5th IEEE International Conference on Recent Advances and Innovations in Engineering- ICRAIE 2020 (IEEE Record#51050)

Face Mask Detection Classifier and Model Pruning with Keras-Surgeon

Alok Negi
Department of Computer Science and Engineering
National Institute of Technology, Uttarakhand
Srinagar (Garhwal), India
alokitn@gmail.com

Prachi Chauhan
Department of Information Technology
G.B. Pant University of Agriculture and Technology
Pantnagar, India
prachi3apr@gmail.com

Krishan Kumar
Department of Computer Science and Engineering
National Institute of Technology, Uttarakhand
Srinagar (Garhwal), India
kkberwal10@gmail.com

R.S. Rajput
Department of Mathematics, Statistics and Computer Science
G.B. Pant University of Agriculture and Technology
Pantnagar, India
rajpoot.rs@gmail.com

2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE) | 978-1-7281-8867-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICRAIE51050.2020.9358337

Abstract— **Multidisciplinary initiatives in the new world of coronavirus were combined to limit the spread of the pandemic. Interestingly, the AI group was a part of those efforts. This result-based approach is used to help scan, assess, predict and track current patients and possibly potential patients. Developments for tracking social distances or recognizing face masks have made headlines in particular. Most current advanced approaches to face mask recognition are built based on deep learning which is dependent on a large number of face samples. Nearly everybody wears a mask during corona virus outbreak in order to effectively avoid the spread of COVID-19 virus. Our goal is to train a customized deep learning model that helps to detect even if or not a person wears a mask and study the concept of model pruning with Keras-Surgeon. Model pruning can be efficient in reducing model size, so that it can be easily implemented and inferred on embedded systems.**

Keywords— CNN, Data Augmentation, Deep learning, Face Mask, Model pruning

I. INTRODUCTION

Covid-19 pandemic is a far broader social, economic and political phenomenon than just a biological phenomenon. The pandemic is turning into one of the worst humanitarian disasters in modern human history, due to its ill-effects on life and livelihood. Furthermore, the mismanagement of the disease has made world reach the population transmission stage on the verge of Covid-19. Despite the state of public health facilities, the emerging pandemic poses a significant threat for both the government and the people.

Coronaviruses [1] are a wide family of viruses common to animals of different types, including bats, cattle, cats and camels etc. They inflict illnesses that range of cold with SARS. Just like influenza epidemic the coronavirus actively and im-plicitly spreads. Direct interaction occurs by close contact with oral secretions via a physical transfer of the microorganism. Explicit interaction happens while a person carrying the virus is sneezing or coughing that disperse droplets of the virus to surfaces. Globally, there were 28,918,900 confirmed cases of COVID-19 registered to the WHO at 3:28pm CEST, 14 September 2020, including 922,252 deaths.

Looking at the current issue of coronavirus disease, the World Health Organization (WHO) has recommended that preventive steps should be taken to protect ourselves. One of the government and WHO's a key preventive measure is to wear facemask when going outside as well as social distancing. Therefore, it has become important to create automated applications to find out whether or not anyone is wearing a mask so measures can be taken accordingly.

Deep learning algorithms (DL) allowed the development of highly precise systems and became a standard option for the analysis of different complex applications [2], [3] in different areas. DL methods have been extremely effective over the last decade and have been widely used to build artificial intelligence [6-8] in nearly every domain. Deep convolutional neural networks have recently been successfully applied to face masks detection. By encouraging this, we suggested a system which would classify those among the crowd who did not wear a mask. This mechanism can be applied in various areas like airports, railway stations, malls and all other crowded places as a preventive measure which has a considerable importance in the current scenario.

The rest of the work is structured as follows: Section 1.2 and Section 1.3 define related work and the proposed work respectively. Section 1.4 describes results and analysis followed by conclusion in Section 1.5 and references.

