Real-Time Face Mask Detection

1] Meenal Jain, Department of Computer Science, IIITD
2] Saurabh Pandey, Department of Computer Science, IIITD
3] Utsav Baghela, Department of Computer Science, IIITD

Abstract

The COVID-19 coronavirus pandemic is wreaking havoc on the world's health. Many preventive steps have been taken by WHO to prevent the spread of this disease, where the usage of a mask is a must. Usage of masks is mandatorily recommended by the World Health Organization (WHO). Every day, a huge amount of people become contaminated with the disease and suffer as a result of it. To implement basic safety principles, proper precautions need to be taken. A face mask detector system can be implemented to strictly check this. Applications of Face Mask detection are many, which includes areas like security, safety, law enforcement and more. In this project, a face mask detector is developed that is able to distinguish between a masked face and a non-masked face i.e. whether a mask is on or not. In this report,

The implementation of the project will be on images. The dataset used by us consists of 7553 images in total, out of which 3828 are unmasked and 3725 are masked Images and used a webcam for real-time facemask implementation using a webcam.

We have used both Machine Learning and Deep Learning methods for face mask detection,

We have proposed a face mask detector that employs ML methods such as a Decision tree, Random Forest classifier, Support Vector Machine to detect the presence of a face mask.

We implemented the Deep learning method Convolution Neural Network (CNN) which is implemented with Max pooling and showed the accuracy of the method.

The CNN with Max pooling achieved 99.61% training accuracy and validation accuracy is 90.24%. Furthermore, our task included detecting in a real-time video whether the people appearing have facemasks on or off.

Keywords *— face mask; social-distancing; max-pooling; SVM, Decision Tree; Random Forest; covid-19; Max pooling; mask detection; CNN; real-time.*

1. Introduction

The term "novel coronavirus" refers to a modern type of coronavirus that has never been observed in humans before. Since the fall of 2019, the world is facing a global health crisis due to the Covid-19 pandemic. It is a respiratory infectious disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). COVID-19 has been a significant healthcare challenge, especially in the 2nd wave. Many Industries were affected by this pandemic. Lockdowns were imposed by the government in many countries. This virus can trigger a variety of illnesses, such as cold, cough, decrease in oxygen levels of patients, loss of taste and smell, severe lung infections to life-threatening infections including Respiratory Syndrome to Severe Acute Respiratory Syndrome, which have collapsed the healthcare system of various countries. By now, the virus has spread to many countries worldwide. Different mutations have been shown by the virus. Deceased caused by Corona Virus as called Covid-19. A major reason for Covid-19 infection is the transmission of respiratory droplets produced when people breathe, cough, cough, or sneeze. Many steps were taken to stop the spread of this covid-19 such as social distancing, lockdowns by the government etc. Therefore, social distances are the way to stop the spreading of this disease.

Before coronavirus, some people used face masks to protect themselves from air and dust pollution. Protection against coronavirus is mandatory. 80 percent of all respiratory

infections can be stopped or prevented by using a good quality mask, according to the WHO. Indeed, wearing a mask is an effective method of blocking. Therefore, wearing a mask and social distancing is a must to take preventive measures against covid-19. Many governments have enforced face mask rules for personal protection. In this context, the deep learning-based model is used, named CNN, to prevent human-to-human transmissions of the Covid-19 for face detection with or without masks.

CNN is a kind of deep neural network which is typically used in deep learning to train models consisting of images. A CNN is a Deep Learning algorithm that would take an image as input, extracts different parts of images using filters and kernels, and differentiate between them. Because of their high precision, CNNs are used for image detection and identification.

We implemented the Deep learning method Convolution Neural Network (CNN) which is implemented with Max pooling and showed the accuracy of the method.
In the project, we showed the performance of various machine learning algorithms and Deep Learning algorithms in detecting face masks in real-time video. In this study, datasets are used for feature extraction using haar cascade. For the classification process, the decision tree algorithm, support vector machine, and random forest are used that give high detection accuracy on each dataset. We measured the performance of models using accuracy after completing the training and testing phase.

The main objective of the project is to detect a person with a face mask on or off in a real-time video. The image used in the process is captured by the web camera. Preprocessing is done on the training datasets and feature extraction and classification are done using CNN. The trained model shows an accuracy of 91.04% by CNN. We designed a binary face classifier in which it says the mask is on or the mask is off after detecting faces.

