International Journal of Information System Modeling and Design
Volume 13 • Issue 7

EfficientNet-B0 Model for Face Mask Detection Based on Social Information Retrieval

Moolchand Sharma, Maharaja Agrasen Institute of Technology, India*
https://orcid.org/0000-0001-6198-5999

Harsh Gunwant, Maharaja Agrasen Institute of Technology, India
Pranay Saggar, GTBIT, India
Luv Gupta, Sharda University, India
Deepak Gupta, Maharaja Agrasen Institute of Technology, India
https://orcid.org/0000-0002-3019-7161

ABSTRACT

The world was introduced to the term coronavirus at the end of 2019, following which everyone was thrown into stress and anxiety. The pandemic has been a complete disaster, wreaking devastation and resulting in a significant loss of human life throughout the world. The governments of various countries have issued guidelines and protocols to be followed for stopping the surge in cases (i.e., wearing masks). Amidst all this chaos, the only weapon is technology. So, the detection of face masks is important. The authors utilized a dataset that included images of individuals in society wearing and not wearing masks. They gathered the information required to train a model by using deep networks like EfficientNetB0, MobileNetV2, ResNet50, and InceptionV3. With EfficientNet-B0, they have been able to achieve an accuracy of 99.70% on a two-class classification issue. These methods make face mask detection easier and help in knowledge discovery. These technological breakthroughs may aid in information retrieval as well as help society and guarantee that such a healthcare disaster does not occur again.

KEYWORDS

COVID-19, Face Mask, Healthcare, Information Retrieval, Knowledge Discovery

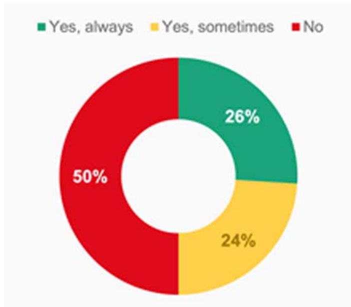
1. INTRODUCTION

Covid-19 presents an unprecedented and unparalleled challenge to the food systems, public health, and disrupted work by creating unemployment. As the number of cases has increased, the healthcare system has been overwhelmed and has largely failed. WHO has announced this as a pandemic (Denison, 2004). One of the prevention techniques is wearing a mask? It is for your safety as well as for those around you. The virus that causes COVID-19 can be communicated by sneezing, coughing, or even speaking at close range, therefore wearing a mask is necessary. However, if people aren't following the protocol and not wearing masks then that increases the risk of a community spreading which can

DOI: 10.4018/IJISMD.313444 *Corresponding Author

become deathly. Many nations have begun the vaccine procedure, but not every country in the globe has access to it. So, until the virus is eradicated, wearing masks daily is a must-do to assist prevent the spread of infection.

Figure 1. A poll was conducted by Morning Consult for people who wear masks.



The figure 1 Poll shows that more than 50% of the population are reluctant to wear masks which is a serious concern and poses a huge threat. Thus, it is very crucial for people to wear masks in public and for the concerned authorities to ensure that people follow the guidelines and the protocols for their safety. Figure 2 below shows the person wearing a mask (A) and the person not wearing a Mask (B).

The primary cause of the infection being intensified was the attention to detail on the part of the population, as well as their lack of awareness. The world is flooded with ordinary individuals, who are either in offices or other residences without wearing any face masks. It is almost impossible to cover everything all the time and on top of that, quite time-consuming. This study is mostly aimed at finding a solution to this problem and helping individuals to protect themselves. This concept is especially relevant in COVID-19, as it is imperative that we preserve ourselves from other people. In addition to offering an effective means to reduce airborne virus infections, the masks will effectively disrupt these infections so that airborne viruses cannot reach a human being's respiratory system, making it an inexpensive way to minimize mortality and respiratory infection disorders. Wearing a facemask can prevent a global pandemic. To avoid this, it is essential to develop an automatic detection for wearing a facemask that will protect each player and prevent the pandemic.

With the recent advancement in deep learning that incorporates computer vision, we are seeing a multitude of breakthroughs in various fields of technology. Deep neural networks' innate capacity to extract information from input photos has recently led to their astounding success in automatic image analysis. Deep learning methods aid in picture classification, quantification, and pattern recognition have identified and emphasized numerous uses of deep neural networks in image processing(Deng et al., 2009; Howard et al., 2017). Convolutional neural networks (CNNs) are the most researched and widely utilized among the several forms of deep neural networks in image processing(Du et al., 2019; Brown et al., 2018; Shin et al., 2016). Although CNN's have a vast number of parameters, large size annotated datasets are required. Large-scale datasets are often difficult to get by. As a result, researchers are combining CNNs with transfer learning to address this problem. The image representations learned with CNNs on large-scale datasets are efficiently and successfully transferred to other tasks that need small-scale datasets in transfer learning. Given the significant time and computational resources required to develop neural network models from the root of these problems, as well as the significant improvements in a skill that they provide on related problems, using pre-trained models as the starting point in natural language processing and computer vision tasks is an accepted and popular approach in deep learning(Pan & Yang, 2010).