II. RELATED WORK

Loey et al. [4] have introduced the hybrid design which uses deep learning for face mask detection which has two parts. Resnet50 is used as a first part for extraction of the feature while second part used the concept of support Vector Machine, ensemble algorithm and decision trees for classification and recorded testing accuracy 99.64 percent using SVM on RMFD dataset, 100 percent on LFW dataset and 99.49 percent on SMFD dataset.

Sabbir et al. [5] constructed the Principal Component Anal-ysis (PCA) to identify the person in a masked and unmasked face. By using extremity of the PCA they observed that wearing masks had an effect on accuracy of face resonance.

Grassi et al. [6] proposed Data preprocessing of a quite smooth image by applying a sequentially shaded elliptical mask focused over the ears. Used during classification in combination with DCT, for extracting functions, and RBF Neural Networks and MPL, it enables system output to be increased without altering the overall computation intensity and also decreases learning time of neural networks with MLP.

Li et al. [7] developed an HGL methodology to overcome the major issue of head pose specification with masks mostly during endemic problem COVID-19. This method utilizes an analysis of image colour distortion as well as a line representation.

Ramachandra et al. [8] presented An observational research on vulnerability identification and appearance attack detection employing custom 3D silicone masks leading to actual targets for commercial face recognition systems (FRS).

Mahore et al. [9] illustrated a modern method for detecting the appearance of an anti-spoofing-based 3D face mask by using intensity and texture attribute descriptors. The proposed solution derives texture-dependent functionality from isolated wavelet processed images dependent on Local Binary Se-quence against Dataset 3D mask attack.

Meenpal et al. [10] introduced a system for producing exact face mask feature vectors of any random image input dimension. Starting with the RGB image of any dimension, the method employed VGG's Predefined Training Weights-16 Architecture for extracting of the function. Training is carried out by complete convolutional networks to separate the faces represented in the image semi-semantically. Gradient Descent is being used to prepare while the Binomial Cross Entropy factor was introduced as a disadvantage. Experimentation for segmented face masks were conducted on Multi Parsing Human Dataset achieving mean pixel level precision of 93.884 percent.

Li et al. [11] suggested an innovative 3D face mask attack detection approach based on visual refractive analysis. The face picture was first analyzed in the proposed approach using an inherent picture decomposition technique to measure the image reflectance. Then, the histograms of the pixel intensities are derived from three orthogonal planes to point out the differences in intensity of reflectance images between the actual face and the 3D face mask. Afterwards, provided that a seamless surface's reflectance image was more susceptible to changes in illumination, the 1D convolutional neural network has been used to describe how various components or surfaces respond differently to illumination variations.

III. PROPOSED WORK

One of the Government and WHO's a key protective step is to wear facemask when moving outside with social distance. Therefore, we developed a image based CNN model to find out if anybody wears a mask or not in this proposed work. Further we incorporated a concept for model pruning using with Keras-Surgeon.

A. Face Mask Detection

We used python script, tensor flow and CNN as deep learning architecture to create an efficient network for the detection of facemasks. Our purpose is to train a custom CNN model to detect if a person is wearing a mask or not. The Flow diagram of the proposed work is shown in Figure 1.

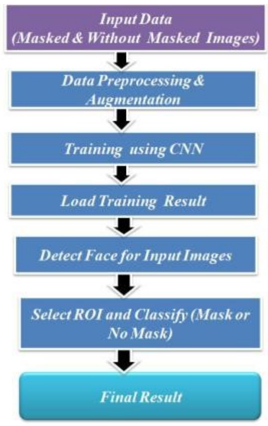


Fig. 1: Proposed Work Flow Diagram

B. Haar Cascade Classifier

For face detection we used the OpenCV based Haar cascade classifier which was developed by the Paul Viola and Michael Jones [12]. It is a machine learning based approach which uses a lot of positive and negative images for training the classifier. Positive images contain images that we want to classify for our classifier and negative images contain of anything else that doesn't involve the entity that we want to find. OpenCV also includes several pre-trained classifiers for eyes, face, smile etc. Firstly, we loaded haarcascade frontalface default xml classifiers and after that loaded An Grayscale mode of input image. Then we have the faces throughout the image. If features are identified, the coordinates of the face region are retrieved as Rect(x, y, w, h). If we have those positions, we will be able to build a face ROI. Finally, on detected face ROI, a trained classifier is applied to determine if a person wears a mask or not.