2. Dataset and its Analysis/Preprocessing

Our dataset has been sourced from Kaggle datasets The dataset used by us consists of 7553 images in total, out of which 3828 are unmasked and 3725 are masked Images. Diverse races images are present in the dataset. There is a suitable proportion of masked and unmasked images to maintain the balanced dataset. We have split our dataset into 2 parts, one is the training dataset other is the test dataset. The splitting of data depends on the type of model we need to develop. Therefore, if a model needs a large training dataset the split is done likewise, which is the case in this scenario

In our approach, we have dedicated 80% of the dataset as the training data and the remaining 20% as The testing data, which makes the split ratio 0.8:0.2 of the train to test set. Before splitting, we have applied to pre-process to the images such as resizing, grayscale etc. To train the model, the training set (which is the subset of the dataset) is used. The model learns from this data.

To test the model, the Testing set (which is the remaining subset of the dataset) is used. The model Test on this data and evaluate a final model fit on the training dataset.

Initially, we converted coloured images to grayscale images. Implemented haar cascade to detect faces present in the images.
After detecting faces, process each face to get face boundaries. Convert images to a 1-D array by flattening them. In images whose face detection was failed by haar-cascade were trained directly converting to grayscale and flatten. We have used the following collections/libraries: os, cv2, NumPy, pickle, matplotlib pyplot, train_test_split from sklearn.model_selection, DecisionTreeClassifier from sklearn.tree, SVC from sklearn.SVM, KNeighborsClassifier from sklearn. neighbours and RandomForestClassifier from sklearn. ensemble.
We have also performed data augmentation to generate the data with faces of varying angles and shapes.

3. Literature Review

There are several methods to solve the classification problem. For example **[1]** solve the face mask detection by CNN and different traditional ML methods. CNN comes out as the best model compared to other models. Among SVM, Decision Trees and Proposed method CNN, CNN is predicting the optimal hyperplane to detect mask or no mask by minimizing the empirical training error and generalization error by finding the largest margin between two separating hyperplanes.
To detect the mask detection problem, we first need to detect faces as a subproblem.
Face Detection involves three steps including face detection, face recognition, and face extraction. As per **[3]** AdamBoost, HaarClassifier are the various algorithms to implement face recognition. Finding faces is the most important part of face detection and it can be found using different techniques. **[3]** used various algorithms implementing and analyzing face detection. Algorithms such as Finding faces via colour, motion, finding faces in limited areas of pixels, Camshaft algorithms and haar cascades can be used to detect faces. They have found haar cascades as a very efficient algorithm and also gives better accuracy in facial expressions.
[2] has different techniques to find the faces from the given image. Locating facial features in an image is an important part of face recognition, face tracking, eye recognition **[2]** uses CMU PIE with 106 images they have 87% efficiency of face detection. The major challenge for face detection comes out with real-time scenarios.**[4]** uses KNN, SVM and MobileNet to find the best algorithm for face detection in a real-time situation. And MobileNet model is found to be the best accuracy (88.7%) both for input images and input video from a camera (real-time). But the camera quality and illumination is the major challenge in implementing a real-time system.
Deep learning has the powerful ability to learn features and implement classification problems to improve performance. For instance **[5]** Madhura Inamdar has proposed Facemasknet architecture to implement face mask detection with an accuracy of 98.6%. In Fact, this was the first paper to perform three-class classification: With Mask, Improperly worn Mask and Without Mask. And it can be implemented to both live stream video or give input images as input. Sometimes during live streaming of video, to remove background and focus only on the image, we can extract the Region of Interest and this is also implemented in **[5]** this paper.
In **[6]** they have proposed a transfer learning using MobileNetV2 pre-trained model with some hyperparameters tuning. Hyper Parameters such as learning rate, batch size, epochs have been used and among these, the learning rate parameter was found to be a very important parameter that results in either convergence or overshoots the model. They have trained their model on the real dataset and tested their model using live video streaming results in good accuracy.
They have also sent an AI alert through the mail to those people who are found with no mask. The present model is working quite well with single faces and their future work is to improve the model for multiple faces and also implement a model to classify three categories: Mask, Without Mask and Improper Mask.

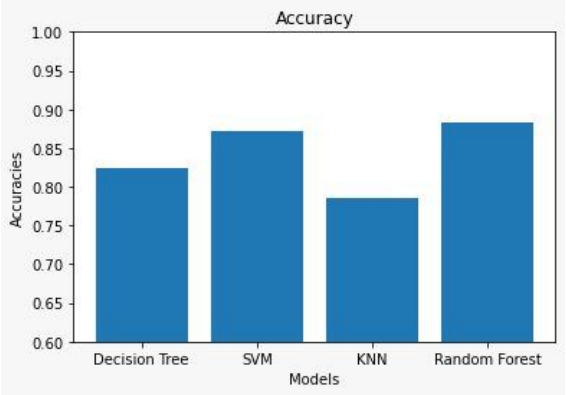
This task has already been implemented by **[5]** with good accuracy.
They **[7]** have proposed three different deep learning methods for face mask detection. CNN with MaxPooling, CNN with Average Pooling and MobileNetV2 architecture to train the models. Max pooling achieved 96.49% training accuracy and validation accuracy is 98.67%. Besides, the Average pooling achieved 95.19% training accuracy and validation accuracy is 96.23%. MobileNetV2 architecture gained the highest accuracy 99.72% for training and 99.82% for validation. Further, their future work is to integrate this model with IoT [32-35] to detect humans without masks automatically.

