Review: Single attribute and multi attribute facial gender and age estimation

Sandeep Kumar Gupta[1] [ID] **· Neeta Nain[1]**

Abstract

Facial age and gender recognition have vital applications as consumer profile prediction, social media advertisement, human-computer interaction, image retrieval system, demographic profiling, customized advertisement systems, security and surveillance. This paper presents a study on Single Attribute (Attribute: either Gender or Age) and Multi-Attribute (both Gender and Age) prediction model. We present a review for facial age estimation and gender classification methods based on conventional as well as deep learning approaches developed so far with analysis of their pros, cons and insights for future research. Moreover, this study also enlists the databases used for benchmarking results with their properties for both constrained and unconstrained environment.

Keywords Multi attribute · Face · Age estimation · Gender classification · Regression

1 Introduction

The face of a person provides different attributes information for recognition as gender, age, emotion and ethnicity. Facial gender and age prediction has received considerable attention among these attributes due to their wide application and use cases. Facial Gender recognition is defined as classifying the person's sex based on facial pattern into its labelled class (male, female). Facial age prediction is known for automatically predicting the person's biological age or its age group as a child, adult and senior citizen etc.

Facial Gender prediction and age estimation have real-time commercial applications as non-invasive forensic determination of victim/criminal's profile, surveillance of specific gender and age group, human-computer interaction, law enforcement, access control and interactive systems etc. For surveillance and access control, it can be used to limit the entry

✉ Sandeep Kumar Gupta
  sandeepmbm@gmail.com

  Neeta Nain
  nnain.cse@mnit.ac.in

[1] Department of Computer Science & Engineering, Malaviya National Institute of Technology, Jaipur-302017, Rajasthan, India

of person belonging to a specific sex or age group into prohibited areas; to permit website access based on specific age group or specific gender; in web-application access; in access control of physical zones (smoking zone, washroom) and risky zones as theme park etc. [76]. For example, in Japan, vending machines are used to recommend beverages (alcohol, smoke packets) based on facial adult estimation (age) of customers [1].

For commercial CCTV applications, It can be used in demographic analysis or any access violation in the crowd. For example, train compartments (or seats), metros, buses, washrooms and hostels have restricted access to a certain gender, the passengers or visitors can be auto directed and monitored for any law violation. Moreover, these predictions can be used for sales and marketing strategy, business planning as finding the number of visitors with profiles (male, female, juvenile, young, adult) in specific zones as public places, malls, banks etc. It can be used for targeted advertising on electronic boards specific to dynamically changing gender and age groups [36]. Facial gender and age estimation prediction systems can also be used for customizing services like auto health care systems (Robotic nurse) in healthcare units [29]. Recently gender and age estimation is added into smartphones as entertainment features. These can be used for automatic album reorganization for managing features as rearrange, retrieve, delete the captured photos according to the selection of age and gender. In human-computer interaction based systems as an auto-HR interview system to recognize a person face attribute as gender, age during physiological behaviour analysis. Gender recognition can also be used for reducing the search index of the database in bio-metric systems. It also increases the accuracy of person identification with age and gender as face attributes [71]. These estimations can also be used for information retrieval as forensic art for predicting the best match of lost people with face recognition application [84]. Moreover, it can be used to generate the updated facial image of an outdated family member or missing children using facial age synthesis.

There are two approaches for facial attribute prediction as shown in Fig. 1. (a) Single Attribute Learning (SAL) or Single Task Learning (STL) (b) Multi-Attribute Learning (MAL) or Multi-Task Learning (MTL). In the SAL/STL based approach, each attribute/class (gender or age) is trained or predicted separately without any correlations between the different attributes. While MAL (MTL) approach includes learning of multiple attributes (for gender and age prediction) using a shared parallel model.

The gender can be predicted with face, voice data, gait analysis (running and jogging etc.), facial images, fingerprints, hand skin and handwriting. While age can be predicted using anthropology study of bones or face. The face is the best suitable attribute due to its easy visibility (not covered clothes), collectability, acceptability and universality. The existing state of the art methods for facial age and gender recognition can be categorized into two categories: (a)Conventional hand-crafted feature engineering approach and (b)Deep learning based approach.