C. Convolutional Neural Network (CNN) Implementation

CNN is analogous to "standard" neural networks, through-out the context that they have been completely invented of hidden layers with "learnable" specifications of neurons. These neurons obtain inputs, perform a dot product and then proceed with non-linearity. The entire network communicates the correlation between the raw pixels of images and their classification ratings. The CNN is a deep neural network that typically takes images as input, trains features based on bias and weights, the value of which is randomly selected technically, Every input image passes through a sequence of kernels, pooling layers, convolution layers, fully connected layers (FCs) and uses Softmax to define an image with stochastic values between 0 and 1. As in Figure 2, CNN architecture has the following layers:

- Convolution is very first layer where information is derived from an input image. Convolution preserves the relationship among pixels through the use of tiny squares in data input to acquire image properties.

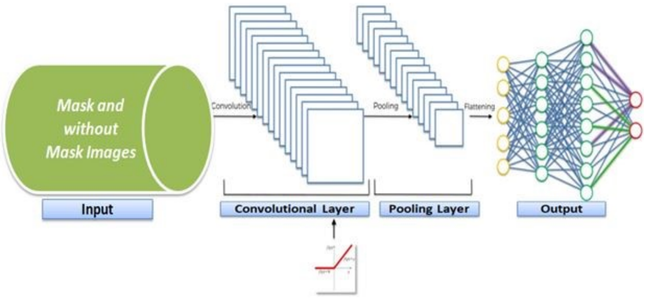


Fig. 2: CNN architecture

- Pooling layers can optimize parameter counts whenever the images are all too large. The Max pooling extracted with the largest factor from the rectified function diagram. It could put on the average pooling even the largest portion. List of all Map Elements feature names as Sum pooling.
- We condensed our matrix into a vector and fed it as a neural network into a fully linked layer, the layer that we label as FC.

In this proposed model, ten convolution layers are implemented using 16, 16, 32, 32, 64, 64, 96, 96, 128 and 128 filters respectively with size 3 x 3 and Relu is used as an activation function which is measured by $f(x) = \max(0,x)$. The model used five maxpooling layer with stride 2 followed by flatten layer. Then four dense layer are constructed using relu activation function in which first, second and third dense have 512, 128 and 64 hidden nodes respectively. Finally fourth dense layer with two hidden nodes are used for the output using softmax activation function. Layered architecture for this classification is shown in Figure 3.

D. Model Pruning with Keras-Surgeon

Pruning is a deep learning technique which helps to make neural nets simpler and more effective. It is a system optimization based strategy that requires eliminating unwanted weight vector values. It results in smoother processing of compact neural networks, which would reduce the computing costs associated in network training.

Keras-surgeon offers simple methods for modifications of trained Keras models. Keras-surgeon suits every model architecture. Any number of layers may be altered within a single network traversal. In this work, ten convolution layers have 16, 16, 32, 32, 64, 64, 96, 96, 128 and 128 filters respectively and feature maps are produced by the each filter. In the specific case of CNNs, the majority of approaches concentrate on removing whole filters and their corresponding feature maps from convolutional layers instead of removing individual weights. The key advantage of this approach is that it does not introduce any sparsity in weight matrices of the network.

