HybridFaceMaskNet: A Novel Face-Mask Detection Framework Using Hybrid Approach

Mrinmoy sadhukhan (✉ mrinmoy.sadhukhan1996@gmail.com)

IGNOU: Indira Gandhi National Open University

Indrajit Bhattacharya

Kalyani Government Engineering College

HybridFaceMaskNet: A Novel Face-mask Detection Framework Using Hybrid Approach

Mrinmoy Sadhukhan ·
Indrajit Bhattacharya

Abstract Coronavirus disease 2019 (covid-19) is a contiguous disease which is caused by severe acute respiratory syndrome coronavirus2(SARAS-2) started from Wuhan, china, and spread all over the world within a few months in 2019. Government of all countries had to apply lockdown to decrease the number of affected patient as mortality rate of many countries became very high at that time. In the awake of 2nd wave of COVID 19 WHO has made mandatory to use mask in largely crowded areas, health centers, communities and in different places to prevent spread of virus. Many countries have invented the vacancies but it will firstly available for corona front line warriors only, not for general people, So, people have to wear mask when they are going out from home. But In recent days it can be followed that people are reluctant to wear mask when they are entering in offices, departmental stores or local shops where, gathering might happen anytime. This could lead to spread of COVID-19 among the communities. With the help of computer-vision, people who are not wearing mask can be detected by generating an alarm signal. To achieve this challenging task, a face mask detector 'HybridFaceMaskNet' is proposed, which is a combination of classical Machine Learning and deep learning algorithm. 'HybridFaceMaskNet' can achieve state-of-art accuracy on public faces. The real challenges are the low-quality images, different distances of people from camera and dynamic lighting on the faces at daylight or in artificial light.This problem can be overcome by using different noise removal techniques. HybridFaceMaskNet is trained with three different classification of images 'proper-mask', 'incorrect-mask' and 'no-mask' which are collected

F. Author
Belur-Math Howrah
Tel.: +91-9831643884
E-mail: mrinmoy.sadhukhan1996@gmail.com

S. Author
Dum Dum Kolkata
E-mail: indra51276@gmail.com

from real life images and some synthetic data , to generate alarm for different scenario .This HybridFaceMaskNet is trained on Google Colab and is compared with different existing face mask detector model. There is a possibility of deploy the model in IOT devices as it is light weight compare to other existing models.

Keywords Random Forest · PCA · XGBOOST · Machine Learning · AI · CNN

1 Introduction

COVID-19 is an airborne infectious disease which causes respiratory illness among people. It mostly affects the aged people who have co-morbidity such as diabetes, high-pressure, cardiovascular disease etc, but sometimes young people who are directly connected to serve the people such as doctors, nurses, police persons might also get affected. It spreads through fluid particles which are less than 0.0002 inches (5 microns) in diameter, called as aerosols, which are usually emitted when people speak, So, to slow down the transmission of corona virus and reduce the numbers of affected the people in hospital WHO has provided many do's and don'ts guide lines, one of them is wearing mask in the public places strictly. Wearing face mask help people to protect themselves from affecting in different kind of airborne infections.

In the rise of COVID-19 different scientist from worldwide make cooperation with each other to stop this epidemic. Artificial Intelligence (AI) based on Machine learning and Deep Learning can help to fight COVID-19 in many ways. Here a face mask detection system is introduced which can be embedded in surveillance system in mall, offices and public places to detect the people who are not wearing face masks and alarm them. But due to unavailability of public data set in a wide variety it becomes more challenging task to detect faces with mask even for state of art detection model. Faces with mask in real life are of different color, orientation and sizes. Hence it is difficult for different existing models to figure out accurate features and train them accurately. Sometimes inspite of obtaining getting 97% accuracy in training, the model cannot produce accurate results in real life testing.

In deep learning there are different types of object detectors, some of them uses one neural network for object detection like SSD,YOLO and some of them uses two neural networks such as R-CNN, where picture data set is required in huge amount for training purpose. To overcome this problem some famous machine learning technique such as HOG histogram-oriented gradient, Scale-invariant feature transformation, Hue moments, Canny edge detection etc. can be used to figure out features where data set is less.

In this paper, the proposed face mask detector can classify the people faces with three types of classification with_mask, without_mask or incorrect_mask. This model is a Hybrid model which is the integration between deep learning, Handcrafted Feature extractor and classical machine learning classifiers. Due

to unavailability of huge number of public data set deep learning and hand-crafted feature extraction, technique is applied to get out more features. In machine learning based feature extraction three handcrafted feature extractor such as LBP, textural Harlick feature and Hue moments are used. In Deep learning based feature extraction, a CNN model is used as a feature extractor. A hybrid feature vector is constructed based on this two type feature extractor. On this Hybrid feature vector, PCA is applied for feature selection. A comparison between 6 machine learning classifiers is drawn upon this selected feature vector to find the best algorithm in terms of the highest accuracy, reduced model size and least time in the process of training and detection.

The rest of the paper is organized as follows: In Section 2, literature survey is presented. Section 3 explains the proposed architecture of the model, followed by explanation and discussion. Section 4 focuses on the novelty of proposed strategy with experimental results. The paper ends with conclusions drawn from the work, as presented in Section 5.