4. Baseline Models

Facemask detection is a classification problem, there are different classification techniques of Machine learning. The given dataset has non-linear separation and to make the classifier we have used different traditional ML methods
To measure the performance of each model we have taken the accuracy parameter.
Below are the results for each model:

Base Model	Accuracy
Decision Tree	0.824542518
SVM	0.871367061
KNN	0.786329386
Random Forest	0.882758620

Fig: Testing Accuracy



Further, we have used the well known Deep Learning model Computational Neural Network (CNN) for image-related tasks.

Final Model	Accuracy
CNN	0.9124

5. Final Model

Proposed model: CNN with MaxPooling
Convolutional Neural Network is a deep learning model which uses convolution as a base technique.
Dataset Images are not all the same size, so preprocessing was required for this Dataset to be used in CNN. Initially, we converted all images to grayscale images to reduce the dimensions of images and better edge detection. We augmented the dataset also.

Architecture of CNN used:

- **Conv Layer 1 : Filter size : 200 @ 3*3**
- **Max-Pool Layer 1 : 2*2**
- **Conv Layer 2 : Filter size : 100 @ 3*3**
- **Max-Pool Layer 2 : 2*2**

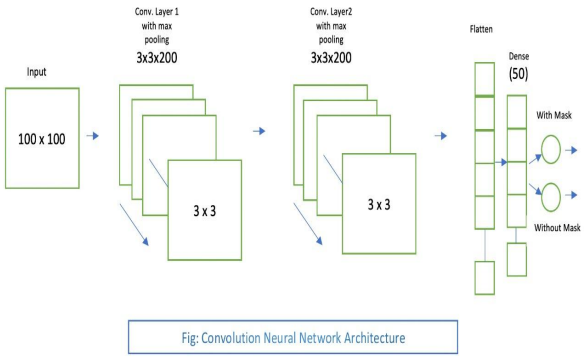


Fig: Convolution Neural Network Architecture

We have used two convolution layers followed by relu activation function and max-pooling layer as CNN architecture. And then have flattened the second layer. Further adding two dense layers, one dense layer with 50 features (nodes) and second dense layer is the output layer with two nodes representing mask and non-mask labels.

We have applied binary cross-entropy as a loss function as we have only two outputs and used adam optimizer to compile our model.

6. Results, Analysis, and comparison of the model with baselines

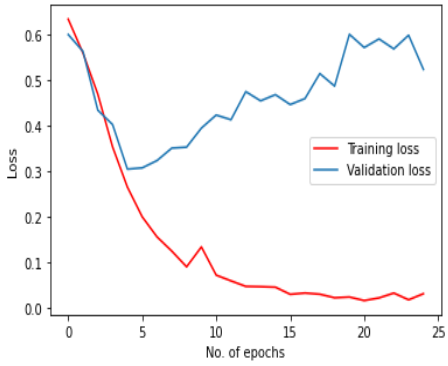
Training and Validation Accuracy:

Epoch	Training accuracy	Validation Accuracy
1	63%	66%
2	71%	68%
3	77%	82%
4	84%	83%
5	88%	88%
6	91%	89%
7	93%	89%
8	95%	88%
9	96%	89%
10	94%	88%
11	97%	88%
12	97%	89%
13	98%	88%
14	98%	89%
15	98%	89%
16	99%	91%

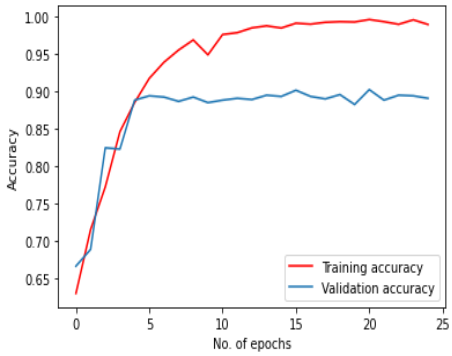
Training and Validation Loss:

Epoch	Training Loss	Validation Loss
1	0.63	0.59
2	0.55	0.56
3	0.46	0.43
4	0.35	0.49
5	0.26	0.30
6	0.19	0.30
7	0.19	0.32
8	0.12	0.34
9	0.08	0.35
10	0.13	0.39
11	0.07	0.42
12	0.05	0.41
13	0.04	0.47
14	0.04	0.45
15	0.04	0.46
16	0.02	0.44

