporated. For low precision applications, RELU6 [116] is more robust than normal RELU. There is no activation function applied to the projection layer as it produces low-dimensional data, and using any non-linearity may destroy the useful information. The MobileNet architecture as shown in figure 7.10, consists of 17 layers of building blocks consecutively which are followed by 1×1 convolutional layer, global average pooling layer, fully connected layer, SoftMax layer and a classification layer. The global average pooling layer generates one feature map for each corresponding category of classification. Finally, the classification layer segregates the data into specific classes to which they belong.

The finely extracted features of MobileNetV2 are transferred to the SVM classifier. The SVM is powerful and flexible supervised machine learning algorithms which are used for classification as well as regression analysis. It is a discriminative classifier which is formally defined by separating hyperplane. The SVM draws a hyperplane for classifying the data points. In this work, the SVM is trained using the transfer learning approach. The MobileNet V2 extracted the features of covid-19 from CT as well as X-Ray images. Both types of image features extracted by MobileNet V2 were fed to the SVM classifier for training. The SVM effectively classified the covid-19 infected images for both the modalities, i.e. CT and X-ray.

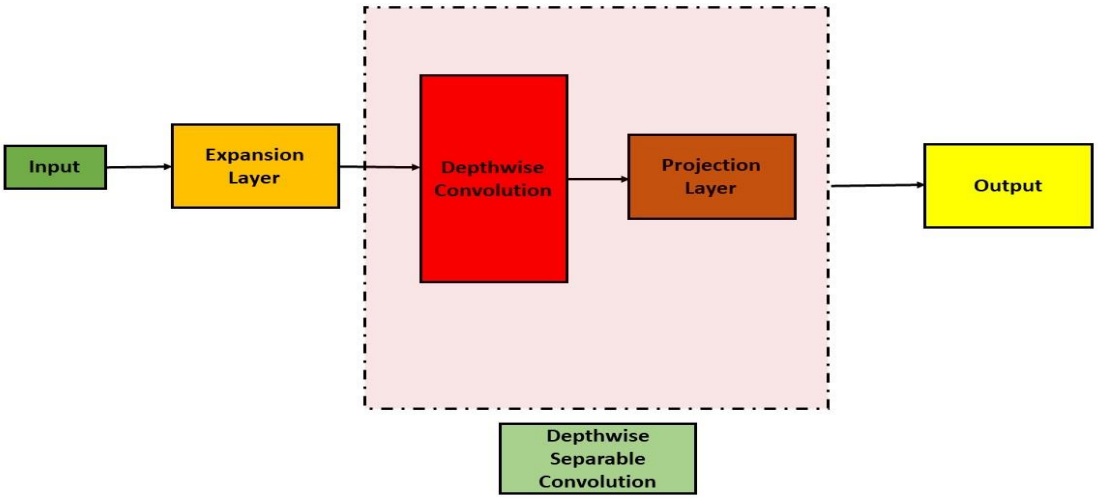

Figure 7.9: Functioning of layers in depthwise separable convolution block.

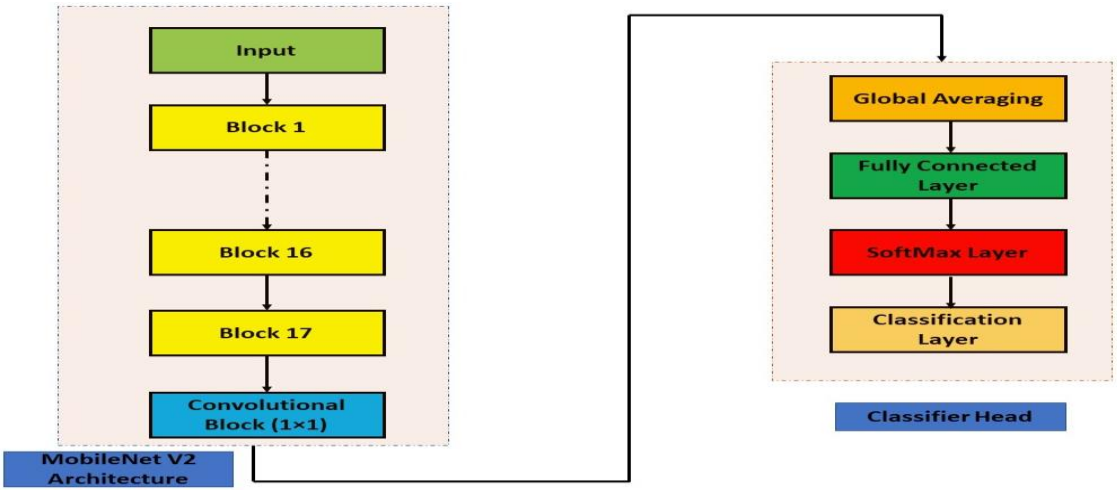

Figure 7.10: Mobile Net V2 architecture for classification.