The lack of a big dataset for training was a problem for the authors, so they devised a workaround by using transfer learning and made a custom dataset to achieve this task.

Figure 2. (A) A person wearing a mask and (B) A person not wearing a mask

The key contribution of our proposed methodology is as follows:

(i) By looking at the facial features, the system can detect whether or not a person is wearing a mask.
(ii) On the dataset, multiple pre-trained models (EfficientNetB0, MobileNetV2, ResNet50, and InceptionV3) are trained individually and produce results for new images.
(iii)We analyzed the performance and accuracy of the four models, and EfficientNet-B0 came out on top with a 99.70 percent accuracy.

The primary concern of this proposed paper is to compare the various pre-trained deep networks like EfficientNetB0, ResNet50, MobileNetV2, and InceptionV3 and choose the best model for future applications that extracts the best features and has the highest accuracy. EfficientNet, originally published by Tan and Le in 2019, is one of the most efficient models for inference, using the fewest FLOPS and achieving State-of-the-Art accuracy on both typical image classification transfer learning tasks and ImageNet. MobileNets are low-latency, low-power models that may be adjusted to fit the needs of different use cases. Similar to how various popular large-scale models are used, they can be built upon for classification, segmentation, detection, and embedding. ResNet, or Residual Networks, is a type of neural network that is used to assist a variety of computer vision applications. We were able to effectively train complicated and incredibly deep neural networks with 140+ layers using the ResNet, which is a major advance. One of the numerous pre-trained image classifiers is InceptionV3. The ImageNet 2012 dataset was used to train it.

The paper is arranged as follows: Section 2 discusses important ideas related to the deployment of previously introduced systems and the use of residual neural networks for picture categorization. Section 3 describes the suggested approach as well as a timeline of the system's development. In Section 4, the system's implementation and outcomes are explained and discussed. Finally, the rest of the part contains the paper and concluding remarks.

2. LITERATURE REVIEW

Many public service providers will need clients to wear masks correctly in the future to receive general services. Therefore, face mask detection has become a crucial task to help the global society. The usage of face masks to postpone the transmission of covid-19 has acquired favour among the world's population(Matuschek et al., 2020). Many scholars have suggested numerous expert systems for image processing in the past. Fast R-CNN is a technique that expands on prior work to categorize object suggestions efficiently using deep convolutional networks. Fast R-CNN leverages numerous advancements over prior work to boost training and testing efficiency while simultaneously enhancing detection accuracy(Girshick,2015). A strategy for video frame generators employing an Encoder-ConvLSTM combination in the paper(Mukherjee et al.,2019). A number of remarkable articles in deep networks, recognizing plant illnesses and detecting fights in hockey videos were also introduced(Mukherjee et al.,2017, Mukherjee et al.,2017). An Area Proposal Network (RPN) that shares full-image convolutional features with the detection network and allows for practically cost-free region suggestions. The RPN is trained from start to finish to create high-quality region suggestions, which Fast R-CNN uses for detection(Ren et al.,2017) .

A deep convolutional neural network architecture was suggested. It was responsible for creating a new benchmark for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014. The architecture's key distinguishing feature is the better usage of computational resources inside the network(Szegedy et al.,2015). Later it was suggested a poorly supervised learning strategy that can learn a deep convolutional neural network (DCNN) that might be used in industrial applications(Sun et al.,2018). People must wear masks as the number of instances of covid increases, and technologists must retain and maintain a watch on this big data. A study that focuses on Big Data Processing Concepts and Techniques since processing a large quantity of data in a single system is extremely difficult(Suvarnamukhi & Seshashayee, 2018).

Training Inception networks with residual connections significantly speeds up the training process. There's additional evidence that residual Inception networks outperform similarly priced Inception networks with no residual connections by a small margin(Szegedy et al.,2016). We choose to take both the ResNet50 and the InceptionV3.

