Soft Biometrics and Deep Learning: Detecting Facial Soft Biometrics Features Using Ocular and Forehead Region for Masked Face Images

Urja Banati (✉ urjab.ic18@nsut.ac.in)

NSIT: Netaji Subhas University of Technology

Vamika Prakash

Netaji Subhas University of Technology

Rashi Verma

Netaji Subhas University of Technology

Smriti Srivastava

Netaji Subhas University of Technology

Research Article

Keywords: Soft biometric, face mask, deep learning, ocular, forehead, transfer learning

Posted Date: January 11th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1174842/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Soft Biometrics and Deep Learning: Detecting Facial Soft Biometrics Features Using Ocular and Forehead Region for Masked Face Images

Urja Banati

Vamika Prakash

Rashi Verma

Smriti Srivastava

Netaji Subhas University
of Technology
(Erstwhile NSIT)
New Delhi, India.
urjab.ic18@nsut.ac.in

Netaji Subhas University
of Technology
(Erstwhile NSIT)
New Delhi, India.
vamikap.ic18@nsut.ac.in

Netaji Subhas University
of Technology,
(Erstwhile NSIT)
New Delhi, India.
rashiv.ic18@nsut.ac.in

Netaji Subhas University
of Technology,
(Erstwhile NSIT)
New Delhi, India.
smriti@nsut.ac.in

Abstract— Soft Biometrics is a growing field that has been known to improve the recognition system as witnessed in the past decade. When combined with hard biometrics like iris, gait, fingerprint recognition etc. it has been seen that the efficiency of the system increases many folds. With the Pandemic came the need to recognise faces covered with mask in an efficient way- soft biometrics proved to be an aid in this. While recent advances in computer vision have helped in the estimation of age and gender - the system could be improved by extending the scope and detecting quite a few other soft biometric attributes that helps us in identifying a person, including but not limited to - eyeglasses, hair type and color, mustache, eyebrows etc. In this paper we propose a system of identification that uses the ocular and forehead part of the face as modalities to train our models that uses transfer learning techniques to help in the detection of 12 soft biometric attributes (FFHQ dataset) and 25 soft biometric attributes (CelebA dataset) for masked faces. We compare the results with the unmasked faces in order to see the variation of efficiency using these data-sets. Throughout the paper we have implemented 4 enhanced models namely - enhanced Alexnet, enhanced Resnet50, enhanced MobilenetV2 and enhanced Squeezenet. The enhanced models apply transfer learning to the normal models and aids in improving accuracy. In the end we compare the results and see how the accuracy varies according to the model used and whether the images are masked or unmasked. We conclude that for images containing facial masks - using enhanced Mobilenet would give a splendid accuracy of 92.5% (for FFHQ dataset) and 87% (for CelebA dataset).

Keywords- Soft biometric · face mask · deep learning · ocular · forehead · transfer learning

I. INTRODUCTION

The concept of biometric for identification is not new, as a matter of fact it dates back to the 1970s when it was used to identify people on the basis of their fingerprints. Over the years many new forms of biometrics were developed ranging from facial recognition to recognition on the basis of DNA, gait, hand geometry, keystroke etc. All of these constitute the hard biometrics. In the past 2 decades there has been seen an increase in the research in the area of soft biometrics as well. Soft biometrics recognition is based on the concept of how humans recognise each other i.e. on the basis of their physical and behavioral characteristics. So traits like height, weight, body

geometry, scars, facial hair, hair color, baldness all constitute as soft biometric attributes. With all the new research over the years it has been evident that the integration of these soft biometric features along with the hard biometric features proves to significantly improve the system of recognition. While the existing system of hard biometrics has reached a level of saturation, soft biometrics still remains a new field of identification that is capable of unraveling the solution to many problems that exist in biometrics currently.

With the pandemic, came the need to introduce a system that was able to identify people while they wore facial masks i.e. when only the area above their nose is accessible for recognition purposes. This posed a difficulty to the existing system of recognition that needed to be resolved. To tackle this problem we propose the use of ocular as well as forehead region of the face to train our models. Using haar-cascade, we process the images present in dataset in order to retrieve only the portion that is important to us and train our models on those images to recognise the soft biometric attributes of the person.

Result: ['Bags_Under_Eyes', 'Black_Hair', 'Bushy_Eyebrows', 'Male', 'Young']

Fig.1 Images containing facial masks before and after preprocessing along with the resulting soft biometrics

Apart from this, another problem that has been identified by researchers is working on low resolution devices. With the increasing use of mobile phones in order to take pictures and selfies and expansion of such pictures on social media, it has become quite evident that there needs to be a system which is able to recognise faces captured with the help of low power devices. These kinds of images can be blurry, have improper lighting and poor resolution in general that provides hindrance to the recognition system. We worked with different models (Table2), fine tuned them to increase the accuracy and tested on images captured via mobile phones in order to identify which enhanced model would give the best accuracy for low powered devices or even for blurry images.

```
[ 'Bangs', 'Heavy makeup',  
'Black_Hair', 'Oval Face',  
'Female', 'Young', 'Smiling' ]
```

```
[ 'Arched eyebrows', 'Wavy hair',  
'Smiling', 'Young',  
'High_cheekbones' ]
```

Fig.2 Pictures taken from low power devices after preprocessing along with their resulting soft biometrics

Soft Biometrics has found a niche in the facial recognition system[37] used nowadays. It is increasingly being used for situations where knowing the exact identifying detail isn't crucial, rather finding a particular segment of people is considered more important. For example websites that aim at a targeted portion of the population for selling their products (say according to age, gender etc.). Sometimes for studying the demographics of an area, knowing the ethnicity, age, gender of communities living in that area can prove to be more beneficial than finding the identity of each individual. Looking for only certain semantics in a video surveillance, say an old man with black hair and beard, can reduce the search to quite an extent and even using a low quality camera for surveillance purposes can satisfy the need to identify the person (or simply narrow down along with other information available) without knowing the exact identity of the person and serve the necessary purpose.

