MATERIALS AND METHODS

Materials

A. *Dataset description*

A face expression recognition dataset nameds 'mask type' was collected from the public database kaggle to evaluate the performance of the proposed hybrid model. The images of this dataset are split into a train, test, and validation. The dataset information is shown in Table I.

Dataset	***Number of image***
Train	2133
Test	439
Validation	533
Total	3100

The train dataset contained four classes of facial expressions (ClothMask, N95Mask, SurgicalMask, WithoutMask), with a minimum of 3000 images per class except for the Disgust class. Table II shows the number of images collected per class.

Dataset	***Per class Image***
ClothMask	878
N95Mask	708
SurgicalMask	680
WithoutMask	400

Here a sample of each class's image is shown in Fig. 1 from the mask dataset.

Cloth-mask

Surgical-mask

N95-mask

No-mask

B. *Data pre-processing*

The images collected from the dataset are of random size. So, all collected images are resized using OpenCV to the required image size to be fed into the model.

Method:

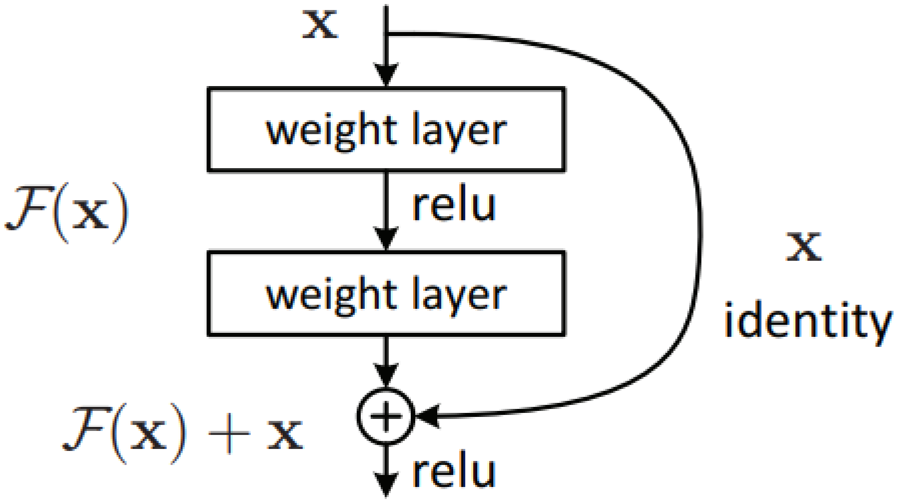
This work introduces a novel pretrained Convolutional neural network (CNN) to recognize types of mask from an image. In the feature extraction layers of the proposed architecture, the CNN network automatically extracts invariant features from images. The extracted part is then fed to fully connected layer for recognizing unseen images.

A. Feature Extraction

CNN has shown outstanding performance in various tasks of feature extractions due to its deep architecture that contains multiple types of filters. CNN's deep architecture enables it to automatically learn mid-level and high-level representative and hierarchical image features from an image. The learned representations provide better classification performance than hand-crafted feature extraction in computer vision tasks. Here the principal advantage of CNN is that we don't need to extract features from images manually.

A pre-trained Resenet deep CNN architecture is used for feature extraction from images. The Resenet deep CNN architecture was pre-trained on an image dataset with more than 1 million images to achieve state-of-the-art accuracy for recognizing general objects with 1000 classes. The interesting concept of Residual blocks .In order to solve the problem of the vanishing/exploding gradient, this architecture introduced the concept called Residual Blocks. In this network, we use a technique called **skip connections**. The skip connection connects activations of a layer to further layers by skipping some layers in between. This forms a residual block. Resnets are made by stacking these Residual blocks together.