7.6.4 Dataset Used:

7.6.4.1 X-ray Dataset:

The data that supports the findings of the study are available in Covid-19 radiography database of Kaggle which consists of 219 chest X-ray images of COVID-19 positive patients,1341 images of normal chest X-ray and 1345 images of viral pneumonia infection [197]. In addition to this, 301 images were taken from the GitHub repository of Dr Joseph Paul Cohen [209]. To increase the data for training, we augmented the data using data augmentation techniques as the deep learning networks need a large amount of dataset to be trained. The data was rotated at the angles of 30^0, 60^0, 90^0,120^0 ,270^0 and 300^0. In total, we created 3120 COVID-19 infected images, 8046 normal images and 8070 viral pneumonia infected images.

7.6.4.2 CT Dataset:

The CT dataset is acquired from CT scans for covid-19 classification database of Kaggle. The database consists of images collected from Union Hospital (HUST-UH) and Liyuan hospital (HUST-LH) [210]. The database consists of 5705 non-informative CT (NiCT) images, 4001 positive CT (pCT) images and 9979 negative CT (nCT).

7.6.5 Results:

This section presents the analysis of the comparative results obtained by the proposed and other existing classifiers models. The models used for comparison are Alexnet, Resnet101, Inception V3, Darknet and ShuffleNet. The results obtained are shown in three categories 1) The comparison of accuracy, F scores and MCC, 2) The parameters whose value should be minimum for achieving good performance and, 3) other parameters for firm confirmation of results obtained. The second category is of those parameters which should attain minimum value to confirm the effectiveness of classification. Six more evaluation metrics are used for

the firm affirmation of results. Table 7-6 and Table 7-7 shows the accuracy, F scores and MCC values obtained by the classifiers for CT and X-Ray images, respectively. Table 7-8 and 7-9 show the value of parameters whose minimum value is desired. Figure 7.11 and Figure 7.12 show the results obtained by other parameters which confirm the robustness of results. The statistical paired t-test was performed to validate the statistical significance of the proposed method. The null hypothesis was rejected as the value of p was less than 0.05 (with α =0.05) and a confidence level of 95% was achieved.

Table 7-6: Evaluation metric comparison of CT images.

Networks	**Accuracy**	**F1**	**F2**	**MCC**
Mobilenet V2	0.9942	0.99	0.9890	0.9852
Alexnet	0.9611	0.9391	0.9643	0.9125
Darknet	0.9805	0.9716	0.9834	0.9606
InceptionV3	0.9761	0.9635	0.9740	0.9461
Shufflenet	0.9817	0.9702	0.9777	0.9587
Resnet	0.9756	0.9629	0.9707	0.9448

Table 7-7: Evaluation metric comparison of X-Ray images.

Networks	**Accuracy**	**F1**	**F2**	**MCC**
Mobilenet V2	0.9854	0.9770	0.9815	0.9657
Alexnet	0.9556	0.9298	0.9601	0.9001
Darknet	0.9705	0.9587	0.9540	0.9379
InceptionV3	0.9667	0.9483	0.9683	0.9249
Shufflenet	0.9702	0.9588	0.9542	0.9380
Resnet	0.9611	0.9391	0.9643	0.9125

Table 7-8: Evaluation metric comparison of CT images with expected low values.

Networks	**FNR**	**FPR**
Mobilenet V2	0.0084	0.0059
Alexnet	0.1	0.0083
Darknet	0.0267	0.0069
InceptionV3	0.0534	0.0091
Shufflenet	0.0367	0.0091
Resnet	0.05	0.0117

Table 7-9: Evaluation metric comparison of X-Ray images with expected low values.