The conventional features depends on feature descriptors crafted by algorithm developers for a unique representation of age and gender pattern from facial images. These features are effective for meaningful representation of age and gender recognition in controlled settings but performance varies for uncontrolled cases which are not taken into account during designing. The different methods of conventional feature engineering include texture based methods as Local Binary Pattern (LBP), Histogram of Gradient (HOG); Haar based features; dimension reduction techniques as PCA, ICA; feature separation techniques as Discrete Cosine Transform (DCT); Scale Invariant Feature Transform (SIFT) features from facial landmarks (distances or statistics). These methods and their different improved versions with different classification techniques are proposed so far.
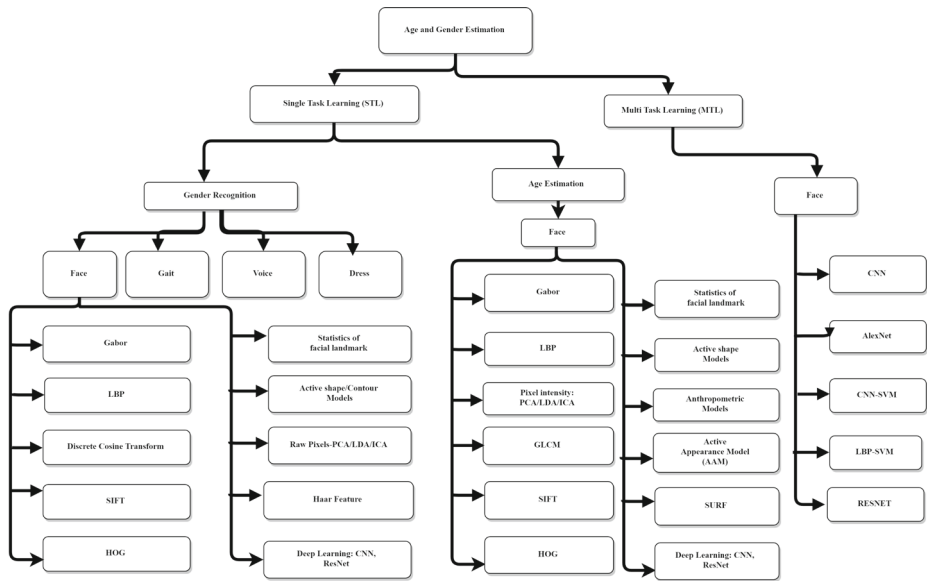
Fig. 1 Age and gender classification approaches

Recently, the deep learning based approach has attracted immense research interest due to the easy availability of multimedia data and improved computational systems like GPUs. Deep learning approach considers Convolutional Neural Network (CNN) for extracting features for determination of age/gender from large facial image data sets by statistical training with powerful nonlinear modelling ability. Over-fitting is possible whenever the network becomes complex and deep and underfitting in simple models if facial images are insufficient in numbers.

Multi-attribute prediction (MAL or MTL) as age estimation and gender recognition through face increase the dimension of use cases and acceptability of application. Moreover, MTL increases the accuracy of the system with co-variate combined prediction where a single attribute does not provide prediction correctly. The deep learning approach is good for MAL/MTL; so different methods are proposed for multi-attribute learning (age and gender) with CNN.

In this paper, the main contributions are the following:

– Comparative analysis and benchmarking of constrained and unconstrained data- sets for facial gender and age estimation.
– Study and analysis of different Single Attribute Learning (SAL) and Multi- Attribute Learning (MAL) approaches for gender recognition and age estimation.
– Inference the advantages and disadvantages of different models including conventional learning and deep learning approaches on all age groups including a juvenile, teens, adults up to senior citizens.

In the remaining paper, Section 2 reviews different datasets which are used widely for age and gender recognition with different real-time challenges. Section 3 gives a review of the development of SAL/STL based age/ gender estimation; Section 4 describes MAL/MTL based age, gender estimation; Section 5 discuss the analysis with pros and cons. Finally, Section 6 provides the conclusion of this study and review.