2 Literature Survey

Toshanlal Meenpal et.al.[1] proposes a binary face classifier method,in which using VGG16 neural network(with pretrained weight) feature is extracted. After that using FCN, features are trained to generate face segmentation mask to detect faces in image with bounding box and it is also capable for detect faces with mask on training with face mask data set. Mingjie Jiang et.al.[2] proposed a one stage detector where high-level semantic information are fused by building feature pyramid network to extract features , then the features are trained in Mobile Net network then the Neural Network model is used to detect faces with mask. The state of art results is achieved by proposed model with public face mask data set which is 2.3% more accurate than baseline result in the face and mask detection precision. C. Jagadeeswari et.al.[3] uses different classifier like MOBILENETV2, Resnet50 and VGG16 and train the classifier with face mask dataset nearly about 1300 images, also usage the different optimizer function such as Adam, SGD etc. to achieve better accuracy. After that trained classifier is used to predict images. Shiming Ge et.al.[4] proposes a KNN-CNN based mask face detection where the proposed model used two CNN. First CNN extracts the feature and then send to second CNN VGGFace to characterise them with noisy descriptors. A KNN module is to refine such descriptors with respect to it nearest neighbours from huge collections of faces and non faces and jointly performs classification and identify correct facial regions and identify it is with mask or not.Madhura Inamdar[5] proposed facemask net 2D-cnn model with one input layer and 8 convolutional layers. The input layer takes image size in 227*227*3 dimension. Here RELU is used as a activation function for output in CNN. It uses SoftMax layer to identify a face and a mask. Model is trained with three types of dataset with-face, improperly-worn-mask and without mask. It archives accuracy up to 98.6% and can be used to detect face mask from CCTV footage. Kaihan

Lin et.al.[6] proposed G-Mask with two branches one is for face detection and other for face and background image segmentation. Here ResNet-101 is used to extract facial features. Region proposal network is used for generate Region of interest. 'Region of interest aligns' is used to faithfully preserve exact spatial locations for generating face mask through FCN and output the feature map to a fixed size. This feature is used to detect faces with mask using bounding box. Vinitha.V et.al.[7] uses MOBILENETV2 as a backbone with pretrained ImageNet weight for classifing images with mask or no mask after training. Here pre-processing of images is done with following step resizing the input image, apply colour filter, scaling and normalizing images with standard mean, centre cropping by 224*224*3 dimension and convert it into pytroch tensors. Mohamed Loey et.al.[8] proposes a model using deep and classical machine learning components, one is feature extraction with RESNET50 neural network and second part is classification process for face mask using decision trees and SVM with accuracy 99.64%. Here three types of dataset is used like real-world masked face, simulated masked face and labelled masked face. Yang et al.[9] proposed a face detector that is composed by concatenating two individual components, the CNN based feature extraction part and the boosting forest part. The feature extraction part extracts feature representation, using only the first few layers from a pre-trained CNN model and the boosting forest model used directly on the feature maps to detect faces. Zhu et al.[10] presented Contextual Multi-Scale Region-based Convolution Neural Network (CMS-RCNN) for face detection in dynamic conditions, containing the region proposal component and the region-of-interest (RoI) detection component. The multi-scale information is grouped both in region proposal and RoI detection to deal with tiny face regions. Secondly, the proposed network allows explicit body contextual reasoning in the network inspired from the intuition of human vision system. This approach is bench marked on WIDER FACE data set which contains high degree of variability. The model receives great level of accuracy of 90% after training. Opitz et al.[11] proposed a face detection system with a novel grid loss layers of CNN's to deal with partial occlusion of faces, which minimized the error rate on sub-blocks of a convolution layer independently rather than over the whole feature map. By mapping loss layer back to a regular fully connected layer, no additional computational cost is required at run time compared to standard CNNs. Li et al.[12] utilized 6 number of cascaded CNN to perform face detection, it operates on multiple resolutions, also rejects the background regions in low resolution images. Here a CNN-based calibration stage is used after each of the detection stages in the cascade to improve localization. There are 6 CNNs in the cascade among them 3 CNNs are for face vs. non-face binary classification and 3 CNNs for bounding box calibration, also known as multiclass classification of discretized displacement pattern. Here Alex Net is used. ReLU non-linearity function is applied after the pooling layer and fully-connected layer. Bell et al.[13] suggested a RNN induced object detector method (ION) that is combined with contextual data and it showed promising results. Zagoruyko et al.[14] have demonstrated a Multi-path network that enhances the R-CNN object detector, with the

access to characteristics at multiple network stack. This multipath model is grouped with "Deep Mask object" in order to get rid of the object detection problem. Ghiasi and Fowlkes.[15] have been able to use hierarchical DPMs for face detection and landmark localization despite there being occlusion in image. Yahia Said[16] proposed a face mask detector called Pynq-YOLO-Net, that implement lightweight YOLO framework as a backbone.They trained the model with ADAM optimizer, after training they placed their model on Pynq Z1 board via SSD card.In the board Model achieves 16FPS processing speed which is considered as real time processing speed.

Some of the existing face mask detector techniques have been discussed above. But most of them do not provide a proper end to end solution for classification of face images with mask using embedded device. Bedsides, they classify face image with classifier with_mask or without_mask only using transfer learning, which are time consuming and may cause trouble with heavy model size at the time of implementation in embedded devices having limited resources.

