The Computational Approach for Human Face Mask Detection

Mr. Prateek Dutta[1], Miss Riddhi Mandal[2]

[1,2]B.Tech Artificial Intelligence, G.H. Raisoni College of Engineering, Maharashtra, India

---

Abstract: *The whole world is suffering from SARS-CoV-2 and its very crucial time for all living beings and it's been impacted in almost every development sector. Since, December 2019, the coronavirus spreading throughout the world. The only prevention to this virus is to use of face mask and sanitizer regularly. Government has mandatory the use of face mask in public area like, shopping complex, offices, schools, colleges etc. But unconventionally few people don't use to take the precautions which needed to be monitor. There are large number of population and physically monitoring of such mischief is very difficult. In order to prevent it, this paper focus of intelligent approach for detection of face mask among human being in public area. Deep Learning algorithm and architecture has been used over it. Convolution Neural Network (CNN) model and ResNet 50 as architectural model has been well trained over the image dataset and achieved a measurable accuracy. Many people lose their life due to the SARS-CoV-2 and now further, it can't be given any chance of excuse as Face Mask detection using CCTV camera has been allocated in several public area.*

1. INTRODUCTION

The practice of wearing masks in public is growing as a result of the COVID-19 corona virus epidemic worldwide. Before Covid-19, people used to wear a mask to protect their health from air pollution. While some people are self-conscious about their appearance, they hide their feelings in public by hiding their faces. Scientists have confirmed that wearing face masks is effective in preventing the transmission of COVID-19. COVID19 (also known as the corona virus) is the latest pandemic that has plagued human life in the last century. By 2020, the rapid spread of COVID-19 has forced the World Health Organization to declare COVID-19 a global epidemic. According to WHO, more than five million cases were infected by COVID-19 in less than 6 months across 188 countries. The virus spreads through close contact and in crowded and overcrowded areas [1].

Since, WHO announce COVID-19 as a pandemic, various efforts have been made to prevent the spread of virus over the community. Use of facial mask, frequently use of sanitizer and maintain social distancing in public area can only solution to prevent the virus into community spread. Further, government has made the rules and regulation for it, but unconventionally some people didn't follow it and it needed to be monitored. CCTV camera has been installed in several public place to monitor and regular surveillance. Capturing the facial recognition and detection of object into it is quite difficult task where Artificial Intelligence (AI) plays important role. Artificial Intelligence plays a vital role in current modernizing world and considered to be a transformation factor. In the medical sector, Artificial Intelligence have empowered the key strength and technology. AI has various segments as Machine learning, Deep Learning and etc which has been classifies in below figure: 1. Deep learning is a subset of machine learning that deals with algorithms that promote the structure and function of the brain called artificial neural networks [2].

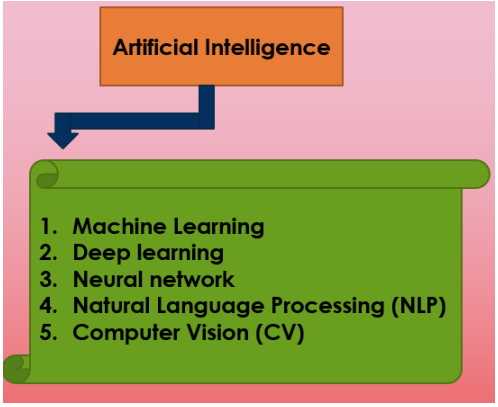

Figure: 1, (Sub-parts of AI)

Here we introduce a mask face detection model that is based on computer vision and deep learning. The proposed model can be integrated with surveillance cameras to impede the COVID-19 transmission by allowing the detection of people who are wearing masks not wearing face masks.

2. Related Works

Building an AI based face detection model is necessary for this COVID 19 situation. Generic mask detection is one of the early projects for mask detection. In this paper, Viola-Jones proposed the boosted-based cascade architecture with simple yet fast Hear features [3]. Manually monitoring of people whether they are wearing mask on face in public and crowded areas, from research paper[4] they had built a real time automated model integrated with surveillance cameras in public which will helps to monitor people that they were wearing mask or they were maintaining social distancing in public areas and report to the respective authorities using computer vision and implementing raspberry pi4. In [5] research paper, made an essential face detection model using classical and deep machine learning with two main components i.e. feature extraction by using Resnet50 and for classification processing of masks using ensemble, support vector machine (SVM) algorithms and decision tree. In this they were using three different datasets Simulated Masked Face Dataset (SMFD), Labelled Faces in the Wild (LFW) and Real-World Masked Face Dataset (RMFD). By using SVM algorithms SMFD, LFW and RMFD datasets have achieved 99.49%, 99.69% and 100% accuracy respectively.

3. Dataset

The dataset considered for this task has been taken from research prospective from online source. The dataset is of Image type containing several images of human faces wearing mask and not wearing mask. The dataset contains classified 3 folder of Test, Train & the validation data. Each contains two sub-folders named as mask and without mask. The dataset illustration with sample image is shown in table:1 and figure: 2.