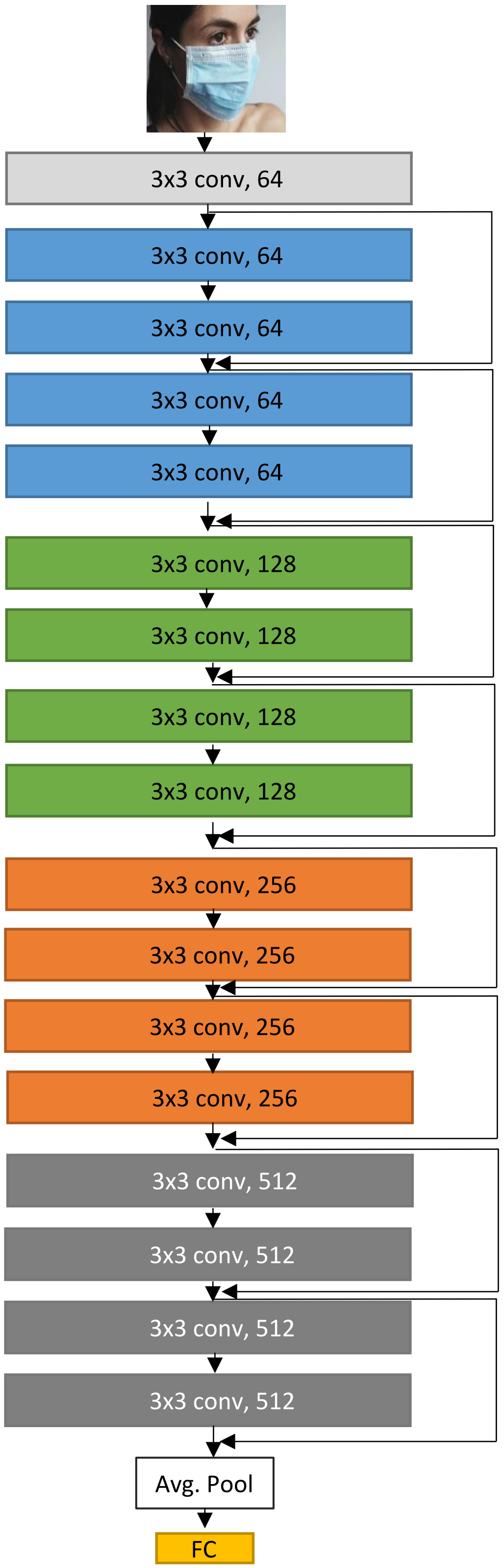
The approach behind this network is instead of layers learning the underlying mapping, we allow the network to fit the residual mapping.


x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Skip (Shortcut) connection

The pre-trained network can partially share its weights to extract discriminative features from images based on transfer learning theory, making the learning process faster and easier. We need to feed an image to the CNN model. Then the

CNNs can automatically learn to extract complex hierarchical features from an image by using different filters during training time. The whole architecture for extracting features from an image is shown in the figure below.



RESULT

A. Experimental Setup

It is required to optimize the parameter and hyperparameter value of the model to measure the efficacy of the proposed Deep CNN-SVM model. In this work, an Adam optimizer is used to optimize the deep CNN model with a learning rate of 0.0001 and choose optimal gamma and 'C' values to optimize the hyperparameters of SVM. Finally, the proposed model is trained by dividing it into training, test, and validation sets in the ratio of 60:20:20. The proposed model was trained on a single NVIDIA GeForce GTX 1080 GPU with a PyTorch environment

B. Evaluation matrices

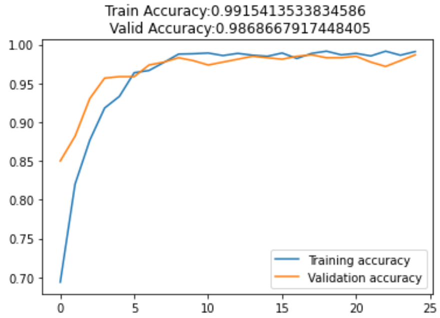
Normal accuracy as an evaluation metrics makes sense only if the class labels are uniformly distributed for automatic image using deep learning. But for imbalanced classes confusion-matrix is good technique to summarizing the classification performance of a classification algorithm.

