Face Gender Recognition in the Wild: An Extensive Performance Comparison of Deep-Learned, Hand-Crafted, and Fused Features with Deep and Traditional Models

Alhanoof Althnian [1] , Nourah Aloboud [2], Norah Alkharashi [3] , Faten Alduwaish [4] , Mead Alrshoud [5] and Heba Kurdi [5,6,*]

1   Information Technology Department, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia; aalthnian@ksu.edu.sa
2   Center of Excellence of Decision Support Center, King Abdulaziz City for Science and Technology, Riyadh 12354, Saudi Arabia; naloboud@kacst.edu.sa
3   Computer Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia; 435204246@student.ksu.edu.sa
4   Saudi Information Technology Company, Riyadh 12382, Saudi Arabia; faldawish@site.sa
5   Computer Science Department, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia; 434201516@student.ksu.edu.sa
6   Mechanical Engineering Department, Massachusetts Institute of Technology (MIT), Cambridge, MA 02142-1308, USA
*   Correspondence: hkurdi@ksu.edu.sa

Abstract: Face gender recognition has many useful applications in human–robot interactions as it can improve the overall user experience. Support vector machines (SVM) and convolutional neural networks (CNNs) have been used successfully in this domain. Researchers have shown an increased interest in comparing and combining different feature extraction paradigms, including deep-learned features, hand-crafted features, and the fusion of both features. Related research in face gender recognition has been mostly restricted to limited comparisons of the deep-learned and fused features with the CNN model or only deep-learned features with the CNN and SVM models. In this work, we perform a comprehensive comparative study to analyze the classification performance of two widely used learning models (i.e., CNN and SVM), when they are combined with seven features that include hand-crafted, deep-learned, and fused features. The experiments were performed using two challenging unconstrained datasets, namely, Adience and Labeled Faces in the Wild. Further, we used T-tests to assess the statistical significance of the differences in performances with respect to the accuracy, f-score, and area under the curve. Our results proved that SVMs showed best performance with fused features, whereas CNN showed the best performance with deep-learned features. CNN outperformed SVM significantly at $p < 0.05$.

Keywords: deep learning; gender recognition; CNN; SVM; deep-learned features; hand-crafted features; feature fusion

1. Introduction

Gender recognition is vital in interconnected information societies; it has applications in many domains such as security surveillance, targeted advertising, and human–robot interactions. Face gender recognition plays a key role in the latter domain since it allows robots to adapt their behavior based on the gender of the interacting user, which increase user acceptance and satisfaction [1]. A wide range of contributions exist in literature that present a variety of frameworks [2–7], feature descriptors [8–13], classification model architectures [14–16], and benchmark datasets [17] with state-of-the-art results. Despite the achieved success, face gender recognition is still considered a challenging and unsolved problem; therefore, researchers continue to seek a solution [15,18].

There are numerous reasons for considering face gender recognition an open research problem. First, face images introduced multiple challenges because of the variations in the appearance, pose, lighting, background, and noise. Yet, numerous reported successes in the literature are achieved with easy constrained datasets, such as facial recognition technology (FERET) [19–22] and UND [20]. These datasets contain the frontal face images that were captured under controlled conditions of facial expressions, illumination, and background. Therefore, they do not reflect real-world situations [23]. Second, some proposed approaches (e.g., [22,24,25]) target a specific challenge in the face images; therefore, they may not achieve the same level of performance in real-world scenarios. Third, there does not exist any unified procedure for the task of gender recognition; therefore, authors follow different experimental setups, such as the number of folds in cross validation, used benchmark datasets, and model parameters (e.g., support vector machine (SVM) kernels), which make the comparison of results inapplicable.

Recently, we witnessed the rise of CNN models as not only a classification model but also as a feature extraction method in different domains [26–28]. Unlike hand-crafted features, which are designed beforehand by human experts, deep-learned features are learned directly from the data by using CNNs. Recent evidence suggests that each feature extraction paradigm focuses on extracting information from the images that are neglected by the other paradigms [29]. In the domain of gender recognition using face images, several attempts have been made to compare the performance of the two feature extraction paradigms. For instance, several studies have reported that fusing hand-crafted features with images can improve the CNN performance [30–32]. Despite the variations in experimental setups, certain studies have produced contradictory findings. For example, Wolfshaar et al. [33] compared the performance of deep-learned features with a fine-tuned network and an SVM. Their results proved that the fine-tuned model outperformed the SVM when oversampling was applied on the Adience dataset. In [34], the same dataset was used but the best performance was achieved when deep-learned features were extracted from a fined-tuned VGG model and fed to an SVM model.

Research on the subject has been mostly restricted to limited comparisons of the multiple feature extraction paradigms with one model [32,35] or a single paradigm with multiple models [33,34]. Little attention has been paid to how the different feature extraction paradigms (i.e., hand-crafted, deep-learned, and fused features) would compare when combined with the different models (CNN and SVM). In this research, we seek to fill this gap. We perform a comprehensive comparative analysis of different combinations of features extraction paradigms and models using two challenging unconstrained benchmark datasets, namely, Adience [17] and Labeled Faces in the Wild (LFW) [36]. Moreover, unlike most of the existing contributions, we report the accuracy, f-score, and area under the curve (AUC) for all the experiments and analyze their significance statistically.

The rest of the paper is organized as follows. In Section 2, we discuss the related literature. In Section 3, we describe the methodology, including the feature extraction, the datasets, the classification models, and performance evaluation. In Section 4, we present and discuss the results. Finally, Section 5 concludes our work.

2. Literature Review

Gender recognition is a domain where high state-of-the-art accuracy has been achieved by SVMs and CNNs [21,33,34,37,38]. These results, however, have been attributed to the characteristics of the dataset used [17,21,22,31]. For example, many of the early efforts in gender recognition have used constrained datasets that included frontal face images that were taken under controlled conditions of facial expressions, illumination, and background [19–21], and hence cannot achieve the same performance with images taken in the wild by surveillance or robot cameras. Building a gender recognition model based on face images is similar to other computer vision tasks; the process has three main stages: selecting the benchmark dataset, feature extraction and selection, and classification. In the

text below, we highlight the main efforts made in each stage for progress in the field. Furthermore, we summarize the results of the most relevant works in Table 1.