Training Loss vs Validation Loss



Training Accuracy vs Validation Accuracy



Results (on real-time Video):

Fig: Image with a mask off

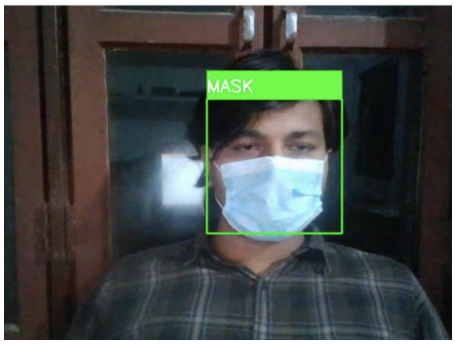


Fig: Image with a mask on

Fig: Image with improperly mask on

- We got the optimal value of validation accuracy for epoch 16: 91.04%.
- We have used two convolution layers followed by relu and max pooling. And then add a dense layer with 50 nodes and after that output layer with 2 nodes which will predict mask or no mask.
- As there are only two target variables to predict we are using binary cross-entropy as a loss function
- We have used Adam optimizer in our CNN model.
- We are using accuracy as our metrics to evaluate the model.
- We have run 25 epochs to train our model but however it can be seen that after epoch 5 the model got suitably fit and gave minimal variations of accuracy after it. The model slowly tends to become more overfit after that.
- The loss follows similar trends in inverse order and thus after epoch 5 the model variance starts increasing.

<u>Comparison with previous models:</u>

An SVM model is used to find the optimal hyperplane that finds the largest margin to separate the hyperplanes. Our model has 87.13% accuracy on the test set.
While in Decision Tree each node corresponds to the feature of the image as we have multiple features so it can lead to overfitting the model

and that's why it does not perform quite good as compared to SVM.
In KNN we have used different k values to find the optimal result but still, it does not perform well as compared to other ML models.
Random Forest is an ensemble learning algorithm and performs well among all other ML models using the criterion as entropy instead of Gini index.
CNN is a deep learning algorithm that is based on the concept of convolution. Convolution is a mathematical model which is used to transform the feature matrix and obtain the best features from it. We have used 2 convolution layers followed by relu and max-pooling and helps in predicting the output more accurately using binary cross-entropy loss function. And therefore when compared to previous ML models the CNN model performed better with proper optimization.

7. Conclusion

We implemented various Ml Models and Deep Learning models on the dataset to see which model gives the best accuracy in real-time facemask detection. Our goal is to come up with a model which gives the best accuracy such that mask identification will be simple throughout the pandemic. We implemented the Deep Learning method CNN in this project as our final model which was able to accurately predict face and mask on the face as on or off in real-time.

A short explanation is added in below table:

Epoch	Training Accuracy	Validation Accuracy
16	0.99	0.91

Epoch	Training Loss	Validation Loss
16	0.02	0.44

8. Contribution of each member

Meenal Jain: Literature Review, Data Analysis and Preprocessing, training and testing baseline models, Final Model Preparation and training, Report, ppt and blog work.
Saurabh Pandey: Literature Review Data Preprocessing, Training baseline models, Training of final model, hyperparameter tuning Report and PPT and Blog work.
Utsav Baghela: Literature Review, Data Selection and preparation, Training baseline models, designing CNN architecture, Preparation of final model, Report and PPT Preparation and Blog work.

10. References:

[1] G K Jakir Hussain, R Priya, S Rajarajeswari, P Prasanth, N Niyazuddeen 2021. The Face Mask Detection Technology for Image Analysis in the Covid-19 Surveillance System
[2] Mehul K Dabhi, Bhavna K Pancholi (2016). Face Detection System Based on Viola-Jones Algorithm
[3] Kruti Goyal, Kartikey Agarwal, Rishi Kumar (2017). Face Detection and Tracking Using OpenCV
[4] Wuttichai Vijitkunsawat, Peerasak Chantngarm, 2021. Study of the Performance of Machine Learning Algorithms for Face Mask Detection
[5] Madhura Inamdar · Ninad Mehendale. Real-time face mask identification using Facemasknet deep learning network.
[6] Riya Chiragkumar Shah ,Rutva Jignesh Shah Department of Computer Science and Engineering Nirma University. "Detection of Face Mask Using Convolutional Neural Network", 2017.
[7] F.M. Javed Mehedi Shamrat, Md. Masum Billah, Md Saidul Islam. Face Mask Detection using Convolutional Neural Network (CNN) to reduce the spread of Covid-19, 2021.

Project Files Link :
https://drive.google.com/drive/folders/10tDZcWZt0BirQwdEzRrcb-6OAw69x2vz?usp=sharing