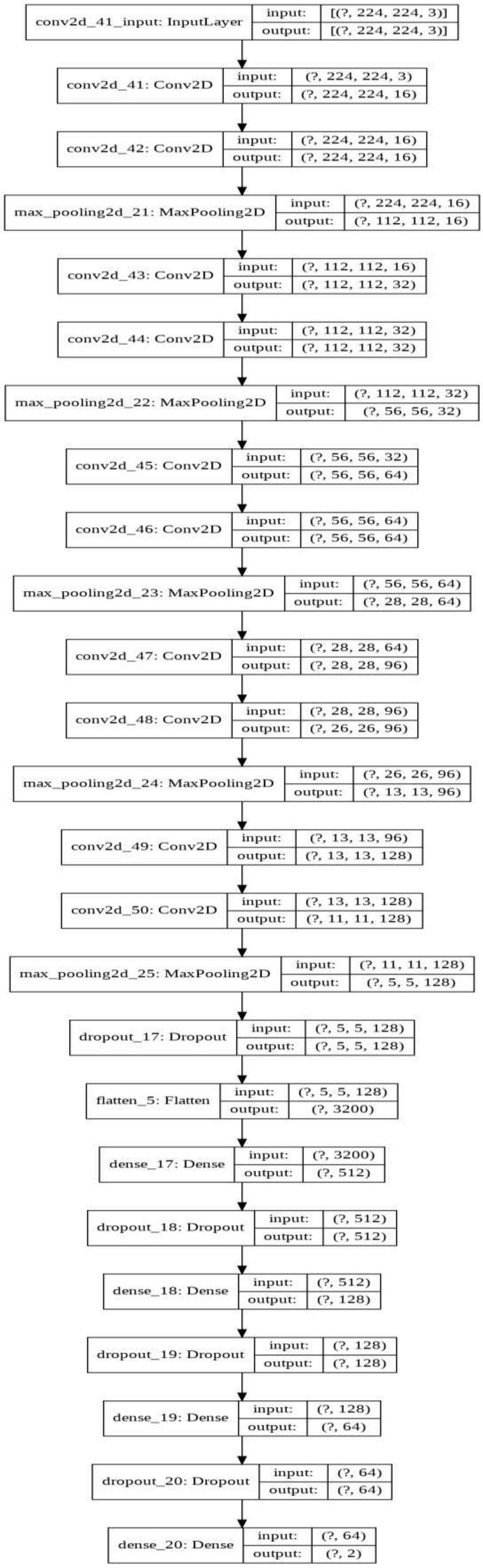


Fig. 3: Proposed CNN Layered Architecture

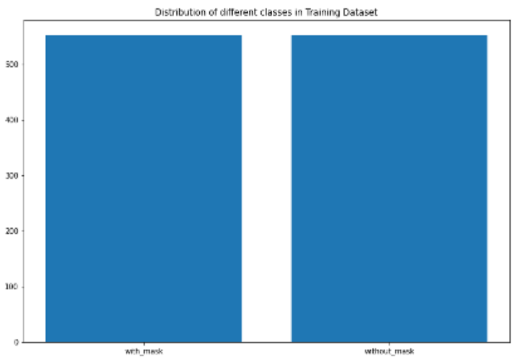
There are various methods of determining filters significance, but the easiest pruning technique is to measure the L1 norm of the weight of filters (take the absolute weight of the filters and add it up) and eliminate those with the lowest L1 standard. While using L1 norm is a simplistic heuristic to rank the importance of filters, we may conclude that pruning convolution filters of low importance away from the network will have a lesser effect than others.