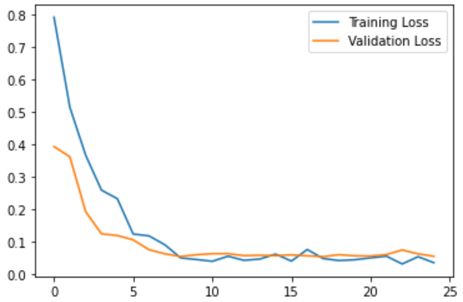
Precision-Recall is a useful measure of success of prediction when the classes are very imbalanced. Precision is a measure of the ability of a classification model to identify only the relevant data points, while recall is a measure of the ability of a model to find all the relevant cases within a dataset.

The precision-recall curve shows the trade-off between precision and recall for different threshold. A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate.

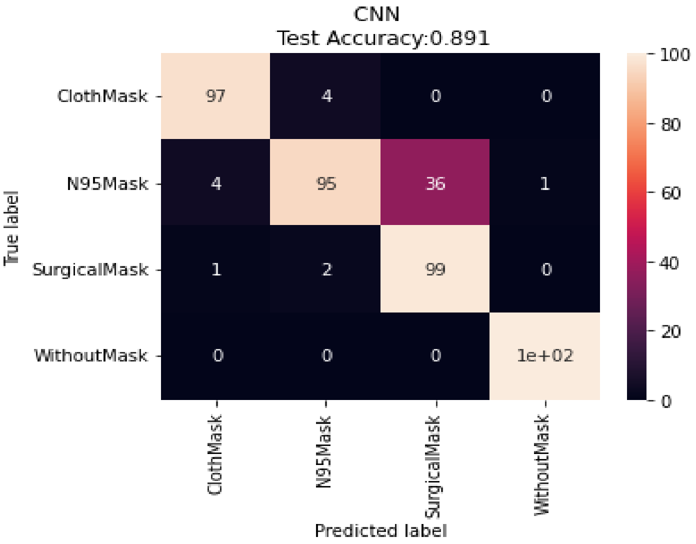
C. Model performance

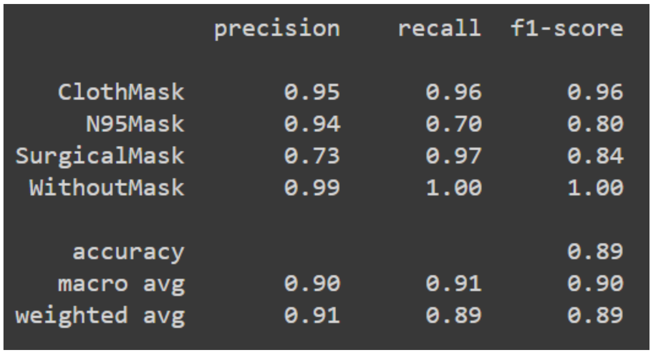
The proposed model is trained to recognize types of mask with hyperparameter, and parameter values achieve an accuracy near 99% on the mask dataset. Fig. 4 and Fig.5 show the training and validation accuracy and loss graphs for training data regarding the number of epochs for the mask dataset





The proposed system has used CNN as a feature selection extractor with fully connected (FC) classifier. The image features were extracted from CNN feature selection algorithms and fit them into the FC classifier. Then the accuracy of this algorithm is observed in terms of precision and recall score. And the confusion matrix is observed in the test images that are correctly classified by the FC classifier. The accuracy of precision and recall score and confusion matrix for the CNN model is depicted below.




```
              precision    recall  f1-score

   ClothMask       0.95      0.96      0.96
     N95Mask       0.94      0.70      0.80
SurgicalMask       0.73      0.97      0.84
 WithoutMask       0.99      1.00      1.00

    accuracy                           0.89
   macro avg       0.90      0.91      0.90
weighted avg       0.91      0.89      0.89
```

Finally, the image classifies to their correspond class using CNN. Here we show some prediction of our CNN model.

predicted: SurgicalMask

predicted: SurgicalMask

predicted: ClothMask

predicted: N95Mask

predicted: N95Mask