Received May 28, 2020, accepted June 20, 2020, date of publication June 30, 2020, date of current version July 28, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3006051*

# Investigating Bias in Facial Analysis Systems: A Systematic Review

**ASHRAF KHALIL[1], SOHA GLAL AHMED[1], ASAD MASOOD KHATTAK[2], (Member, IEEE), AND NABEEL AL-QIRIM[3], (Senior Member, IEEE)**

[1]College of Engineering, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates
[2]College of Technological Innovation, Zayed University, Abu Dhabi 144534, United Arab Emirates
[3]College of Information Technology, United Arab Emirates University, Al Ain 15551, United Arab Emirates

Corresponding author: Ashraf Khalil (ashraf.khalil@adu.ac.ae)

This work was supported by the Abu Dhabi Award for Research Excellence (AARE) 2019 under Grant 274.

**ABSTRACT** Recent studies have demonstrated that most commercial facial analysis systems are biased against certain categories of race, ethnicity, culture, age and gender. The bias can be traced in some cases to the algorithms used and in other cases to insufficient training of algorithms, while in still other cases bias can be traced to insufficient databases. To date, no comprehensive literature review exists which systematically investigates bias and discrimination in the currently available facial analysis software. To address the gap, this study conducts a systematic literature review (SLR) in which the context of facial analysis system bias is investigated in detail. The review, involving 24 studies, additionally aims to identify (a) facial analysis databases that were created to alleviate bias, (b) the full range of bias in facial analysis software and (c) algorithms and techniques implemented to mitigate bias in facial analysis.

**INDEX TERMS** Algorithmic discrimination, classification bias, facial analysis, bias, unfairness.

## I. INTRODUCTION

Artificial Intelligence (AI) is steadily invading every aspect of our lives. Decisions that have traditionally been executed by humans are increasingly performed by algorithms, ranging from trivial decisions, such as judging in a beauty pageant [1] or classifying sentiment of online hotels and restaurants reviews [2] to much more critical ones like identifying criminal suspects [3] or bail decisions in courtrooms [4]. It has even been used for rating a country's citizens [5]. Automatic facial analysis, a branch of AI, has been utilized in various domains, and its use is expected to increase in coming years. In medicine, it has been used for clinical assessment of depression [6], estimation of pain intensity in non-communicative individuals [7], monitoring progression of motor neuron disease [6] and in computer-aided diagnosis, such as for pre-surgical epilepsy evaluation [8]. It has also been used to monitor the effects of prenatal alcohol exposure on face morphology [9] and to differentiate between normal individuals and those who suffered childhood cancer [10]. In social sciences, it has been utilized to analyze emotions including aggressive feelings [11] and happiness [12] and in recognizing deceptive facial expressions, including estimating smile genuineness [13].

Furthermore, this technology has been utilized to study the connection between subjective evaluation of facial aesthetics and selected objective parameters based on photo quality and facial soft biometrics [14]. In law, facial analysis software can be exploited to monitor adolescent alcohol consumption using selfie photos [15] and to help identify suspects and recognize criminals [4]. In marketing and commerce, it has been employed in targeted marketing by recognizing the race, gender and age of individuals. Other areas of established usage include security and video surveillance, human-computer/robot interaction, communication, entertainment and assistive technologies for education [16]. The premise is that this technology will facilitate standardization, mitigate bias and efficiently serve key purposes, especially in medicine [8], [17].

However, the majority of facial analysis software has been found to be biased against a specific group or category [18]. For instance, in an international beauty contest launched in 2016, Beauty.AI, an automatic face analysis system was used to identify the most attractive contestants based on

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Zia Uddin.

objective factors, such as facial symmetry and wrinkles. The contest received roughly 6000 entries from more than 100 countries. Surprisingly, out of 44 winners, nearly all were white, a handful were Asian, and only one had dark skin. Beauty.AI, which is supported by Microsoft, relied on large datasets of photos to build an algorithm that could assess beauty. The chief science officer of Beauty.AI explained that there could be a number of reasons why the algorithm favoured white people, but the main problem was that the data the project used to establish standards of attractiveness clearly did not include enough darker skinned faces [1]. Although the group did not build the algorithm to promote light skin as a marker of beauty, the input data effectively led the robot judges to reach that conclusion [1]. Even in academic settings, researchers working in facial analysis technology build, train and test their models using open source collections of images [16], [19], [20]. Nonetheless, open source collections are often limited in diversity, and creating a dataset can be a time-consuming and costly option.

The often limited and possibly misrepresentative range of public facial photo databases is a well-known problem [16]. The source material for these databases reflects very poor variability in race, ethnicity, and cultural details. A prime example of this is the MORPH database [21]. It is widely used in the fields of face recognition and age estimation, but this dataset is extremely skewed [16].

The second problem is that large companies and AI service providers are pushing facial analysis technology to become mainstream, without taking the necessary measures to ensure the adequacy, fairness and reliability of the results produced by these systems. For example, the face surveillance technology portion of Amazon's Rekognition program has been marketed aggressively for US law enforcement, and a sheriff's department in Oregon has already started using it [3]. Yet, in a test of the program by the American Civil Liberties Union (ACLU), 28 US Congress members were falsely matched with mugshots of people who have been arrested. Nearly 40 percent of Rekognition's false matches involved people of color, even though that demographic represents only 20 percent of Congress [3]. In another case, a software used across the US for rating a defendant's risk for future crime was found to be biased against blacks [4]. Such risk assessment software has previously been used in conjunction with evaluation of a defendant's rehabilitation needs or determining bail amount, but newer applications of the software have brought its inherent bias to light.

To the best of our knowledge, there is no systematic literature review to date investigating bias and discrimination in facial analysis software. The most comprehensive study investigating this issue comes from the empirical work conducted by Das *et al.* [22]. In their study, they explored the joint classification of gender, age and race and found some facial analysis algorithms and systems to be biased in regard to these categories. Yet, no systematic procedure was followed to review the existing literature. This paper attempts to address these shortcomings by building upon relevant literature. Specifically, the aim of this systematic review is to investigate bias in facial analysis systems/software, algorithms and databases based on published scientific studies. Following are specific objectives of this study:

- To identify facial analysis databases founded to alleviate bias
- To identify the various aspects of discrimination in facial analysis technology
- To recognize algorithms and techniques employed to mitigate bias in facial analysis

The rest of this paper is organized as follows. Section I outlines the methodology employed to conduct the systematic review. Section II then presents the results. In Section III, we discuss the results and Section IV concludes the research study.

