

Article

An Intelligent Gender Classification System in the Era of Pandemic Chaos with Veiled Faces

Jawad Rasheed ^{1,*} , Sadaf Waziry ² , Shtwai Alsubai ³ and Adnan M. Abu-Mahfouz ^{4,5} ¹ Department of Software Engineering, Nisantasi University, Istanbul 34398, Turkey² Department of Software Engineering, Istanbul Aydin University, Istanbul 34295, Turkey; sadafwaziry@stu.aydin.edu.tr³ Department of Computer Science, College of Computer Engineering and Sciences in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia; sa.alsubai@psau.edu.sa⁴ Council for Scientific and Industrial Research (CSIR), Pretoria 0184, South Africa; a.abumahfouz@ieee.org⁵ Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg 2006, South Africa

* Correspondence: jawad.rasheed@nisantasi.edu.tr

Abstract: In the world of chaos, the pandemic has driven individuals around the globe to wear face masks for preventing the virus's transmission, however, this has made it difficult to determine the gender of the person wearing a mask. Gender information is part of soft biometrics, which provides extra information about a person's identification, thus, identifying a gender based on a veiled face is among the urgent challenges that must be advocated for in the next decade. Therefore, this study exploited various pre-trained deep learning networks (DenseNet121, DenseNet169, ResNet50, ResNet101, Xception, InceptionV3, MobileNetV2, EfficientNetB0, and VGG16) to analyze the effect of the mask while identifying the gender using facial images of human beings. The study comprises two strategies. First, the experimental part involves the training of models using facial images with and without masks, while the second strategy considers images with masks only, to train the pre-trained models. Experimental results reveal that DenseNet121 and Xception networks performed well for both strategies. Besides this, the Inception network outperformed all others by attaining 98.75% accuracy for the first strategy, whereas EfficientNetB0 performed well for the second strategy by securing 97.27%. Moreover, results suggest that facemasks evidently impact the performance of state-of-the-art pre-trained networks for gender classification.

Keywords: deep learning; facemasks; facial images; gender identification; pre-trained networks

Citation: Rasheed, J.; Waziry, S.; Alsubai, S.; Abu-Mahfouz, A.M. An Intelligent Gender Classification System in the Era of Pandemic Chaos with Veiled Faces. *Processes* **2022**, *10*, 1427. <https://doi.org/10.3390/pr10071427>

Academic Editor: Chien-Chih Wang

Received: 27 May 2022

Accepted: 14 July 2022

Published: 21 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gender classification is significant in several contexts. Gender information is part of soft biometrics, which provides extra information about a person's identification. Furthermore, it can increase facial recognition performance, which is considered one of the most useful biometric features and has more benefits than other biometric systems. As a result, it is frequently employed to deliver advanced analysis in human–computer interaction in several applications. Gender classification has been researched for decades and has attracted substantial attention from researchers and expanded fast owing to its usefulness in providing secure and dependable security for enterprises, organizations, face monitoring, airports, etc.

Gender detection can be easily applied in different areas while using different types of data such as voice [1], text data [2], speech [3], and images [4]. For instance, the authors of [1] explored gender classification based on a person's speech using audio pre-processing approaches that can be offered for the creation of efficient software systems for gender detection based on audio recordings. To benefit legal investigation, marketing analysis, advertising, and recommendation areas, Ref. [2] investigated an effective method

to classify gender from Twitter's tweet texts using natural language processing (NLP), bag of words, word embedding language processing techniques, and traditional machine learning algorithms such as support vector machine (SVM), naive Bayes (NB), and logistic regression (LG). To improve the accuracy of an emotion recognition system employing speech data, Ref [3] proposed a hybrid scheme by combining the random forest recursive feature elimination (RF-RFE) technique for the selection of features with the gradient boosting machine (GBM) approach for gender categorization. Moreover, studies such as [5,6] where facial recognition is performed using various machine learning classifiers also encourage scientists to expand the models to gender classification.

Coronavirus infection (COVID-19), which is a contagious disease that mostly spreads via direct or indirect contact with an affected individual [7], has driven individuals all over the world to wear face masks to prevent the virus's transmission [8], making it difficult to determine the gender of the person wearing a mask. The analysis of a human face in images without a facial mask is a tedious procedure since the human face in an image might vary owing to changes in position, orientation, and many additional factors such as photo resolution, lighting conditions, etc. To overcome this issue, this study employed deep neural network (DNN) models that produce accurate gender prediction results. Figure 1 shows the workflow of the proposed scheme for gender classification. The pre-processing phase removes duplicate and unnecessary data, eliminates the outliers, and resizes the images to a 299×299 fixed ratio. Later, the pre-trained networks utilize this imaging dataset to determine the gender of each individual wearing a face mask.

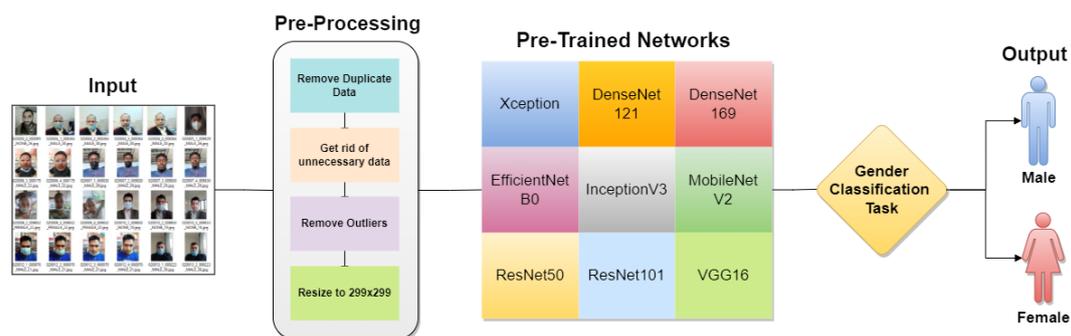


Figure 1. Workflow of proposed scheme.

Besides extracting relevant features, DNN models have excellent computer vision capabilities to perform image recognition tasks [9], for example, convolutional neural networks (CNNs) are often employed for visual image evaluation [10]. Deep learning (DL) algorithms, considered as subset of artificial intelligence (AI), focus on computer learning and improving on their own by analyzing several algorithms [11]. The algorithm takes an input image and assigns priority, learnable weights, and biases to distinct areas of the image in order to differentiate it from other images. Compared to other classification algorithms, a ConvNet requires substantially less pre-processing, thus this study exploits DL-based pre-trained models that use facial images of humans with different types of face masks. Broadly, the major contributions are outlined as follows:

- An extensive review to show that gender detection using face images with masks is in its infancy.
- To ensure smoothness, this work employs nine various deep-learning pre-trained networks.
- To analyze the effect of wearing facemasks on gender classification, the study deploys two strategies: models trained with images where humans wear face masks, and an extended version that incorporates facial images without masks.
- To lessen the false positivity rate, it exploits a technique to remove outliers.
- To check the robustness, this work is applied to unseen images and the performance is measured through several performance metrics.