A dataset is an integral part of gender recognition research. Selecting an appropriate dataset to benchmark the proposed approach is a crucial decision because datasets introduce different challenges, such as pose variations, illumination variations, and occlusions. Gender recognition datasets can be broadly categorized into constrained and unconstrained datasets. The former includes frontal face images that were taken under controlled conditions of facial expressions, illumination, and background. Numerous early studies have been criticized for benchmarking their works with constrained datasets, such as FERET [19–22] and UND [20] because they do not reflect real-world situations [23,39]. Therefore, many studies were aimed at the challenges posed by the images taken under uncontrolled conditions, for example, LFW [20,22] and Adience [17,23,32,40,41] datasets and datasets with occlusions (e.g., sunglasses and hats), such as AR [20,22], Gallagher [32], and MORPH [40]. The authors in [17] offered a unique unconstrained and unfiltered dataset. Torralba and Efros [42] argued that the most popular datasets were biased, and they emphasized that using a single dataset for training and testing was not representative of the variations that exist in the real world. Therefore, to simulate real-life situations, recent studies [17,21,22,31] have adopted a cross-data approach, where one dataset is trained, and another dataset is tested. Other contributions [38,41] used a fusion of multiple constrained and unconstrained datasets for testing purposes. Moreover, some efforts have targeted a specific type of image, such as low-resolution thumbnail faces [43] and low-frequency components of the mosaic 8×8 images [44].

A fundamental problem was to determine what features in a person's face can help determine the person's gender. A wide range of studies have been devoted to improving the extraction and selection of features [45–47]. In recent years, there has been an increasing amount of computer vision literature that distinguishes between hand-crafted features and deep-learned features [45]. The hand-crafted features are designed beforehand by human experts, whereas deep-learned features are learned directly from the data using CNN. Furthermore, some studies reported performance improvements when the two features were combined [31,32]. One of the early works on the hand-crafted features is [48], where the authors combined the 3D distances with multiple measurements (such as the distance between the key points in the face, the ratios, and the angles between the key points) into a single function. Tamura et al. [44] divided the human face into four parts to determine which part contributed the most to identifying the gender. The results revealed that the face shape and cheek bone shape are the most important aspects. Further, the authors of [49] identified nine facial features that varied and hence could be used to distinguish males from females, namely, the hairline, eyebrows, eyes, distance between the eye and eyebrows, nose, lips, chin, cheeks, and face shape.

The hand-crafted features could be extracted from the facial features including the face shape by using the histogram of oriented gradients (HOGs) [50], texture using the local binary pattern (LBP) [51], and intensity features using the gray level of each pixel [20]. The geometric features can also be extracted, such as scale invariant feature transform (SIFT) [52] and Haar-Like features [21]. Jabid et al. [19] presented face images using a novel texture descriptor local directional pattern (LDP), and Shobeirinejad and Gao [10] proposed interlaced derivative patterns, which outperformed the LBP and LDP features. A number of authors have reported performance improvements when different types of hand-crafted features are fused, such as domain-specific and trainable features [18], trainable shapes, and color features [53], LBP and local phase quantization features [8], shape and texture features [54], LBP and radii spatial scales features [20], appearance-based and geometric-based features [55], appearance and geometry features [12], gradient and Gabor wavelets features [13], and LBP, SIFT, and color histogram [52]. In contrast, Alexandre [11] showed that a single feature from different scales could outperform multiple features at a single scale. In [9], adaptive features were proposed, which resulted in accuracy improvements in

the SVM model. The research in [31] showed that the hand-crafted features fusion could improve the SVM performance.

A growing body of literature has investigated deep-learned features and investigated how gender recognition accuracy differs when compared and combined with hand-crafted features. Nanni et al. [29] proposed a generic computer vision system that extracted, compared, and combined hand-crafted features with deep-learned features to train an SVM model using several datasets from different domains. The authors showed that a fusion of both hand-crafted and deep-learned features provided the best performance with SVM. Ozbulak et al. [34] explored transfer learning using generic and domain-specific models to extract deep-learned features to train different CNN and SVM models. Their results proved that the use of deep-learned features extracted using domain-specific models could improve the accuracy of all the models. In [56], the authors proposed the joint features learning deep neural networks, which could learn from the joint high-level and low-level features. The proposed architecture outperformed CNNs, SVM with face pixels, and SVM with LBP features. In [35], the authors compared hand-crafted and deep-learned features by training a CNN model for pedestrian gender recognition. Their results showed that hand-crafted and deep-learned features performed comparably on small-sized homogenous datasets, but the latter performed significantly better on heterogeneous data. In [57], the authors showed that feeding deep-learned features into an SVM rather than Softmax in VGGNet-16 provided better results. The fusion of deep-learned and hand-crafted features achieved better results than using only deep-learned features with ensemble learning [58].

SVM is a widely used model in the gender recognition domain [9,17,19–21,25,43,47,54,59,60]. Lately, deep learning is being used in many computer vision applications [61–64]. Therefore, studies have proposed varying architectures and experimental setups for CNNs to improve gender recognition [5,14,16,21,23,24,30,40,44,49,62,65–80]. Other authors have used ensemble learning [58] and K-nearest neighbor (KNN) [63] methods. The studies [21,32–34,37,38] are most similar to our study, wherein the main aim is to compare the performances of different features and machine learning models for the task of gender recognition. The studies [33] and [34] investigated the use of deep-learned features with the CNN and SVM models; they report contradictory findings. While the result in [33] proved that the fine-tuned model outperformed SVM, when oversampling was applied, the best performance was achieved in [34] when the deep-learned features were extracted from a fine-tuned VGG model and fed into an SVM model. Hosseini et al. [32] showed that feeding hand-crafted features to CNN can improve their performance. The SVM performances with hand-crafted features and CNNs with deep-learned features are explored in [21,37,38,81]; CNNs with deep-learned features achieved the best results.

Studies in the field of gender recognition have only focused on comparing the two feature extraction paradigms with one model [32], a single paradigm with multiple models [33,34], or limited feature extraction methods and models [21,37,38,81]. Because of the variations in the experimental setups, the results from different studies cannot be compared. Therefore, it is not clear yet how the different feature extraction paradigms (i.e., hand-crafted, deep-learned, and fused features) would perform when combined with different models (including CNN and SVM); this concern is addressed in this research.

Table 1. Related works with their reported results.