## II. METHODS

This review was conducted according to a predetermined protocol and was reported following the PRISMA Statement [23].

### A. DATA SOURCE AND SEARCH STRATEGY

A computerized database search was performed to identify abstracts relevant to the research topic. The strategy was applied to the following databases: Scopus, IEEE Xplore, ACM digital library, the International Prospective Register of Systematic Reviews (PROSPERO), Cochrane Database of Systematic Reviews (CDSR) and Scientific Electronic Library Online (Scielo). There were no restrictions with regard to language, date or status of publication. The initial search was conducted on July 2019 and updated on 2 November 2019 using a computerized database search. One reviewer (SGA) of our study developed the search strategy and conducted the initial search using our proposed keywords listed below.

The search terms were developed using controlled vocabularies and keywords. Two groups of words constituted the search strategy: (1) the method and the body area of interest (facial analysis); and (2) the factor of interest (bias). The Boolean search strings in the article title or abstracts were as follows: ((''face analysis'' OR ''facial analysis'') AND (bias OR discrimination OR unfairness OR disparities)).

#### 1) FIRST ROUND OF SCREENING: TITLE AND ABSTRACT SCREENING

A comprehensive search of the six electronic databases was performed using the pre-defined search strategy and the records retrieved were imported to Excel software. Duplications were then removed. Any additional records identified through manual search were added to the Excel sheet.

Titles and abstracts of identified records were screened according to set criteria by two trained reviewers (AFK and SGA), and the records were then classified into one of the following categories using predefined inclusion and exclusion criteria:

- Potentially eligible, full-text will be accessed
- Exclude
- Unclear

Discrepant opinions between the two reviewers were resolved by discussion and further consultation with a third reviewer (SAI). Reasons for exclusion were documented.

#### 2) SECOND ROUND OF SCREENING: FULL-TEXT SCREENING

Full-texts of potentially eligible studies and those "Unclear" studies were accessed and screened in the same way as in the first round. Reasons for exclusion were documented. Inter-reviewer agreement on study selection was assessed using the $\kappa$ statistics for both rounds of screening.

### B. SELECTION AND VALIDITY ASSESSMENT

We considered studies for inclusion if they involved a facial analysis dataset or an algorithm or software known for handling or introducing a bias in automatic facial analysis.

Specifically, eligible studies should have met all the following criteria: 1) Subjects' age: no restrictions were imposed; 2) Ethnicity/Race: no restrictions were imposed; 3) Technique: only automatic (computerized) facial analysis algorithm, software or database were considered; 4) Reporting of results: standard error could be estimated from the reported values, the reported values should be accurate to one decimal place; 5) Focus: bias; 6) Language: no restrictions were imposed; 7) Date: no restrictions were imposed; 8) Status of publication: no restrictions were imposed; 9) Publication type: conference proceedings or journal articles.

Studies were excluded if one or more cases from the below list were present: 1) Introduced a new facial analysis algorithm, tool or application without directly addressing the problem of bias in automatic facial analysis; 2) Exclusively reported the board and general ethical consequences of AI and Big Data use in society without directly or indirectly tackling facial analysis bias; 3) Focused on the ethical and privacy concerns of facial analysis algorithms and technology; 4) Called for algorithmic transparency as a mechanism to fight bias and discrimination without directly addressing facial analysis algorithms; 5) Studied face analysis bias where the analysis and recognition was not performed by algorithms or machines.

Eligibility of the selected studies was determined by two experts (reviewers) in this domain. In the first-round screening, the reviewers independently read the title and abstracts of articles identified by the search. All the articles that met the inclusion criteria of the systematic review topic were selected and the actual articles collected. The reference lists of the retrieved articles were also hand searched and assessed to identify other potentially eligible studies that may have been missed in the database searches. In the second-round screening, full texts of those records judged to be potentially eligible were assessed for inclusion.

### C. DATA EXTRACTION

Study characteristics and demographics, such as name of the first author, year of publication, study location, origin of the subjects, sample source, sample size, age range, and gender were extracted. Data extraction was performed by one reviewer (SGA) using a predefined piloted spreadsheet in Microsoft Excel 2019 and the results of extraction were then verified by a senior author (AFK).

### D. ASSESSMENT OF RISK BIAS

Since the goal of our study was to identify and highlight bias in facial analysis algorithms, databases and technology, we did not assess eligible studies' risk of bias. Specifically, we were interested in collecting both biased studies and studies trying to identify and alleviate bias in facial analysis technology.

#### 1) ASSESSMENT OF STUDY QUALITY

Risk of bias of eligible studies was assessed by two independent reviewers (AFK AND SGA). Biasness was assessed in four domains of eligible studies:

- Study design
- Appropriateness of statistical analysis

A score of 0, 0.5 or 1 was assigned to each item indicating free of bias, partially free of bias and subject to bias, respectively. Inapplicable items were not scored. A percentage score was calculated for each study by dividing the sum of item scores by the total number of applicable items. the lower the score of a study, the lower the risk that its findings will be biased. A score of 0.40 was used as the cut-off value to differentiate studies with high and low risk of bias.

#### 2) DATA ANALYSIS

Pertaining to the aim of this review, the extracted data was organised into a hierarchical structure in which individual studies were nested within ethnicities/races that in turn were nested within the total population.