2 Review analysis and discussion of facial dataset for gender recognition and age prediction

A facial dataset is essential for covariate research study and benchmarking on different challenges of gender recognition and age prediction. We studied the most used dataset in gender and age estimation systems. A face dataset must have sufficient images, subjects, age variations, gender, race distribution and real-time environmental variability for exploration of Facial Attribute Recognition(FAR) systems. FAR Systems perform better on highly constrained images but in a real scenario, they may misbehave due to unconstrained challenges. There are huge real-life factors of face imaging which affect the performance of a system as Resolution(R), Sharpness(S), Illumination(I), Expression(E), Occlusion(O), Profile(P), Frontal View(F), Constrained Environment(C), Unconstrained Environment(U), Longitudinal(L), Race(R), Hair(H) and Scale(Sc) etc. Different pro-perties of datasets are shown in Table 1. This table considers both available datasets.

The review divides the study of the datasets into two parts (1) controlled dataset and (2) uncontrolled dataset. A controlled dataset is prepared in a particular controlled environment with limited variability during data capture, while uncontrolled dataset includes different variability of real-life challenges. Controlled Dataset as FG-NET [28], UIUC-IFP-Y Internal Aging [30], MORPH [80] facial datasets are captured in a constrained environment; labelled with exact age and gender attributes. FGNET dataset [28] contains face image series with various age progressions, it has limited number of images (1, 000), of 82 subjects prepared by scanning photos. Some factors make it a challenging dataset due to variations in illumination, background, resolution and noise of scanner during scanning. Unsurprisingly, performance in terms of error (mean average age) has saturated upto 5% approximately on FG-NET [28]. CLF dataset [23] contains longitudinal images of 10000 child faces from 3000 children with age progression, 40% of the subjects are females. This dataset is used to study ageing effects in children, but it is not publicly available. MORPH dataset [80] is another most used dataset which is divided into various albums or subsets. It includes information about the birth date, gender, ethnicity and date of acquisition. Academic version consist of 55,000 images of approximately 13,000 persons captured under controlled conditions in which 42,589 images are of African ethnicity, and others are Asian, European and Hispanic [80]. UIUC-IFP-Y dataset for Internal Aging [30] contains 8,000 images acquired under lab-controlled conditions, related to 1,600 subjects of Asian origin labelled with gender [30].

VADANA data set [88] consists of 2,298 images of 43 subjects and permit study of age progression of the same face by providing same subject's multiple images at different ages. Large number (approximately 168000) of intra-personal pairs make it distinguish from other datasets. The Cross-Age Celebrity dataset (CACD) consist of 163,446 frames of 2000 persons collected from Internet and labelled with their date of birth [15]. FERET dataset [78] is very widely used for facial gender recognition in a controlled environment, having a better resolution as compared to MORPH and FG-NET. It has adult gender information, profile variation and more representative information of texture which makes it better for extracting local descriptors.

Uncontrolled Dataset as LFW [48] and IMDB-WIKI [82] are publicly available datasets for age and gender estimation of people in the wild (unconstrained). The ChaLearn Looking at People (LAP) [27] is collected via crowd-sourcing; used for apparent age estimation in which 4699 images are labelled with age. Adience dataset [26] is cross-sectional that contains non-adult subjects and label them for different age groups but does not contain

Table 1 Available dataset for gender or age prediction: Table explain different datasets number of images, number of subject in dataset, class labelled as age group or exact age in numerical, gender; environmental variability/challenges. The environmental variability/challenges are defined as Resolution(R), Sharpness(S), Illumination(I), Expression(E), Occlusion(O), Profile(P), Frontal View(F), Constrained Environment(C), Unconstrained Environment(U), Longitudinal(L), Race(R), Hair(H) and Scale(Sc) etc