The following is the structure of the paper: Section 2 examines the numerous studies that have been conducted on the subject. Section 3 outlines the suggested technique. Section 4 summarizes the experiment's findings and provides a comparison/discussion, while Section 5 discusses the conclusion.

2. Literature Review

Gender categorization has been studied extensively using a variety of methodologies and approaches. For instance, the authors of [4] presented gender prediction for a facial images or real-time video using CNNs. They carried out face detection, cropping, and resizing as pre-processing steps in this study, and proposed three CNN-based models with different architectures created. They found that using a CNN model with a deeper network (more layers) produces the best results. The authors of [12] aimed to classify a person's gender and emotions in real time or by using the person's image on a smartphone or a hard copy of the picture. Gender detection was achieved by obtaining the softmax value to identify gender using real-time CNNs. It is underlined that during gender identification, the cardinal and essential item to consider is face detection, as well as appropriate and relevant classification of facial features in low lighting and imperfect situations.

Similarly, the authors of [13] also proposed a CNN to classify gender and age by training on ten thousand grayscale human facial images. They exploited a pre-trained ResNet model for training, while the leading process of their study identified faces using the Haar cascade frontal face as the default classifier that then classified gender based on those faces. To find the best features of the iris for classifying gender using NIR images, [14] has proposed five different experiments: using whole features from normalized images, using a transfer learning approach with a VGG19 model, selecting the most predominant blocks using a genetic algorithm (GA), selecting the most predominant pixels using p -values, and encoding the images using a quaternionic code with 4 bits per pixel. Based on the experimental results, they claimed that the quaternionic code-based scheme outperformed others by securing the highest accuracy, 95%, for the right iris and 93% for the left iris, with 2400 selected features.

The authors of [15] examined the performance of gender classification using deeper CNNs trained on different facial components. The results demonstrated that their proposed strategy worked well with larger crop sizes as it achieved promising efficiency. In comparison to eyes, their research showed that the proposed approach can accurately determine gender from the mouth, nose, and face. The authors of [16] provide a gender classification technique that combines image processing techniques and data mining methodologies. The system performs typical image processing procedures such as acquisition, pre-processing, feature extraction using the LBG vector quantization approach, and classification using data mining methods such as naive Bayes, SVM poly kernel, SVM radial basis function (RDF) kernel, and k-nearest neighboring (kNN). All of the classifiers' classification findings demonstrate that the male classification rate is higher than the female rate.

A fast gender classification method from frontal facial images using features selected from the mouth and chin has been proposed by [17] using two standard classifiers: SVM and a probabilistic neural network (PNN). The method performs gender classification by extracting the lower part of frontal face images using the geometric model method, builds a gray level co-occurrence matrix (GLCM) from the retrieved image, extracted features from the GLCM, and classified the face based on gender. The result shows that SVM outperforms the PNN with 94.34% accuracy. Likewise, the authors of [18] investigated a gender classification method based on multi-level local phase quantization (ML-LPQ) features derived from normalized face images using an SVM model with a non-linear kernel (RBF) classifier. As a result, their strategy outperformed other state-of-the-art approaches.

The local directional pattern (LDP) is a distinct texture description that provides a robust feature to characterize facial appearance. For instance, the authors of [19] introduced the LDP to explain a gender recognition facial picture by dividing the facial regions into tiny parts to collect LDP histograms and later combined them into a single feature vector

from various sections. They used SVM to perform classification, which outperformed many traditional pattern classifiers in gender classification tasks. In addition, the authors of [20] provided experimental research that applied wavelet transform for gender categorization for the first time. It decomposed facial images using a 2-D discrete wavelet transform (DWT). Moreover, they incorporated fisher linear discriminant (FLD) and principal component analysis (PCA) to decompose coefficients for feature reduction and gender classification. They determined the accuracy rate of their approaches by employing 10-fold cross-validation methodology. As a result, they discovered that the nose on the face is the most distinguishing feature.

The authors of [21] introduced a unique technique based on the spectral angle mapper (SAM) that can efficiently gather spectral information across many spectral bands and classify it using the linear SVM. By measuring the photometric property of the acquired image, they investigated the feasibility of extended multi-spectral imaging for gender categorization. They also tested the approach on the extended multi-spectral face database, made using six different illuminations. A strategy based on the multi-scale facial fusion feature (MS3F) has been presented in [22] to predict gender from faces using SVM as a base classifier and local phase quantization (LPQ) and local binary pattern (LBP) as feature descriptors to extract the feature from facial images. In terms of accuracy, they compared their work to state-of-the-art approaches and attained a superior result. To develop an architecture that can be implemented on mobile devices, smart device developers presented a lightweight multi-task CNN (LMTCNN) architecture in [23] for simultaneous gender classification.

Besides these, [24] offered a single picture gender categorization technique that contained characteristics based on appearance and geometry, such as the LBP, discrete cosine transform (DCT), and the geometrical distance feature (GDF). They tested the approach on two datasets and produced extremely high accuracy in both. The authors of [25] developed a new method for classifying gender from face images by using the LBP as a binary quantization and GLCMs to extract the geometric structure of the faces. Furthermore, they utilized a histogram equalization technique to adjust the contrast of the input image, and SVM as the classifier for gender classification. As an outcome, the use of both LBP and GLCM features showed high classification performance. Similarly, [26] exploited an image processing and AI-based technique to perform age and gender determination using dental X-ray images. They pre-processed the tooth images followed by binary conversion using the M-1, M-2, and M-3 approaches. The dynamic structure divides the images into segments, extracts the features to have vectors, and then feeds them to a multi-layer neural network to determine age and gender.

Generally, people find gender classification to be an easy procedure, however it is still a tough assignment for computers, especially when assessing a human face with a facial mask, as most of the main facial features such as the nose, chin, and mouth are not visible. Additionally, as per our knowledge and research, the challenge of detecting a person's gender while wearing a face mask has not been solved to date. As a result, we offer a technique for performing gender classification using DL networks in this research that not only determines gender using facial images but it also figures out the gender even when a human is wearing a mask.

3. Proposed Scheme and Dataset Details

In this study, the gender detection technique is divided into three phases: data pre-processing, model training, and gender classification. During the pre-processing step, approaches such as removing duplicate data, eliminating irrelevant data, removing outliers, and resizing images into dimensions of 299×299 are exploited. The data, which consist of images of people wearing face masks, are fed into several deep learning-based gender classifiers that have been pre-trained. These models include DenseNet121, DenseNet169, Xception, InceptionV3, ResNet50, ResNet101, VGG16, MobileNetV2, and EfficientNetB0 that

are fine-tuned and trained to determine a person's gender based on facial images with and without face masks.