In the paper, the following information can be found in each section- Section II., introduces the readers to the datasets used in the project. In Section III., we have compiled a literature study of the work done on this particular area of soft biometrics and deep learning techniques. Further, in Section IV., we show our contributions to the area of research, and how it is different from the work done earlier in this field. Section V., discusses the methodology applied in our study and introduces the various models being used and how they are implemented in our paper to get the results. Section VI., is a compilation of the results obtained from running our models on the datasets introduced and discusses its significance. As we move ahead, in Section VII., we conclude our paper by bringing out the key points covered, its significance and its impact on the real world. We also discuss the future scope of the project and how it could be applied to real world scenarios to improve existing systems or be applied to new systems. We have also included the Compliance with Ethical Standards (Section VIII.) followed by Authorship Contribution (Section IX.). Lastly, we include in Section X., the references to previous studies that have been used throughout the paper.

II. DATASETS USED

There are two datasets that have been used in this paper on which our models are trained.

The first dataset used is CelebFaces Attributes (CelebA), characterized by the following features- The CelebA dataset is usually used in order to train and test models for face detection, especially for facial recognition of soft biometric attributes like hair, smile, eyeglasses. Images consists of different soft attributes, diverse faces, and a large number of pictures. It consists of -

- 202,599 face images
- 10,177 individuals
- 40 labels per image to give the soft biometric attributes

We have used CelebA images and reduced it to images containing ocular and forehead parts of the face using image preprocessing in order to represent images containing facial masks. For unmasked images we preprocess the images to contain the whole frontal face. Further all the enhanced models were applied to these processed images to find out the accuracy of detection of 25 facial soft biometric attributes selected out of the 40 available in the dataset.

The other dataset used in the paper is the Flickr-Faces-HQ (FFHQ) dataset. It consists of PNG images of human faces, generally used for the work revolving around GANs. It has the following features -

- 70,000 high quality images at 1024×1024 resolution that have considerable variation when looking at age, ethnicity and image background.
- Including many secondary attributes like eyeglasses

We have used FFHQ images and done a preprocessing similar to CelebA. All the enhanced models were applied to the processed images here as well for masked and unmasked images in order to detect 12 facial soft biometric attributes that are a subset of those found using celeba dataset so that the two results are comparable with each other. One of the added advantage of using this dataset is that it contains images of not only adults but also infants and children, a feature that isn't available in other datasets including CelebA.

The preprocessing so as to extract required facial features for both the datasets is done using OpenCV and Haar Cascade Classification. The required feature is the frontal face in case of unmasked images and ocular and forehead region for masked images. Haar Cascade Classification is divided into two parts, the first is the classification task that takes images and gives an output in the form of a binary value of 1 or 0, showing us whether the features present in the image or not. The second part is a face localization that includes an image as input and

gives the location of the facial feature within that image as an output in a bounding box with dimensions as width and height.

III. RELATED WORK

The majority of the research work in the field of soft biometric focuses on age[29,30,31], gender[32,33], and race[34] estimation from face images to date. Earlier, techniques like infrared iris images for finding out the age of people were used, which was possible by using the already existing iris datasets available[20,21]. Geometric or textual information was used in these studies, obtaining an accuracy of ~64%.

Meanwhile, in recent years facial age estimation approaches have shifted to the computationally efficient CNN design and Multitask learning.

Numerous recent research works are focused on the development of small and efficient neural networks suitable for systems with limited resources, for instance mobile devices. A common approach is reducing the amount of parameters in the convolutions, with the MobileNet[24,25], Shufflenet[26,27], and Xception[28] models utilizing depth-wise separable convolutions. **Rattani et al. [16, 18, 19]** was one of the pioneers who carried out recognition of age or gender from RGB ocular mobile devices images .

Fernando Alonso-Fernandez et al.(2021)[1] addressed how using images captured as a selfie with mobile phones contained the required ocular images to estimate soft biometrics such as age and gender, they used light-weight CNNs (of a some megabytes [53–56] to work with low quality images of the eyes) proposed within the context of the ImageNet Challenge, and extra architectures that were given for face recognition using mobile devices. 11,299 images of the Adience benchmark were employed to perform the experiments, which includes images clicked in the wild using smartphones that were uploaded to Flickr. **M Vasileiadis et al.[5]** proposed a computationally efficient CNN architecture, appropriate for real-time implementation on low power devices, which concurrently performs gender, age, race, eyes state, eyewear, smile, beard, and mustache estimation from unconstrained face images. The architecture employs Mobilenet and exploits the correlation between the individual biometric features. **Shervin Minaee et al.(2021)[2]** presented a comprehensive review of the recent advances that appeared between 2014 and 2019 employing deep learning networks in biometric recognition. For each step, they provided a comprehensive overview of latest contributions, loss functions and network architecture, evolved to achieve state-of-the-art performance in biometric recognition. It provides a survey of greater than one hundred twenty promising works on biometric recognition (which include fingerprint, face, iris, palmprint, ear, signature, gait recognition and voice). **Fernando Alonso-Fernandez et al.(2020)[4]** suggested the use of CNN's prebuilt architectures, previously trained in the context of the ImageNet Large Scale Visual Recognition Challenge, for the task of soft-biometrics classification with facial images. 12007 images were employed from the Labeled Faces in the Wild (LFW) database for the experiment. **Yiming Lin et al. (2021)[3]** propose an effective approach of exploiting face parsing semantics for age estimation and design a