In order to get more refined classification of face mask, three types of classification is proposed viz. mask, with_mask and incorrect_mask. To deal with noisy and low quality images pre-processing techniques image-sharpening and brightness adjusting methods are introduced in the current work. For implementing the model in embedded device with less resources and to make model size small, machine learning classifier is used on the proposed hybrid feature vector that might reduce the size of the application.

3 Proposed Technique

3.1 Outline of Proposed Methodology

In the Figure 1 an overview of the proposed architecture is presented. In this proposed architecture face images are used. Images are classified by with_mask or incorrect_mask or without_mask as input of face mask recognition system for training. The first step is to convert the images in 224*224 pixel[17] as the images taken from the repository are not of same size and dimension. To recognize the mask from faces, this proposed system uses machine learning classifier (Random Forest) on the Hybrid Feature set produces by CNN and Handcrafted feature extractor from the input images. The detailed discussion of image pre-processing is done in Section 3.2. Feature extraction based on hand crafted feature extractor technique is demonstrated in section 3.3. Section 3.4 discusses about image feature extraction by deep learning. Feature concatenation technique is explained in section 3.5. In section 3.6 feature selection using PCA is discussed. Classification using random Forest Classifier is proposed in section 3.7.
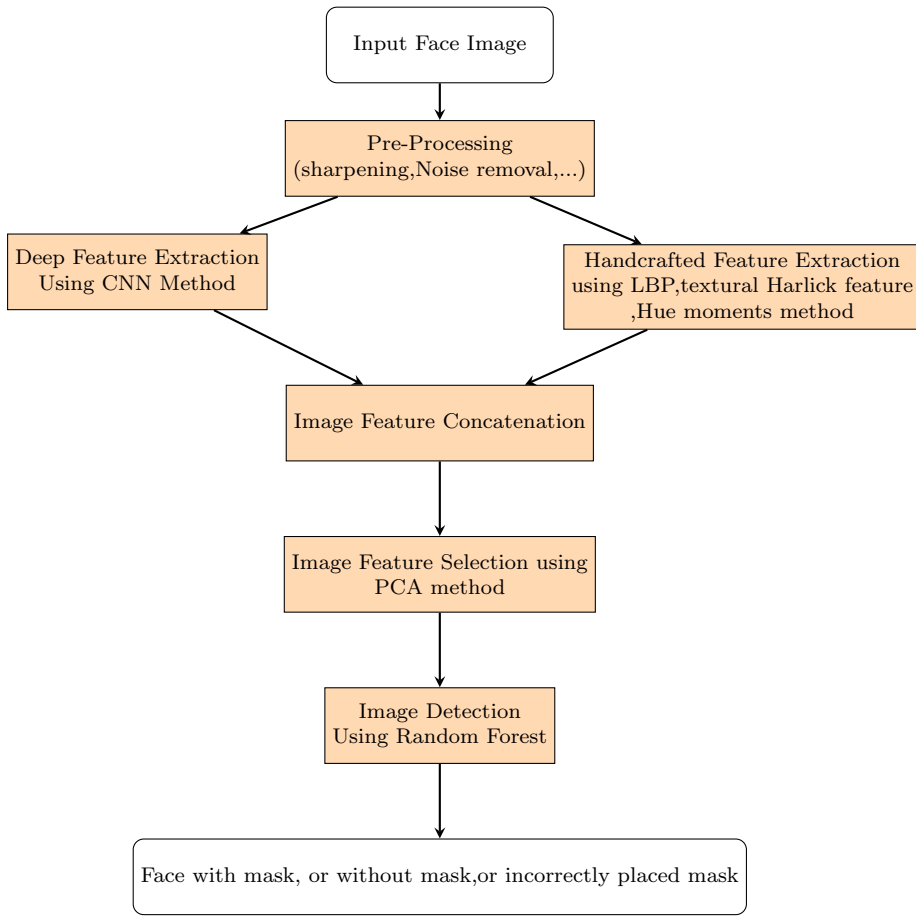
Fig. 1: Flow chart of proposed architecture

;

3.2 Pre Processing of Image

Images taken from heterogeneous sources are not of same quality. To preserve the details of images and enhance the picture quality some image preprocessing technique is proposed. In order to adjust the contrast of an image by modifying the intensity distribution, here constant limited adaptive histogram equalization is applied. Then to make images more visible, brightness is increased. After that, to increase luminance level of images Gama filter is applied. Finally to decreased noise and increase the sharpness of image Gaussian blur and a 3X3 kernel is applied respectively. Now the image is ready to feed the CNN and Hand crafted Feature extractor.

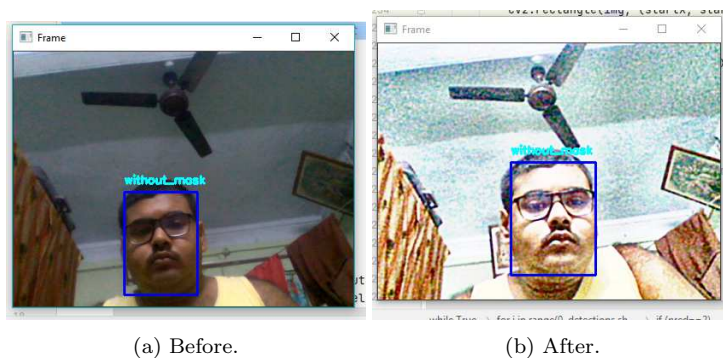(a) Before.   (b) After.