3.1. Dataset

The dataset is downloaded from a publicly available Kaggle repository [27]. The dataset used to train our gender detection models contains 40,000 images, which after removing duplicates and unnecessary images contained 11,536 total images for four ways of masks wearing; the types are: the mask is properly worn and covers the nose and mouth, the mask covers the mouth but not the nose, the mask is on but does not cover the nose or mouth, and there is no mask on the face (see Figure 2a). There are 8691 total images for 3 types of masks worn which include all the above mentioned types, except the type with no mask on the face (see Figure 2b). Each item comprises the following information: image size, image type, person's age, gender, and user ID. All photos were gathered using the crowdsourcing site Toloka.ai and confirmed by TrainingData.ru. We use the 4 and 3 types of these images separately with a ratio of 70/30 (see Table 1) for training the classifiers in order to precisely evaluate the accuracy of each model.

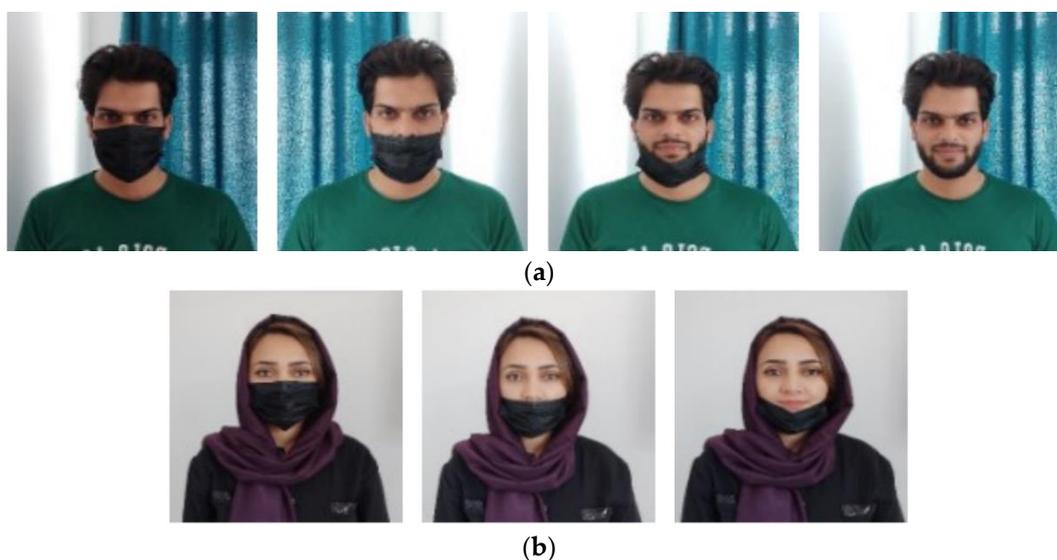


Figure 2. Samples of facial images of individuals wearing masks in: (a) 4 different ways/styles (wearing the mask fully, half, partially, or not wearing one at all); (b) 3 different ways/styles (wearing the mask fully, half, or partially).

Table 1. Face mask image dataset training and testing split.

Dataset Type/No. of Classes	Training Set	Test Set	Total
4 ways of wearing mask	8075	3461	11,536
3 ways of wearing mask	6083	2608	8691

3.2. Exploratory Data Analysis and Pre-Processing

Before feeding data to DL models, the data samples need to be in order for better performance. Thus, the study performs 4 steps as data pre-processing; the first step includes removing the duplicate images and filtering out the data that have more than 4 images of a single person with the same type of mask wearing. In the second step, we removed unnecessary data, as there were some records that the gender was neither male nor female and this was indicated as NONE so we used the pandas built-in query feature for this step. The third step includes removing outliers of some fields of data samples, for instance, the size of images and age of each person by finding the median value of these fields and removing values that exceed the median. Figure 3 depicts the visualization of removing

outliers from the size of images by finding the median value which is 4 in this case and removing those records from dataset that exceed the median value. The last step comprises resizing the images to the ratio of 299×299 in order to feed them into the pre-trained models of fixed width and height.

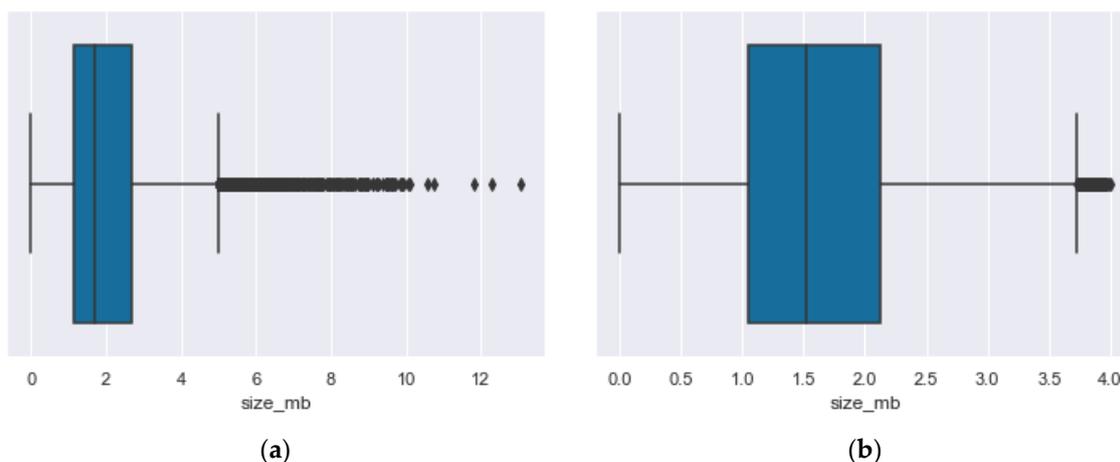


Figure 3. This visualization of removing outliers from the dataset: (a) 4-ways of wearing masks; (b) 3-ways of wearing masks.