Dataset	Images, subjects	Age type Age Range	Gender	Environment Challenges
UIUC-IFP-Y [30]	8000, 1600	Exact age	Yes	–
MORPH-II [80]	55134 13618	Exact Age 16 to 77	Yes	C,L,R,H
VADANA [88]	2298 43	Age Group	Yes	C,P,I,E
CACD [15]	163446 2000	Exact age 16 to 62	Yes	–
LAP [27]	4699	Age Group	Yes	–
CLF [23]	3682 919	absolute Age 2 to 18	Yes	C,I,E,Sc L,O,F
CelebA [68]	202599 10,177	absolute Age 2 to 18	Yes	C,I,E L,O,F,Sc
FG-NET [28]	1002, 82	Exact Age 0 to 69	Yes	R,I,E,O,S,Sc
UTK Face [105]	20000	absolute Age 0 to 116	Yes	C,I,E L,O,F,Sc
FERET [78]	14051 1199	– 10 to 70	Yes	C,P
LFW [48]	13233 5749	–	Yes	U
Adience [26]	26580 2284	Age Group 0 to 100	Yes	U
PbFig [55]	58797 200	Age Group	Yes	U
IMDB WIKI [82]	523,051 20284	Exact Age 0 to 100	Yes	U
Gallagher Image group [32]	5080 28231	Age Group	Yes	U

longitudinal information. It includes all the variations for a real-world scenario like appearance, lighting, pose and noise etc. IMDB-WIKI [82] has half a million (approximately) collection of age labelled images from 10 to 90 years which makes it the largest publicly available dataset for facial attribute recognition. It is a joint collection of IMDB (460723 images of 20284 celebrities) and WIKI (62238 images). IMDB-WIKI includes the real-life challenges as rotation, pose variation, illumination, poor qualities, sketch faces and human comic faces etc. It contains blank images which affect the network prediction in an adverse manner [82]. In [32], Gallagher and Chen introduced a dataset for group photo study. This

dataset includes low-resolution images of frontal faces with multiple subjects that make it difficult to recognize correctly. Public Figures dataset (PubFig) [55] is created for facial attribute recognition that consists of 60,000 high-quality images of only 200 celebrity faces from media and news websites associating camera centered on the face of a person in an uncontrolled environment.

3 Review, analysis and discussion for single attribute learning based gender recognition and age estimation

Single Task Learning (STL) based Facial Attribute Recognition (gender and age) is divided into two parts:- (1) Gender prediction and (2) Age estimation.

3.1 Facial gender prediction

From the literature, it is found that gender can be predicted using gait image, voice data, dress image and facial image of a person while biological age can be predicted mostly from facial images. The most used techniques are based on different key methods as shown in Figure 1. This study review and analyze the different STL based gender attribute classification methods using facial data, developed so far from 1991 to 2020. These important different state of the art methods are presented in Table 2.

The facial gender recognition systems are studied in two different categories based on learning approach used as (1) Conventional learning with handcrafted features and (2) Deep learning based approach as shown in Table 2. These techniques, shown in Table 2, are discussed with detail analysis separately in following subsections.

3.1.1 Conventional learning based facial gender recognition

Facial gender recognition is classified into two approaches based on feature space and feature extraction approach: appearance based feature extraction (Global Features) and geometry (Local Features) based feature extraction. In appearance based approach, the whole face is considered as feature space while geometric-based methods perform feature extraction from prominent facial parts as eyebrow, nose, lip etc. Cottrell [20] proposed gender recognition in 1991 on constraint environment with autoencoder and backpropagation on private frontal view dataset. [10, 42, 74] have extended gender recognition on profile variation facial data as FERET dataset using raw pixels for feature engineering with different classifiers like decision tree, support vector machine and AdaBoost. SVM-RBF classifier outperforms as RBF can better classify in high dimensions compared to linear plane with raw features generated through facial images. Kim [54] proposed appearance based approach with Gaussian kernel using raw pixels on AR dataset for gender recognition.

Facial gender is also classified based on feature dimension reduction techniques as Independent Component Analysis (ICA) with Linear Discriminant Analysis(LDA) classification [49], Principal Component Analysis (PCA) with neural network classification [52], 2D-PCA with SVM classification [69] and PCA with LDA classification [11]. The best results are achieved by Independent Component Analysis (ICA) with LDA classifier on FERET dataset. PCA and ICA both are used for dimension reduction but feature vectors of ICA are known as spatially independent basis vectors which can better distinguish the inter class variations compared to PCA.