Fig. 2: Pre-processing of images during live testing.

3.3 Feature extraction based on Hand crafted Feature extractor

In the world of machine learning there are many handcrafted filters such as Gabor filters, local phase quantization, LBP, quality testing, and DLTP for extracting image features. Among these methods Hue moments,Haralick texture feature and LBP method are used in this paper. The LBP (Local Binary Pattern) feature extraction method yielded high detection accuracy and offers several advantages to the output image features, including robustness to lighting intensity, variation of rotation and computation of local representation of texture by comparing each pixel with its surrounding neighborhood of pixels. In the case of face mask detection, it is required to extract the local features like type and pattern of masks. The Haralick texture features are used to compute a global representation of texture based on the Gray Level Co-occurrence Matrix, a matrix that counts the co-occurrence of neighboring gray levels in the image. Here it is used to extract features from the position of the mask with respect to entire face globally. Hu Moments (or rather Hu moment invariants) are a set of 7 numbers calculated using central moments from gray scale image that are invariant to image transformations. Hue moment is used here for shape matching of mask in different lighting variation and facial orientation. After extracting the features for one image by using above said method, features are concatenated to make a single feature vector. It is about 46 component feature vector.

3.4 Image Feature Extraction By Deep Learning

Although the handcrafted feature extraction technique is proved sufficient but many cases it fails when there is a situation to learn complex relation from the pixels of images . The feature vector size that has been produced by earlier technique is limited to 46 component but to increase training efficiency it is also required to enhance the size of feature vectors in order to explain the finer details of the objects. This will in turn increase the efficiency of the model. In

this scenario deep learning based technique should be used. CNN has proven as a powerful method by different studies in field of computer vision. CNN is used in real world to combact with many real time challenges gender detection, face detection, vechile Number-plate detection, handwriting detection and many more.

In CNN there are two part one convolution layers and another is fully-connected layer. Convolution part is the main backbone of CNN, it extracts the complex features from images which is called feature map. After flattening them it sends the array to fully connected layer to training and testing and validating the data with the given labels. In CNN there are different filters/kernels consisting of trainable parameters which can be convolve on given images to detect the spatial features, edges and shapes. The stacked layered of CNN is used to detect complex spatial shapes from the spatial features at every subsequent levels. Hence after the end layer CNN turn image into a highly abstract representation which is easy for prediction.
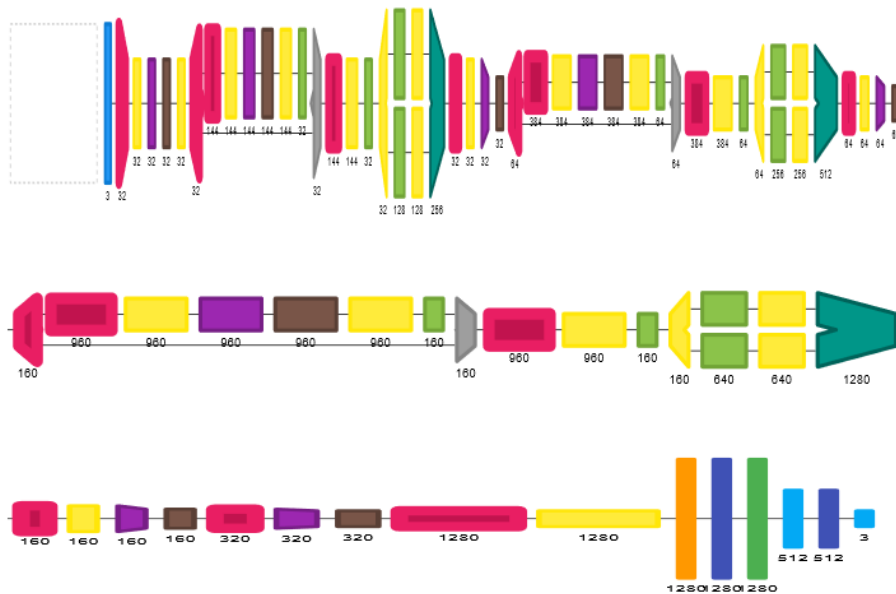

Fig. 3: The Proposed Structure of CNN.

In Present study,a CNN is constructed from scratch for extracting image features by deep learning method, for the face mask detection problem. Figure 3 depicted the proposed Neural Network structure.

In CNN architecture(Figure 5), it has two types of block, one is residual bottleneck block with stride 1 (combination of 1X1 convolution with relu6 activation, depth wise convolution and 1X1 convolution without any relu activation), Second is fire block combination of squeeze module (1X1 filter) and expand
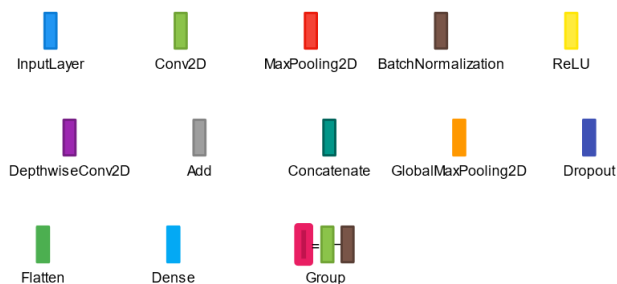
Fig. 4: The Legends Of Proposed Structure of CNN.