3.3. Deep Neural Network Models

3.3.1. DenseNet

In a feed-forward approach, dense convolutional network [28] links each layer to every other layer. Each layer generates a feature map that is an input to the next layers. It is composed of two critical components: blocks of dense and transition layers. Each DenseNet topology has four dense blocks, each having a different number of layers. DenseNet-121 contains four dense blocks with 6, 12, 24, and 16 layers, respectively, whereas DenseNet-169 has four dense blocks with 6, 12, 32, and 32 layers, respectively, and more than 20 million parameters. DenseNets address the vanishing-gradient issue, increase feature propagation, and improve feature reuse while using fewer parameters than typical CNNs since they do not need to learn unnecessary feature mappings. Equation (1) shows the DenseNet architecture output where $[z_0, z_1, \dots, z_{l-1}]$ is the combination of the feature maps generated by the $[0, 1, \dots, lth]$ layers.

$$z_l = H_l([z_0, z_1, \dots, z_{l-1}]) \quad (1)$$

3.3.2. ResNet

Residual networks are a type of traditional neural network, utilized as a foundation for many computer vision applications. In 2015, the model won the ImageNet challenge [29]. This innovation empowered experts to effectively train incredibly DNNs having more than 150 layers. Prior to ResNet, building DNN models with a huge number of hidden layers was a challenging task due to vanishing gradients. However, ResNet established the notion of skip connections, which alleviated the problem of vanishing gradients by permitting the gradient to flow through a substitute shortcut direction and allowing the model to learn an identity mapping that ensures the higher layer performs well at the lower layer, if not better. Without the skip connection, the input, y , is multiplied by the layers' weights, w , followed by addition of a bias term, b . The activation function, $f(y)$, is then utilized, to obtain the resultant, $H(y)$, as in (2).

$$H(y) = f(wy + b) \text{ or } H(y) = f(y) \quad (2)$$

However, the output has been altered from $H(y)$ to (3) following the advent of the skip connection technique.

$$H(y) = f(y) + y \quad (3)$$

Furthermore, when using a convolutional layer or pooling layers, the input dimension may differ from the output dimension. Therefore, padding a zero using the skip connection to expand its dimensions, and appending 1×1 convolutional layer to the input to meet the dimensions, can handle the problem at hand. Thus, a resultant can be enhanced by adding w_1 as an extra parameter, given in (4).

$$H(y) = f(y) + w_1 \cdot y \quad (4)$$

ResNet50 refers to the variation that can function with 50 neural network layers, whereas ResNet101 refers to a CNN with 101 layers.

3.3.3. InceptionV3

The InceptionV3 is a 48-layer DL model based on CNNs, used for image classification [30]. The InceptionV3 is an improved version of the fundamental model InceptionV1, which was launched in 2014 as GoogLeNet. It is a widely used image recognition model that has shown promising results by achieving more than 78.1% accuracy on the ImageNet dataset. It is also intended to function well even under stringent memory and computational budget limitations. The inception layer is a mixture of all 1×1 , 3×3 , and 5×5 convolutional layers, with their output filter banks concatenated into a unified output vector that serves as the following stage's input.

3.3.4. Xception

Xception is an Inception architecture enhancement that substitutes ordinary Inception modules with depth-wise invertible convolutions [31]. It forms a feature extraction basis of the network with 71 convolutional layers. Except for the first and last modules, the convolutional layers in Xception are organized into modules surrounded by linear residual connections, thus, it is a depth-wise separate convolution layer stack with residual connections. As opposed to InceptionV2 or V3, which are significantly more complicated to specify, this makes the architecture relatively straightforward to define and adapt. Xception's overall architecture consists of three flows: entry, middle, and exit flow. The data first pass via the input flow, then through the middle flow, where they repeat themselves 8 times, and lastly through the exit flow. The limit of convolutions for a single kernel is set according to (5), while the limit for N kernels is set according to (6), where K is the resultant dimension after convolution, which depends on the padding employed. C denotes the number of channels, while d reflects the size of the convolution filter.

$$K^2 \times d^2 \times C \quad (5)$$

$$K^2 \times d^2 \times C \times N \quad (6)$$

3.3.5. VGG16

VGG16 is a CNN-based architecture model that won the 2014 ILSVR (ImageNet) competition [32]. It consists of a 16-layer DNN with around 138 million parameters employing 3×3 filters. Rather than a huge number of hyper-parameters, VGG16 concentrates on 3×3 filtered convolution layers with a stride of 1 and a 2×2 filtered maxpool layer with 2 strides having the same padding. The overall architecture is composed of convolutional layers having different depths followed by three fully connected layers; 4096 channels in first two layers, 1000 channels in the third layer, and the last layer is the soft-max layer.

3.3.6. MobileNetV2

MobileNetV2 is a 53-layer deep network that is particularly good at object recognition and segmentation. It improves the state-of-the-art effectiveness of mobile models in a number of activities and benchmarks, as well as across a range of model sizes [33]. Its foundation is an inverted residual structure with residual connection between bottleneck levels. It enables real-time categorization under processing restrictions in devices such as smartphones. This methodology allows the use of ImageNet transfer learning on our dataset. More detail about the overall architecture of MobileNetV2 can be found in [33]. Each layer has n -times repeating sequences of one or more identical (modulo stride) layers. The number of output channels is the same for all layers in the same sequence. Each sequence's initial layer has a stride, while the rest utilize stride of 1. Moreover, 3×3 kernels are used in all spatial convolutions.

3.3.7. EfficientNetB0

EfficientNet is a scaling approach that uses a compound coefficient to consistently scale all depth, resolution, and width dimensions [34]. The base EfficientNetB0 network is built on the inverted bottleneck residual blocks of MobileNetV2, as well as squeeze-and-excite blocks, and has 237 layers made up of 5 modules. It significantly outperformed other convolutional networks, in fact, EfficientNetB7 obtains a new state-of-the-art 84.3% top accuracy. In a nutshell, ϕ is a user-defined coefficient that affects the number of additional resources accessible. The variables α , β , and γ define how these extra resources are distributed among networks in terms of depth d , width ω , and resolution of the input r as shown in (7)–(9), where $s \cdot t \cdot \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$ and $\alpha \geq 1$, $\beta \geq 1$, $\gamma \geq 1$.

$$d = \alpha^\phi \quad (7)$$

$$\omega = \beta^\phi \quad (8)$$

$$r = \gamma^\phi \quad (9)$$

To conduct gender classification, we feed the final output tensor from the convolutional base into a dense layer to fine-tune these pre-trained models for training with the dataset at hand. The dense layer accepts one-dimensional vectors as input and produces a three-dimensional tensor as output. We begin by flattening (or unrolling) the 3D output to 1D, and then, because our data will be categorized into two classes, male and female, we add a final dense layer with two outputs and a softmax activation function. To train and evaluate the models, the initial dataset is separated into training and test sets with a 70/30 ratio. After successful training, the accuracy is computed using all images from the test dataset in each iteration.

4. Results and Discussion

As this paper emphasizes a solution to gender classification with face masks, it is necessary to evaluate processing and classification performance. The training and testing of models are executed on an AMD Ryzen 5 5600G processor having a RAM of 64 GB. Software includes: Jupyter Notebook having Keras packages with Python 3.6. In addition, the system contains NVIDIA GFORCE GTX 1050. The network was implemented with a TensorFlow framework, fine-tuning nine Keras applications utilizing two different datasets which contain four types and three types of facial images with and without face masks per person, respectively. To evaluate the models' accuracy, we tested each model with a dataset containing 40,000 images. Moreover, we determined the following six metrics to analyze the effectiveness and performance of the exploited model.