framework to aggregate features from different levels of the face parsing network. **Peng Peng et al. (2020)[6]** proposed a software framework for PCA (principal component analysis) based face recognition [39] aiming at assisting software developers to customize their own applications efficiently due to its property of reducing data dimensionality without losing important information. **Jafri et al. (2020)** identified a deep neural networks-based face recognition approach to test human face detection with livenessNet. It searches for the position and dimensions of all features belonging to a class called a face. **Rondik J. Hassan & Adnan Mohsin Abdulazeez (2021)[7]** aimed to review a significant number of papers to cover recent innovations in the field of facial recognition. **Gao et al., (2020)[8]** suggested the Chinese facial ethnicity recognition (CFER) model based on the deep convolution network's transfer learning form. **Ahmed et al. (2020)[9]** proposed a comparative analysis using CNN-based models such as VGG16, VGG19, MobileNet, and AlexNet to capture the size of faces from a personalized dataset of (10) identities of various celebrities. **Liu et al. (2020)[40]** propose a network FER architecture based on a Sobel-enhanced CNN operator and a combined L2-SVM network. **Yao (2020)[42]** suggested a compact convolutional neural network for face recognition. Their work not only offers a compact basis for Deep learning techniques but also greatly increases the efficiency of face recognition. They use the Caffe toolkit for preparation and rollout in the implementation process. **Xu et al. (2020)[10]** designed a robust generalized deep convolutional neural network system face recognition. The structure has proven to increase the precision of training speed and face recognition and can be used on small-scale data sets. **Perdana & Prahara (2019)[17]** suggested a light-CNN to identify faces with a small dataset based on an updated VGG16 model. VGG16 has very deep-layers with many narrow convolution-layers with separate kernel numbers followed by max-pooling optimized for large-scale classification. The planned architecture uses (120 X 120) pixels as the input-image size and has two types of convolutional-layers, followed by max-pooling. Every Convolutionary layer is preceded by the activation feature of the rectified linear unit (ReLU). The proposed light-CNN is small but delivers good efficiency with 94.4% accuracy. **Zadeh et al. (2019)[43]** suggested a human emotion recognition system which uses the Gabor feature extraction filters and then the Classification Convolutional Neural Network (CNN). They used the JAFFE database containing 213 Japanese female model images containing seven face emotional states. The experimental findings indicate that the approach suggested improves both the CNN pace training mechanism and the precision of identification. **Khan et al. (2019)[14]** presented a face-recognition system using a ConvNet which is AlexNet. **Liu (2019)[15]** suggested Facial Expression Recognition (FER) system by using the fer2013 dataset and an effective deep convolution neural network to train an efficient model and then use the Tkinter name tool used by the Graphical User Interface (GUI) to evaluate the image of expression and achieve realistic performance. Also, AlexNet, VGG16, VGG19, and ResNet152 are used. **Yingying Wang(2019)[22]** proposes a method that combines multiple sub-regions and the entire face image by weighting, which can capture more important feature information that is conducive for improving the recognition accuracy. The proposed method was evaluated based on four well-known publicly available facial expression

databases: JAFFE, CK+, FER2013 and SFEW. **Eskandari et al. (2019)[12]** developed a multimodal system that used features from the face and ocular regions. They employ three hundred near sight infrared (NIR) images of CASIA-Iris-Distance as well as 405 VIS images of the MBGC database(divided into three equal parts for face,left eye and right eye images). As features, they used ULBP (with interlocking blocks), and Backtracking Search Algorithm (BSA) to reduce dimensionality. **Li & Zhu (2016)[13]** proposed a framework for smartphone face recognition. **Surya Prasad Rao Bora et al. (2020)[11]** suggest a cascade object detector based face detection and convolutional neural network alexnet based face recognition that can recognize the faces. **Jin et al.[41]** proposed the use of topological connections for reducing computational requirements. **Musab Coskun et. al. [23]** suggested a CNN for recognition of face with a number of convolutional layers. They have used the Georgia Tech Database and showed that the approach has improved the face recognition performance with better recognition results.

IV. CONTRIBUTIONS

As seen from above, the research done previously on soft biometrics is extensively being used in facial recognition systems. We can further notice that works related to identification of people wearing mask was required on an urgent basis after the pandemic and soft biometrics has been playing a substantial role in tending to this need.

The most recent studies ongoing is the use of soft biometric attributes recognition using the ocular data provided i.e. images of eyes are evaluated in order to identify the characteristics of people- mainly limited to age, gender and ethnicity. In addition to this, most of the earlier work makes use of the classic models available for use like CNN, Mobilenet etc. to identify the facial feature, including identifying soft biometric attributes of images taken from low powered devices.

- In our paper we aim at extending this idea further by introducing an additional modality. We preprocess the images available in the datasets (namely CelebA and FFHQ) to contain not only the ocular region but also the forehead region of the face. We are not using databases containing only these parts rather the full image so that we are able to understand if we need the face in its entirety or just the selected portions and the difference it would make. This helps us in expanding the scope of recognition and training our fine tuned models on these preprocessed images. Using large datasets containing as high as 200,000 images proved to be an additional advantage in training and testing our models.
- Further we have applied transfer learning techniques to the basic models available via python libraries like CNN(Alexnet), Resnet-50, MobileNet, SqueezeNet etc. We can see from the results obtained that having an enhanced model which contains 2 extra fully connected layers(after which the results get more or less stable) can significantly improve the values of accuracy and fbeta, which makes our overall results better.
- In the recent years we notice that many researchers have worked on finding a method to

identify soft biometric attributes however it has mostly been restricted to 3 labels namely-age,gender and ethnicity.Additionally in works that contain a few more attributes,the images are taken in unconstrained environments i.e. without facial masks being considered.We aim to devise a method for identifying around 25 soft biometric attributes (CelebA dataset) and 12 soft biometric attributes (FFHQ dataset) for identifying people that are wearing a facial mask.We further compare the accuracy and fbeta results with images that don't contain facial masks to get an idea of how the system is working.