Ref.	Feature Descriptor	Classifier	Dataset	Result (Accuracy)
[21]	Image pixels	Support vector machine (SVM)	FERET	78.65%
			WWW	76.71%
	Image pixels	Neural network (NN)	FERET	86.98%
			WWW	66.94%
	Local binary patterns (LBP)	SVM	FERET	81.40%
			WWW	76.01%
[30]	Deep neural network (DNN)	DNN	LFW	92.60%
			Gallagher	84.28%
	Deep convolutional neural network (DCNN)	DCNN	LFW	94.09%
			Gallagher	86.04%
	Local-DNN	Local-DNN	LFW	96.25%
			Gallagher	90.58%
[31]	Histogram of oriented gradients (HOG)	SVM	GROUPS	88.23%
	Principal component analysis (PCA)			77.91%
	LBP			86.74%
	Local oriented statistics information booster (LOSIB)			86.65%
	Local salient patterns (LSP)			85.58%
	HOG + LBP + LOSIB			94.28%
	CNN + HOG + LBP + LOSIB			97.23%
[32]	Gabor response	CNN	Adience	89.20%
			Webface	91.00%
	Fused Gabor response	CNN	Adience	90.10%
			Webface	92.10%
[33]	Convolutional neural network (CNN)	CNN	Adience	87.20%
	CNN	SVM	Adience	81.40%
[34]	CNN	SVM	Adience	92.00%
	CNN	CNN	Adience	91.90%
[37]	LBP	SVM	FaceScrub	75.32%
	HOG	SVM	FaceScrub	80.58%
	CNN	CNN	FaceScrub	94.76%
[38]	CNN	CNN	Adience	96.10%
			FERET	97.90%
	PCA	SVM	Adience	77.40%
			FERET	90.20%
	Image pixels	SVM	Adience	77.30%
			FERET	87.10%
	HOG	SVM	Adience	75.80%
			FERET	85.60%
	Double tree complex wavelet transform (DTCWT)	SVM	Adience	68.50%
			FERET	90.70%
[41]	CNN	CNN	Adience	84.00%
[81]	CNN	CNN	LFW	98.90%
			GROUPS	96.10%

3. Methodology

We adopted an experimental methodology to compare the performances of two classification methods and seven feature extraction methods in the domain of gender recognition with respect to three performance measures. In addition, we performed a statistical analysis of the obtained results using T-tests to assess the statistical significance of the differences in performance.

3.1. Features Extraction

We applied seven features extraction methods, which can be divided into three main categories: hand-crafted features, deep-learned features, and fused features.

3.1.1. Hand-Crafted Features

Hand-crafted features can be categorized into global features, pixel-based features, and appearance-based features. A feature extraction method was selected from each category based on the previous usage by the community in the gender recognition domain. All the methods are well-known and widely used in many domains. We briefly explain each method below.

Local Binary Pattern (LBP): This is a simple yet effective pixel-based texture descriptor that was originally proposed by Ojala et al. [51] LBP is one of the most commonly used hand-crafted feature extraction methods in gender recognition [31,34,69,71,82–84]. The original descriptor assigns a binary digit for each pixel in a 3×3 neighborhood by comparing their pixel intensity values with the central pixel, which acts as a threshold. One digit is assigned to the pixel if its value is greater than or equal to the central pixel; otherwise, the pixel value is zero. The binary value for the central pixel is then computed by concatenating the eight binary digits of the neighboring pixels in a clockwise direction. LBP was later improved by using flexible neighborhood sizes [85]. The descriptor has two main parameters, which are the parameters of the circular neighborhood (P, R). This parameter determines the neighborhood size, where P is the number of sampling points in a circle of radius R. In our experiments, we used P = 24 and R = 3. The resulting LBP features are of size 26.

Histogram of Oriented Gradient (HOG): This is an appearance-based descriptor that extracts the gradients and orientations of edges in an image to describe the structure or shape of the object. It was promoted by Dalal and Triggs in 2005 [50] and has been applied successfully for face gender recognition [71]. The HOG features are extracted as follows. First, we compute the gradient of each pixel in both the x and y directions. Second, using the gradients, we calculate the magnitude and direction of each pixel. Third, we divide the image into small cells, and we compute the histogram of the gradients for each cell. Next, multiple cells are combined to form a block, and normalization is applied. Lastly, the normalized histograms of the blocks are combined to form the HOG features. Multiple parameters can be tuned to improve the accuracy of this descriptor including the cell size, the overlap between cells, block normalization, and types of blocks (either rectangular R-HOG blocks or circular C-HOG blocks). The following values were used in our experiments with R-HOG blocks: cell size = (8, 8), block size = (16, 16), and number of orientation pins = 9. The resulting features are of size 1764.

Principal Component Analysis (PCA): This is a global feature extraction method that uses linear transformation to map the features space into lower dimensions while maximizing their variance. PCA can be applied to images' raw pixel values or to other hand-crafted features, resulting in second-order uncorrelated features. To extract the PCA features, the dataset must be standardized. Then, we identify the relationships between the features by computing a covariance matrix for the dataset. Next, we perform eigen decomposition to obtain the eigenvalues and eigenvectors of the matrix. The principle components of the dataset are the eigenvectors with the greatest eigenvalues. The user may decide to keep all or only a subset of the principle components. Lastly, the selected principle components are transposed and multiplied to the transpose of the original dataset,

which yields the PCA features. In this work, PCA has been applied on the images' raw pixel values, where the first two components were used.