Networks	FNR	FPR
Mobilenet V2	0.0304	0.0076
Alexnet	0.1166	0.0083
Darknet	0.0433	0.0250
InceptionV3	0.0833	0.0083
Shufflenet	0.0434	0.0249
Resnet	0.1	0.0083

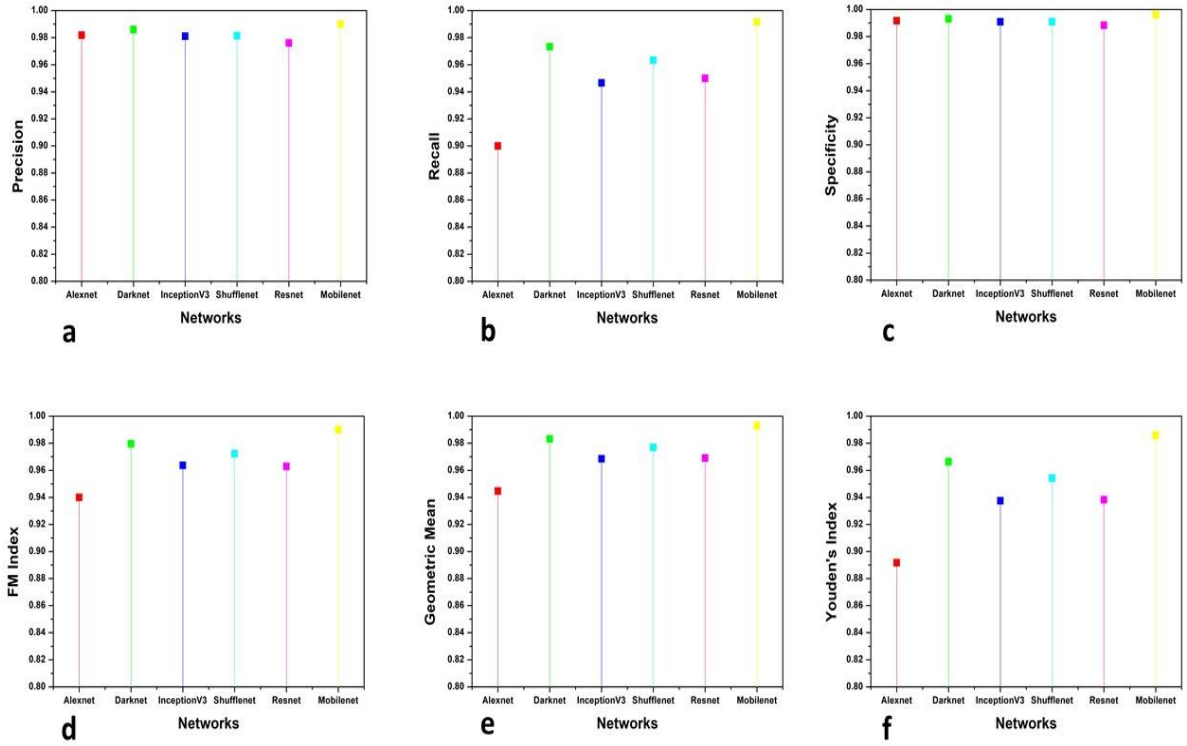

Figure 7.11: Evaluation Metrics comparison of CT images.