## III. RESULTS

### A. LITERATURE SEARCH

The literature search resulted in 70 abstracts distributed as follows: Scopus identified 51, IEEE Xplore 11, ACM digital library 4, PROSPERO 3, CDSR 1 (Issue 7 of 12, July 2019) and Scientific Electronic Library Online (Scielo) 0 references. PROSPERO results were ongoing reviews (not published yet) that are not relevant to the review topic. In CDSR, one result was also irrelevant. Seven studies were indexed in two databases resulting in duplication. Specifically, one article from ACM and six articles from IEEE were indexed in Scopus. Hand searching Google Scholar and reference lists of the relevant articles resulted in the identification of 40 other records. After the first screening round (selection based on titles and abstracts) 43 potentially eligible articles were accessed for full-texts

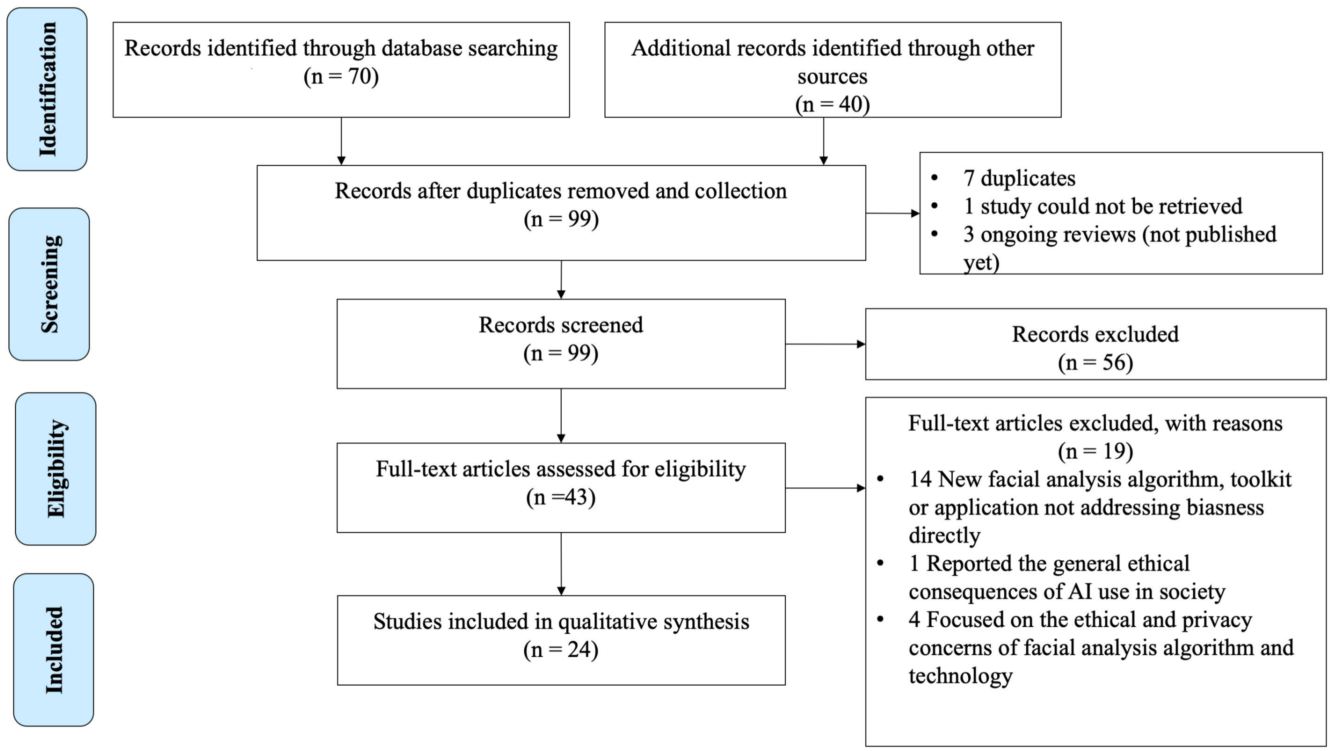


**FIGURE 1.** **Flow diagram of the studies selection [23].**

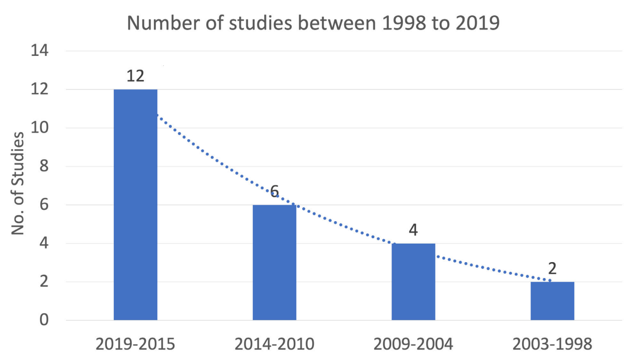


**FIGURE 2.** **Distribution of selected research studies over the years.**

and underwent the second screening round. Of these, 24 eligible articles [6], [16], [24]–[33], [18], [34]–[37], [19], [20], [21], [22], [38], [39], and [40] were included. Fig. 1, summarizes the process of study identification and selection.

### B. STUDIES' CHARACTERISTICS

The studies' year of publication ranged between 1998 and 2019. Fig. 2 shows the distribution of the studies over the years. All studies were in English. For this review, studies were classified into three categories: facial analysis databases (thirteen studies), algorithmic auditing (eight studies) and suggested solutions for bias in facial analysis (three studies). The first category presents public face databases while highlighting bias as a problem. The second category presents algorithmic auditing research studies. An algorithmic audit involves the collection and analysis of outcomes from a fixed algorithm or defined model within a system. Through the stimulation of a mock user population, these audits can uncover problematic patterns in models of interest [38]. Algorithmic audits can play a key role in increasing algorithmic fairness and transparency in commercial facial analysis systems. The third category highlights studies performed with the goal of mitigating bias in facial analysis algorithms and databases. This group of studies provides a specific algorithmic solution for the problem or suggests specific techniques.

### C. CATEGORY 1: FACIAL ANALYSIS DATABASES

Table 1 presents the thirteen studies on public face databases included in the review and their reported demographic distribution. Unconstrained face detection and recognition databases have the largest number of subjects and images, followed by databases created for multiple facial analysis tasks. Facial stimuli (detecting emotions from facial expression) and social cognition databases have the lowest number of subjects and images. Aside from FotW database [16], all the remaining databases' demographic distribution is extremely skewed. This is partly because many research groups have created face databases to represent a specific ethnicity or group, such as the Indian Movie Face Database [34],