Since this type of work is new and there is no previous work done of using ocular and forehead region and identifying people wearing facial masks by using models fine tuned by transfer learning, we have compared the results obtained on running our models on two datasets,both for masked and unmasked faces.Furthermore we have used 4 kinds of enhanced models in order to compare how the results vary and which model is best suited for our purpose.

V. PROPOSED METHODOLOGY

We worked on using Convolutional Neural Network (CNNs) to extract the presence of soft biometric features such as gender, bald, beard, blurry, eyeglasses, smile etc.. We have trained the CNNs for two different types of dataset, the first is original(unmasked) CelebA and FFHQ, and the second is modified(masked) CelebA and FFHQ. The modified CelebA and FFHQ have the same labels as the original, however the images are cropped to have only the ocular and forehead region. We have worked on extracting the required facial features using OpenCV and Haar Cascade Classification. Our work employs deep network architectures (Alexnet, Resnet50, MobilenetV2, Squeezenet), combined with two fully connected multi-part classification layers, in order to identify the presence of a number of facial soft biometric traits. We have used pre-trained models applying transfer learning of above mentioned models. The basic premise of transfer learning is simple: take a model trained on a large dataset and transfer its knowledge to a smaller dataset.

- AlexNet is a convolutional neural network, it contains eight layers, the first five are convolutional layers, these convolutional layers are sometimes followed by max-pooling layers. The last three layers are fully connected layers. The activation function used is ReLu, which has been proven to show better performance over tanh and sigmoid activation functions. The inputs are RGB images ($h \times w \times 3$, where h and w are the height and width of the images) , which go through five layers of filters, each of the filters producing an image as its output. After each filter, the image sizes are changed. The images produced by the last layer of filters can be seen as a 4,096-element one dimensional vector and used to classify and produce up to 1,000 output probabilities, for each of the required output classes.

Fig. 3 Architecture of Alexnet and its convoluted layers

- Residual Network (ResNet) is a Convolutional Neural Network (CNN) architecture that has made it possible to have up to thousands of convolutional layers, as it solved the “vanishing gradient” problem. These convolutional neural networks with large number of layers, outperforms the networks with lesser number of layers.

ResNet50 is one of its variants. It has

- 1) 48 Convolution layers
- 2) 1 Max-Pool and
- 3) 1 Average Pool layer
- 4) Fully Connected Layers

The building blocks of such models are residual blocks or identity blocks. A residual block is shown in Fig. 4, the activation of a layer is fast-forwarded to a deeper layer using skip connections in the neural network.

Fig. 4 Architecture of ResNet50 model

- SqueezeNet is a convolutional neural network made for reducing the parameters, markedly with the help of fire modules. These fire modules "squeeze" parameters using 1x1 Convolutions. It is a small and compact network. The number of parameters are 50 times lesser than AlexNet, and still its speed is 3x compared to that of AlexNet. Fire module consists of a squeeze convolution layer (which has only 1x1 convolution layers), feeding into an expand layer that has a mix of 1x1 and 3x3 convolution layers. Since these have a lesser number of parameters they can easily be fitted inside a computer's memory and be transmitted over a computer network.

Fig. 5 Architecture of Squeezenet along with its layers

- MobileNetV2 is a convolutional neural network architecture, based on an inverted residual structure. The residual connections exist between the bottleneck layers. The intermediate expansion layer uses lightweight depthwise convolutions (Dwise) to filter features and act as a source of non-linearity. The overall architecture of MobileNetV2 contains the initial fully convolution layers and residual bottleneck layers. It consists of two kinds of blocks, the first one is a residual block which has a stride of 1 and the next one with a stride of 2 which is present for the purpose of downsizing. There are 3 layers for both kinds of blocks-
 - Firstly, there exists a layer which provides 1×1 convolution with the help of ReLU6
 - The next layer provides a depthwise (Dwise) convolution and
 - The final layer gives one more 1×1 convolution which does not contain any non-linearity like the modified rectified linear unit (ReLU6) in the previous layer.

Fig. 6 Architecture of MobileNetV2 along with its layers

Fig. 7 (Description of the Model employed to identify the labels from ocular and forehead region of face. The exact configuration of convolutional model depends on which model out of the four is used)

Fig.8 (Description of the Model employed to identify the labels from frontal face region. The exact configuration of convolutional model depends on which model out of the four is used)

Network	Layers	Model Size	Time taken (per epoch)
AlexNet	8	12.6 MB	4 min 30 sec
SqueezeNet	18	9.1 MB	2 minute 40 sec
MobileNetV2	53	40 MB	1 minute 22 sec
Resnet50	50	114.8 MB	85 minutes

Table 1 Comparison of the layers,size and time taken per epoch for each model used

Network Architecture

Pretrained Convolutional Neural Network (CNN) models are available in various python libraries. These are trained on large scale image datasets. They can be further used to add more layers, fine-tune and predict results for other datasets.

We have added two fully connected layers to the above mentioned types of Convolutional Neural Network which are AlexNet, SqueezeNet, MobileNetV2 and ResNet50.

The input given to our models are either frontal face image or ocular and forehead images. The ocular and forehead images are used to check for accuracy in case of masked images. Our models require input images normalized in the same way, i.e. 3-channel RGB images of shape (3 x 128 x 128). The data loaders used convert the input image to tensors of size [3, 128, 128] pixels for our neural network model to use.