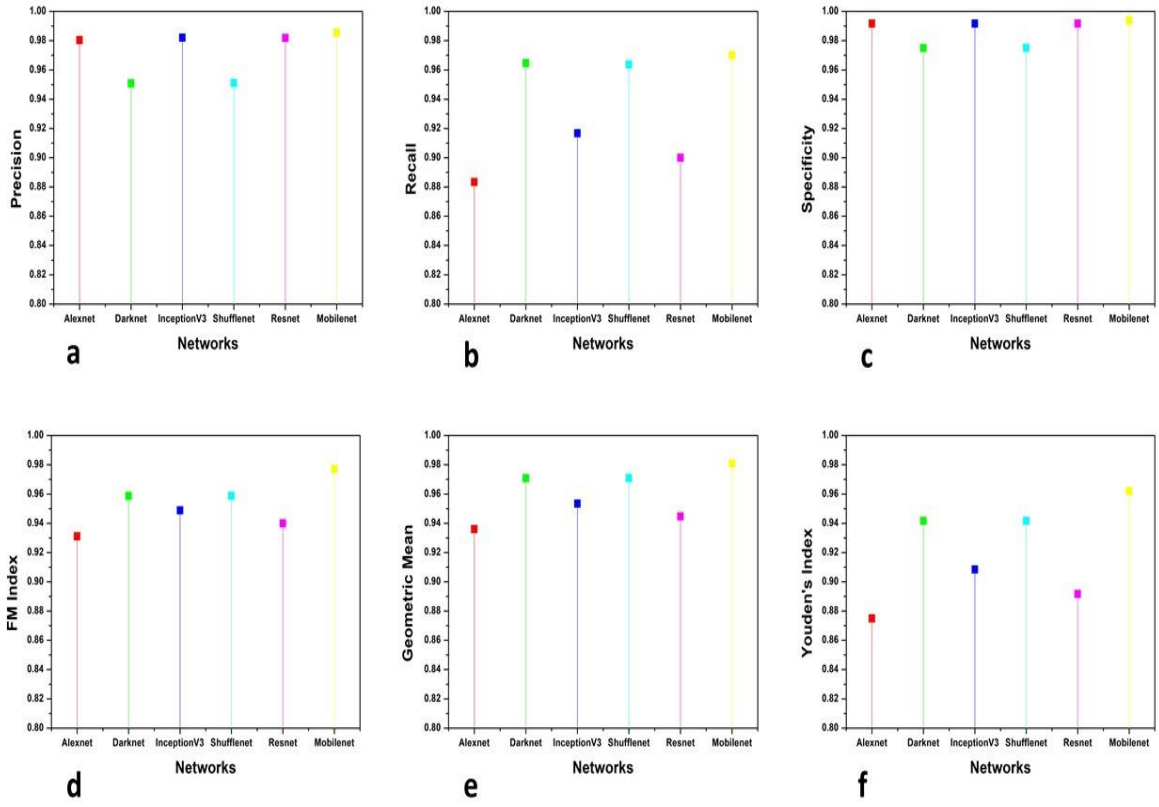

Figure 7.12: Evaluation Metrics comparison of X-Ray images.

7.6.6 Discussion

The accurate analysis and detection of covid-19 infection are quite necessary in the current scenario as the pandemic has placed the whole world in an unprecedented situation by its life-threatening effects. The proposed classifiers segregate the data into covid (+) and covid (-) classes where covid (+) is the patient infected by SARS-CoV-2 and covid (-) is the patient not having the symptoms of infection. The prediction of infection is made on the basis of the four-parameter values of the confusion matrix, i.e. True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The TP and TN values identify covid (+) and covid (-) patients. FP represents the outcomes which are incorrectly recognised as covid (+) while FN depicts the outcome which is incorrectly classified as covid (-).

7.6.6.1 Quantitative Analysis:

Accuracy is a measure of accurate identification of the class of a pixel. The evaluation of correctly identified pixels is made using accuracy. Accuracy is computed using the following equation:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7.14)$$

The accuracy for both the modalities is showing improvement of more than 1.5%. The distribution of data is not taken into account by the accuracy so the F- score [198] is a better parameter for judging the classification efficiency. It is given by the following equation:

$$F1\ Score = \frac{2PR}{P+R} \quad (7.15)$$

and

$$F0.5\ Score = \frac{1.25PR}{0.25P+R} \quad (7.16)$$

Where P and R are the Precision and Recall values. The F1-score is defined as the harmonic mean of the precision and recall values with a distance error tolerance. It is a measure of a model's accuracy on a dataset. F0.5 score has the effect of raising the importance of precision and lowering the importance of recall. The F1 score maintains the balance of precision and recall values has shown an improvement of 1.8% for CT images and 1.9% for X-ray images. The f 0.5 score, which gives more importance to precision as desired for medical analysis is also improved by 1.25 % and 1.4% for CT and X-ray images respectively. Accuracy and F1 measure can sometimes mislead the result interpretation because they fail to consider the ratio of positive and negative elements. The Matthews correlation coefficient (MCC) is used to firmly quantitative the analysis of classification as the mathematical properties of MCC handles both dataset imbalance and their invariants effectively. It gives correct predictions for both majorities of the negative cases, and positive cases, independently of their ratios in the overall dataset [199]. The following equation describes MCC mathematically:

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7.17)$$

MCC has shown a remarkable improvement of respectively 2.6% and 3.8% for CT and X-ray images. This signifies the better classification ability of the proposed model as the MCC has achieved the highest score along with the accuracy and F-scores.

False Positive Rate (FPR) and False Negative Rate (FNR) gives the frequency with which the classifier makes a mistake by classifying normal state as infection and infection state as normal respectively. FPR and FNR are given by the following equations:

$$FPR = \frac{FP}{FP+TN} \quad (7.18)$$

and

$$FNR = \frac{FN}{FN+TP} \quad (7.19)$$

The values of FPR and FNR achieved by the proposed classifier is minimum among the compared methods, which once again proves the effectiveness of the classification results of the proposed classifier.