**TABLE 1.** ***Public face analysis databases and their subjects' reported distribution.***

| Dataset | Subj. | Images | Purpose | Subjects' Reported Distribution |
|---|---|---|---|---|
| IJB-A [20] | 500 | 5712 plus (2085 video) | unconstrained face detection and recognition | Asia 89, Oceania 7, North America 135, South America 50, Europe 149, Middle East 29, Africa 41 |
| LFW [19] | 5749 | 13233 | | Estimated to be 77.5% male and 83.5% Caucasians [18] |
| PubFig [36] | 200 | 60k | | Ethnicity distribution was not reported by after manually examine the dataset. |
| MS-Celeb-1M [34] | 100K | 10M | | Top 5 represented countries by sample size: America, Great Britain, Germany, Canada, France. Gender: female 80% male 20% |
| FACES [31] | 171 | 2052 | facial stimuli | Strictly Caucasian |
| CMU-Pittsburgh AU-CodedFace Expression [26] | 182 | 1917 | | Gender: 69% female, 31% male, Ethnicity: 81% Euro-American, 13% Afro-American, and 6% other groups |
| RaFD [37] | 67 | 8040 | | Caucasian (20 male adults,19 female adults, 4 male children, 6 female children), Moroccan (18 male adults) |
| Japanese Female Facial Expression (JAFFE) Database [32] | 10 | 213 | | Japanese female models |
| *MORPH* [21] | 13k | 55k | adult age-progression | Ethnicity: African 42589, European 10559, Asian 154, Hispanic 1769, Other 63. Gender: 46645 males, 8489 females |
| FotW [16] | | 25k | Multiple facial analysis tasks (age, gender, face detection and recognition, accessories classification) | 25% Asian, 25% Black, 25% Hispanic and 25% White. 50% men. All ages, from new- born babies to elderly people. |
| Indian Movie Face Database [34] | 100 | 34512 | | Indian actors, 67 male and 33 female actors with at least 200 images for each actor. |
| CAS-PEAL Chinese Face Database [25] | 1040 | 99594 | | 595 males and 445 females |
| CaNAFF [24] | 147 | 441 | social cognition | Male Caucasian faces 75, Male North African faces 72, No females |

Japanese Female Facial Expression (JAFFE) Database [32], CAS-PEAL Chinese Face Database [25] and CaNAFF with 72 Male North African faces [24].

### D. CATEGORY 2: ALGORYTHMIC AUDITING

The eight algorithmic auditing studies included in this review are presented in Table 2. Gender Shades, the first algorithmic audit of performance disparities related to gender and skin type in commercial facial analysis models, developed their own dataset (Pilot Parliaments Benchmark) and used the benchmark to evaluate three commercial gender classification systems (Microsoft, Face ++ and IBM) [18]. The results showed that darker-skinned females are the most misclassified group (with error rates of up to 34.7%) while the maximum error rate for lighter-skinned males is 0.8%. In addition, to confirming the bias and low accuracy in classifying females other algorithmic audit reported low classification accuracy for children and young populations in general [39], [29], [30].

#### 1) THE "OTHER-RACE" EFFECT IN ALGORITHMS

Our ability to perceive the unique identity of other-race faces is limited relative to our ability to perceive the unique identity of faces of our own race. Causes of this phenomenon can be partially attributed to social prejudices, but perceptual factors that begin to develop early in infancy were found to be the primary cause. Specifically, optimizing the encoding of unique features for the types of faces we encounter most frequently—usually faces of our own race, e.g., family members—result in a perceptual filter that limits the quality of representations that can be formed for faces that are not well described by these features. Facial analysis algorithms suffer from this effect as well. Face Recognition Vendor Test (FRVT) 2006 reported results supporting this assumption [40]. In the test, the results of eight algorithms from Western countries were fused together as were, separately, the outcomes of five algorithms from East Asian countries. The false acceptance rate, or FAR, is the measure of the likelihood that the biometric security system will incorrectly accept an access attempt by an unauthorized user. Therefore, for security reasons a low acceptance rate is usually set or chosen before utilizing a certain biometric security system. If the system does not pass the low false acceptance rate, the system will not be utilized. At the low false acceptance rates required for most security applications, the Western algorithms recognized Caucasian faces more accurately than

**TABLE 2.** *Algorithmic auditing studies: purpose and reported results.*

| Study | Goal | systems/models | Benchmark (testing set) | Training sets. | Results |
|---|---|---|---|---|---|
| [18] | To evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups. | Microsoft, Face++, and IBM | Pilot Parliaments Benchmark (PPB) [18] | black-box testing | All classifiers perform better on male faces than female faces and on lighter faces than darker faces. All classifiers perform worst on darker female faces. |
| [38] | To analyze the impact of publicly naming and disclosing performance results of biased AI system. | Kairos, Amazon, Microsoft, Face++, and IBM | Pilot Parliaments Benchmark (PPB) [18] | black-box testing | Reduction in overall error by 5.7%, 8.3% and 7.7% respectively for Microsoft, Face++ and IBM. |
| [6] | To demonstrate a bias affecting the performance of common facial landmark detection and expression recognition algorithms on the faces of older adults with dementia. | Active appearance models (AAM) [41] as implemented in the Menpo project [42], the face alignment network (FAN) [43], face alignment by coarse- to-fine shape searching (CFSS) [44], mnemonic descent method (MDM) [45], and OpenFace [46], Affdex SDK [47] | 688 frontal images from 86 older adults (participant with dementia=42, female =61) | 300W-LP-2D [43]<br><br>300-W [48]<br><br>Multi-PIE [49] | Large training datasets do not capture the variability of a clinical population |
| [40] | To examine algorithm performance as a function of the interaction between the demographic origin of the algorithm (i.e., where it was developed) and the demographics of the population to be recognized. | Algorithms participating in the FRVT 2006, Five algorithms from East Asia and eight algorithms by Western countries [50] | FRVT 2006 database [50] | participants data set(s) [51] | The performance of state-of-the-art face recognition algorithms varies as a joint function of the demographic origin of the algorithm and the demographic structure of the test population. This result is analogous to findings for human face recognition. |
| [39] | To evaluate the age and gender bias in the performance of state-of-the-art pedestrian detection algorithms. | Algorithms included in the Caltech Pedestrian Detection Benchmark | INRIA Person Dataset extended with child, adult, male and female labels | test-set of the INRIA dataset. | The results showed a clear bias in performance: children had higher miss rates on 82% of the algorithms, and on 100% of the 24 top- performing algorithms. The differences were statistically significant. Female pedestrians also had higher miss rates than male on 72% of the algorithms but the differences were not statistically significant. |
| [28] | To investigate whether gender and ethnicity affect each other during classification. | SVM Classifier with three types of features: Pixel-Based, Local Binary Patterns and Histogram of Oriented Gradients | Same as the training using Leave One Out cross validation | FERET [52], TRECVID [53] | The machines, to classify the gender, use characteristics that are shared by all the ethnicity and, vice versa, to classify the ethnicity, use discriminative characteristics present in both male and female. Therefore, using the specified algorithms and set of features, gender and ethnicity classification tasks can be solved separately. |