3.1.2. Deep-Learned Features

We applied deep transfer learning by using a CNN as a fixed feature extractor (see the upper part of Figure 1). Similar to the methods used in [34,75,86], we used a pre-trained VGG-16 on ImageNet [87] and removed the last fully connected layer. We treated the rest of ConvNet as a fixed feature extractor for our datasets. The input layer accepted images of the size 224×224 and had three channels: red, green, and blue. The input images went through a series of hidden convolution layers, which used the rectified linear unit activation function. Some layers were followed by a max-pool layer, which was performed over non-overlapping max-pool windows of the size 2×2 with the stride equal to two. The dimension of the deep-learned features was $7 \times 7 \times 512$.
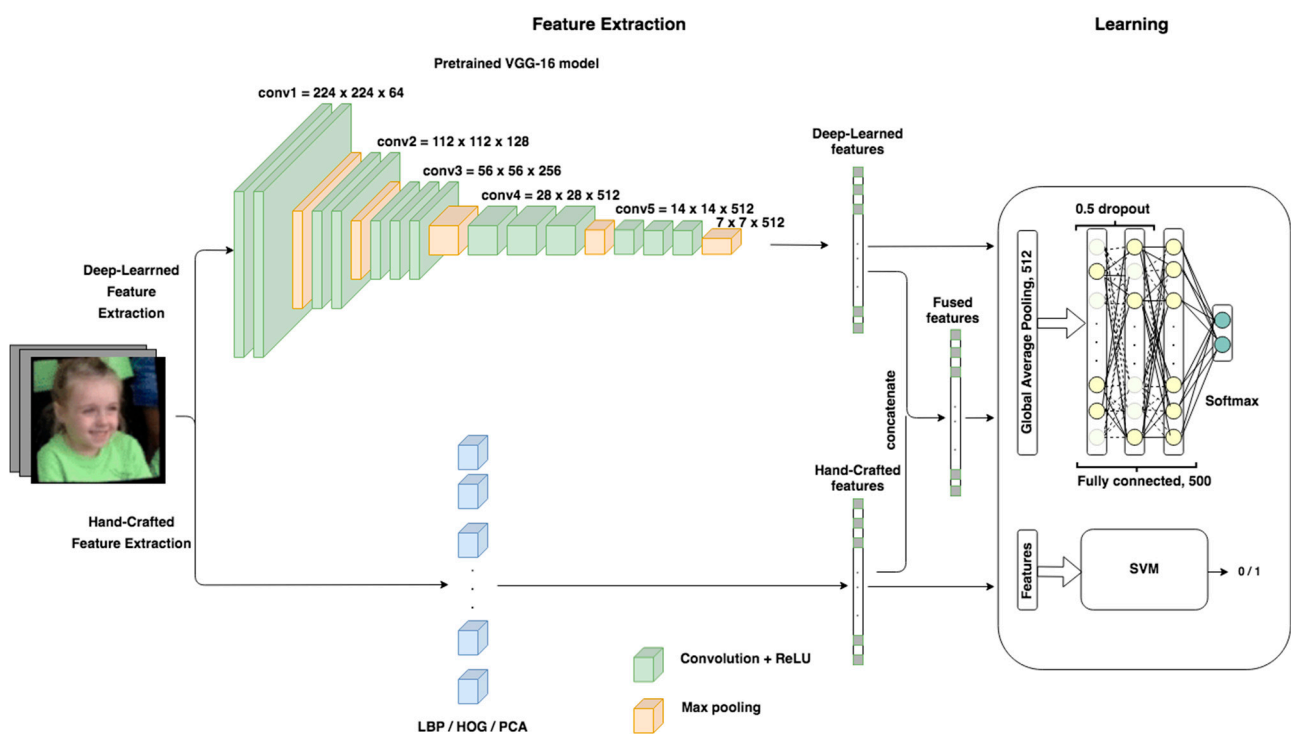

Figure 1. Illustration of the methodology. Deep-learned features are extracted using a pre-trained vgg-16 model, and the hand-crafted features are extracted using the local binary pattern (LBP), histogram of oriented gradient (HOG), and principal component analysis (PCA) methods. The two types of features are fused. CNN and SVM models are trained using the three types of features.

3.1.3. Fused Features

The fusion of deep-learned and hand-crafted features was aimed to provide a holistic description of the images. As mentioned previously, several studies have reported that fusing specific hand-crafted features with images can improve the performance of CNNs [30,32]. For this purpose, the extracted deep-learned features were concatenated with the hand-crafted features, namely LBP, HOG, and PCA, yielding fusion of HOG and deep-learned, fusion of LBP and deep-learned, and a fusion of PCA and deep-learned features. The fused features are then fed to the classification model, as shown in Figure 1.

3.2. Dataset

There are mainly two types of benchmark datasets that have been used in literature. The first type is the constrained dataset, in which images were taken under controlled

conditions. The second type was unconstrained datasets, where images are taken under uncontrolled conditions. In this study, we used two challenging and commonly used unconstrained benchmark datasets, which are briefly described below.

3.2.1. Labeled Faces in the Wild

We used LFW deep funneled images dataset [36]. LFW consists of over 13,000 face images of real people from both genders collected from the web. The face images are of varying conditions of image quality, facial expressions, head poses, illuminations, and occlusions. The samples are shown in Figure 2. We used the deep funneled version of the dataset because it is the best version available in terms of achieved accuracy. In this version, the face images were aligned using deep learning [36]. Similar to [20] and [39], we used a subset of the dataset. The original dataset was unbalanced; therefore, we performed undersampling of the majority class to create a balanced dataset having a size of 6000 images. Further, following [86], we performed preprocessing to resize all the images to 224×224 to be processed by the VGG-16 model. The dataset was divided into balanced five folds to perform cross validation.

Figure 2. Samples from the face images in the used datasets (top row: Adience dataset [17], bottom row: LFW dataset [36]).

3.2.2. Adience

This dataset is one of the most challenging available datasets because it includes more images and subjects than other available datasets, such as Gallagher and PubFig [17]. It contains more than 26,000 images of over 2000 people uploaded to the Flicker.com public albums. According to the authors, the faces in the images were first detected using a Viola and Jones face detector [88], and the facial feature points were then identified by a modified version of the study in [89]. In this research, we used the whole dataset of the aligned and cropped face image version, which was already divided into five folds for cross validation [17].

3.3. Classification Methods

3.3.1. SVM

SVM is a widely used learning model, which is applied for classification and regression. The basic idea of SVM is to separate the data by finding a hyperplane that maximizes the margin between the two classes of data. The margin represents the distance between the data points from each class that lies closest to the hyperplane, known as support vectors. SVM uses a kernel function to map non-linearly separable data into a higher dimensional feature space, where it becomes linearly separable. SVM performance can be optimized by tuning the parameters kernel, C, and gamma. The kernel variations used include linear,

RBF, and polynomial kernel. The parameter C is used for regularization; if C is set to a large value, a small margin will be used for optimization and vice versa. Gamma is set when a Gaussian RBF kernel is used. Features are fed directly to SVM, but in case of the deep-learned features, they are first flattened from $7 \times 7 \times 512$ to a one-dimension vector of size 25,088. In this work, we used SVM with an RBF kernel. The values of the parameters are C = 10 and gamma = 0.001.

3.3.2. CNN

As explained previously, the deep-learned features were extracted using a pre-trained VGG-16 model. The last maximum pooling layer in the model was connected to a global average pooling to convert the image features from a $7 \times 7 \times 512$ vector to a $1 \times 1 \times 512$ vector. Then, we trained three dense layers for our dataset with two dropout layers with 0.5 probability to avoid overfitting. The Softmax function was used on the last layer to convert the layer output to a vector that represented the probability distribution of a list of possible outcomes for two classes. In our experiments, CNN was trained with 2000 epochs, 128 batch sizes, an Adam optimizer, and the binary cross-entropy loss function.