- **Accuracy:** determines how many observations, both positively and negatively, are properly categorized; it represents the proportion of true predictions obtained by each model.

- **AUCRAC:** the term AUC refers to the area under the curve, which is a threshold and scale invariant that determines the rank correlation of predictions and targets.
- **PRAUC:** the mean of the accuracy scores computed for each recall threshold.
- **Precision:** calculates the proportion of cases that are properly categorized.
- **Recall:** calculates the quantity of positive class predictions produced from all positive instances in the dataset.
- **F1-score:** computes the harmonic mean between precision and recall by integrating these into a single measure.

Table 2 presents the comparative performance analysis of exploited state-of-the-art DL models for gender classification on the test dataset when trained with facial images of individuals wearing facemasks in four different ways. Besides just testing models trained on a dataset that also contains facial images without facemasks, identical models are also trained with images that contain facial images with masks only (three ways of wearing masks). Table 3 compares the performance of each model for gender classification on the test dataset when trained with images with veiled faces. Among these predicted accuracies using CNN pre-trained models for gender classification, evidently, the most efficient ones are DenseNet121, Xception, EfficientNetB0, and InceptionV3 as these four models show the highest accuracy and lowest loss among all others in both datasets.

Table 2. The performance analysis of various models for gender classification on test dataset when trained with facial images where individuals either wear the mask fully, half, partially, or do not wear one at all (4 ways/styles of wearing).

Model	Loss	Accuracy	AUCRAC	PRAUC	Precision	Recall	F1-Score
DenseNet121	0.01	98.50	97.93	97.13	98.58	99.24	98.91
DenseNet169	0.08	97.58	96.47	95.45	97.83	98.62	98.22
ResNet50	0.08	97.58	96.60	95.28	97.69	98.76	98.22
ResNet101	0.10	97.08	96.27	94.66	97.86	97.81	97.83
Xception	0.05	98.33	97.69	96.00	98.58	98.95	98.76
InceptionV3	0.04	98.75	98.18	97.60	98.95	99.19	99.07
MobileNetV2	0.31	94.83	92.30	90.00	95.71	96.71	96.21
EfficientNetB0	0.15	97.50	96.42	95.39	98.10	98.24	98.17
VGG16	0.36	84.75	77.83	72.86	87.55	90.38	88.94

Table 3. The performance analysis of various models for gender classification on test dataset when trained on images with veiled faces (faces with masks in three different ways/styles of wearing).

Model	Loss	Accuracy	AUCRAC	PRAUC	Precision	Recall	F1-Score
DenseNet121	0.23	96.35	95.68	93.01	96.17	98.90	97.52
DenseNet169	0.47	93.78	91.01	88.32	94.85	96.69	95.77
ResNet50	0.25	94.44	91.37	89.27	95.79	96.59	96.19
ResNet101	0.33	93.78	90.42	88.23	95.09	96.58	95.83
Xception	0.17	95.59	92.84	91.16	95.34	98.79	97.04
InceptionV3	0.21	94.05	90.86	88.44	94.69	97.32	95.99
MobileNetV2	0.36	93.44	92.41	87.85	94.37	96.80	95.57
EfficientNetB0	0.13	97.27	96.02	94.59	96.98	99.37	98.16
VGG16	0.56	74.50	54.92	53.85	80.40	86.10	83.15

After successful training, we also computed accuracy and loss using all images from the test dataset in each iteration. Figures 4 and 5 show the visualization of both training and validation accuracies with the increase in the number of training iterations with both type of datasets (four and three types of mask images) for each of our pre-trained models. To gain optimal performance, the experimental work exploited an early stopping, thus the models stop after 10 epochs.

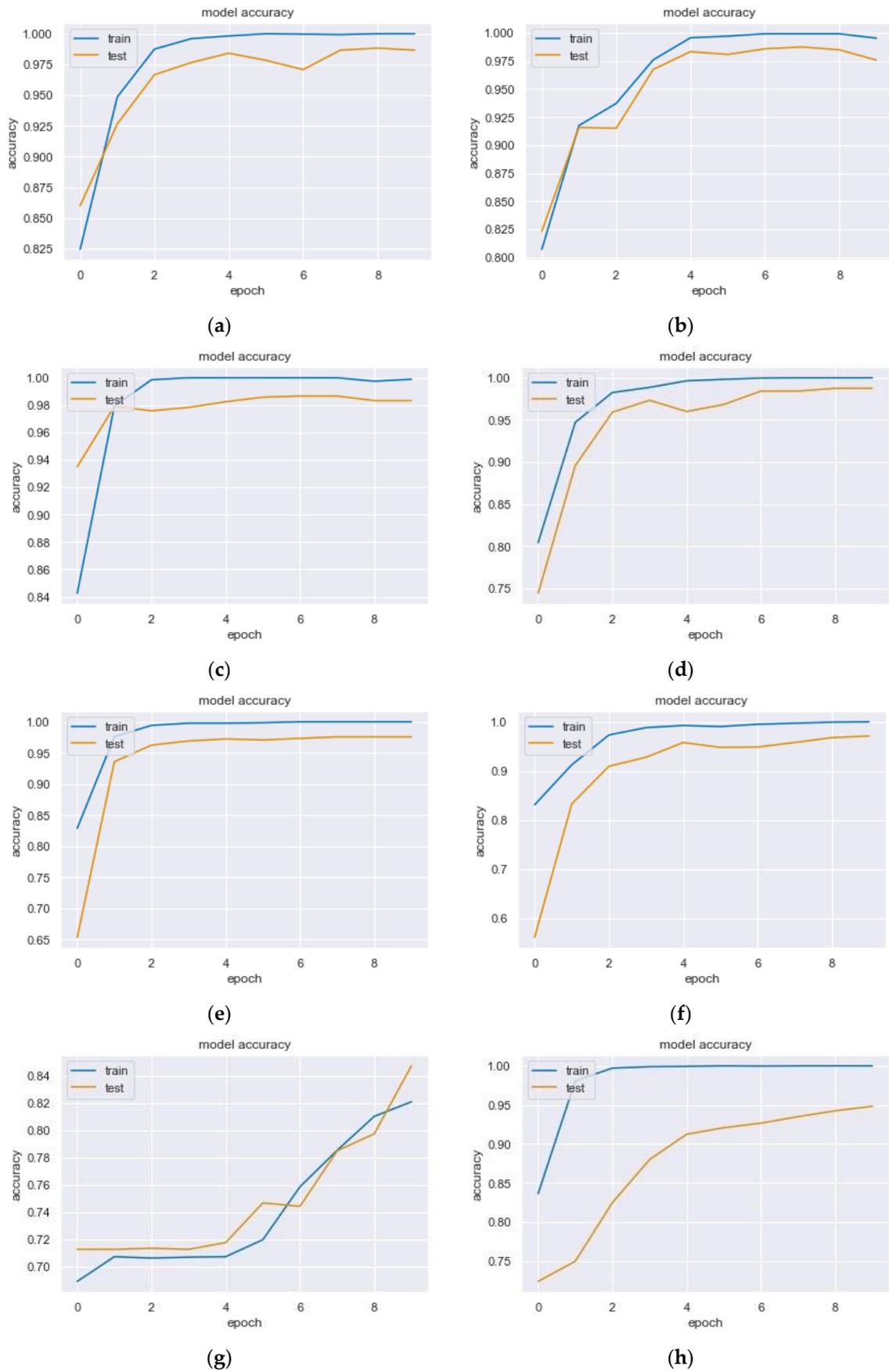
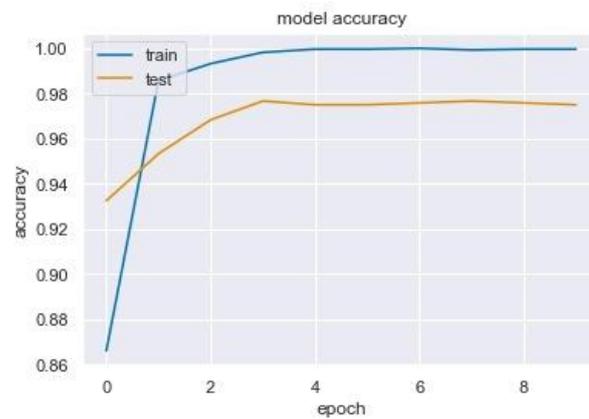
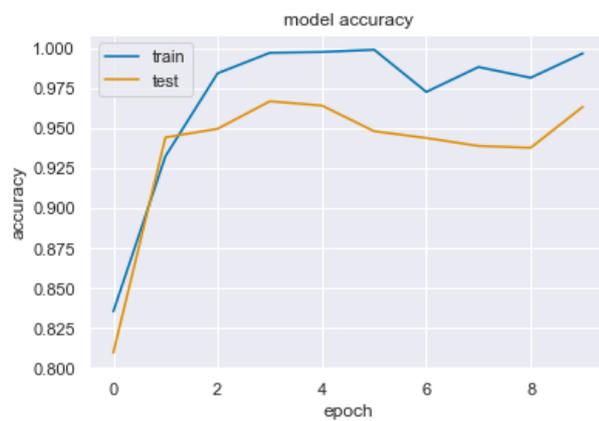