IV. Result And Analysis

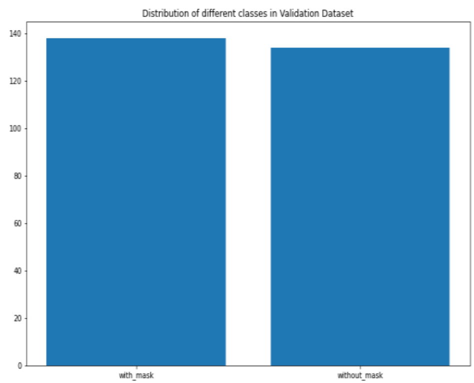
In this work, the model is trained on Google colab using python script with batch size 32 and Adam optimizer for just 15 epochs. The number of training examples utilized in one iteration refers to batch size while epochs refers to number of passes made by the learning algorithm over the entire training dataset. As if batch size of the whole training set of data therefore the number of steps equal epochs numbers. Relu and Softmax are used as a activation functions. ReLU activation function is a non-linear operation of the rectified linear unit with a max (0, x) output. Softmax has the ability to handle multiple classes, so it is used for classifying inputs into multiple classes on output layer. During training, the total parameter is 2,181,778, of which 2,181,778 are parameter trainable and there is no non-trainable.

A. Dataset Description

The dataset includes 1376 images of both masked and without masked real time faces. This dataset does not have a validation set, so it was split into a training set and a validation set for assessment. Training set includes 1104 images (552 mask and 552 without mask) while validation set includes 272 images (138 mask and 134 without mask). Distribution of dataset is plotted in Figure 4a and Figure 4b.



(a) Distribution of Classes in Training Set

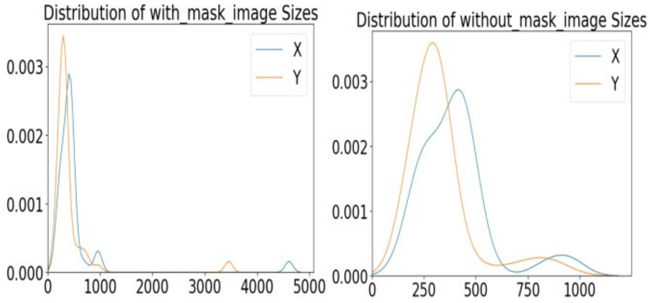


(b) Distribution of Classes in Validation Set

Fig. 4: Dataset Distribution

B. Kernel Density Plot

The kernels are the free parameters of the kernel density estimate which defines the distribution shape placed at each point and the kernel bandwidth determines the kernel size at each point. The Kernel Density Estimation is used to measure a continuous variable's Probability Density. It represents the density of probabilities in a continuous variable at various values. We can also plot a single graph for multiple samples which helps to visualize data more efficiently. Figure 5a and Figure 5b shows the kernel density plot for the dataset.