**TABLE 2.** *(Continued.)Algorithmic auditing studies: purpose and reported results.*

| Study | Goal | systems/models | Benchmark (testing set) | Training sets. | Results |
|---|---|---|---|---|---|
| [30] | To provide an objective, independent, open, and free assessment of current automated gender classification technology. Also to investigate gender classification accuracy across various factors, including age, ethnicity, and constrained versus 'in the wild' facial images. | Five commercial providers (Cognitec , Neurotechnology , NEC , MITRE, Zhuhai-Yisheng) and one university (Tsinghua University ) | Around one million images consisting of: LFW [19], GROUPS, Visa images, Mugshot images, FERET Sketch images | black-box testing | 1. Gender classification is more accurate in males than females. 2. Gender classification accuracy decreases significantly as age increases for adult females. 3. Gender classification is more accurate in adult males (ages 21-60) than young boys (ages 0-10) 4. For females, gender classification is the most accurate in young adults (ages 21-30). 5. Most of the algorithms demonstrate the lowest gender classification accuracy on subjects from Taiwan and Japan, with males being more often misclassified than females. |

East Asian faces and the East Asian algorithms recognized East Asian faces more accurately than Caucasian faces. The FRVT 2006 measured performance with sequestered data (data not previously seen by the researchers or developers) [54]. Researchers concluded that the underlying causes of the ''other-race'' effect in humans applies to algorithms as well.

#### 2) DEEP LEARNING vs. TRADITIONAL FEATURE-BASED ML

Three out of the eight algorithmic auditing studies examined for this survey evaluated commercial facial analysis systems as a black box with no access to the underlying algorithm. However, some of the remaining studies reported performance discrepancies between the different types of algorithms. For instance, the inclusion of healthy older adults in the training data considerably enhanced the performance of both the AAM (feature-based ML) and FAN (deep convolutional neural network) models [6]. Yet, the inclusion of faces of people with dementia in the training data did not improve FAN model results while AAM model demonstrated significant improvement, making its performance comparable to FAN and even far surpassing it in another instance [6].

### *E. CATEGORY 3: SOLUTIONS FOR BIAS IN FACIAL ANALYSIS*

Several studies have proposed solutions to the problem of biased performance across different race and gender subgroups [22], [27], [33]. Klare *et al.* in their algorithmic auditing study stated that the problem of biased data can be solved either by creating datasets that are uniformly distributed across demographics or using a technique called dynamic face matcher selection [29]. These suggested solutions are already known in the scientific community and rather obvious. The other three techniques, however, were built specifically to address the problem of biased demographic distribution algorithmically without creating a balance or uniformly distributed dataset.

#### 1) TRANSFER LEARNING APPROACH

Dwork *et al.* proposed the use of a decoupled classifier [27]. Specifically, they proposed utilization of transfer learning to mitigate the problem of having too little data on any one group. Their decoupling technique can be added on top of any black-box machine learning algorithm, to learn different classifiers for different groups. They argue that the learning of sensitive attributes can be separated from a downstream task in order to maximize both fairness and accuracy. Ryu *et al.* follow a similar approach, using transfer learning. But unlike Dwork *et al.*, who utilized the transfer learning sub-type of domain adaptation, they used the sub-type of task transfer learning. Another difference is that, in their settings, both gender and race are learned as independent, coupled classifiers, not inferred at run time. They have demonstrated the feasibility of using transfer learning with demographics to improve performance across demographic categories [33].

Das *et al.* proposed a Multi-Task Convolution Neural Network (MTCNN) employing joint dynamic loss weight adjustment for the classification of gender, race and age [22]. They added to Facenet network architecture [55] three fully connected layers: one for race classification, a second for gender classification and a third for age classification. They optimized the effect of multi-task facial attribute classification (i.e., gender, age and race) by learning them jointly and dynamically, depending on the degree of relevance of the feature present to each classification task. Precisely, the MTCNN directly learned classification task relations from data instead of subjective task grouping, thus determining the weight of the task sharing. Therefore, they proposed a joint dynamic weighting scheme to automatically assign the loss weights for each task during training.

### 2) THE ROLE OF TRAINING DATA

In an experiment investigating the impact of the demographic distribution in the training set on the performance of a trainable face recognition algorithm, Klare *et al.* [29] showed that the matching accuracy for race/ethnicity and age cohorts can be improved by training exclusively on that specific cohort. In practice, this result can be utilized in a scenario, called dynamic face matcher selection, where several face recognition algorithms (each trained on different demographic cohorts) are accessible by a system operator. The selection of the algorithm to be utilized for classification depends on the demographic information extracted from the probe image. The group also demonstrated that training face recognition algorithms on datasets that are uniformly distributed across demographics provide consistently high accuracy across all cohorts [29].

## IV. DISCUSSION

In this study, we have highlighted the significantly skewed demographic distribution of public face databases (refer to Table 1). The majority of the popular open databases of facial photos were created as part of open challenges [16]. The goal of the competitions was to promote rigorous scientific analysis of face recognition, fair comparison of face recognition technologies, and advances in face recognition research [19]. However, the use of over-simplified datasets has led many people, including large tech companies, to be over-confident in lower-level face analysis capabilities, such as face detection and facial point localization [16]. The popular face databases feature only mildly non-frontal views with decent and constant illumination and little to no occlusion. Additionally, very often these images contain only one face to be analyzed, which greatly simplifies the analysis task [16].

Faces of the World (FotW) dataset has been created with the aim of overcoming these issues [16]. The dataset was created by collecting over 25,000 publicly available images from the Internet, with special emphasis on gathering diverse and balanced data (refer to Table 1). It purposely includes classification for face-related accessories (earrings, hat, glasses, necklace, necktie, headband and neckscarf). Therefore, FotW is a very diverse and technically challenging dataset. Still, these accessories are for the most part associated with Western cultures. Accessories relating to other cultures, to the best of our knowledge, have not yet been featured in any database.

The importance of facial accessories and non-facial clues classification is clearly highlighted by one experiment [32] that was conducted to uncover the mental mechanism humans follow for gender classification when presented with a face image. The study concluded that visual information surrounding the face (clothing, hair, accessories, etc.) in an image are of high importance in classifying the gender of the face, especially when the facial features (eyes, nose, ears, etc.) are ambiguous.