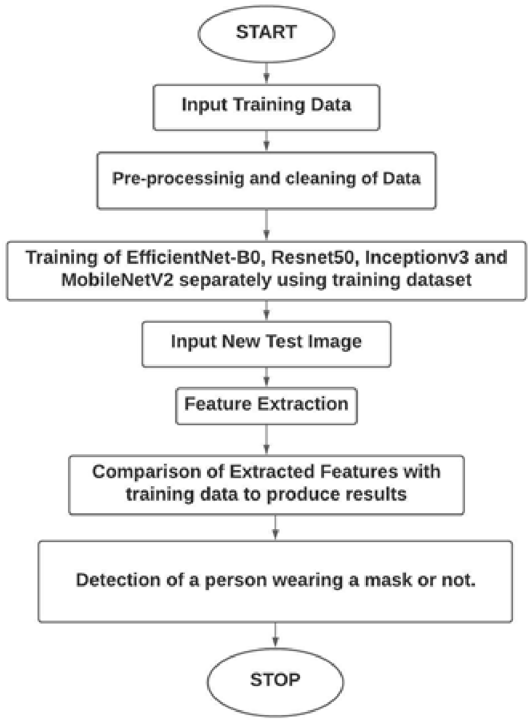
All of these developments in Convolutional Neural Networks, Pattern Recognition, and Image Processing have to be applied to the current COVID-19 situation. A method was suggested that takes the use of Retina Face Mask, a high-accuracy, and efficient face mask detector(Jiang et al.,2020). They also looked into developing Retina Face Mask for embedded or mobile devices using MobileNet, a light-weighted neural network. To solve the challenge of masked face recognition. A study suggested a Convolutional Neural Network (CNN) that was evaluated using three well-known image identification techniques: Principal Component Analysis (PCA), Local Binary Patterns Histograms (LBPH), and K–Nearest Neighbour (KNN). They achieved the best recognition accuracy of 98.3 percent with the suggested CNN (Kamencay et al.,2017). Mask identification and levying penalties on those who do not wear them require high accuracy. The symptoms related to the coronavirus and its distinction between flu and this virus are mentioned. Masks should be used as part of a broader plan to stop the spread of disease and save lives. Some of the well-published articles have all the necessary information and queries regarding masks in the context of COVID-19(Singhal, 2020; Cheng et al.,2020).

In recent years, face mask detection and identification has become one of the most important study fields. The major issue here is to correctly detect the face in the image and then determine whether or not it is covered by a mask.

3. PROPOSED METHODOLOGY

This section has been divided into five major subsections including the description of EfficientNetB0, description of ResNet50, description of InceptionV3, description of MobileNetV2, and the description of the dataset used. Figure 3 below depicts the flow method employed in the article, and further, we will examine the structure of the several expert systems which were used for the simulation.

Figure 3. The Flow Diagram of the pre-trained deep network Used



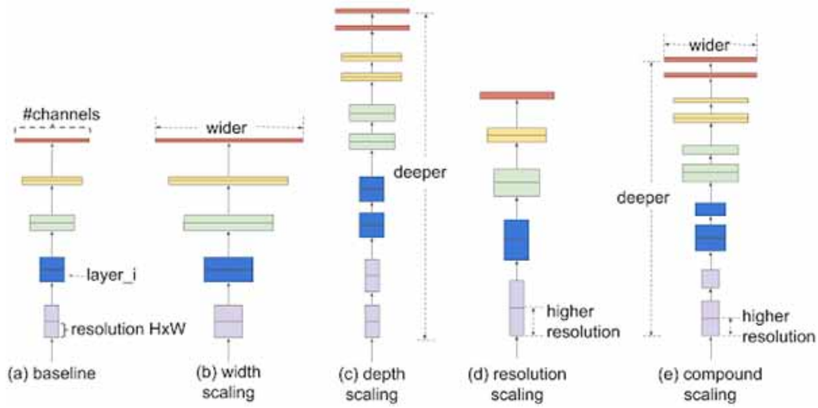
3.1 EfficientNetB0

EfficientNet was first introduced by engineers from the Google brain team(Tan & Le,2020). Their main idea was to scale up the Deep Neural Networks but it also introduced a new category known as EfficientNets. There are eight models of EfficientNets (EfficientNet-B0 - EfficientNet-B7), the baseline version is EfficientNet-B0, the rest EfficientNet-B1 - EfficientNet-B7 are scaled versions.

The conventional method to increase the accuracy of a model was to first create a base model with a fixed cost and then scale up the base model when you have more resources. Some of the methods for scaling up are increasing or decreasing the depth or width of the model. One can also scale up the resolution of the image to increase the accuracy as shown in figure 4. Instead of scaling the width, depth, or resolution at random, a fresh way of evenly scaling the dimensions with a set of scaling coefficients was introduced. With this innovative strategy, a new family of models known as EfficientNets was established, which boosted accuracy and was ten times more efficient than prior models while being smaller and quicker.