The output of the last convolutional layer passes through an average pooling layer in order to reduce the spatial dimensions of the feature maps to 1×1 , and transform them into one dimensional feature vectors. The classification stage encodes the soft biometric and image specific data. It receives as input the feature vector generated by the feature extraction stage and passes it to 12 labels for FFHQ Dataset and 25 labels for CelebA Dataset, one for each of the target facial soft biometrics. Hence, the output vector is one dimensional vector of size 12 or 25 depending on the dataset used, and has value one or zero signifying presence or absence of the label.

The loss function used in our CNNs is Flattened Loss. Flattened loss is a type of loss function where the input of the previous layer is flattened. Flatten function flattens the multi-dimensional input tensors into one dimensional outputs.

The metrics used in our paper are accuracy and fbeta. Let us understand these further -

Accuracy : Accuracy is a very useful metric for assessing classification models. Accuracy can be thought of as the portion of predictions our model got right. It is expressed as:

$$\text{Accuracy} = \text{True Positive} / (\text{True Positive} + \text{True Negative}) * 100$$

Fbeta: It is an extraction of the F-measure which has the balance of precision and recall in the calculation of the harmonic mean. The coefficient controlling it is called beta, which in our case is 0.5

$$\text{Fbeta} = ((1 + \text{beta}^2) * \text{Precision} * \text{Recall}) / (\text{beta}^2 * \text{Precision} + \text{Recall})$$

These metrics help us understand the system a bit better and act as a basis to compare the different models used. Along with comparison of accuracies of individual labels present in our multilabel classification we get to understand the role of each label and how it gets us the final overall accuracy.

VI. RESULTS AND DISCUSSION

Accuracy (in percentage) Metric comparison

Models	Celeba dataset		FFHQ dataset	
	Masked	Unmasked	Masked	Unmasked
Enhanced Alexnet	86.1681	90.3567	92.3476	94.0628
Enhanced Resnet	86.2178	91.0164	92.0406	94.4987
Enhanced Squeezenet	85.9956	88.0020	92.3448	94.0952
Enhanced MobilenetV2	87.0949	91.2958	92.5206	94.6991

Table 2 Comparison of accuracies using different models on both dataset

Fbeta comparison (beta = 0.5)

Models	Celeba dataset		FFHQ dataset	
	Masked	Unmasked	Masked	Unmasked
Enhanced Alexnet	0.776892	0.830623	0.914402	0.933433
Enhanced Resnet	0.777263	0.841470	0.913085	0.939176
Enhanced Squeezenet	0.776648	0.794722	0.914245	0.934025
Enhanced MobilenetV2	0.787939	0.844636	0.916279	0.940595

Table 3 Comparison of fbeta values using different models on both dataset

We can see in the above table the overall performance of the system based on accuracy and fbeta values and how it varies along with the model that is used. Moreover it is visible that there exists quite a gap between the values of the two datasets. However the values of accuracies and fbeta for the masked and unmasked images are found to be quite similar with minimal difference between the two which indicated the recognition system would work comparably well for both masked as well as unmasked face images.

Further we can evaluate the accuracies of individual labels that make up the overall accuracies of the recognition system and how they contribute individually to the results in the following tables for all the models implemented.

	squeezenet	alexnet	resnet4	mobilenet
Arched_Eyebro	74.56%	76.18%	74.48%	74.76%
Bald	98.37%	98.39%	98.39%	98.39%
Bangs	94.69%	95.38%	94.14%	94.19%
Black_Hair	76.89%	77.92%	77.40%	75.98%
Blond_Hair	87.74%	88.39%	87.61%	87.85%
Blurry	96.86%	97.10%	97.08%	96.86%
Brown_Hair	79.81%	79.88%	79.22%	79.38%
Bushy_Eyebrow	85.93%	88.39%	85.45%	85.52%
Chubby	95.62%	95.60%	95.56%	95.56%
Eyeglasses	99.52%	99.65%	99.22%	99.32%
Goatee	94.14%	94.19%	94.23%	94.23%
Gray_Hair	98.19%	98.19%	98.28%	98.30%
Heavy_Makeup	85.06%	87.13%	84.95%	84.21%
High_Cheekbon	65.22%	71.56%	68.45%	67.60%
Male	89.85%	94.14%	90.51%	90.16%
Mustache	96.60%	96.56%	96.58%	96.58%
Narrow_Eyes	92.75%	92.88%	92.84%	92.77%
No_Beard	86.13%	86.76%	86.17%	86.11%
Oval_Face	66.46%	67.40%	66.01%	65.77%
Pale_Skin	95.64%	95.54%	95.71%	95.47%
Sideburns	95.01%	95.03%	95.03%	95.03%
Smiling	64.61%	69.12%	67.31%	67.42%
Straight_Hair	78.27%	78.20%	78.07%	78.18%
Wavy_Hair	66.79%	67.05%	66.77%	65.46%
Young	86.32%	86.74%	86.00%	85.84%