According to such evidence, it is clear that algorithmic audits play a fundamental role in informing strategies for engaging both researchers and corporations in effectively addressing algorithmic bias. Mitigating algorithmic bias is a difficult task that requires a systemic, sociotechnical and holistic perspective [56]. The publication of the Gender Shades study [18], for instance, not only played a significant role in sparking interest in gender and ethnicity classification but also motivated the research community to investigate bias and discrimination in facial analysis algorithms and systems in other domains, for instance, detecting bias against older adults in clinical settings [6] and bias against children in state-of-the-art pedestrian detection algorithms. In addition, it had a major positive effect on commercial gender classification services. Precisely, it motivated target companies to prioritize addressing classification bias in their systems and yielded significant improvements within seven months. In addition to technical updates, organizational and systemic changes took place. Specifically, Microsoft created an "AI and Ethics in Engineering and Research" (AETHER) Committee, investing in strategies and tools for detecting and addressing bias in AI systems, on March 29, 2018, followed by IBM's "Principles for Trust and Transparency" two months later. This practice should continue and flourish over the upcoming years to ensure all groups and cultures will be fairly and accurately represented in a future where AI systems are expected to take main focus in automated decision making in all fields.

We believe bias can be eliminated from facial analysis technology by following the lead of some research groups who have created face databases to represent a particular ethnicity or other demographic, for example, the Indian Movie Face Database [34], Japanese Female Facial Expression (JAFFE) Database [32] and CAS-PEAL Chinese Face Database [25]. Another solution can be realized via the implementation of specific algorithms which utilize transfer learning, as in [22], [27], [33]. We applaud these research groups and their proactive role in the advancement of facial analysis technology while taking into consideration factors, such as race and ethnicity.

Lastly, this survey experience taught us several lessons about the research process that we would like to share. When conducting a systematic literature review, researchers usually have to choose between database searches or backward snowballing. In this survey, we chose database search but the majority of eligible studies were found in the reference list of selected research papers. Therefore, we recommend backward snowballing whenever the initial database search retrieves few research papers (less than 100 studies). Also, following a clear predetermined written research protocol speed up the review process and helped the research team to focus on the research goals and minimize distractions.

## V. CONCLUSION AND FUTURE RESEARCH

Machine learning algorithms and the data that feed them are, at their core, a result of human-provided data and calculations, therefore they are not exempt from reflecting human biases. This is especially true in the case of facial analysis. Recent studies have demonstrated that the majority of commercial facial analysis software and algorithms are biased against certain categories of race, ethnicity, culture, age and gender. Studies have further identified that the main reason for the bias is that open source material for facial image databases, utilized in commerce and academia for training facial analysis algorithms, reflects very poor variability in these categories. This study identified and highlighted the bias existing in public face databases by way of a systematic review and categorization of scientific approaches to understanding and eliminating such bias. Furthermore, some algorithmic solutions to this issue were presented.

Finally, the systematic review has helped us to establish a series of lines of research in the area of facial analysis and detection. A future line of research would be the definition of a formal guide or process model for conducting facial analysis algorithmic auditing. This will help establish a standardized process that will encourage further research in this direction. Also, it would be interesting to investigate the effect of biased facial analysis datasets on different types of machine learning algorithms (for instance deep learning vs. traditional machine learning). Another line of research could be the development of multiple face benchmark databases that include diverse racial, ethnic and cultural differences.

## REFERENCES

[1] S. Levin. (2020). A beauty contest was judged by AI and the robots didn't like dark skin. The Guardian. Accessed: Aug. 28, 2019. [Online]. Available: https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people

[2] M. Thelwall, "Gender bias in sentiment analysis," *Online Inf. Rev.*, vol. 42, no. 1, pp. 45–57, 2018, doi: 10.1108/oir-05-2017-0139.

[3] J. Snow. (2018). Amazon's face recognition falsely matched 28 members of congress with Mugshots. American Civil Liberties Union. Accessed: Aug. 28, 2019. [Online]. Available: https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

[4] J. Angwin, J. Larson, S. Mattu, and L. Kirchner. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. Accessed: Aug. 28, 2019. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[5] R. Botsman. (2017). Big data meets Big Brother as China moves to rate its citizens. Wired. Accessed: Mar. 2, 2020. [Online]. Available: https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion

[6] B. Taati, S. Zhao, A. B. Ashraf, A. Asgarian, M. E. Browne, K. M. Prkachin, A. Mihailidis, and T. Hadjistavropoulos, "Algorithmic bias in clinical populations–evaluating and improving facial analysis technology in older adults with dementia," *IEEE Access*, vol. 7, pp. 25527–25534, 2019, doi: 10.1109/ACCESS.2019.2900022.

[7] C. Florea, L. Florea, and C. Vertan, "Learning pain from emotion: Transferred HoT data representation for pain intensity estimation," in *Computer Vision—ECCV 2014* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), L. Agapito, M. Bronstein, and C. Rother, Eds. Zurich, Switzerland: Springer, 2015, pp. 778–790, doi: 10.1007/978-3-319-16199-0_54.

[8] D. Ahmedt-Aristizabal, C. Fookes, K. Nguyen, S. Denman, S. Sridharan, and S. Dionisio, "Deep facial analysis: A new phase i epilepsy evaluation using computer vision," *Epilepsy Behav.*, vol. 82, pp. 17–24, May 2018.

[9] M. Suttie, J. R. Wozniak, S. E. Parnell, L. Wetherill, S. N. Mattson, E. R. Sowell, E. Kan, E. P. Riley, K. L. Jones, C. Coles, T. Foroud, P. Hammond, and T. Cifasd, "Combined face-brain morphology and associated neurocognitive correlates in fetal alcohol spectrum disorders," *Alcoholism, Clin. Exp. Res.*, vol. 42, no. 9, pp. 1769–1782, Sep. 2018, doi: 10.1111/acer.13820.

[10] S. M. J. Hopman, J. H. M. Merks, M. Suttie, R. C. M. Hennekam, and P. Hammond, "3D morphometry aids facial analysis of individuals with a childhood cancer," *Amer. J. Med. Genet. Part A*, vol. 170, no. 11, pp. 2905–2915, Nov. 2016, doi: 10.1002/ajmg.a.37850.