Input	Operator	T	C	N	S
224^2 X 3	Conv2d 3X3	-	32	1	2
112^2 X 32	Bottleneck	6	32	1	1
112^2 X 32	Fire	4	64	1	2
56^2 X 64	Bottleneck	6	64	1	1
56^2 X 64	Fire	4	160	1	2
28^2 X 160	Bottleneck	6	160	1	1
28^2 X 160	Fire	4	320	1	2
14^2 X 320	Depth conv3X3	-	320	1	2
7^2 X 320	Conv 2d 1X1	-	1280	1	1
7^2 X 1280	Global Max Pooling	-	-	1	-

Fig. 5: The Proposed Structure of CNN in tabular view where T is expansion factor, C is no of output channels, N is repeating factor,S is stride

module (combination of 1X1 and 3X3 filter). To extract features from image, height and width of image decreases by half after every time a image pass through bottleneck block followed by a fire block.The number of convolution filters are used 32, 64, 160, 320, 1280 respectively. To keep intact the spatial size of the output volumes, zero-padding is added around the borders in all layers. After the last convolution layer a Global Max Pooling layer followed by flatten layer with 50% dropout is used to fed the feature vector to first dense layer with 512 neurons.the output of first dense layer is again fed to dropout layer followed by a second dense layer with 3 neurons. Here dropout layer is used to overcome the problem of over fitting in the time of classification. Finally from the output of second dense layer we get out-put in 3 class

as with_mask, without_mask, incorrect_mask with probability.

To train our model the pre-processed face data set is used. For training the CNN the learning rate is kept at 0.0001 and the optimizer function Adam is applied. As Adam is an adaptive learning rate method, which means, it computes individual learning rates for different parameters and it shows huge performance gains in terms of speed of training.

In the proposed architecture of CNN, the number of layers has increased by trial and error method in such a way that CNN does not produce biased result to one kind of class only. The increased number of layers helps to take out more number features. Which in turn helps to get nearly same accuracy in all types image classes in time of training.

After training, the trained CNN model is again loaded to take out feature vector from last convolution layer in the shape of ('train_data_size', 7*7*1280) that is near about 62720 component feature vector.

3.5 Feature Concatenation

In the previous section we have discussed the hand crafted feature extractor and CNN methods, by which the features are extracted from images.This two feature vectors are concatinated to make Hybrid feature vector.This Hybrid feature vector contains richer information of an image than the single source feature vector. Let fb is defined as feature vector produced by handcrafted feature extractor method and fh is defined as feature vector produced by CNN. After concatenation of fb and fh it will be

$$f = [fb.fh]$$

This composite feature vector increases the size of image features. In detail, it can be said that feature vector fh has 62720 component which is extracted by CNN and feature vector fb has 46 component which is extracted by hand-crafted feature extractor. As a result ,the hybrid feature vector is in 62766-dimensional space (62720 from CNN and 46 from handcrafted feature extractor).

3.6 Feature selection Using PCA

The high dimensional feature vector of images always require huge processing power and time to solve the classification with machine learning classifier. For this,to reduce the size of hybrid feature vecto, PCA method is applied before classifying the features into three labels. This technique randomly select reduced number of features from the feature set and those features are highly capable of detecting or classifying objects among pool of objects. This will intern help to reduce the size of the feature vector in an efficient way. After applying PCA method the size of feature vector has reduced to 1663-dimensional space.

3.7 Classification Using Random Forest Classifier

For many image-based system CNN method is sufficient but it faces the over-fitting problem caused by the use of a large volume of network parameter(3,395,587). As a result, to train its parameter, the training process requires a large volume of data. However, such a large volume of data is difficult to collect. To reduce the effect of over fitting problem with small data set, Random Forest Classifier with 100 trees is used. It is trained with selected feature vector(which has been generated by PCA) to categorize the input image as with_mask, without_mask or incorrect_mask. Random Forest Classifier produce best result in the real life scenario where randomness and dynamic behaviour has to be encountered. Using 100 trees of random forest we can reach the optimum decision at a very fast rate at the time of image classification with minimal hardware usage.
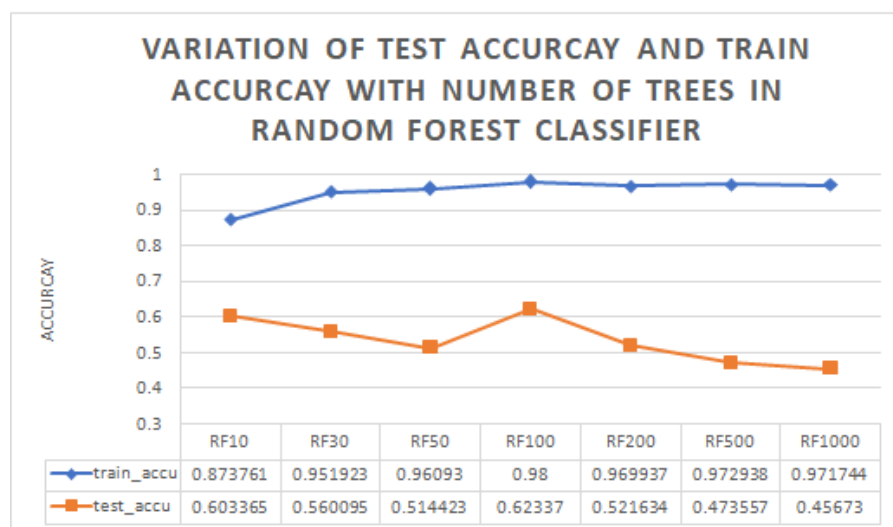