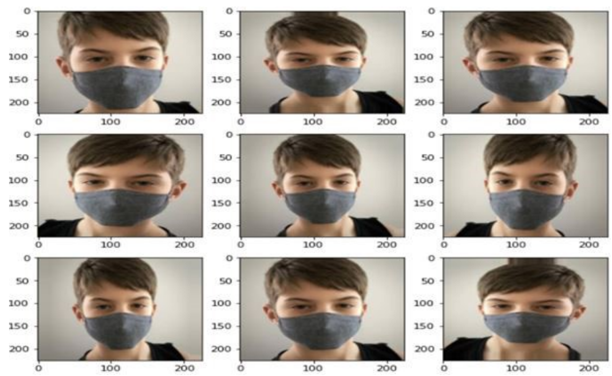
(a) Mask Images (b) Without Mask Images

Fig. 5: Kernel Density Plot

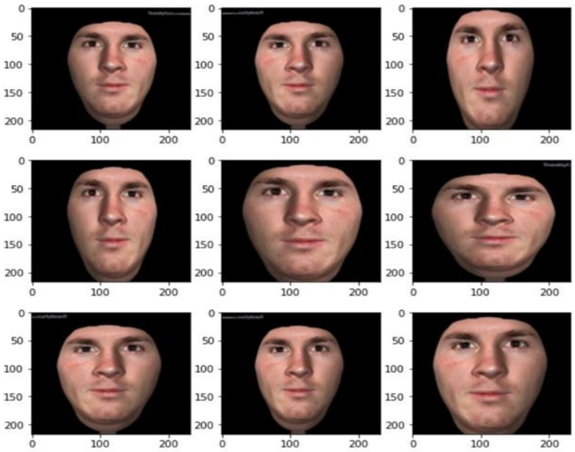
C. Data Preprocessing and Augmentation

Input images for this dataset often come in different sizes and resolutions so they were resized to 224 x 224 x 3 to reduce scale. Modern machine learning models, such as deep neural networks, may have trillions of parameters and require large, frequently inaccessible, labelled training datasets. The technique of expanding artificially labelled training datasets known as data augmentation has quickly become critical to tackling this data scarcity problem. Data augmentation is currently used as a secret sauce in almost every state-of-the-art model for image classification, and is becoming increasingly popular in other modalities such as understanding the natural language. The most commonly data augmentation operations are Zooming, Flipping, Cropping, Shearing, Rotation etc. Data augmentation is used for the proposed work that allows practitioners to increase dramatically the data available for training models without actually collecting new data. The sample augmented images are shown in Figure 6a and Figure 6b and the parameter used for data augmentation for proposed work is below:

- Shear : 0.2
- Rescale : 1./255
- Zoom : 0.2
- Horizontal Flip : True



(a) Mask Images

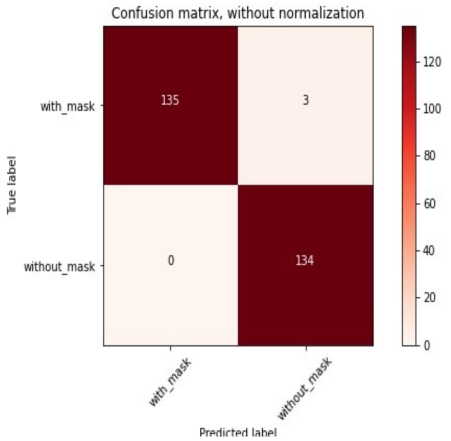


(b) Without Mask Images
Fig. 6: Random Transformation Using Data Augmentation

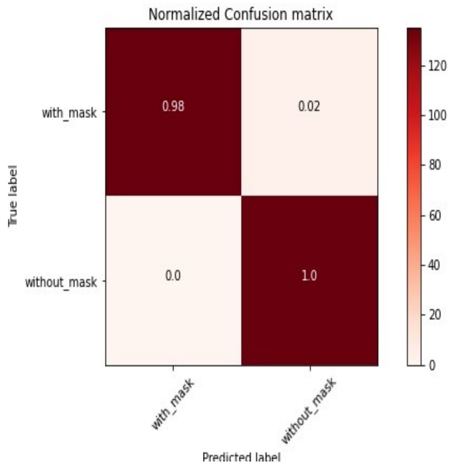
D. *Accuracy, Loss, Confusion matrix, Precision, Recall and F1 Score*

Accuracy curve, loss curve, confusion matrix, precision, recall, f1 score and specificity-based analysis are performed for this work. Validation set recorded the following results as shown in Figure 7a and Figure 7b using confusion matrix.

- True Negatives (TN): 135
- False Positives (FP): 3
- False Negatives (FN): 0
- True Positives ((TP): 134



(a) Without Normalization



(b) With Normalization
Fig. 7: Confusion matrix

The mathematics behind Accuracy, loss, precision, recall, f1 score and specificity are shown by the below Equation 1, 2, 3, 4, 5 and 6 respectively
.

$$\text{Accuracy} = (TP + TN)/ (FN + TP + TN + FP) \quad (1)$$

$$\text{logloss} = \frac{-1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} * \log(p)_{ij} \quad (2)$$

$$\text{Precision} = TP/(TP + FP) \quad (3)$$

$$\text{Recall} = TP/ (FN + TP) \quad (4)$$

$$\text{f1 Score} = 2 \text{ x } (\text{Precision Recall}) /(\text{Precision} + \text{Recall}) \quad (5)$$

$$\text{specificity} = T N/(T N + F P) \quad (6)$$

Based on the above equation, we documented the following results and classification report for the individual classes is shown in Table 1.1.