Figure 4. Cont.

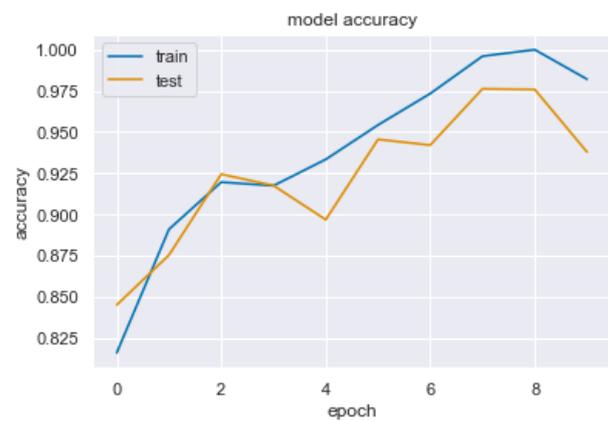


(i)

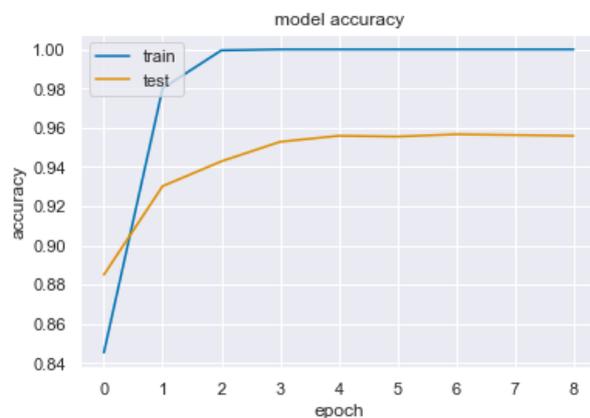
Figure 4. Performance curves of various models for gender classification when trained with facemask images where individuals either wear the mask fully, half, partially, or do not wear one at all (4 ways/styles of wearing). Accuracy curves for (a) DenseNet121; (b) DenseNet169; (c) Xception; (d) InceptionV3; (e) ResNet50; (f) ResNet101; (g) VGG16; (h) MobileNetV2; (i) EfficientNetB0.



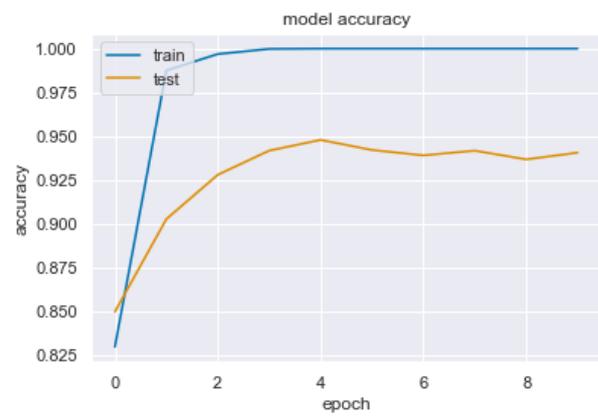
(a)



(b)



(c)



(d)

Figure 5. Cont.