VARIATION OF TEST ACCURCAY AND TRAIN ACCURCAY WITH NUMBER OF TREES IN RANDOM FOREST CLASSIFIER

	RF10	RF30	RF50	RF100	RF200	RF500	RF1000
train_accu	0.873761	0.951923	0.96093	0.98	0.969937	0.972938	0.971744
test_accu	0.603365	0.560095	0.514423	0.62337	0.521634	0.473557	0.45673

Fig. 6: Accuracy vs No of Trees

4 Experimental Result

To evaluate proposed architecture, public data set is used. It is collected from kaggle face-mask-detector (mask, not-mask, incorrect-mask). There are total 2079 files which are classified into 3 classes. Each class having 680 images in average, with different dimensions. To fit images after pre-processing in CNN, train-test-split method is used to divide the total images in 80 percent for training purpose and 20 percent for testing. As CNN requires large amount of

data for training purpose, it might cause over-fitting problem in the time of training. To reduce this problem data set is enlarged by Image Data Generator by rotation, scaling, shifting, cropping etc of the original image in both horizontal and vertical direction. Here f1-score of this scratch model for incorrect_mask, with_mask, without_mask is 0.98, 1.00, 0.96 respectively, and accuracy of model is 0.99. In fig 7 loss vs accuracy graph of CNN model is presented.
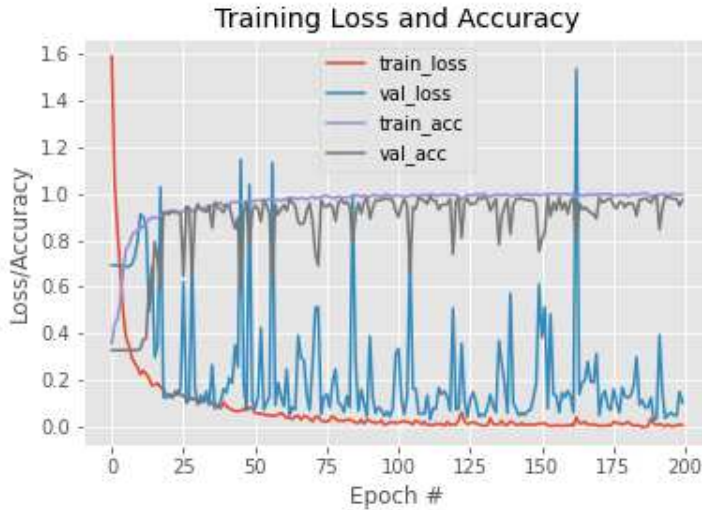

Fig. 7: Loss vs Accuracy Graph

After extracting feature from trained CNN model and Hand-crafted Feature vector, they are concatenated and Principal Component Analysis is applied on the feature vector to reduce the feature vector size. To choose best model for classification a 10-fold Cross validation is applied among different state-of-art machine learning classifier like Random Forest,Gaussian Naive Bayes, XGBOOST and SVM. Here other machine learning classifier such as linear regression and logistic regression is not used as they perform worst in case of dynamic data set.In Figure 8 a accuracy vs machine learning classifier graph is presented. From this Figure we can observe that XGBOOST and Random Forest classifier produce best accuracy of 99.21% and 98% respectively. Here Random Forest Classifier is choosen for proposed model of prediction as it perform very fast and accurate prediction in compare with than XGBOOST. It can be observed from the above result that XGBOOST outperforms other existing algorithm like RF and SVM, in terms of test accuracy which is nearly 80%. But it should be noted that as the processing time of XGBOOST algorithm is very high, it might not be applicable in real life scenario for dynamic face detection (with_mask, without_mask, incorrect_mask) though it might