Individually scaling model dimensions enhanced performance, however, it was shown that balancing all network dimensions (width, depth, and picture resolution) against available resources enhanced model performance the most(Tan & Le,2020).

Figure 4. Different scaling methods. (b),(c),(d) are the conventional methods with single-dimensional scaling.



The success of model scaling is also determined by the baseline network. EfficientNets basic architecture is EfficientNet-B0. It was created by applying neural architecture search to the AutoML MNAS framework, which enhanced both efficiency and accuracy. Mobile inverted bottleneck convolution is used in the resultant design. The Architecture of EfficientNetB0 is shown in figure 5.

Figure 5. Architecture of EfficientNet-B0

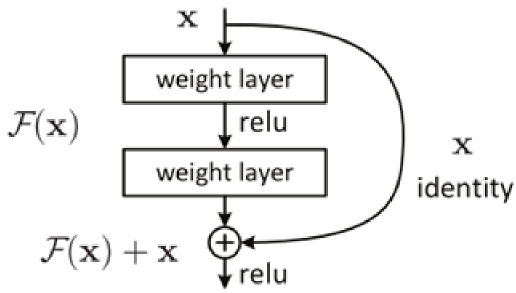
Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

3.2 ResNet50

ResNet was introduced in 2016. On the tasks of ImageNet detection and ImageNet localization which received first place(He et al., 2016).

ResNet50 uses the method of Residual learning. Many layers of neurons are layered and trained in Deep Convolutional Neural Networks in general. After each layer, the network learns features. However, with residual learning, the learning that takes place on the leftovers is shown below I figure 6. The difference in features learned from that layer's input is called residual. ResNet employs "identity shortcut connections," which connect the Nth layer's input to the (N+X)th layer immediately. It is less difficult to train than traditional Deep Convolutional Neural Networks. It also solves the problem of degrading accuracy.

Figure 6. Residual Identity Mapping



Previously the model used to skip two layers as shown in fig.6; but for ResNet50, they skipped three layers, and also 1x1 convolution layers were added which can be seen in the ResNet50 Architecture in figure 7 below:

Figure 7. Architecture of ResNet50

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times23$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

3.3 InceptionV3

One of the numerous pre-trained image classifiers is InceptionV3. The ImageNet 2012 dataset, which comprises roughly 14 million pictures and 1000 classes, was used to train it. At a cost of 5.6 billion multiply-adds per inference and fewer than 25 million parameters, InceptionV3 achieves a 21.2 percent top-1 and a 5.6 percent top-5 error rate.

In the InceptionV3 model, there are 48 layers. Some calculations, such as 1x1, 3x3, 5x5 convolutions, and max-pooling, are chained together(Szegedy et al.,2016). Inception Layer is the name given to this concatenation. The 5x5 convolutions are replaced by two continuous 3x3 convolutions in this approach. The Architecture of InceptionV3 is shown below in fig.8.

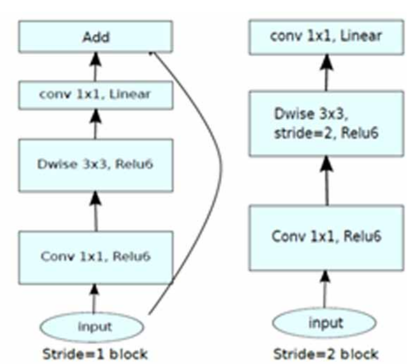
Figure 8. The architecture of InceptionV3

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

3.4 MobileNetV2

MobileNetV2 was introduced in 2018. For mobile devices, it’s a Convolution Neural Network Architecture (Sandler et al.,2018). In MobileNetV1, “Depth wise Separable Convolution” was introduced. It minimizes the model’s complexity, cost, and size, making it acceptable for small devices or devices with minimal processing capacity. An inverted residual structure module was introduced in MobileNetV2 which was better than the MobileNetV1. Non-linearities in thin layers are also taken off in MobileNetV2. For object detection and semantic segmentation, state-of-the-art performances are achieved when MobileNetV2 is used for feature extraction.

Figure 9. MobileNetV2



The block with stride 1 is the residual block, whereas the block with stride 2 is utilized for shrinking in MobileNetV2 as shown in fig.9.

Figure 10. The architecture of MobileNetV2

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

The architecture of MobileNetV2 is shown in fig.10; where t is the expansion factor, n is the repeating number, c is the number of output channels, and s is the stride of each sequence's initial layer 33% kernel are used in spatial convolution.

Hyperparameter optimization was used to improve the accuracy and performance of the above-mentioned models by first comparing the accuracy of the models and then tuning hyperparameters within them to select the appropriate set of hyperparameters for the model, which was very useful because there was a significant difference in the model accuracy and prediction before and after hyperparameters were tuned.