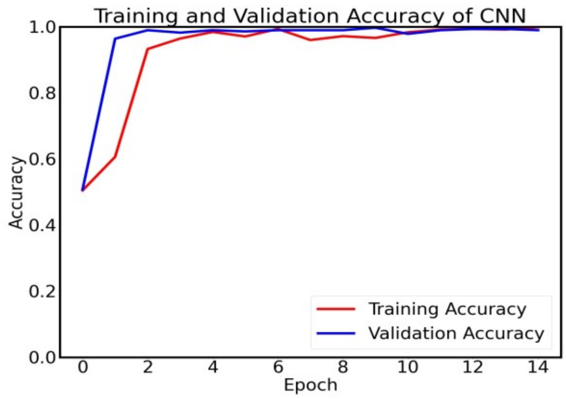
- Accuracy: 98.90 percent
- Loss : 0.12
- Precision: 97.81 percent
- Recall: 100.00 percent
- F1 Score: 98.89 percent
- Specificity: 98 percent

As shown in Table I, with mask class recorded precision, recall and f1 score are recorded 100, 98 and 99 percent respectively while without mask classes recorded 98, 100 and 99 percent respectively which are quite good and comparable with other models.

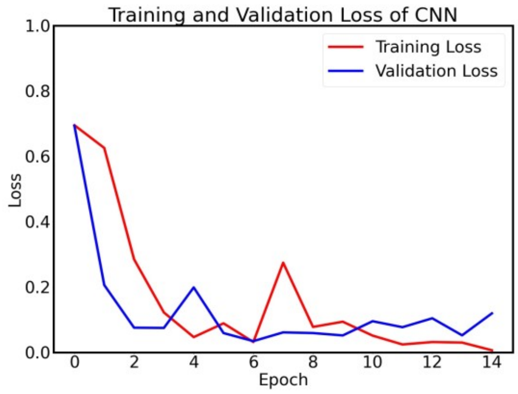
TABLE I: Precision, Recall, f1 Score (In Percent) and Support

Class and Avg Type	Precision	Recall	f1 score	Support
with mask	100	98	99	138
without mask	98	100	99	134
macro avg	99	99	99	272
weighted avg	99	99	99	272

The proposed work recorded the accuracy of the model and its loss curve per epoch as shown in Figure 8a and Figure 8b. The model achieved the training accuracy 100.00 percent with 0.00 loss while validation accuracy 98.90 percent with loss 0.12 just in 15 epochs.



(a) Accuracy Curve



(b) Loss Curve

Fig. 8: Accuracy and Loss Curve

As the Figure 9 shows some random input images and also represents faces marked within a bounding rectangle with their respective pixel-level image followed by proposed model prediction.