Table 4 Masked CelebA - accuracy per label

	ResNet50	AlexNet	SqueezeNet	MobileNetV2
Arched_Eyebrows	74.48%	74.76%	74.56%	76.18%
Bald	98.39%	98.39%	98.37%	98.39%
Bangs	94.14%	94.19%	94.69%	95.38%
Black_Hair	77.40%	75.98%	76.89%	77.92%
Blond_Hair	87.61%	87.85%	87.74%	88.39%
Blurry	97.08%	96.86%	96.86%	97.10%
Brown_Hair	79.22%	79.38%	79.81%	79.88%
Bushy_Eyebrows	85.45%	85.52%	85.93%	88.39%
Chubby	95.56%	95.56%	95.62%	95.60%
Eyeglasses	99.22%	99.32%	99.52%	99.65%
Goatee	94.23%	94.23%	94.14%	94.19%
Gray_Hair	98.28%	98.30%	98.19%	98.19%
Heavy_Makeup	84.95%	84.21%	85.06%	87.13%
High_Cheekbones	68.45%	67.60%	65.22%	71.56%
Male	90.51%	90.16%	89.85%	94.14%
Mustache	96.58%	96.58%	96.60%	96.56%
Narrow_Eyes	92.84%	92.77%	92.75%	92.88%
No_Beard	86.17%	86.11%	86.13%	86.76%
Oval_Face	66.01%	65.77%	66.46%	67.40%
Pale_Skin	95.71%	95.47%	95.64%	95.54%
Sideburns	95.03%	95.03%	95.01%	95.03%
Smiling	67.31%	67.42%	64.61%	69.12%
Straight_Hair	78.07%	78.18%	78.27%	78.20%
Wavy_Hair	66.77%	65.46%	66.79%	67.05%
Young	86.00%	85.84%	86.32%	86.74%

Table 5 Unmasked CelebA - accuracy per label

	MobileNetV2	AlexNet	ResNet50	SqueezeNet
bald	97.96%	97.96%	97.96%	97.96%
beard	99.87%	99.87%	99.87%	99.87%
blur	79.34%	79.77%	79.77%	79.34%
female	85.77%	86.70%	84.90%	85.77%
forehead Occluded	95.91%	95.45%	95.45%	95.91%
glasses	99.16%	99.03%	98.93%	99.16%
smile	76.05%	74.41%	73.11%	76.05%
young	94.94%	95.11%	95.11%	94.94%
eyeOcclud	99.93%	99.93%	99.93%	99.93%
moustache	96.00%	96.00%	96.00%	96.00%
sideburns	99.97%	99.97%	99.97%	99.97%
invisible	79.77%	79.74%	79.77%	79.34%

Table 6 FFHQ Masked-accuracy per label

	AlexNet	MobileNetV2	ResNet50	SqueezeNet
bald	98.75%	98.57%	98.70%	97.96%
beard	99.66%	99.66%	99.58%	99.87%
blur	79.61%	79.90%	79.40%	79.77%
female	92.31%	94.78%	93.40%	84.90%
forehead Occluded	96.96%	96.78%	96.57%	95.45%
glasses	98.81%	98.70%	98.96%	98.93%
smile	89.40%	92.86%	91.71%	73.11%
young	94.44%	95.82%	95.53%	95.11%
eye_occlud	99.93%	99.93%	99.93%	99.93%
moustache	98.90%	98.90%	98.90%	98.90%
sideburns	99.71%	99.77%	99.74%	99.64%
invisible	79.61%	79.90%	79.40%	79.25%

Table 7 FFHQ Unmasked-accuracy per label

Our results indicate the possibility of providing soft-biometrics labels for masked images, that is using images containing only the ocular and forehead regions. The difference in overall accuracy for unmasked and masked data is not beyond 5.27% for any of the models. The maximum difference for CelebA Dataset is 5.27% and in case of FFHQ Dataset is 2.60%. The difference in FBeta (for beta = 0.5) between masked or unmasked and masked data is not beyond 6.47% for any of the models. The maximum difference for CelebA Dataset is 6.47% and in the case of FFHQ Dataset is 2.78%.

These overall accuracy are calculated by taking average of all accuracy per label, that is for 25 labels in case of CelebA Dataset. These labels are mentioned in table 8. For FFHQ Dataset, the labels are mentioned in Table 9.

'Arched_Eyebrows'	'Bald'	'Bangs'	'Black_Hair'	'Blond_Hair'
'Bushy_Eyebrows'	'Brown_Hair'	'Blurry'	'Chubby'	'Eyeglasses'
'Heavy_Makeup'	'Gray_Hair'	'Goatee'	'Sideburns'	'Smiling'
'High_Cheekbones'	'Mustache'	'Young'	'Male'	'Oval_Face'
'Straight_Hair'	'Narrow_Eyes'	'Pale_Skin'	'Wavy_Hair'	'No_Beard'

Table 8 Labels used for CelebA Dataset

'bald'	'young'	'sideburns'	'foreheadOccluded'
'glasses'	'beard'	'female'	'eye_occluded'
'moustache'	'invisible'	'blur'	'smile'

Table 9 Labels used for FFHQ Dataset

(a)

MobileNetV2, AlexNet, ResNet50 and SqueezeNet

(b)

Fig. 9 Accuracy of each label in Masked CelebA(a) and FFHQ (b)

(a)

(b)

Fig. 10 Accuracy of each label in Unmasked CelebA(a) and FFHQ (b)

The above figure helps us understand the difference between accuracies for different labels and understand which model and dataset work better for each other and for what label. For both the dataset and both kinds of input images, the model which has provided the best results is MobileNetV2. The maximum value for both FBeta and accuracy has been obtained for MobileNetV2.

As the number of soft biometrics attributes predicted in one go increases, the overall accuracy tends to decrease. This is expected as the overall result is the average of all the accuracies. A fair trade off for detecting more number of soft biometrics is decreasing the accuracy or vice versa.

Coming to the comparison of the number of parameters used in different models that we used we can see a pictorial representation in order to notice the differences between the four.

Fig. 11 Comparison of the parameters in the models used

We have compared the computational requirements and time for different models by comparing the number of layers, parameters and time taken per epoch for each of the models by referring to Fig 10 and Table 1.

However, the other models used in our paper should not be disregarded, as even if these do not provide the best accuracy or FBeta score value, the difference in accuracy is not beyond experimental limits and is limited to under 1.5% in all cases. Hence, we can use other models too depending on time and system constraints by compromising only a little on the efficiency of the model.