3.4. Performance Evaluation

Unlike most of the existing efforts in literature, which adopt the classification rate as the only performance measure, we recognize the importance of looking at the performance of a classifier from different angles [90]. Therefore, we evaluate the performance of the classification models with respect to three important metrics, namely, accuracy, F-score, and AUC. Therefore, investigating the performance with respect to different metrics can help the community improve the performance of classifiers in this domain. Further, the k-fold cross-validated paired t-test is applied to assess the statistical significance between two models A and B according to Equation (1) below.

$$t = \frac{\overline{p}\ \sqrt{k}}{\sqrt{\frac{\sum_{i=1}^{k}\left(p^i - \overline{p}\right)^2}{k-1}}} \tag{1}$$

where k is the number of folds, p^i is the difference between the model performances in the ith iteration $p^i = p^i_A - p^i_B$ and $\overline{p}$ computes the average difference between the model performances $\overline{p} = \frac{1}{k} \sum_{i=1}^{k} p^i$.

4. Results and Discussion

The experimental results are shown in Tables 2–4 for both Adience and LFW datasets. The tables show the performance of CNN and SVM models with different types of features. We trained SVM with seven types of features, namely, HOG, LBP, PCA, deep-learned, fusion of HOG and deep-learned, fusion of LBP and deep-learned, and fusion of PCA and deep-learned features. Moreover, we trained CNN with four features, namely, deep-learned, fusion of HOG and deep-learned, fusion of LBP and deep-learned, and fusion of PCA and deep-learned features. The parameters of the methods were instantiated based on the empirical experiments and by following the recommendations from the literature. All the reported results are the average of five-fold cross validation. T-tests were used to analyze the relationship between the performances of different combinations of features and classifiers.

Table 2 is quite revealing in several ways. First, we can observe that, on average, SVM performs comparably with HOG and LBP features, whereas it has slightly less accuracy using the PCA features. Yet, when deep-learned features are used, SVM performance with respect to the accuracy increases by 12.95% as compared with the best performance with hand-crafted features. However, what is interesting in our result is that the best SVM performance is achieved when fused features are used because the classifier achieves at

least 22.40% and 9.45% increase in accuracy as compared with hand-crafted and deep-learned features, respectively. Our SVM results with deep-learned features outperform those reported in [33] when SVM with dropout and oversampling is trained on the Adience dataset.

Table 2. Performance evaluation with respect to accuracy on the Adience and LFW datasets.

Features		Classifiers	Datasets		Average over All Datasets
			Adience	LFW	
Hand-Crafted	HOG	SVM	65.5%	64.4%	64.95%
	LBP	SVM	62.5%	67.3%	64.90%
	PCA	SVM	60.9%	65%	62.95%
Deep-Learned	CNN features	SVM	83.3%	72.5%	77.90%
	CNN features	CNN	89.2%	84%	86.60%
Fusion	HOG-DL	SVM	84.1%	90.6%	87.35%
		CNN	81.7%	80.2%	80.95%
	LBP-DL	SVM	84.9%	91.3%	88.10%
		CNN	71.4%	89.7%	80.55%
	PCA-DL	SVM	84.8%	91.1%	87.95%
		CNN	54.3%	57.2%	55.75%

Table 3. Performance evaluation with respect to f-score on the Adience and LFW datasets.

Features		Classifiers	Datasets		Average over All Datasets
			Adience	LFW	
Hand-Crafted	HOG	SVM	66.5%	66.4%	66.45%
	LBP	SVM	65.0%	67.1%	66.05%
	PCA	SVM	65.7%	64.5%	65.10%
Deep-Learned	CNN features	SVM	82.3%	62.6%	72.45%
	CNN features	CNN	88.7%	81.4%	85.05%
Fusion	HOG-DL	SVM	85%	90.7%	87.85%
		CNN	81.7%	69.6%	75.65%
	LBP-DL	SVM	84.8%	91.3%	88.05%
		CNN	76.2%	89.5%	82.85%
	PCA-DL	SVM	85.7%	91.1%	88.40%
		CNN	65.5%	62.2%	63.85%

Next, we considered the CNN model. We observed that the CNN model had the best performance with deep-learned features. Table 2 shows that the model accuracy is 86.60% with deep-learned features; however, this accuracy drops by at least 5.65% when fused features are used. These results contradict earlier findings by [32], which showed that feeding hand-crafted features to CNN can improve their performance. This difference can be explained by the fact that only Gabor filters were used in [32] as hand-crafted features. Furthermore, the CNN accuracy achieved in this research is higher than that reported in [41], where a CNN model trained on the Adience dataset achieved 84% accuracy.

On comparing the SVM and CNN performances with different types of features, we can see that the CNN model with deep-learned features outperforms the best SVM result with fused features on the Adience dataset. However, opposite results were obtained on the LFW dataset. Our T-test shows that the result with the Adience dataset ($p = 0.0002$) is

significant whereas the result with LFW ($p = 0.093$) is insignificant at $p < 0.05$. These results suggest that CNN with deep-learned features is superior to SVM using any type of feature. These results further support the observations from earlier studies [33].

Table 4. Performance evaluation with respect to AUC on the Adience and LFW datasets.

Features		Classifiers	Datasets		Average over All Datasets
			Adience	LFW	
Hand-Crafted	HOG	SVM	65.6%	64.4%	65.00%
	LBP	SVM	62.3%	67.3%	64.80%
	PCA	SVM	60.1%	65.3%	62.70%
Deep-Learned	CNN features	SVM	83.2%	72.6%	77.90%
	CNN features	CNN	89.1%	84%	86.55%
Fusion	HOG-DL	SVM	84.1%	90.6%	87.35%
		CNN	82%	80.2%	81.10%
	LBP-DL	SVM	84.7%	91.3%	88.00%
		CNN	69.3%	89.5%	79.40%
	PCA-DL	SVM	84.6%	91.1%	87.85%
		CNN	51.4%	57.2%	54.30%

Similar trends can be observed in Tables 3 and 4, where the performances are presented with respect to the F-score and AUC, respectively. In both the tables, SVM exhibits the worst average performance with hand-crafted features. The model's average performance improves when deep-learned features are used, whereas further improvement is achieved with fused features. For CNNs, the latter features yield the worst performance as compared with the performance with deep-learned features. In addition, similar to the observations in Table 2, the CNN model performs significantly better at $p < 0.05$ than the best-performing SVM with fused features with $p = 0.002$ on the Adience dataset; however, the difference in performances between the SVM with the fused features and CNN with deep-learned features on the LFW dataset is insignificant ($p = 0.123$). Similar observations apply on the AUC with $p = 0.00003$ on the Adience dataset and $p = 0.098$ on the LFW dataset.