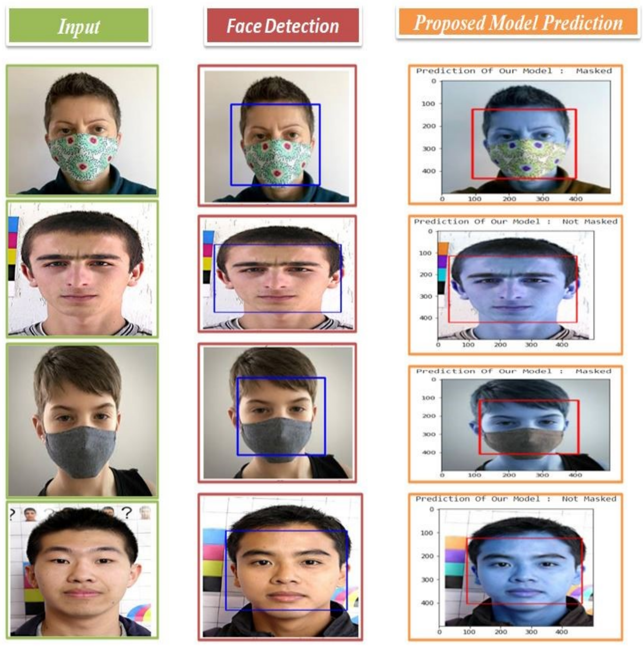


Fig. 9: Sample Images, Face Detection and Model Prediction

Although proposed work recorded 100.00 percent training accuracy with 0.00 loss and 98.90 validation accuracy percent with loss 0.12 but number of experiments can be increased with advanced convolution neural network models on larger dataset for better analysis.

V. Conclusion

Along with depending on medical interventions, it is crucial to alleviate individuals' socio-economic conditions to avoid infection with Covid-19 and to prevent transmission. One of the preventive measures that can restrict the spread of such respiratory viral diseases like COVID-19 is wearing a medical mask. The proposed model recorded 100.00 percent training accuracy with 0.00 loss and 98.90 percent validation accuracy with loss 0.12 in just 15 epochs which shows the goodness of the work. The proposed face mask detection study contains a preventive measure of covid-19 which shows the great significance in the present scenario.

References

[1] Y.-R. Guo, Q.-D. Cao, Z.-S. Hong, Y.-Y. Tan, S.-D. Chen, H.-J. Jin, K.S.Tan, D.-Y. Wang, and Y. Yan, "The origin, transmission and clinical therapies on coronavirus disease 2019 (covid-19) outbreak–an update on the status," Military Medical Research, vol. 7, no. 1, pp. 1–10, 2020.

[2] K. Kumar, "Evs-dk: Event video skimming using deep keyframe," Journal of Visual Communication and Image Representation, vol. 58, pp. 345–352, 2019.

[3] K. Kumar and D. D. Shrimankar, "F-des: Fast and deep event summarization," IEEE Transactions on Multimedia, vol. 20, no. 2, pp. 323–334, 2017.

[4] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic," Measurement, p. 108288, 2020.

[5] M. S. Ejaz, M. R. Islam, M. Sifatullah, and A. Sarker, "Implementation of principal component analysis on masked and non-masked face recog-nition," in 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). IEEE, 2019, pp. 1–5.

[6] M. Grassi and M. Faundez-Zanuy, "Face recognition with facial mask application and neural networks," in International Work-Conference on Artificial Neural Networks. Springer, 2007, pp. 709–716.

[7] S. Li, X. Ning, L. Yu, L. Zhang, X. Dong, Y. Shi, and W. He, "Multi-angle head pose classification when wearing the mask for face recogni-tion under the covid-19 coronavirus epidemic," in 2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS). IEEE, 2020, pp. 1–5.

[8] R. Ramachandra, S. Venkatesh, K. B. Raja, S. Bhattacharjee, P. Wasnik, S.Marcel, and C. Busch, "Custom silicone face masks: Vulnerability of commercial face recognition systems & presentation attack detection," in 2019 7th International Workshop on Biometrics and Forensics (IWBF). IEEE, 2019, pp. 1–6.

[9] A. Mahore and M. Tripathi, "Detection of 3d mask in 2d face recognition system using dwt and lbp," in 2018 IEEE 3rd International Conference on Communication and Information Systems (ICCIS). IEEE, 2018, pp. 18–22.

[10] T. Meenpal, A. Balakrishnan, and A. Verma, "Facial mask detection using semantic segmentation," in 2019 4th International Conference on Computing, Communications and Security (ICCCS).

[11] L. Li, Z. Xia, X. Jiang, Y. Ma, F. Roli, and X. Feng, "3d face mask presentation attack detection based on intrinsic image analysis," IET Biometrics, vol. 9, no. 3, pp. 100–108, 2020.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1. IEEE, 2001, pp. I–I.