VII. CONCLUSION

In this study we have seen that enhanced MobileNetV2 gives the best accuracy in detection of soft biometric facial features for identifying people donning a facial mask. We compared our results for different models that have been enhanced by using transfer learning techniques for both images containing facial masks and without them as well. We calculated the accuracy and fbeta values (the fbeta value closest to 1 is considered to give the best results) for all the cases and conclude the following -

- MobileNetV2 gives the highest accuracy for FFHQ dataset with an accuracy of 92.52% and fbeta value of 0.91 for images containing facial masks
- MobileNetV2 gives the highest accuracy for FFHQ dataset with an accuracy of 94.69% and fbeta value of 0.94 for images not containing facial masks (Unmasked images)
- MobileNetV2 gives the highest accuracy for CelebA dataset with an accuracy of 87.09% and fbeta value of 0.78 for images containing facial masks
- MobileNetV2 gives the highest accuracy for CelebA dataset with an accuracy of 91.29% and fbeta value of 0.84 for images not containing facial masks (Unmasked images)

These results show that MobileNetV2 containing two extra fully connected layers (enhanced) has proved to give the best accuracy for both masked and unmasked images as well as for both the datasets. This can be attributed to the model's ability to consist of fewer parameters yet have a very high processing power and thus giving good accuracy. These kinds of models have an added advantage due to the use of techniques such as point-wise convolution, bottleneck layers, residual connections and even depth-wise separable convolution.

The difference in the results obtained on the two datasets can be attributed to the fact that the number of labels present in both and the type and number of images also vary. For the FFHQ dataset that gave a better result of 92 percent only 12 labels have been taken into account while on the other hand CelebA dataset which gives the accuracy of 87 percent is being used to predict 25 soft biometric attributes. Hence, we can conclude that there needs to be a tradeoff between accuracy and number of labels.

We also notice that the accuracy of unmasked faces is higher than the masked faces which is expected since there is a larger region of interest in unconstrained frontal face images used for unmasked face images as compared to only the eyes and forehead portion in masked face images. However it is noteworthy that the difference between the accuracies of the two is a minimal $\pm 2\%$, hence we can see that the system would get approximately similar results for both masked as well as unmasked images.

Therefore, we can conclude that this methodology could prove to be quite beneficial when compared to the current context where face engines are forced to work with images of people wearing masks, we evaluate the feasibility of using partial images containing only the ocular or mouth regions (Figure 1).

This work can have a wide array of implementation, especially due to the fact that it takes into account low power devices images, in the real world - it can serve to be beneficial in security surveillance systems[38], census systems to study the demographic of an area, identifying a particular segment of the people (e.g. according to age group) for a targeted commercial product etc.

Further we have considered the overall accuracy of the models which takes into account many sub classes (multilabel classification), however when considering the individual accuracies the accuracies could vary for every subclass due to class imbalance, that is the number of per label may vary. As a future scope of this project we can envision a number of areas where the scope of this project could further extend, for example creating clusters based on the labels, integrating it with a hard biometric such as gait, iris recognition etc. which would further enhance the face recognition system.

VIII. COMPLIANCE WITH ETHICAL STANDARDS

I. Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors

II. Funding details (In case of Funding)

This study is not funded by any agency.

III. Conflict of interest

The authors declare that they have no conflict of interest..

IV. Informed Consent

The work has been done completely on datasets available publicly and does not involve any other research on humans and/or animals.

IX. AUTHORSHIP CONTRIBUTION

- **Urja Banati:** Conceptualization, Methodology, Software, Writing– original draft, Writing –Review & Editing
- **Vamika Prakash:** Conceptualization, Methodology, Software, Writing– original draft, Writing –Review & Editing
- **Rashi Verma:** Conceptualization, Methodology, Writing– original draft, Writing –Review & Editing
- **Smriti Srivastava:** Guidance, Conceptualization, Supervision, Editing, Validation

X. REFERENCES

- [1] Fernandez,Alonso Fernando - et al.(2021) :Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images,In :International Conference of the Biometrics Special Interest Group (BIOSIG),1-19 ,IEEE,Germany.
- [2] Minaee, Shervin et al. (2021) : Biometrics Recognition Using Deep Learning: A Survey, 1-32.
- [3] Lin, Yiming et al. (2021):FP-Age: Leveraging Face Parsing Attention for Facial Age Estimation in the Wild,1-10.
- [4] Abdulazeez, Mohsin Adnan & Rondik, J. Hassan(2021): Deep Learning Convolutional Neural Network for Face Recognition: A Review.In : International Journal of Science and Business, IJSAB International, vol. 5(2), pages 114-127.
- [5] Vasileiadis, Manolis et al. (2020):Facial Soft Biometrics Detection on Low Power Devices In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 1-11,USA
- [6] Peng, Peng et al.(2020) :A face recognition software framework based on principal component analysis,1-46.
- [7] Fernando, Alonso-Fernandez1 et al. (2020) :Soft-Biometrics Estimation In the Era of Facial Masks , In: 2020 International Conference of the Biometrics Special Interest Group (BIOSIG),pp. 1-9, Germany .
- [8] Gao, S. et al. (2020):Facial Ethnicity Recognition Based on Transfer Learning from Deep Convolutional Networks. In:3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), 310–314.
- [9] Ahmed, T. et al. (2020):A Comparative Study on Convolutional Neural Network Based Face Recognition. In: 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1–5.
- [10] Xu et al.(2020):Research on Inception Module Incorporated Siamese Convolutional Neural Networks to Realize Face Recognition. IEEE Access, 8, 12168–12178.
- [11] Borra, Surya Prasad Rao et al.(2020): Face Recognition based on some Neural Network.In:International Journal of Engineering and Advanced Technology (IJEAT), 2249–8958, Volume 9,1-5.