5. Conclusions

Face gender recognition plays a key role in robot–human interactions since it allows robots to adapt their behavior based on the gender of the interacting user, which increases user acceptance and satisfaction. The main goal of the current study was to comprehensively assess the performance of the most successful machine learning model in gender recognition, namely CNN and SVM, when combined with seven common feature extraction methods that included hand-crafted, deep-learned, and fused features. Previous studies on the subject have been mostly restricted to making limited comparisons of hand-crafted and deep-learned features with one model [27,46] or deep-learned features with multiple models [16,21]. Furthermore, contradictory findings have been reported about the best-performing combination in the latter category. For this purpose, we performed a comparative analysis of the CNN and SVM models when trained using three hand-crafted features (HOG, LBP, and PCA), deep-learned features (using transfer learning to extract features from a pre-trained VGG-16 model), and a fusion of both features; this analysis yielded seven sets of features. We used the most challenging datasets available, namely, Adience and LFW, and we presented the performance with respect to the accuracy, f-score, and AUC.

The most significant findings from this study are that (1) SVM performs the best when trained on a fusion of hand-crafted and deep-learned features, followed by deep-learned

features. The worst performance is exhibited when trained on hand-crafted features. (2) CNN performance decreases when the features are fused with hand-crafted features, including HOG, LBP, and PCA. (3) The CNN model outperforms SVM with all three feature extraction paradigms. The results of this study prove that although deep-learned features can enhance the performance of SVM, CNN still exhibits superior performance in the gender recognition domain. The reported results are possibly influenced by the fact that the Adience dataset is much larger in size than LFW (26,000 vs. 6000) but is more challenging dataset since, unlike LFW, it contains images of individuals from eight age groups [17]. A natural progression of this research would be to analyze the performance using other hand-crafted features, such as SIFT and Gabor filters and with deep-learned features extracted by CNNs of varying architectures and with fine tuning. Another possible area for future research would in investigating whether the findings of this research would hold with cross-data training, where a model is trained on a dataset and tested on a different dataset.

Author Contributions: Conceptualization, A.A. and H.K.; Data curation, A.A. and N.A. (Nourah Aloboud); Formal analysis, A.A.; Funding acquisition, H.K.; Investigation, A.A., N.A. (Nourah Aloboud) and H.K.; Methodology, A.A. and N.A. (Nourah Aloboud); Resources, H.K.; Software, N.A. (Norah Alkharashi), F.A. and M.A.; Supervision, H.K.; Validation, N.A. (Nourah Aloboud), N.A. (Norah Alkharashi), F.A. and M.A.; Writing—original draft, A.A.; Writing—review and editing, A.A. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available data were analyzed in this study. This data can be found here: [Adience: https://talhassner.github.io/home/projects/Adience/Adience-data.html], [LFW: http://vis-www.cs.umass.edu/lfw/].

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Scheuerman, M.K.; Paul, J.M.; Brubaker, J.R. How computers see gender: An evaluation of gender classification in commercial facial analysis services. *Proc. ACM Hum. Comput. Interact.* **2019**, *3*, 1–33. [CrossRef]
2. Carcagnì, P.; Cazzato, D.; Del Coco, M.; Leo, M.; Pioggia, G.; Distante, C. Real-Time Gender Based Behavior System for Human-Robot Interaction. In Proceedings of the International Conference on Social Robotics, Sydney, NSW, Australia, 27–29 October 2014; Springer: Cham, Switzerland, 2014.
3. Foggia, P.; Greco, A.; Percannella, G.; Vento, M.; Vigilante, V. A system for gender recognition on mobile robots. In Proceedings of the 2nd International Conference on Applications of Intelligent Systems, Las Palmas de Gran Canaria, Spain, 7–12 January 2019.
4. Carletti, V.; Greco, A.; Saggese, A.; Vento, M. An effective real time gender recognition system for smart cameras. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *11*, 2407–2419. [CrossRef]
5. Ranjan, R.; Patel, V.M.; Chellappa, R. HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *41*, 121–135. [CrossRef] [PubMed]
6. Greco, A.; Saggese, A.; Vento, M. Digital Signage by Real-Time Gender Recognition from Face Images. In Proceedings of the 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Rome, Italy, 3–5 June 2020; Volume 2020, pp. 309–313.
7. Khan, K.; Attique, M.; Syed, I.; Gul, A. Automatic Gender Classification through Face Segmentation. *Symmetry* **2019**, *11*, 770. [CrossRef]
8. 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]
9. Shmaglit, L.; Khryashchev, V. Gender classification of human face images based on adaptive features and support vector machines. *Opt. Mem. Neural Netw.* **2013**, *22*, 228–235. [CrossRef]
10. Shobeirinejad, A.; Gao, Y. Gender Classification Using Interlaced Derivative Patterns. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 1509–1512.
11. Alexandre, L.A. Gender recognition: A multiscale decision fusion approach. *Pattern Recognit. Lett.* **2010**, *31*, 1422–1427. [CrossRef]

12.   Xu, Z.; Lu, L.; Shi, P. A hybrid approach to gender classification from face images. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.

13.   Ren, H.; Li, Z.-N. Gender Recognition Using Complexity-Aware Local Features. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; Volume 2014, pp. 2389–2394.

14.   Lin, C.-J.; Li, Y.-C.; Lin, H.-Y. Using Convolutional Neural Networks Based on a Taguchi Method for Face Gender Recognition. *Electronics* **2020**, *9*, 1227. [CrossRef]

15.   Greco, A.; Saggese, A.; Vento, M.; Vigilante, V. A Convolutional Neural Network for Gender Recognition Optimizing the Accuracy/Speed Tradeoff. *IEEE Access* **2020**, *8*, 130771–130781. [CrossRef]

16.   Rafique, I.; Hamid, A.; Naseer, S.; Asad, M.; Awais, M.; Yasir, T. Age and Gender Prediction using Deep Convolutional Neural Networks. In Proceedings of the 2019 International Conference on Innovative Computing (ICIC), Seoul, Korea, 26–29 August 2019; Volume 2019, pp. 1–6.