Figure: 2, (Dataset sample images)

Test	With Mask	483 Images
	Without Mask	509 Images
Train	With Mask	5000 Images
	Without Mask	5000 Images
Validation	With Mask	400 Images
	Without Mask	400 Images

Table: 1, (Dataset overview)

4. Methodology

For the proposed task, we have used the intelligence approach called as Artificial Intelligence (AI). As, AI various sub-parts which is required by specific tasks so, in this paper Convolution Neural Network and architectural model ResNet-50 has been implemented. Activation layer "Relu", along with Max pooling layer with pool size of (4,4) has been

applied over input data. Dropout function has been utilized which is a technique where randomly selected neurons are ignored during training. They are "dropped-out" randomly. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass. While compiling the Neural Network, optimizer Adam, Binary Cross entropy as a loss and accuracy as a metrics has been used, where optimizer plays a vital role as it is a method that used to change the attributes of the neural network such as learning rate to reduce the losses and weights. Further description is mentioned in below figure: 3.

```
input_data = Input(shape=(128, 128, 3))
#Convolution
x = Conv2D(32, (3, 3), activation="relu")(input_data)
#Pooling
x = MaxPooling2D(pool_size = (4, 4), strides=(4, 4))(x)
#Dropout
x = Dropout(0.25)(x)
# 2nd Convolution
x = Conv2D(32, (3, 3), activation="relu")(x)
# 2nd Pooling layer
x = MaxPooling2D(pool_size = (2, 2))(x)
#Dropout
x = Dropout(0.3)(x)
#3rd Convolution
x = Conv2D(32, (3, 3), activation='relu')(x)
#3rd Pooling Layer
x = MaxPooling2D(pool_size=(2, 2))(x)
#Dropout
x = Dropout(0.3)(x)
# Flatten the layer
x = Flatten()(x)
# Fully Connected Layers
x = Dense(128, activation = 'relu')(x)
output = Dense(1, activation = 'sigmoid')(x)
cnn = Model(inputs=input_data, outputs=output)
# Compile the Neural network
cnn.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics = ['accuracy'])
```

```
Model: "model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 128, 128, 3)]     0
_________________________________________________________________
conv2d (Conv2D)              (None, 126, 126, 32)      896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 31, 31, 32)        0
_________________________________________________________________
dropout (Dropout)            (None, 31, 31, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 29, 29, 32)        9248
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 14, 14, 32)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 14, 14, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 12, 12, 32)        9248
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 6, 6, 32)          0
_________________________________________________________________
dropout_2 (Dropout)          (None, 6, 6, 32)          0
_________________________________________________________________
flatten (Flatten)            (None, 1152)              0
_________________________________________________________________
dense (Dense)                (None, 128)               147584
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 129
=================================================================
Total params: 167,105
Trainable params: 167,105
Non-trainable params: 0
_________________________________________________________________
```

Figure: 3, (Model Detailed Description)

In Convolution Neural networks (CNN), while training the model up to 5 epochs where each epoch contains 313 instances the resultant training accuracy of 98.45% and validation accuracy of 99.50%. Whereas, while training the model, the training loss of 4.43% and validation loss of 2.32% was observed. The figure: 4 shows the Loss, Accuracy, val. Accuracy & val. loss of CNN model.

```
Epoch 1/5
313/313 [==============================] - 161s 500ms/step - loss: 0.2861 - accuracy: 0.8745 - val_loss: 0.1025 - val_accurac
y: 0.9775
Epoch 2/5
313/313 [==============================] - 141s 450ms/step - loss: 0.0934 - accuracy: 0.9664 - val_loss: 0.0835 - val_accurac
y: 0.9850
Epoch 3/5
313/313 [==============================] - 140s 446ms/step - loss: 0.0704 - accuracy: 0.9755 - val_loss: 0.0438 - val_accurac
y: 0.9862
Epoch 4/5
313/313 [==============================] - 139s 445ms/step - loss: 0.0482 - accuracy: 0.9825 - val_loss: 0.0580 - val_accurac
y: 0.9800
Epoch 5/5
313/313 [==============================] - 137s 437ms/step - loss: 0.0443 - accuracy: 0.9845 - val_loss: 0.0232 - val_accurac
y: 0.9950
```

Figure: 4, (Loss, Accuracy, val. Accuracy & val. loss of CNN model)

The evaluation of CNN model has been done on the basis of Accuracy, Precision, Recall, F-1 score and support and further confusion matrix has been constructed which is illustrated in figure: 5.

```
31/31 - 5s - loss: 0.0374 - accuracy: 0.9899
              precision    recall  f1-score   support

         0.0       0.99      0.99      0.99       509
         1.0       0.99      0.99      0.99       483

    accuracy                           0.99       992
   macro avg       0.99      0.99      0.99       992
weighted avg       0.99      0.99      0.99       992

[[505   4]
 [  6 477]]
```

Figure: 5, (CNN Model Evaluation)