3.5 Dataset

The dataset (https://www.kaggle.com/omkargurav/face-mask-dataset) that was used for testing comprises 7553 photos, 3725 of which are of people wearing masks and 3828 are of persons who are not wearing a mask also the people not wearing a proper mask would be considered as "not wearing a mask" because of their similar features. The dataset is entirely open-source and it is freely accessible on Kaggle. For increasing the accuracies of our models, we have performed image augmentation on the dataset. Without actually gathering additional data, this improves the size of our dataset and the diversity of data accessible for training models. Padding, horizontal flipping, and cropping are standard data augmentation techniques used to train massive neural networks. Some of the images from the dataset can be seen in figure 11.

Figure 11. Samples from Data including faces with mask and without a mask.

3.6 Experimental Setup

The entire system was built on mobile computing infrastructure. A computer having an 8th generation Intel i5 processor clocked at 2.1 GHz with 3MB of cache memory combined with 8 GB of RAM and a 1 TB hard drive, running on the Microsoft Windows operating system (ver. 20H2, 64 bit). Included with the software configuration was Google Collaboratory with Python 3.2 with TensorFlow, NumPy, Scikit-learn, and Pandas.

4. RESULTS AND DISCUSSION

4.1 Evaluation Parameters

When doing classifications in Machine Learning, we utilize specialized measures such as F1 score, precision, and recall since accuracy is insufficient. Machine learning has a distinctive viewpoint in this sector, and in order to persuade the field that AI can be applied, we need a good assessment approach to confirm our results.

Positive predictive value(PPV): It is defined as the total of True Positives and False Positives, plus a ratio of True Positives, as shown by equation (1). It's a metric for determining how important a positive result is. It informs us how many of the forecasts were correct out of all of them.

$$\text{PPV} = \frac{True\,Positives\left(TP\right)}{True\,Positives\left(TP\right) + False\,Positives\left(FP\right)} \tag{1}$$

Sensitivity is a measure of the percent of positive real-world cases that were predicted to be positive. The recall is another name for sensitivity. The sensitivity score reveals what proportion of actual face masks were properly predicted by the model. For example, if the test set had 100 photos of individuals wearing masks, the sensitivity score for the Face mask class reflects what proportion of those 100 persons were properly predicted by our model (EfficientNet-B0). It is a mathematical ratio of True Positives to the sum of True Positives + False Negatives. Sensitivity is described in Equation (2).

$$\text{Sensitivity} = \frac{True\,Positives\left(TP\right)}{True\,Positives\left(TP\right) + False\,Negatives\left(FN\right)} \tag{2}$$

4.2 Results

The system can recognize partly obscured faces with a mask, hand, or hair with efficiency and effectiveness. To distinguish between a face covered by a hand and an annotated mask, it considers the degree of occlusion in four regions: the chin, eye, nose, and mouth. As a result, the model will only regard a mask that completely covers the face, including the nose and chin, as "with a mask". Table 1 is the Evaluation metrics for the pre-trained deep networks as shown below:

Table 1. Top-1 Accuracy, Sensitivity, and PPV of the Architectures

Architecture	Top-1 Accuracy(%)	Sensitivity(%)	PPV(%)
EfficientNet-B0	99.70	97.77	96.54
MobileNetV2	91.86	93.69	90.28
InceptionV3	84.38	91.53	93.21
ResNet50	84.45	84.36	95.82

The evaluation and comparison for the deep neural networks are accuracies, positive predictive value (PPV), and sensitivity for two types of faces were tested on the Face mask dataset. The suggested proposal's metrics models are shown in Table 1.The final-test accuracy is at 91.99% for MobileNetV2, 99.70% for EfficientNet-B0,84.38% for InceptionV3 and 84.45% for ResNet50 as shown below in figure 12, 13, 14 and figure 15. This demonstrates how the suggested models' design and architecture fulfil the task, data, and operational requirements of the Face Mask detection. The higher the true positive and true negative more precisely and effectively it is detecting masks. Lower False positives and false negatives indicate its failure in detection. EfficientNet-B0 turned out to be the best Algorithm among all because it achieved the highest score of Top-1 accuracy, Sensitivity, and PPV.