Table 2 Comparative analysis of performance based on different handcrafted feature engineering with conventional learning as well as deep learning approach for facial gender recognition techniques. The table includes different state of art methods with author, dataset used, feature extraction technique, classification method, and performance in terms of accuracy(%)

	Authors	Dataset	Feature Extraction	Classification	Accuacy(%)
Conventional learning	Cottrell [20]	Private	Autoencoder	Backpropagation	99.0
	Gutta et al. [42]	FERET	Raw pixels	RBF + Decision tree	96.0
	Moghaddam[74]	FERET	Raw pixels	SVM-RBF	96.62
	Shakhnarovic[85]	Web	Haar feature	Adaboost	79.0
	Jain et al. [49]	FERET	ICA	LDA	99.30
	Khan et al. [52]	Private	PCA	NN	88.70
	Kim et al. [54]	AR	Raw pixels	GPC	97.0
	Baluja et al. [10]	FERET	Raw Pixels	Adaboost	94.3
	Yang et al. [99]	Private	LBP	Adaboost	96.3
	Leng et al. [59]	FERET	Gabor+Fuzzy	SVM	98.0
	Makinen [71]	FERET	LBP, Haar	ANN,SVM	92.86
	Li et al. [63]	YGA	DCT	Spatial GMM	92.5
	Lu et al. [69]	FERET	2D PCA	SVM-RBF	94.85
	Wang et al. [95]	FERET	SIFT,context	Adaboost	95.0
	Alexandre et al. [7]	FERET	LBP	SVM-linear	99.07
	Guo et al. [36]	YGA	LBP,HOG,BIF	NonlinearSVM	89.28
	Alamri et al. [6]	FERET	LBP, WLD	N. Neighbor	98.82
	Bissoon [11]	FERET	PCA	LDA	85
	Hassner [45]	Adience	LBP+FPLBP	SVM	79.3
	Yildirim [101]	Private	HOG	Random Forest	92.3
				Adaboost	85.6
	Abbas et al. [2]	ALSPAC	Geodesic Path	LDA	87.3
Deep learning	Van et al. [93]	FERET	CNN	Oversampling	97.30
	Antipov [8]	LFW	CNN	Softmax	97.31
	Mansanet [72]	LFW	DCNN	Class-posterior	94.48
	Jia et al. [50]	LFW	CNN	Softmax	98.90
	Cirne et al. [18]	AR Face	Geometric Descriptor	CNN	97.50
	Aslam et al. [9]	FERET	CNN	VGG-16	98.90
	Simanjuntak [87]	LFW	COSFIRE	VGG Face	99.28
	Afifi et al. [3]	Adience	Isolated+Hol-	DCNN+	90.43
		FERET	istic features	Adaboost	99.28
		LFW	(Foggy face)	score fusion	95.98
	DAmelio [21]	LFW	DCNN	Sparse dictionary learning	95.13
	Moeini et al. [73]	FERET	Gray and LBP	Separate dictionary learning	99.90
		LFW	pixel values		99.0

Local Binary Pattern (LBP) and Histogram of Gradient (HOG) both are used for generating texture descriptors. Yildirim et al. [101] achieved 85.6% and 92.3% accuracy by HOG features with Adaboost and Random Forest classifiers respectively. LBP with adaboost classifier [99] achieved 96.3% accuracy and performs better compared to HOG with adaboost achieving 85.6% accuracy [101] and SIFT with adaboost achieving 95% accuracy [96] . The SVM linear outperforms with LBP features. Alexandre [7] achieved 99.07% accuracy using LBP-SVM-Linear kernel on adult face (FERET dataset) which is captured in constraint environment while [45] achieved 79.3% using LBP, FPLBP with SVM on Adience data which is captured in unconstrained environment in all age group. These results show that LBP is only suitable for constrained images. The Adience dataset also includes images of children faces which have limited gender distinguishing features. The uncontrolled environment further makes it harder to recognize.