A residual neural network (ResNet) is a kind of deep transfer learning based on residual learning [6]. ResNet-101, ResNet-50, and ResNet-18 are the versions of ResNet. ResNet-50 is a 50-layers deep convolution layer, starts with convolution layer and end with a fully-connected layer, and in between the layers, it is followed by 16 residual bottleneck blocks with each block containing three layers of convolution layer. On achieving unsatisfactory result from CNN model, architectural model ResNet-50 has been applied, which is considered to be best for image classification task. Activation layer "Relu", and optimizer Adam, Binary Cross entropy as a loss and accuracy as a metrics has been used over compilation of ResNet model which is illustrated in figure: 6.

```
Model: "model_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_3 (InputLayer)         [(None, 128, 128, 3)]     0
_________________________________________________________________
resnet50 (Functional)        (None, 4, 4, 2048)        23587712
_________________________________________________________________
global_max_pooling2d (Global (None, 2048)              0
_________________________________________________________________
batch_normalization (BatchNo (None, 2048)              8192
_________________________________________________________________
dropout_3 (Dropout)          (None, 2048)              0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               262272
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 129
=================================================================
Total params: 23,858,305
Trainable params: 266,497
Non-trainable params: 23,591,808
_________________________________________________________________
```

Figure: 6, (ResNet Model Illustration)

In ResNet-50 architectural model, while training the model up to 5 epochs where each epoch contains 313 instances the resultant training accuracy of 89.66% and validation accuracy of 91.25%. Whereas, while training the model, the training loss of 24.17% and validation loss of 22.82% was observed. The figure: 7 shows the Loss, Accuracy, val. Accuracy & val. loss of ResNet model.

```
Epoch 1/5
313/313 [==============================] - 840s 3s/step - loss: 0.4364 - accuracy: 0.7923 - val_loss: 0.3585 - val_accuracy:
0.8375
Epoch 2/5
313/313 [==============================] - 818s 3s/step - loss: 0.3063 - accuracy: 0.8715 - val_loss: 0.2481 - val_accuracy:
0.8925
Epoch 3/5
313/313 [==============================] - 807s 3s/step - loss: 0.2819 - accuracy: 0.8782 - val_loss: 0.2390 - val_accuracy:
0.8988
Epoch 4/5
313/313 [==============================] - 812s 3s/step - loss: 0.2579 - accuracy: 0.8897 - val_loss: 0.2406 - val_accuracy:
0.9025
Epoch 5/5
313/313 [==============================] - 804s 3s/step - loss: 0.2417 - accuracy: 0.8966 - val_loss: 0.2282 - val_accuracy:
0.9125
```

Figure: 7, (Loss, Accuracy, val. Accuracy & val. loss of ResNet model)

The evaluation of ResNet architectural model has been done on the basis of Accuracy, Precision, Recall, F-1 score and support and further confusion matrix has been constructed which is illustrated in figure: 8.

```
31/31 - 65s - loss: 0.1863 - accuracy: 0.9405
              precision    recall  f1-score   support

         0.0       0.95      0.93      0.94       509
         1.0       0.93      0.95      0.94       483

    accuracy                           0.94       992
   macro avg       0.94      0.94      0.94       992
weighted avg       0.94      0.94      0.94       992

[[474  35]
 [ 24 459]]
```

Figure: 8, (ResNet Model Evaluation)

5. Conclusion

Hence, During the COVID pandemic, many people have lost their lives and many families have lost their near and dear once. Its now our responsibility to stop the evolutionary pandemic with modern and equipped technologies. With the help of this face mask detection technology, we can easily handle the people who are knowingly or unknowingly unaware of it. Throughout the task, Convolution Neural Network (CNN) and architectural model ResNet-50 has been implemented. On comparing both the model (CNN & ResNet-50), its been concluded that CNN model performed well in such object identification task.

6. Conflict of Interest

Mr. Prateek Dutta and Miss Riddhi Mandal, the author of the research work entitled "The Computational Approach for Human Face Mask Detection" states that there is no conflict in the proposed work. All the figures and table has been constructed by author while the observation has been performed technically with the help of Python programming. No living being has been impacted with the research work. This Research work states for the beneficial of human beings during the COVID pandemic.

References:

[1] "WHO Coronavirus Disease (COVID-19) Dashboard." https://covid19.who.int/ .

[2] Prateek Dutta. "A Deep Learning approach for Animal Breed Classification-Sheep". International Journal for Research in Applied Science & Engineering Technology (IJRASET). Volume 9 Issue V May 2021. ISSN: 2321-9653. https://doi.org/10.22214/ijraset.2021.34050.

[3] P. Viola and M. Jones, ''Rapid object detection using a boosted cascade of simple features''. ACCEPTED CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001. https://web.iitd.ac.in/~sumeet/viola-cvpr-01.pdf

[4] S. Yadav, "Deep learning based safe social distancing and facemask detection in public areas for COVID-19 safety guidelines adherence", International Journal for Research in Applied Science & Engineering Technology (IJRASET). July 2020. https://doi.org/10.22214/ijraset.2020.30560

[5] M Loey, G Manogaran, MHN Taha and NEM. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic". 28 July, 2020. https://doi.org/10.1016/j.measurement.2020.108288.

[6] He K., Zhang X., Ren S., Sun J. 2016 EEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016. Deep residual learning for image recognition; pp. 770–778.