perform well over a single image (frame by frame). But in real life it is found that, it takes considerable amount of time for such face detection where frames of second image might be overlapped with the first, hence results in incorrect face mask detection.
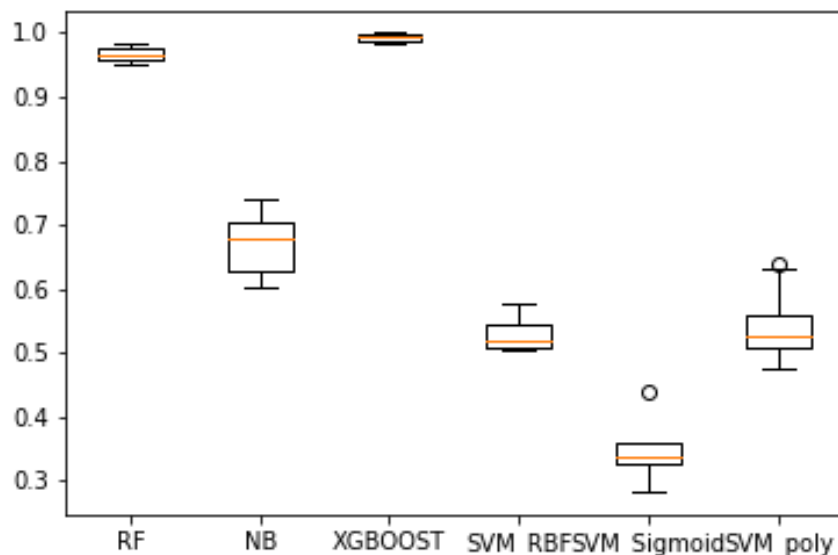

Fig. 8: Accuracy vs Machine learning Classifier

In Figure 9, a performance comparison between FaceMaskNet(FM), Resnet-50 with SVM(RS), RetinaFaceMask(RFM), PynqYOLO_Net(PYN) and proposed HybridFaceMaskNet with Random Forest Classifier(HR), HybridFace-MaskNet with XGBOOST(HX) are plotted.

In Figure 10 and 11, a comparison is made with proposed model HR, HX and state of art face mask detector model RS, FM in terms of space and time complexity.

From the Figure 9, 10 and 11, it might be observed that our proposed hybrid model HR and HX is slightly less accurate than RS and FM. But from the space and time complexity it is shown that our hybrid model HR requires less memory size , low processing time to classify one image and therefore might be best suited for implementation in the real life scenario, that might be integrated with back-end IOT systems.

From the above result it can also be observed that machine learning classifier in Hybrid models might take less time to classify images, where as neural network takes more time to predict one image. Because in the time of prediction, neural network gives the probability of different classes with respect to one image as an output which consumes more time, where as machine learning
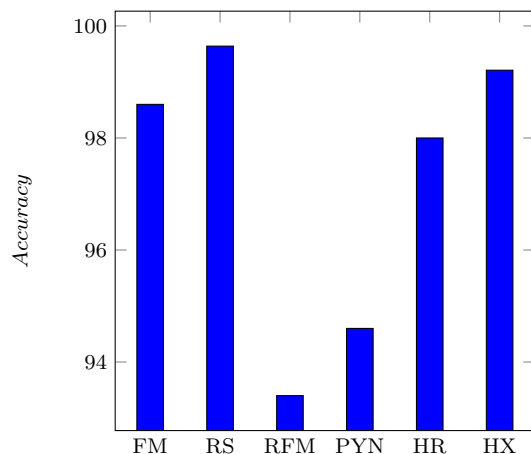
Fig. 9: A Comparison among different Face Mask Detector in terms of accuracy.
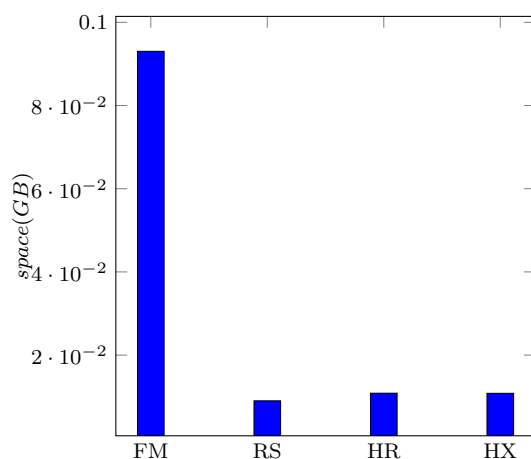

Fig. 10: A Comparison among different Face Mask Detector in terms of Space.

classifier only classify image in one class.

Also we can observe that hybrid model HX is much better than HR in terms of model accuracy. The classification model HR perform better than the proposed hybrid model with XGBOOST(HX). As HX classifier takes lot of time in classification due to generation of huge number of trees during the training phases, which might create bottleneck in the process of classification.

Some screenshot of the classification result obtained by the model HR is presented below at the time of testing using computer webcam.
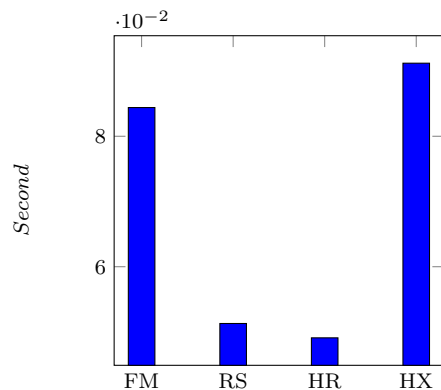
Fig. 11: A Comparison among different Face Mask Detector in terms of classification time.