For more analysis and a better understanding of results, we used six additional evaluation metrics for examining the obtained results. Precision, Recall, Specificity, FM index, Geometric mean and Youden's index values are used to compare the results of classification by various methods. Precision [211] which is also known as Positive prediction value (PPV) is the ratio of positive samples which were correctly classified to the total number of samples predicted as positive. It is given by the following equation:

$$Precision = \frac{TP}{TP+FP} \quad (7.20)$$

Recall or sensitivity is given as the ratio of positive correctly classified samples to the total number of positive samples. It is also known as True positive rate (TPR) and hit rate. The following equation represents Recall:

$$Recall = \frac{TP}{TP+FN} \quad (7.21)$$

The recall is the ability of a model to find all the relevant cases within a dataset while the ability of a classification model to identify only the relevant data points is termed as Precision. Specificity, which is also termed as inverse Recall or True negative rate (TNR) is expressed as the fraction of correctly classified negative samples to the total count of negative samples. Specificity is expressed mathematically as:

$$Specificity = \frac{TN}{FP+TN} \quad (7.22)$$

Fowlkes-Mallows Index (FM Index) [200] gives a more accurate representation of unrelated data. The higher value of FM Index gives more significant similarity between classified and ground truth data. It is given by the following equation:

$$FM = \sqrt{\frac{TP}{TP+FP} * \frac{TP}{TP+FN}} \quad (7.23)$$

The prime goal of the classification process is to improve sensitivity without sacrificing the specificity. This is quite a tough task for imbalanced datasets. Thus Geometric Mean (GM) [212] represents the aggregation of both these metrics using the following equation:

$$GM = \sqrt{\frac{TP}{TP+FN} * \frac{TN}{TN+FP}} \quad (7.24)$$

The values obtained by the Specificity, FM and GM has shown considerable improvement shown in the figure 7.11. The higher values of sensitivity and specificity indicate that the true values of positive and negative classified samples are greater in the results obtained by the classifier. Thus the classifier is said to have predicted the covid (+) and covid (-) samples more accurately than others if these two scores have larger values than other methods in comparison. Fowlkes-Mallows Index (FM), which gives a more accurate description of unrelated data shows a good degree of improvement for classification of COVID-19. The last evaluation

metric Youden's index (YI) or Bookmaker Informedness [213] is an important metrics for medical image analysis as it evaluates the discriminative power of the diagnostic test. It evaluates the probability of informed decision and is more suitable for datasets with imbalanced classes. Youden's index shows an increment of more than 2% for both CT and X-ray images. Mathematically it is represented by the following equation:

$$J = \frac{TP}{TP+FN} + \frac{TN}{TN+FP} - 1 \qquad (7.25)$$

Where J represents Youden's index. In this work, total twelve metrics were used for evaluating the performance of classification of covid-19 infected images. The proposed classifier works equally well for both types of medical imaging modality, i.e. CT and X-ray.

7.6.7 Conclusion:

The correct and timely diagnosis of COVID-19 infection is the demand of current time as the pandemic is a significant threat to human life. The proposed classifier efficiently classified both types of infected images, i.e. viral pneumonia and COVID-19. The results were confirmed using MCC, FM index, along with accuracy and F scores (1 and 0.5). Matthews correlation coefficient (MCC) was used because sometimes accuracy and F score misleads the prediction of classification as mathematically it is a perfect balance between all the four cardinalities of the confusion matrix. The feature maps of both the comparable classes were generated, and the difference between the two classes was learnt efficiently by the proposed classifier. For detecting the COVID-19 in the early stage, the presented study is helpful for medical diagnosis as we have adopted more parameters to confirm the results obtained by our model.

Transfer learning-based approach is presented in this chapter for accurate identification of covid-19 in CT and X-ray images. The features finely extracted by Mobile Net V2 model paved the way for this research. Matthews correlation coefficient (MCC) was used because sometimes accuracy and F score misleads the prediction of classification as mathematically it is a perfect

balance between all the four parameters of the confusion matrix. Youden's Index, which is quite a trustworthy statistical measure for the medical image analysis, shows outstanding improvement, which proves the medical effectiveness of the results obtained by the proposed model. Our proposed method can be helpful in identification of the infection and can be used as an alternative to expensive Covid-19 testing kits. Moreover, our proposed model eliminates the need for a medical practitioner to confirm the infection as it automatically detects the infection precisely using the knowledge gained through training. The availability of resources does not match the preventive measures to cope with the pandemic in the current time. In such a situation, the computer-based analysis may save the time of doctors and life of the patients by early diagnosing the infection.