[11] T. Moriyama, K. Abdelaziz, and N. Shimomura, "Face analysis of aggressive moods in automobile driving using mutual subspace method," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Tsukuba, Japan, Nov. 2012, pp. 2898–2901.

[12] M. Hao, G. Liu, A. Gokhale, Y. Xu, and R. Chen, "Detecting happiness using hyperspectral imaging technology," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–16, Jan. 2019, doi: 10.1155/2019/1965789.

[13] M. Kawulok, J. Nalepa, K. Nurzynska, and B. Smolka, "In search of truth: Analysis of smile intensity dynamics to detect deception," in *Advances in Artificial Intelligence—IBERAMIA 2016* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), M. GómezHugo, H. Escalante, A. Segura, and J. Murillo, Eds. San José, Costa Rica: Springer, 2016, pp. 325–337, doi: 10.1007/978-3-319-47955-2_27.

[14] A. Dantcheva and J. Dugelay, "Female facial aesthetics based on soft biometrics and photo-quality," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2011, pp. 1–6.

[15] R. Pang, A. Baretto, H. Kautz, and J. Luo, "Monitoring adolescent alcohol use via multimodal analysis in social multimedia," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Santa Clara, CA, USA, Oct. 2015, pp. 1509–1518, doi: 10.1109/BigData.2015.7363914.

[16] S. Escalera, M. A. Bagheri, M. Valstar, M. T. Torres, B. Martinez, X. Baro, H. J. Escalante, I. Guyon, G. Tzimiropoulos, C. Corneanu, and M. Oliu, "ChaLearn looking at people and faces of the world: Face AnalysisWorkshop and challenge 2016," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 1–8.

[17] M. Suttie, L. Wetherill, S. W. Jacobson, J. L. Jacobson, H. E. Hoyme, E. R. Sowell, C. Coles, J. R. Wozniak, E. P. Riley, K. L. Jones, T. Foroud, P. Hammond, and t. CIFASD, "Facial curvature detects and explicates ethnic differences in effects of prenatal alcohol exposure," *Alcoholism, Clin. Exp. Res.*, vol. 41, no. 8, pp. 1471–1483, Aug. 2017.

[18] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proc. Conf. Fairness, Accountability Transparency*, 2018, pp. 77–91.

[19] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Proc. Workshop Faces 'Real-Life' Images, Detection, Alignment, Recognit.*, Marseille, France, Oct. 2008.

[20] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus benchmark a," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1931–1939, doi: 10.1109/CVPR.2015.7298803.

[21] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, Southampton, U.K., Apr. 2006, pp. 341–345, doi: 10.1109/FGR.2006.78.

[22] A. Das, A. Dantcheva, and F. Bremond, "Mitigating bias in gender, age and ethnicity classification: A multi-task convolution neural network approach," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), L. Leal-Taixé and S. Roth, Eds. Munich, Germany: Springer, Sep. 2018, pp. 573–585, doi: 10.1007/978-3-030-11009-3_35.

[23] D. Moher, A. Liberati, J. Tetzlaff, and D. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLoS Med.*, vol. 6, no. 7, 2009, Art. no. e1000097, doi: 10.1371/journal.pmed.1000097.

[24] R. Courset, M. Rougier, R. Palluel-Germain, A. Smeding, J. M. Jonte, A. Chauvin, and D. Müller, "The Caucasian and North African French faces (CaNAFF): A face database," *Int. Rev. Social Psychol.*, vol. 31, no. 1, p. 22, Jul. 2018, doi: 10.5334/irsp.179.

[25] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 1, pp. 149–161, Jan. 2008, doi: 10.1109/TSMCA.2007.909557.

[26] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Grenoble, France, Mar. 2000, pp. 46–53, doi: 10.1109/AFGR.2000.840611.

[27] C. Dwork, N. Immorlica, A. T. Kalai, and M. Leiserson, "Decoupled classifiers for group-fair and efficient machine learning," in *Proc. Conf. Fairness, Accountability Transparency*, 2018, pp. 119–133.

[28] G. Farinella and J.-L. Dugelay, "Demographic classification: Do gender and ethnicity affect each other?" in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, Dhaka, Bangladesh, May 2012, pp. 383–390, doi: 10.1109/ICIEV.2012.6317383.

[29] B. F. Klare, M. J. Burge, J. C. Klontz, R. W. Vorder Bruegge, and A. K. Jain, "Face recognition performance: Role of demographic information," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1789–1801, Dec. 2012, doi: 10.1109/TIFS.2012.2214212.

[30] M. Ngan and P. Grother, "Face recognition vendor test (FRVT)-performance of automated gender classification algorithms," NIST Interagency/Internal Report (NISTIR), Tech. Rep. 8052, 2015, doi: 10.6028/nist.ir.8052.

[31] N. Ebner, M. Riediger, and U. Lindenberger, "FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behav. Res. Methods*, vol. 42, no. 1, pp. 351–362, 2010, doi: 10.3758/brm.42.1.351.

[32] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Nara, Japan, Apr. 1998, pp. 200–205, doi: 10.1109/AFGR.1998.670949.

[33] H. J. Ryu, H. Adam, and M. Mitchell, "InclusiveFaceNet: Improving face attribute detection with race and gender diversity," 2017, *arXiv:1712.00193*. [Online]. Available: http://arxiv.org/abs/1712.00193

[34] S. Setty, M. Husain, P. Beham, J. Gudavalli, M. Kandasamy, R. Vaddi, V. Hemadri, J. C. Karure, R. Raju, B. Rajan, V. Kumar, and C. V. Jawahar, "Indian movie face database: A benchmark for face recognition under wide variations," in *Proc. 4th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph. (NCVPRIPG)*, Jodhpur, India, Dec. 2013, pp. 1–5, doi: 10.1109/NCVPRIPG.2013.6776225.

[35] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in *Computer Vision—ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Amsterdam, The Netherlands: Springer, 2016, pp. 87–102, doi: 10.1007/978-3-319-46487-9_6.

[36] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep. 2009, pp. 365–372, doi: 10.1109/ICCV.2009.5459250.

[37] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the Radboud faces database," *Cognition Emotion*, vol. 24, no. 8, pp. 1377–1388, Dec. 2010.

[38] I. Raji and J. Buolamwini, "Actionable auditing," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, 2019, pp. 429–435, doi: 10.1145/3306618.3314244.