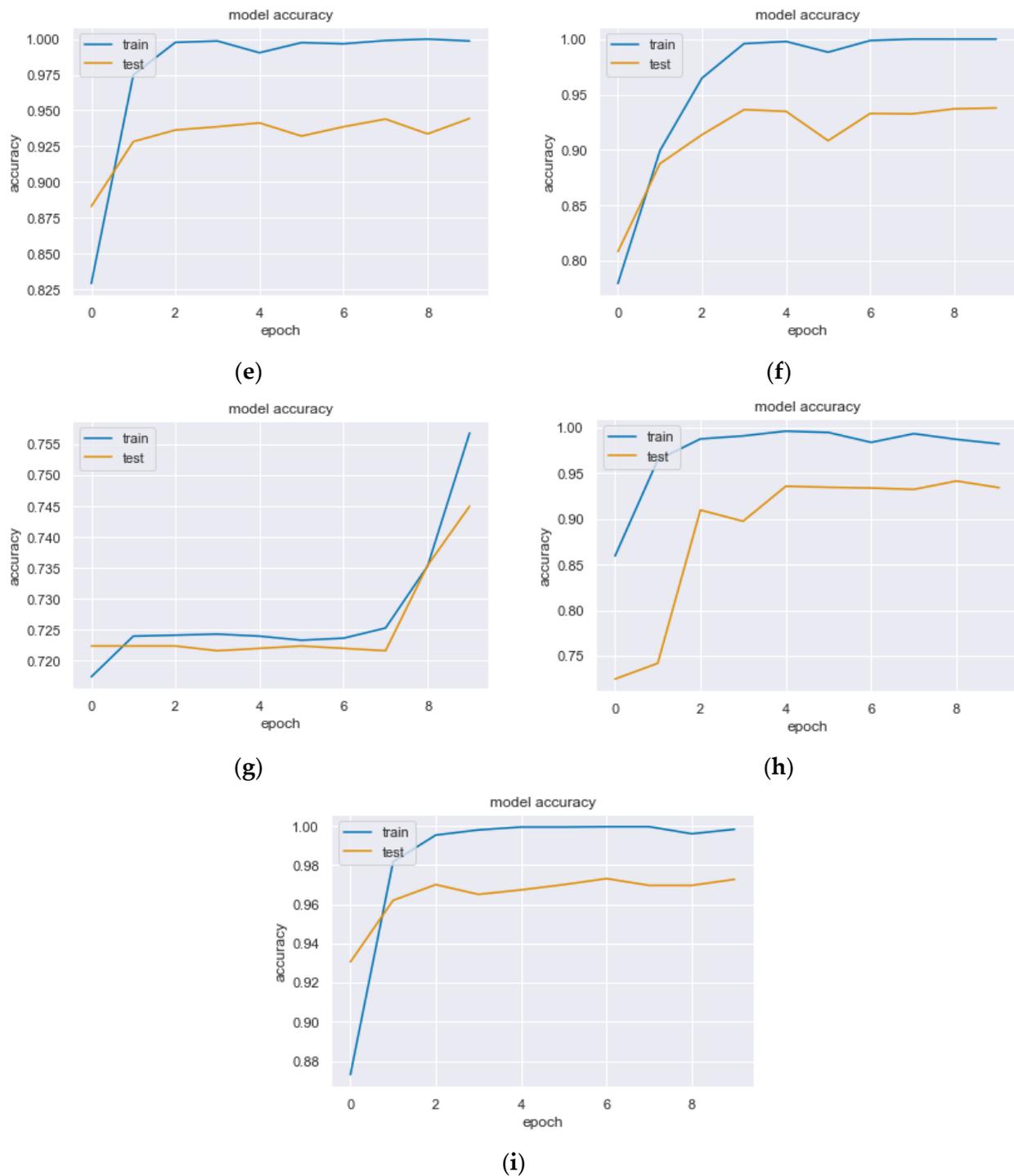


Figure 5. Performance curves of various models for gender classification when trained with facemask images where individuals either wear the mask fully, half, or partially (3 ways/styles of wearing). Accuracy curves for (a) DenseNet121; (b) DenseNet169; (c) Xception; (d) InceptionV3; (e) ResNet50; (f) ResNet101; (g) VGG16; (h) MobileNetV2; (i) EfficientNetB0.

Besides analyzing the trained model on the test dataset, we also tested the exploited models on unseen data. It is noted that the top models (DenseNet121, Xception, EfficientNetB0, and InceptionV3) successfully classified the given unseen facial images (as shown in Figure 6), however, they misclassified a few unseen images where the background was too complex or more facial features were hidden due to wearing a cap or scarf. The pictures show that the applied method gives satisfactory classification results.



Figure 6. Results obtained when different input images were tested on our gender detection system. (a) image of a male, predicted as male; (b) image of a male, predicted as male; (c) image of a female, predicted as female; (d) image of a female, predicted as female.

Moreover, we also measured the computational time of each model for both datasets (4 ways and 3 ways of wearing the masks). Table 4 lists the computational time to train and fit the exploited state-of-the-art DL models for gender classification. It is worth noting that the DenseNet121 model efficiently performed the training in less computational time as compared to other models for both datasets, while EfficientNetB0 was successful in training for the dataset with 3 ways of wearing the masks.

Table 4. The computational time to fit the models for gender classification over two datasets (4 ways of wearing mask, and 3 ways of wearing mask).

Model	Time (Minutes)	
	4 Ways of Wearing Mask	3 Ways of Wearing Mask
DenseNet121	28.3	25.9
DenseNet169	30.9	26.7
ResNet50	29.0	26.3
ResNet101	32.3	29.1
Xception	29.1	26.2
InceptionV3	28.4	27.0
MobileNetV2	30.5	28.7
EfficientNetB0	28.5	25.9
VGG16	36.3	34.4

5. Conclusions

One of the most essential biometric characteristics is the face. One can learn a lot about a person's ethnicity, gender, age, expression, identity, and more by evaluating their face. However, gender classification using facial images with masks is one of the most challenging obstacles in classification tasks in the pandemic era due to its intricacy. As per our knowledge, to disentangle the problem, no work based on deep learning and feature selection has been proposed until today. Therefore, this study proposed a gender classification system that determines a person's gender (male/female) based on the face of the individual wearing a mask in a given image. The study analyzed and compared performance of various state-of-the-art deep learning pre-trained networks to identify gender using two strategies. In the first strategy, the models are trained using human facial images where individuals either wear the mask fully, half, partially, or do not wear one at all (four ways of wearing). On the other hand, in the second strategy, the training phase only includes the images with veiled faces (faces with masks in three different styles). Evidently, experimental results conclude that the models' performances are significantly reduced in the second strategy, however, the EfficientNetB0 model still managed to perform well by

retaining a classification accuracy above 97% in distinguishing male and female identity in both strategies. Therefore, many applications, including smart human–computer interface, can benefit from this gender classification approach. Nevertheless, the proposed scheme can be enhanced in future for more accurate performance by integrating data augmentation techniques, and various computer vision and deep feature extraction schemes.

Author Contributions: Conceptualization, J.R.; methodology, J.R. and S.W.; software, J.R., S.W. and A.M.A.-M.; validation, J.R., S.W., S.A. and A.M.A.-M.; formal analysis, J.R., S.W. and A.M.A.-M.; investigation, J.R., S.W., S.A. and A.M.A.-M.; resources, J.R. and S.W.; data curation, J.R. and S.W.; writing—original draft preparation, J.R. and S.W.; writing—review and editing, J.R. and A.M.A.-M.; visualization, J.R., S.W. and S.A.; supervision, J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: The dataset is publicly available and no image is reproduced/published in this study from this dataset, and thus does not require informed consent. However, some external sample facial images of individuals (placed in this paper) are collected specifically for this study. Therefore, informed consent was obtained from all subjects involved in the study and written informed consent has been obtained from the individuals to publish this paper.