17.   Eidinger, E.; Enbar, R.; Hassner, T. Age and Gender Estimation of Unfiltered Faces. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2170–2179. [CrossRef]

18.   Azzopardi, G.; Greco, A.; Saggese, A.; Vento, M. Fusion of Domain-Specific and Trainable Features for Gender Recognition From Face Images. *IEEE Access* **2018**, *6*, 24171–24183. [CrossRef]

19.   Jabid, T.; Kabir, 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.

20.   Tapia, J.E.; Perez, C.A. Gender Classification Based on Fusion of Different Spatial Scale Features Selected by Mutual Information From Histogram of LBP, Intensity, and Shape. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 488–499. [CrossRef]

21.   Mäkinen, E.; Raisamo, R. An experimental comparison of gender classification methods. *Pattern Recognit. Lett.* **2008**, *29*, 1544–1556. [CrossRef]

22.   Rai, P.; Khanna, P. A gender classification system robust to occlusion using Gabor features based (2D)2PCA. *J. Vis. Commun. Image Represent.* **2014**, *25*, 1118–1129. [CrossRef]

23.   Levi, G.; Hassner, T. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2019; pp. 34–42.

24.   Aslam, A.; Hussain, B.; Cetin, A.E.; Umar, A.I.; Ansari, R. Gender classification based on isolated facial features and foggy faces using jointly trained deep convolutional neural network. *J. Electron. Imaging* **2018**, *27*, 053023. [CrossRef]

25.   Yang, M.-H.; Moghaddam, B. Support vector machines for visual gender classification. In Proceedings of the 15th International Conference on Pattern Recognition, ICPR-2000, Barcelona, Spain, 3–7 September 2002.

26.   Salaken, S.M.; Khosravi, A.; Khatami, A.; Nahavandi, S.; Hosen, M.A. Lung cancer classification using deep learned features on low population dataset. In Proceedings of the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, Canada, 30 April–3 May 2017; Volume 2017, pp. 1–5.

27.   Oh, S.H.; Kim, G.-W.; Lim, K.-S. Compact deep learned feature-based face recognition for Visual Internet of Things. *J. SuperComput.* **2017**, *74*, 6729–6741. [CrossRef]

28.   Egede, J.; Valstar, M.; Martinez, B. Fusing Deep Learned and Hand-Crafted Features of Appearance, Shape, and Dynamics for Automatic Pain Estimation. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; Volume 2017, pp. 689–696.

29.   Nanni, L.; Ghidoni, S.; Brahnam, S. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognit.* **2017**, *71*, 158–172. [CrossRef]

30.   Mansanet, J.; Albiol, A.; Paredes, R. Local Deep Neural Networks for gender recognition. *Pattern Recognit. Lett.* **2016**, *70*, 80–86. [CrossRef]

31.   Castrilln-Santana, M.; Lorenzo-Navarro, J.; Ramn-Balmaseda, E. Descriptors and regions of interest fusion for gender classification in the wild. comparison and combination with convolutional neural networks. *arXiv* **2015**, arXiv:1507.06838v2.

32.   Hosseini, S.; Lee, S.H.; Cho, N.I. Feeding hand-crafted features for enhancing the performance of convolutional neural networks. *arXiv* **2018**, arXiv:1801.07848.

33.   Van De Wolfshaar, J.; Karaaba, M.F.; Wiering, M.A. Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition. In Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, 7–10 December 2015; Volume 2015, pp. 188–195.

34.   Ozbulak, G.; Aytar, Y.; Ekenel, H.K. How Transferable Are CNN-Based Features for Age and Gender Classification? In Proceedings of the 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 21–23 September 2016; Volume 2016, pp. 1–6.

35.   Antipov, G.; Berrani, S.-A.; Ruchaud, N.; Dugelay, J.-L. Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition. In Proceedings of the 23rd ACM international conference on Multimedia—MM '15, Brisbane, Australia, 26 October 2015; pp. 1263–1266.

36.   Huang, G.; Mattar, M.; Lee, H.; Learned-Miller, E. Learning to align from scratch. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 764–772, 2012.

37.   Kabasakal, B.; Sumer, E. Gender recognition using innovative pattern recognition techniques. In Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018; pp. 1–4.

38. Andonie, R. Comparison of recent machine learning techniques for gender recognition from facial images. *MAICS* **2018**, *10*, 97–102.

39. Shan, C. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognit. Lett.* **2012**, *33*, 431–437. [CrossRef]

40. Duan, M.; Li, K.; Yang, C.; Li, K. A hybrid deep learning CNN–ELM for age and gender classification. *Neurocomputing* **2018**, *275*, 448–461. [CrossRef]

41. Nistor, S.C.; Marina, A.-C.; Darabant, A.S.; Borza, D. Automatic gender recognition for "in the wild" facial images using convolutional neural networks. In Proceedings of the 2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 7–9 September 2017; Volume 2017, pp. 287–291.

42. Torralba, A.; Efros, A.A. Unbiased look at dataset bias. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; Volume 2011, pp. 1521–1528.

43. Moghaddam, B.; Yang, M.H. Gender classification with support vector machines. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, 28–30 March 2000; pp. 306–311.

44. Tamura, S.; Kawai, H.; Mitsumoto, H. Male/female identification from 8×6 very low resolution face images by neural network. *Pattern Recognit.* **1996**, *29*, 331–335. [CrossRef]

45. Cai, L.; Zhu, J.; Zeng, H.; Chen, J.; Cai, C. Deep-Learned and Hand-Crafted Features Fusion Network for Pedestrian Gender Recognition. In *Proceedings in Adaptation, Learning and Optimization*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 9, pp. 207–215.

46. Ng, C.-B.; Tay, Y.-H.; Goi, B.-M. Pedestrian gender classification using combined global and local parts-based convolutional neural networks. *Pattern Anal. Appl.* **2018**, *22*, 1469–1480. [CrossRef]

47. Sun, Z.; Bebis, G.; Yuan, X.; Louis, S.J. Genetic feature subset selection for gender classification: A comparison study. In Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision, 2002. (WACV 2002), Orlando, FL, USA, 3–4 December 2002; pp. 165–170.

48. Burton, A.M.; Bruce, V.; Dench, N. What's the Difference between Men and Women? Evidence from Facial Measurement. *Perception* **1993**, *22*, 153–176. [CrossRef]

49. Kalansuriya, T.R.; Dharmaratne, A.T. Neural network based age and gender classification for facial images. *ICTer* **2014**, *7*, 2. [CrossRef]

50. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1.

51. Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [CrossRef]

52. Fazl-Ersi, E.; Mousa-Pasandi, M.E.; Laganière, R.; Awad, M. Age and gender recognition using informative features of various types. *2014 IEEE Int. Conf. Image Process.* **2014**, 5891–5895. [CrossRef]

53. Azzopardi, G.; Foggia, P.; Greco, A.; Saggese, A.; Vento, M. Gender recognition from face images using trainable shape and color features. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; Volume 2018, pp. 1983–1988.

54. Lian, H.-C.; Lu, B.-L. Multi-view Gender Classification Using Local Binary Patterns and Support Vector Machines. In *Computer Vision*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2006; Volume 3972, pp. 202–209.

55. 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.

56. Jiang, Y.; Li, S.; Liu, P.; Dai, Q. Multi-feature deep learning for face gender recognition. In Proceedings of the 2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference, Beijing, China, 20–21 December 2014; Volume 2014, pp. 507–511.