[39] M. Brandao, "Age and gender bias in pedestrian detection algorithms," 2019, *arXiv:1906.10490*. [Online]. Available: https://arxiv.org/abs/1906.10490

[40] P. J. Phillips, F. Jiang, A. Narvekar, J. Ayyad, and A. J. O'Toole, "An other-race effect for face recognition algorithms," *ACM Trans. Appl. Perception*, vol. 8, no. 2, pp. 1–11, Jan. 2011, doi: 10.1145/1870076.1870082.

[41] I. Matthews and S. Baker, "Active appearance models revisited," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 135–164, Nov. 2004, doi: 10.1023/b:visi.0000029666.37597.d3.

[42] J. Alabort-i-Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou, "Menpo: A comprehensive platform for parametric image alignment and visual deformable models," in *Proc. ACM Int. Conf. Multimedia (MM)*, 2014, pp. 679–682, doi: 10.1145/2647868.2654890.

[43] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks)," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1021–1030.

[44] S. Zhu, C. Li, C. C. Loy, and X. Tang, "Face alignment by coarse-to-fine shape searching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4998–5006.

[45] G. Trigeorgis, P. Snape, M. A. Nicolaou, E. Antonakos, and S. Zafeiriou, "Mnemonic descent method: A recurrent process applied for end-to-end face alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4177–4187, doi: 10.1109/cvpr.2016.453.

[46] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," CMU School Comput. Sci., Tech. Rep. CMU-CS-16-118, 2016.

[47] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. E. Kaliouby, "AFFDEX SDK," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst. (CHI EA)*, 2016, pp. 3723–3726, doi: 10.1145/2851581.2890247.

[48] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "A semi-automatic methodology for facial landmark annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 896–903.

[49] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, May 2010, doi: 10.1016/j.imavis.2009.08.002.

[50] P. J. Phillips, K. W. Boyer, P. J. Flynn, A. J. O'Toole, P. J. Phillips, C. L. Schott, W. T. Scruggs, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale results," Tech. Rep. NISTIR 7408, 2007, doi: 10.6028/nist.ir.7408.

[51] *Face Recognition Vendor Test (FRVT) 2006*. NIST. Accessed: 2020. [Online]. Available: https://www.nist.gov/itl/iad/image-group/face-recognition-vendor-test-frvt-2006

[52] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, Apr. 1998, doi: 10.1016/s0262-8856(97)00070-x.

[53] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVid," in *Proc. 8th ACM Int. Workshop Multimedia Inf. Retr. (MIR)*, 2006, pp. 321–330, doi: 10.1145/1178677.1178722.

[54] NIST. (2006). *Face Recognition Vendor Test (FRVT) 2006*. Accessed: Aug. 26, 2019. [Online]. Available: https://www.nist.gov/itl/iad/image-group/face-recognition-vendor-test-frvt-2006

[55] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

[56] C. Draude, G. Klumbyte, P. Lücking, and P. Treusch, "Situated algorithms: A sociotechnical systemic approach to bias," *Online Inf. Rev.*, vol. 44, no. 2, pp. 325–342, 2019, doi: .10.1108/oir-10-2018-0332.

**ASHRAF KHALIL** (Member, IEEE) received the Ph.D. degree in computer science from Indiana University, USA. He is currently the Director of Research and a Professor of computer science and information technology with Abu Dhabi University. His work was featured in public media, including the UAE's the National and the USA's National Public Radio. Specializing in applying and developing usability testing techniques to mobile applications, he is interested in applying diverse technological innovations in addressing pertinent problems. His researches are published in top conferences in the fields, such as CHI, CSCW, and INTERACT. His research interests include ubiquitous computing, social and mobile computing, bioinformatics, persuasive computing, and human-computer interaction. He was a recipient of many awards and research grants, including the Second Place from the National Made in UAE Competition, the Abu Dhabi University Ambassador Award, and the Leadership Award.

**SOHA GLAL AHMED** received the B.S. degree in computer science from Abu Dhabi University, Abu Dhabi, United Arab Emirates, in 2008, and the M.S. degree in computer engineering from the American University of Sharjah, Sharjah, United Arab Emirates, in 2014. For a period of 12 years, she was a Researcher throughout multiple universities in the region. From 2010 to 2012, she was also an Instructor with the American University of Sharjah. Her research interests include natural language processing (NLP), text mining, computational intelligence, digital signal processing, human-computer interaction, and biomedical sensors and systems.

**ASAD MASOOD KHATTAK** (Member, IEEE) received the M.S. degree in information technology from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2008, and the Ph.D. degree in computer engineering from Kyung Hee University, South Korea, in 2012. He was a Postdoctoral Fellow with the Department of Computer Engineering, Kyung Hee University, where he was an Assistant Professor. He joined Zayed University, Abu Dhabi, United Arab Emirates, in August 2014, where he is currently an Associate Professor with the College of Technological Innovation. He has authored/coauthored more than 120 journal and conference articles in highly reputed venues. He has delivered keynote speeches, invited talks, guest lectures, and delivered short courses in many universities. He is leading three research projects, collaborating in four research projects. He has successfully completed five research projects in the fields of data curation, context-aware computing, the IoT, and secure computing. He and his team received several national and international awards in different competitions. He serves as a Reviewer, a program committee member, and a guest editor for many conferences and journals.

**NABEEL AL-QIRIM** (Senior Member, IEEE) was a Business and IT Consultant for a period of ten years with the IT Industry, from 1989 to 1998. He has been in academia with United Arab Emirates University (UAEU) and the Auckland University of Technology, since 1999. More than 30 years of professional and academic experience shaped his teaching and research directions and approaches. He is currently a Professor with the College of Information Technology, UAEU. He is a Passionate Teacher and a Developer with the different Information Systems, Business, and Innovation and Entrepreneurship Courses. He is also a Keen Researcher (h-index 20). He has published more than 150 articles in refereed and highly-impact research outlets. He has lead/participated in several panels, workshops, conferences, and journal's special-issues. His research interests include different information systems topics, innovation and entrepreneurship, and IT-education/pedagogy mostly related to investigating the adoption and diffusion of different technologies in different contexts using both qualitative (case studies and focus group) and quantitative (surveys) methods. He is a member with leading academic and professional associations, such as ACM and AIS, and the editorial advisory board of several journals.