Figure 12. Training and validation accuracy(A), training and validation loss(B) for MobileNetV

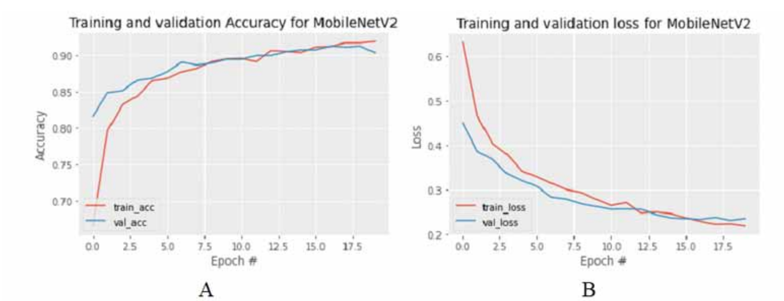


Figure 13. Training and validation accuracy(A), training and validation loss(B) for ResNet50

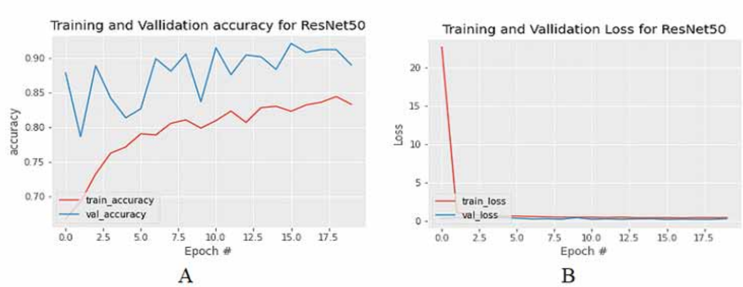


Figure 14. Training and validation accuracy(A), training and validation loss(B) for InceptionV3

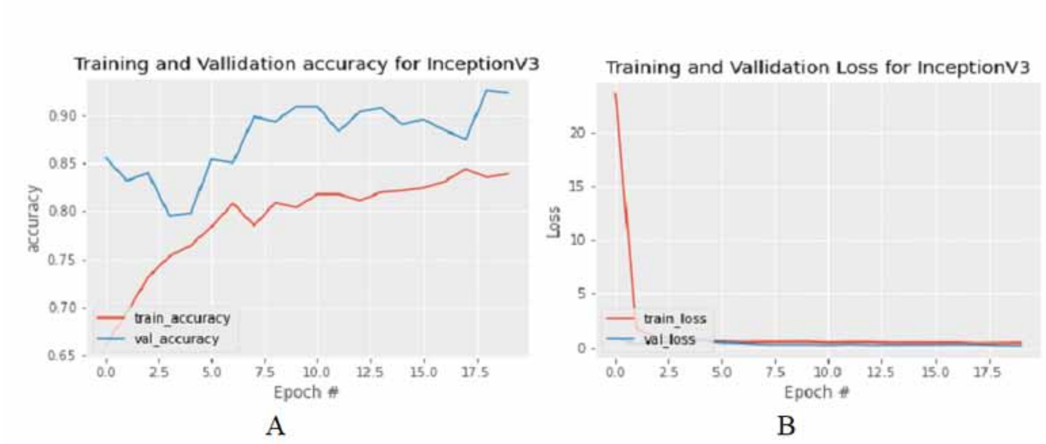
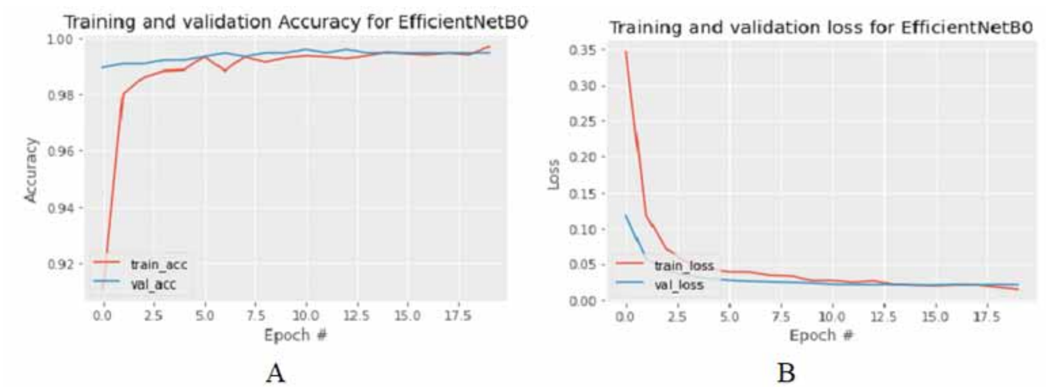


Figure 15. Training and validation accuracy(A), training and validation loss(B) for EfficientNet-B0



According to the findings above, different pre-trained CNN models have various feature extraction characteristics. fulfil the comparative benefit of the four deep neural network models in this investigation is highly encouraging.