57. Liu, T.; Ye, X.; Sun, B. Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; Volume 2018, pp. 3477–3481.

58. Tasci, E.; Uğur, A. Image classification using ensemble algorithms with deep learning and hand-crafted features. In Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018; Volume 2018, pp. 1–4.

59. Rai, P.; Khanna, P. Gender Classification Techniques: A Review. In *Advances in Intelligent and Soft Computing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 51–59.

60. Santarcangelo, V.; Farinella, G.M.; Battiato, S. Gender recognition: Methods, datasets and results. In Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 29 June–3 July 2015; pp. 1–6.

61. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]

62. Antipov, G.; Berrani, S.-A.; Dugelay, J.-L. Minimalistic CNN-based ensemble model for gender prediction from face images. *Pattern Recognit. Lett.* **2016**, *70*, 59–65. [CrossRef]

63. Mualla, N.; Houssein, E.H.; Zayed, H.H. Face Age Estimation Approach based on Deep Learning and Principle Component Analysis. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 152–157. [CrossRef]

64. Rawat, W.; Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.* **2017**, *29*, 2352–2449. [CrossRef] [PubMed]

65. Orozco, C.; Iglesias, F.; Buemi, M.; Berlles, J. Real-Time Gender Recognition from Face Images Using Deep Convolutional Neural Network. In Proceedings of the 7th Latin American Conference on Networked and Electronic Media (LACNEM 2017), Valparaiso, Chile, 6–7 November 2017; Institution of Engineering and Technology (IET): London, UK, 2017; pp. 7–11.

66. Dhomne, A.; Kumar, R.; Bhan, V. Gender Recognition through Face Using Deep Learning. *Procedia Comput. Sci.* **2018**, *132*, 2–10. [CrossRef]

67. Liew, S.S.; Khalil-Hani, M.; Radzi, S.B.A.; Bakhteri, R. Gender classification: A convolutional neural network approach. *Turk. J. Electr. Eng. Comput. Sci.* **2016**, *24*, 1248–1264. [CrossRef]

68. Samek, W.; Binder, A.; Lapuschkin, S.; Müller, K.-R. Understanding and Comparing Deep Neural Networks for Age and Gender Classification. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 1629–1638.

69. Haider, K.Z.; Malik, K.R.; Khalid, S.; Nawaz, T.; Jabbar, S. Deepgender: Real-time gender classification using deep learning for smartphones. *J. Real-Time Image Process.* **2019**, *16*, 15–29. [CrossRef]

70. Zhang, K.; Tan, L.; Li, Z.; Qiao, Y. Gender and Smile Classification Using Deep Convolutional Neural Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 739–743.

71. Qawaqneh, Z.; Abu Mallouh, A.; Barkana, B.D. Age and gender classification from speech and face images by jointly fine-tuned deep neural networks. *Expert Syst. Appl.* **2017**, *85*, 76–86. [CrossRef]

72. Zhang, Y.; Xu, T. Landmark-Guided Local Deep Neural Networks for Age and Gender Classification. *J. Sensors* **2018**, *2018*, 1–10. [CrossRef]

73. Antipov, G.; Baccouche, M.; Berrani, S.-A.; Dugelay, J.-L. Effective training of convolutional neural networks for face-based gender and age prediction. *Pattern Recognit.* **2017**, *72*, 15–26. [CrossRef]

74. Hosseini, S.; Lee, S.H.; Kwon, H.J.; Koo, H.I.; Cho, N.I. Age and gender classification using wide convolutional neural network and Gabor filter. In Proceedings of the 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 7–9 January 2018; Volume 2018, pp. 1–3. [CrossRef]

75. Smith, P.; Chen, C. Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 2564–2571.

76. Arora, S.; Bhatia, M. A Robust Approach for Gender Recognition Using Deep Learning. In Proceedings of the 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bangalore, India, 10–12 July 2018; Volume 2018, pp. 1–6.

77. Akbulut, Y.; Sengur, A.; Ekici, S. Gender recognition from face images with deep learning. In Proceedings of the 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 16–17 September 2017; Volume 2017, pp. 1–4. [CrossRef]

78. Deng, Q.; Xu, Y.; Wang, J.; Sun, K. Deep learning for gender recognition. In Proceedings of the 2015 International Conference on Computers, Communications, and Systems (ICCCS), Mauritius, India, 2–3 November 2015; Volume 2015, pp. 206–209.

79. Liu, X.; Li, J.; Hu, C.; Pan, J.-S. Deep convolutional neural networks-based age and gender classification with facial images. In Proceedings of the 2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS), Harbin, China, 3–5 June 2017; pp. 1–4.

80. Aslam, A.; Hayat, K.; Umar, A.I.; Zohuri, B.; Zarkesh-Ha, P.; Modissette, D.; Khan, S.Z.; Hussian, B. Wavelet-based convolutional neural networks for gender classification. *J. Electron. Imaging* **2019**, *28*. [CrossRef]

81. Jia, S.; Lansdall-Welfare, T.; Cristianini, N. Gender Classification by Deep Learning on Millions of Weakly Labelled Images. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 12–15 December 2016; Volume 2016, pp. 462–467.

82. Castrillón-Santana, M.; Lorenzo-Navarro, J.; Travieso-Gonzalez, C.M.; Freire-Obregón, D.; Alonso-Hernández, J.B. Evaluation of local descriptors and CNNs for non-adult detection in visual content. *Pattern Recognit. Lett.* **2018**, *113*, 10–18. [CrossRef]

83. Zeni, L.F.D.A.; Jung, C.R. Real-Time Gender Detection in the Wild Using Deep Neural Networks. In Proceedings of the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Paraná, Brazil, 29 October–1 November 2018; Volume 10, pp. 118–125.

84. Taheri, S.; Toygar, Ö. Multi-stage age estimation using two level fusions of handcrafted and learned features on facial images. *IET Biom.* **2019**, *8*, 124–133. [CrossRef]

85. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [CrossRef]

86. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

87. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

88. Viola, P.; Jones, M. Robust real-time object detection. *Int. J. Comput. Vis.* **2001**, *4*, 34–47.

89. Zhu, X.; Ramanan, D. Face detection, pose estimation, and landmark localization in the wild. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; Volume 2012, pp. 2879–2886.

90. Japkowicz, N. Why Question Machine Learning Evaluation Methods. 2006, pp. 6–11. Available online: https://www.aaai.org/Papers/Workshops/2006/WS-06-06/WS06-06-003.pdf (accessed on 19 January 2020).