(a) With Mask.   (b) without Mask

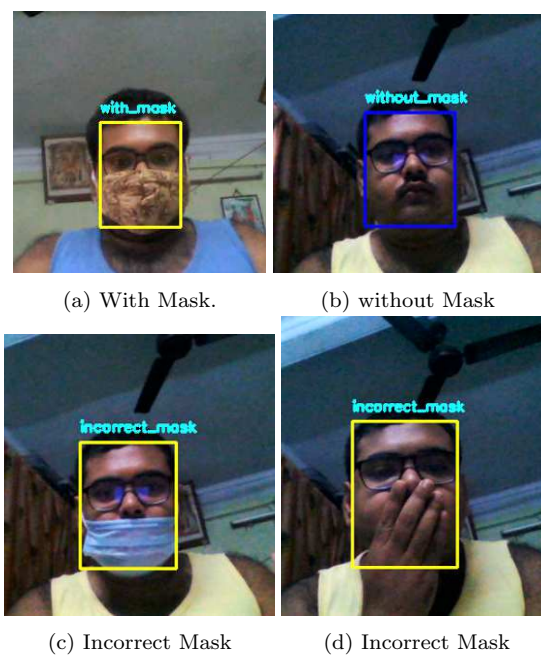(c) Incorrect Mask   (d) Incorrect Mask

Fig. 12: Test image captured by webcam.

5 Conclusion

We have developed a face mask detector HybridFaceMaskNet, which can be applied in variety of application domains. The proposed method uses hybrid feature vector, which is the combination of features from CNN based deep learning method and Hand crafted feature extractor. This hybrid feature vec-

tor set contains richer information than that obtained using single feature extraction method.Hand crafted feature extractor is suitable for low quality image to detect features where as high quality images are suitable for deep learning. Thus by combining this two approaches we can enhance classification accuracy. To classify images we have used a machine learning based classifier Random forest as a result of reduce size of application. Thus it can be used for high and low computation scenarios.

OpenCV, tensor flow, keras and CNN are used to detect whether people wearing mask or not in different public places. The models were tested with images and real-time video streams.Though the system finds the test accuracy nearly 62% in case random forest with 100 trees, but this accuracy might increase by increasing size of training data and incorporating historical data set in the training data set with localized information that might be effective with respect to a particular geographical location.

References

1. Toshanlal Meenpal, Ashutosh Balakrishnan, Amit Verma. *Facial Mask Detection using Semantic Segmentation.* Published in: 2019 4th International Conference on Computing, Communications and Security (ICCCS),Date of Conference:10-12 Oct. 2019,ISBN:978-1-7281-0876-6.
2. Mingjie Jiang, Xinqi Fan, Hong Yan. *RETINAFACEMASK: A Face Mask Detector.* arXiv:2005.03950v2 [cs.CV] 8 Jun 2020.
3. C.Jagadeeswari, M.Uday Theja. *Performance Evaluation of Intelligent FaceMask Detection System with various Deep Learning Classifiers.* International Journal of Advanced Science and Technology Vol. 29, No. 11s, (2020),pp. 3074-3082
4. Shiming Ge, Jia Li, Qiting Ye, Zhao Luo. *Detecting Masked Faces in the Wild with LLE-CNNs CVF.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2682-2690.
5. Inamdar, Madhura, and Ninad Mehendale. *Real-time face mask identification using Facemask Net deep learning network.* SSRN July 29,2020.
6. Kaihan Lin ,Huimin Zhao ,Jujian Lv ,Canyao Li,Xiaoyong Liu,Rongjun Chen,and Ruoyan Zhao. *Face Detection and Segmentation Based on ImprovedMask R-CNN.* hindawi 01 may 2020.
7. Vinitha.V, Velantina.V. *COVID-19 Facemask Detection With Deep Learning and Computer Vision.* International Research Journal of Engineering and Technology Volume: 07 Issue: 08 — Aug 2020.
8. Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and Nour Eldeen M. Khalifad. *A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic.* Measurement Volume 167, 1 January 2021, 108288.
9. B Yang, J Yan, Z Lei, SZ Li. *Convolutional Channel features.* IEEE ICCV, pages 82–90, 2015.
10. Chenchen Zhu, Yutong Zheng, Khoa Luu, Marios Savvides. *CMS-RCNN: Contextual Multi-Scale Region-based CNN for Unconstrained Face Detection.* arXiv:1606.05413.
11. Michael Opitz, Georg Waltner, Georg Poier, Horst Possegger, Horst Bischof. *Grid Loss: Detecting Occluded Faces.* arXiv:1609.
12. Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, Gang Hua. *A Convolutional Neural Network Cascade for Face Detection.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp.5325-5334.
13. Sean Bell, C. Lawrence Zitnick, Kavita Bala, Ross Girshick. *Inside-Outside Net: Detecting Objects inContext with Skip Pooling and Recurrent Neural Networks.* arXiv:1512.0414.

14. Sergey Zagoruyko, Adam Lerer, Tsung-Yi Lin, Pedro O. Pinheiro, Sam Gross, Soumith Chintala, Piotr Dollár. *A Multipath Network for Object Detection.* arXiv:1604.02135.
15. Golnaz Ghiasi, Charless C. Fowlkes. *Occlusion Coherence: Detecting and Localizing Occluded Faces.*arXiv:1506.08347.
16. Yahia Said *Pynq-YOLO-Net: An Embedded Quantized Convolutional Neural Network for Face Mask Detection in COVID-19 Pandemic Era .* International Journal of Advanced Computer Science and Applications, Vol. 11, No. 9, 2020
17. Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen. *MobileNetV2: Inverted Residuals and Linear Bottlenecks* The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520