5. LIMITATIONS

The EfficientNet Model begins at B0 and ends at B7. However, we chose the B0 model for our study due to machine and space constraints, as the input size of EfficientNets goes from 224x224 in B0 to 600x600 in B7. As a result, space and time become more complex. EfficientNetB7 requires a significant amount of time to train on a large dataset. The parameters grow from 5.3 million in B0 to 66 million in B7. Thus, in order to use the B7 model or any other model, we require a high-performance machine.

6. MANAGERIAL AND SOCIAL IMPLICATIONS

There are many implications of the research:

1. The face mask detection may be conducted utilizing CCTV cameras to levy fines on those who do not wear a face mask. This may compel individuals to wear masks, reducing the propagation of the virus by a large margin and thereby lowering the infection's fatality rate.
2. Automatic doors can be connected with face mask recognition in Hospitals, shopping malls, offices, and other places of large social gatherings to prevent those without masks from passing through.
3. Home CCTV cameras can be implanted with this technology and can be wirelessly linked to the mobile phones of people, detecting whether the person is leaving their house without a mask and sending a reminder to the person on his mobile phone to put on a mask before leaving.
4. It may be used to identify people that are not wearing masks on public transit, such as aircraft, trains, and buses. Also, if the driver is not wearing a mask, the ride will not begin.

7. CONCLUSION AND FUTURE SCOPE

Our findings show that the techniques(ResNet50, EfficientNetB0, MobileNetV2, and Inception V3) might be used to classify data from any domain, despite the fact that we used them to categorize data with and without mask photographs. Additionally, it has been discovered that when EfficientNet-B0 is used, mask detection accuracy improves significantly (approximately 8%) when compared to the other three models used. In the future, EfficientNet-B0 may be considered a safe bet for generating higher results in image prediction and classification—accuracy achieved at a substantially lower computational cost in these models. Additionally, we may increase the size of the dataset used for Mask detection. These algorithms/models are applicable to a range of fields, including healthcare and image processing.

8. CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

9. FUNDING AGENCY

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Brown, J. M., Campbell, J. P., Beers, A., Chang, K., Ostmo, S., Chan, R. V. P., Dy, J., Erdogmus, D., Ioannidis, S., Kalpathy-Cramer, J., & Chiang, M. F. (2018). Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks. *JAMA Ophthalmology*, *136*(7), 803. doi:10.1001/jamaophthalmol.2018.1934 PMID:29801159

Cheng, K. K., Lam, T. H., & Leung, C. C. (2020). Wearing face masks in the community during the COVID-19 pandemic: Altruism and solidarity. *Lancet*. Advance online publication. doi:10.1016/S0140-6736(20)30918-1 PMID:32305074

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Li, F.-F. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR.2009.5206848

Denison, M. R. (2004). Coronavirus research: Keys to diagnosis, treatment, and prevention of SARS. In *Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary*. National Academies Press (US). Available from: https://www.ncbi.nlm.nih.gov/books/NBK92477/

Du, W., Rao, N., Liu, D., Jiang, H., Luo, C., Li, Z., Gan, T., & Zeng, B. (2019). Review on the Applications of Deep Learning in the Analysis of Gastrointestinal Endoscopy Images. *IEEE Access*. .10.1109/ACCESS.2019.2944676

Girshick, R. (2015). Fast R-CNN. *Proc. IEEE Int. Conf. Comput. Vis.*, 1440-1448. doi:10.1109/ICCV.2015.169

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Howard, A. G. (2017). *Mobilenets: efficient convolutional neural networks for mobile vision applications*. arXiv preprint arXiv:1704.04861

Jiang, M., Fan, X., & Yan, H (2020). *Retina Mask: A Face Mask Detector*. Available: arXiv.org.

Kamencay, P., Benco, M., Mizdos, T., & Radil, R. (2017). A New Method for Face Recognition Using Convolutional Neural Network. *Advances in Electrical and Electronic Engineering.*, *15*(4). Advance online publication. doi:10.15598/aeee.v15i4.2389

Matuschek, C., Moll, F., Fangerau, H., Fischer, J. C., Zänker, K., van Griensven, M., & Haussmann, J. (2020). Face masks: Benefits and risks during the COVID-19 crisis. *European Journal of Medical Research*, *25*(1), 32. Advance online publication. doi:10.1186/s40001-020-00430-5 PMID:32787926

Mukherjee, S., Ghosh, S., Ghosh, S., Kumar, P., & Roy, P. P. (2019, May). Predicting video-frames using encoder-ConvLSTM combination. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2027-2031). IEEE. doi:10.1109/ICASSP.2019.8682158

Mukherjee, S., Kumar, P., Saini, R., Roy, P. P., Dogra, D. P., & Kim, B. G. (2017). Plant disease identification using deep neural networks. *Journal of Multimedia Information System*, *4*(4), 233–238. doi:10.9717/JMIS.2017.4.4.233