- [12] Eskandari, M., Sharifi, O. (2019): Effect of face and ocular multimodal biometric systems on gender classification. *IET Biom.* 8(4), 243–248 .]
- [11] Li, Haifeng, & Zhu, X. (2016): Face recognition technology research and implementation based on mobile phone systems. In: 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 972–976.
<https://doi.org/10.1109/FSKD.2016.7603310>
- [14] Khan et al. (2019): Transfer Learning of a Neural Network Using Deep Learning to Perform Face Recognition. In: International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 1–5.
- [15] Liu, L. (2019): Human Face Expression Recognition Based on Deep Learning-Deep Convolutional Neural Network. 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA), 221–224.
<https://doi.org/10.1109/ICSGEA.2019.00058>
- [16] Rattani, A., Reddy, N., Derakhshani, R.: Convolutional neural network for age classification from smart-phone based ocular images. In: IEEE International joint conference on biometrics (IJCB), pp. 756–761. IEEE, (2017).
- [17] Perdana, A. B., & Prahara, A. (2019): Face Recognition Using Light-Convolutional Neural Networks Based On Modified Vgg16 Model. In: International Conference of Computer Science and Information Technology (ICoSNIKOM), 1–4.
- [18] Rattani, A., Reddy, N., Derakhshani, R.: Gender prediction from mobile ocular images: a feasibility study. In: IEEE International Symposium on Technologies for Homeland Security (HST), pp. 1–6. IEEE, (2017).
- [19] Rattani et al. : Convolutional neural networks for gender prediction from smartphone-based ocular images. *IET Biom.* 7(5), 423–430 (2018).
- [20] Erbilek et al. : Age prediction from iris biometrics. In: 5th International Conference on Imaging for Crime Detection and Prevention (ICDP), pp. 1–5. IET, (2013).
- [21] Sgroi, A., Bowyer, K.W., Flynn, P.J.: The prediction of old and young subjects from iris texture. In: International Conference on Biometrics (ICB), pp. 1–5. IEEE, (2013).
- [22] Yingying Wang et al. (2019): Facial Expression Recognition Based on Auxiliary Models,

pp. 1-15. Algorithms, 12(11), 227.

[23] Ozal Yildirim et al.(2017): Face recognition based on convolutional neural network. In: International Conference on Modern Electrical and Energy Systems (MEES)-Face recognition based on convolutional neural network. , 376–379,Ukraine
doi:10.1109/MEES.2017.8248937

[24] A. G. Howard et al.(2017) : Efficient convolutional neural networks for mobile vision applications,1-9.
arXiv preprint arXiv:1704.04861

[25] M. Sandler et al. : Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4510–4520, 2018.

[26] X. Zhang et al.: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6848–6856, 2018.

[27] N. Ma et al. : Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), pages 116–131, 2018.

[28] F. Chollet. Xception: Deep learning with depth wise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1251–1258, 2017.

[29] H. Han, C. Otto, and A. K. Jain:Age estimation from face images-Human vs. machine performance. In IEEE International Conference on Biometrics, pages 1–8. IEEE, 2013.

[30] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security, 9(12):2170–2179, 2014.

[31] G. Panis et al. :Overview of research on facial ageing using the fg-net ageing database. IET Biometrics, 5(2),37–46, 2016.

[32] E. Makinen and R. Raisamo: Evaluation of gender classification methods with automatically detected and aligned faces .IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(3):541–547, 2008.

[33] D. A. Reid et al. : Soft biometrics for surveillance: an overview. In Handbook of statistics, volume 31, pages 327–352. Elsevier, 2013.

- [34] H. He, S. Fu and Z.-G. Hou: Learning race from face-A survey. *IEEE transactions on pattern analysis and machine intelligence*, 36(12):2483–2509, 2014.
- [35] Ross,Arun,Chen,Cunjian and Banerjee, Sudipta(2019):Some Research Problems in Biometrics: The Future Beckons.In:International Conference on Biometrics (ICB), 1–8,Greece. doi:10.1109/ICB45273.2019.8987307
- [36] Gonzalez-Sosa, Ester et al. (2018):Facial Soft Biometrics for Recognition in the Wild-Recent Works, Annotation, and COTS Evaluation.*IEEE Transactions on Information Forensics and Security*, 1–14. doi:10.1109/TIFS.2018.2807791
- [37] Wu , Zhengdong et al.(2016):A twice face recognition algorithm,Volume 20 ,Issue 3, pp. 1007–1019,Springer,Heidelberg.
- [38] Gra',Manuel and Marques,Ion(2010):Face Processing for Security: A Short Review *Computational Intelligence in Security for Information Systems* , pp.89-96,Springer,Heidelberg.
- [39] Moon, Hae -Min et al.(2017) :A face recognition system based on convolution neural networks using multiple distance face,Volume 21 , Issue 17,pp 4995–5002,Springer,Heidelberg.
- [40] Liu, S., Tang, X., & Wang, D. (2020): Facial Expression Recognition Based on Sobel Operator and Improved CNN SVM. In: *IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP)*, 236–240. <https://doi.org/10.1109/ICICSP50920.2020.9232063>
- [41] Chan, S.-H.G. , Jin,Xing ,Tu,Wanqing (2008): Scalable and Efficient End-to-End Network Topology Inference. *IEEE Transactions on Parallel and Distributed Systems*, 19(6), 837–850. doi:10.1109/tpds.2007.70771
- [42] Yao, Li (2020):A Compressed Deep Convolutional Neural Networks for Face Recognition In: *IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pp. 144–149,China doi:10.1109/ICCCBDA49378.20
- [43] Taghi Zadeh, M. M., Imani, M., & Majidi, B. (2019): Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters. In:*5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, 577–581. <https://doi.org/10.1109/KBEI.2019.8734943>