Data Availability Statement: This study did not generate any new datasets. The dataset used can be downloaded from a publicly available Kaggle repository [27].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kuchebo, A.V.; Bazanov, V.V.; Kondratev, I.; Kataeva, A.M. Convolution Neural Network Efficiency Research in Gender and Age Classification From Speech. In Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), Moscow, Russia, 26–29 January 2021; pp. 2145–2149. [\[CrossRef\]](#)
2. Vashisth, P.; Meehan, K. Gender Classification using Twitter Text Data. In Proceedings of the 2020 31st Irish Signals and Systems Conference (ISSC), Letterkenny, Ireland, 11–12 June 2020; pp. 1–6. [\[CrossRef\]](#)
3. Zvarevashe, K.; Olugbara, O.O. Gender Voice Recognition Using Random Forest Recursive Feature Elimination with Gradient Boosting Machines. In Proceedings of the 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 6–7 August 2018; pp. 1–6. [\[CrossRef\]](#)
4. Benkaddour, M.K.; Lahlali, S.; Trabelsi, M. Human Age and Gender Classification using Convolutional Neural Network. In Proceedings of the 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being (IHSH), Boumerdes, Algeria, 9–10 February 2021; pp. 215–220. [\[CrossRef\]](#)
5. Salama AbdELminaam, D.; Almansori, A.M.; Taha, M.; Badr, E. A deep facial recognition system using computational intelligent algorithms. *PLoS ONE* **2020**, *15*, e0242269. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Rasheed, J.; Alimovski, E.; Rasheed, A.; Sirin, Y.; Jamil, A.; Yesiltepe, M. Effects of Glow Data Augmentation on Face Recognition System based on Deep Learning. In Proceedings of the 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Istanbul, Turkey, 22–24 October 2020; pp. 1–5. [\[CrossRef\]](#)
7. Rasheed, J.; Jamil, A.; Hameed, A.A.; Aftab, U.; Aftab, J.; Shah, S.A.; Draheim, D. A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. *Chaos Solitons Fractals* **2020**, *141*, 110337. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Rasheed, J.; Hameed, A.A.; Djeddi, C.; Jamil, A.; Al-Turjman, F. A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. *Interdiscip. Sci. Comput. Life Sci.* **2021**, *13*, 103–117. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Arora, D.; Garg, M.; Gupta, M. Diving deep in Deep Convolutional Neural Network. In Proceedings of the 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 18–19 December 2020; pp. 749–751. [\[CrossRef\]](#)
10. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6. [\[CrossRef\]](#)
11. Alakus, T.B.; Turkoglu, I. Comparison of deep learning approaches to predict COVID-19 infection. *Chaos Solitons Fractals* **2020**, *140*, 110120. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Gogate, U.; Parate, A.; Sah, S.; Narayanan, S. Real Time Emotion Recognition and Gender Classification. In Proceedings of the 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), Islamabad, Pakistan, 23–25 November 2021; pp. 138–143. [\[CrossRef\]](#)

13. Mustafa, A.; Meehan, K. Gender Classification and Age Prediction using CNN and ResNet in Real-Time. In Proceedings of the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, 26–27 October 2020; pp. 1–6. [[CrossRef](#)]
14. Tapia, J.E.; Perez, C.A. Gender Classification from NIR Images by Using Quadrature Encoding Filters of the Most Relevant Features. *IEEE Access* **2019**, *7*, 29114–29127. [[CrossRef](#)]
15. Lee, B.; Gilani, S.Z.; Hassan, G.M.; Mian, A. Facial Gender Classification—Analysis using Convolutional Neural Networks. In Proceedings of the 2019 Digital Image Computing: Techniques and Applications (DICTA), Perth, WA, Australia, 2–4 December 2019; pp. 1–8. [[CrossRef](#)]
16. Shinde, S.R.; Thepade, S. Gender Classification from Face Images Using LBG Vector Quantization with Data Mining Algorithms. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–5. [[CrossRef](#)]
17. Hasnat, A.; Haider, S.; Bhattacharjee, D.; Nasipuri, M. A proposed system for gender classification using lower part of face image. In Proceedings of the 2015 International Conference on Information Processing (ICIP), Pune, India, 16–19 December 2015; pp. 581–585. [[CrossRef](#)]
18. Bekhouche, S.E.; Ouafi, A.; Benlamoudi, A.; Taleb-Ahmed, A.; Hadid, A. Facial age estimation and gender classification using multi level local phase quantization. In Proceedings of the 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT), Tlemcen, Algeria, 25–27 May 2015; pp. 1–4. [[CrossRef](#)]
19. Jabid, T.; Kabir, M.H.; Chae, O. Gender Classification Using Local Directional Pattern (LDP). In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2162–2165. [[CrossRef](#)]
20. Ozbudak, O.; Tukel, M.; Seker, S. Fast gender classification. In Proceedings of the 2010 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 28–29 December 2010; pp. 1–5. [[CrossRef](#)]
21. Vetrekar, N.; Ramachandra, R.; Raja, K.B.; Gad, R.S.; Busch, C. Robust Gender Classification Using Multi-Spectral Imaging. In Proceedings of the 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Jaipur, India, 4–7 December 2017; pp. 222–228. [[CrossRef](#)]
22. Zhang, C.; Ding, H.; Shang, Y.; Shao, Z.; Fu, X. Gender Classification Based on Multiscale Facial Fusion Feature. *Math. Probl. Eng.* **2018**, *2018*, 1–6. [[CrossRef](#)]
23. Lee, J.-H.; Chan, Y.-M.; Chen, T.-Y.; Chen, C.-S. Joint Estimation of Age and Gender from Unconstrained Face Images Using Lightweight Multi-Task CNN for Mobile Applications. In Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, 10–12 April 2018; pp. 162–165. [[CrossRef](#)]
24. Mozaffari, S.; Behravan, H.; Akbari, R. Gender Classification Using Single Frontal Image Per Person: Combination of Appearance and Geometric Based Features. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 1192–1195. [[CrossRef](#)]
25. Omer, H.K.; Jalab, H.A.; Hasan, A.M.; Tawfiq, N.E. Combination of Local Binary Pattern and Face Geometric Features for Gender Classification from Face Images. In Proceedings of the 2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSC), Penang, Malaysia, 1 December–29 November 2019; pp. 158–161. [[CrossRef](#)]
26. Avuçlu, E.; Başçiftçi, F. Novel approaches to determine age and gender from dental x-ray images by using multiplayer perceptron neural networks and image processing techniques. *Chaos Solitons Fractals* **2019**, *120*, 127–138. [[CrossRef](#)]
27. 500 GB of Images with People Wearing Masks. Part 3 | Kaggle. Available online: <https://www.kaggle.com/datasets/tapakah68/medical-masks-p3> (accessed on 30 March 2022).
28. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [[CrossRef](#)]
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [[CrossRef](#)]
30. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [[CrossRef](#)]
31. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807. [[CrossRef](#)]
32. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
33. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520. [[CrossRef](#)]
34. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.