Mukherjee, S., Saini, R., Kumar, P., Roy, P. P., Dogra, D. P., & Kim, B. G. (2017). Fight detection in hockey videos using deep network. *Journal of Multimedia Information System*, *4*(4), 225–232. doi:10.9717/JMIS.2017.4.4.225

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. doi:10.1109/TKDE.2009.191

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. doi:10.1109/TPAMI.2016.2577031 PMID:27295650

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR.2018.00474

Singhal, T. (2020). A Review of Coronavirus Disease-2019 (COVID-19). *Indian Journal of Pediatrics*, *87*(4), 281–286. doi:10.1007/s12098-020-03263-6 PMID:32166607

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298. doi:10.1109/TMI.2016.2528162 PMID:26886976

Sun, L., Zhao, C., Yan, Z., Liu, P., Duckett, T., & Stolkin, R. (2018). A novel weakly-supervised approach for RGB-D-based nuclear waste object detection. *IEEE Sensors Journal*, *19*(9), 3487–3500. doi:10.1109/JSEN.2018.2888815

Suvarnamukhi, B., & Seshashayee, M. (2018). Big Data Concepts and Techniques in Data Processing. *International Journal on Computer Science and Engineering*, *6*(10), 712–714.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2015.7298594

Szegedy, C., Ioffe, S., & Vanhoucke, V. (2016). Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning. *Proceedings of the 2016 International Conference on Learning Representations*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2016.308

TanM.LeQ. V. (2020). Efficientnet: Rethinking model scaling for convolutional neural networks. https://arxiv.org/abs/1905.11946

Moolchand Sharma is currently an Assistant Professor in the Department of Computer Science and Engineering at Maharaja Agrasen Institute of Technology, GGSIPU Delhi. He has published scientific research publications in reputed International Journals and Conferences, including SCI indexed and Scopus indexed Journals. The Journals are Cognitive Systems Research (Elsevier), Physical Communication Journal (Elsevier), International Journal of Image & Graphics (World Scientific), International Journal of Innovative Computing and Applications (Inderscience), Cyber-Physical Systems Journal (Taylor & Francis) & Innovative Computing and Communication Journal (Scientific Peer-reviewed Journal). He has authored/co-authored in chapters with international publishers like Elsevier, Wiley, De Gruyter & CRC Press. He is associated with various professional bodies like IAENG, ICSES, UACEE, Internet Society, etc. He possesses teaching experience of more than seven years. He is also the co-convener of the 'ICICC'& 'ICDAM' springer conference series. He is a doctoral researcher at DCR University of Science & Technology, Haryana.

Harsh Gunwant is a student of Computer Science & Engineering 4th year at Maharaja Agrasen Institute of Technology, GGSIPU, Rohini, Delhi.

Pranay Saggar is a 3rd-year student of Information Technology at Guru Teg Bahadur Institute of Technology, which is affiliated with GGSIPU, Delhi.

Luv Gupta is a student of Medical Science & Research at Sharda University, Greater Noida, Uttar Pradesh.

Deepak Gupta received a B.Tech. degree in 2006 from the Guru Gobind Singh Indraprastha University, Delhi, India. He received an M.E. degree in 2010 from Delhi Technological University, India, and Ph. D. degree in 2017 from Dr. APJ Abdul Kalam Technical University (AKTU), Lucknow, India. He has completed his Post-Doc from National Institute of Telecommunications (Inatel), Brazil in 2018. He is the recipient of 2021 IEEE System Council Best Paper Award. He have been featured in the list of top 2% scientist/researcher database in the world [Table-S7-singleyr-2019]. He is the Treasurer of the IEEE ComSoc-Delhi Executive Committee. He has co-authored more than 195 journal articles including 105 SCI papers and 44 conference articles. He has authored/edited 50 books, published by IEEE-Wiley, Elsevier, Springer, Wiley, CRC Press, DeGruyter and Katsons. He has filed four Indian patents. He is convener of ICICC, ICDAM & DoSCI Springer conferences series. Currently he is Associate Editor of Expert Systems (Wiley), and Intelligent Decision Technologies (IOS Press). He is also working towards promoting Startups and also serving as a Startup Consultant. He is also a series editor of "Elsevier Biomedical Engineering" at Academic Press, Elsevier, "Intelligent Biomedical Data Analysis" at De Gruyter, Germany, "Explainable AI (XAI) for Engineering Applications" at CRC Press and "Computational Intelligence for Data Analysis" (Bentham Science). He is appointed as Consulting Editor at Elsevier.