3.1.2 Deep learning based facial gender recognition

As shown in Table 2; Antipov et al. [8] introduced ensemble model based on CNN for facial gender recognition and performance is measured as 97.31% on the LFW dataset. They used 3 CNNs in the ensemble model and optimized the last CNN in case of computation and memory requirements. Mansanet et al. [72] used deep architecture and local feature to design deep neural network for facial gender recognition and performed a experiment on Gallagher, LFW dataset and achieved better performance in cross dataset scenario, in which one dataset is used for training and other for testing. Jia et al. [50] performed experiments on weekly labelled facial images and found that CNN performance differs with different depths. Experiments on LFW dataset achieved 98.90% accuracy for facial gender identification. In [18] authors introduced a gender identification model based on geometric descriptors. Which applied leave one out cross-validation on different datasets and achieved robust results. CNN based approach by [93] and [9] achieved 97.3% and 98.9% accuracy FERET dataset. CNN with Softmax classification [50] achieved better accuracy as compared to deep neural network (DCNN) with class posterior classifier [72] on LFW dataset.

Simanjuntak and Azzopardi [87] used fusion of (combination of) shifted filter responses (COSFIRE) features and CNN. Experimented on FERET dataset for gender recognition and observed that the error rates are dropped by more than 50%. Afifi and Abdelhamed [3] addressed the gender recognition problem by using a combination of isolated facial features and holistic features(foggy face). To classify the individual features separately, four DCNNs are used. for aggregation of prediction scores derived from the CNNs, AdaBoost based score fusion approach is applied. By evaluating method on LFW, Adience and FERET datasets 95.98%, 90.43%, 99.28% accuracy achieved for gender classification.

D Amelio et al. [21] introduced a model for gender classification from real-world face images. In this model, features are extracted through VGG-Face Deep Convolutional Neural Network (DCNN). The model utilizes the effectiveness of the sparse sub-dictionary learning on DCNN features. Characteristics (local and global) of the training and probe facial images are represented by sparse sub-dictionary learning. The experimental results show that with small training samples the model can deal with variations in lighting, pose, facial expressions, occlusions and ethnicity. Accuracy for gender classification on LFW dataset is 95.13% instead of huge cardinality difference between training and test set used. Moeini and Mozaffari [73] proposed gender recognition in wild face images under wide ranges of expression, pose and so on. Initially, to represent gender in face images, two separate dictionaries are defined for male and female genders. Features are automatically extracted by the fusion of gray pixels with LBP features. In the training phase, two dictionary learning

techniques are developed to learn the defined dictionaries and in the testing phase, Sparse Representation Classification (SRC) is used for classification. Then, a probability decision making is used for gender classification from proposed gender formulation and estimated values by SRC. On FERET dataset accuracy was 99.9% which is greatest in comparison to state-of-the-art results and 99.0% on LFW dataset.

- Analysis of Conventional v/s Deep Learning: The above results show that deep learning based approaches achieved better accuracy compared to handcrafted feature engineering (conventional learning) even on unconstrained facial images for gender recognition. The deep learning based approach can recognize the gender better for real-life challenging facial images with variations in scale, rotation, illumination etc., compared to conventional approach of feature engineering. The drawback of CNN based approach is its huge data required for proper regularization.

3.2 Facial age prediction

The problem of facial age estimation can be categorized in (1) age group classification and (2) age regression models, and the performance is evaluated as accuracy and error as Mean Absolute Error(MAE). The classification accuracy is defined as ratio of total correctly classified samples (true positive + true negative) and total test samples. The MAE is defined as the mean value of the absolute differences between predicted age and real age (ground truth) of test samples. These methods include two phases: (a) feature extraction and (b) learning for classification or regression. In the feature extraction process; the unique and distinguishable patterns related to particular classes are generated and extracted. For age estimation, the facial age features are facial appearances due to ageing as texture or edge relationship on facial skin. On the basis of feature space, feature extraction is divided into three categories as global, local and hybrid. The study of facial age estimation is classified as (a) conventional learning and (b) deep learning based approaches. Different models based on these approaches are discussed in the next subsection.

Table 3 describes different state-of-the-art techniques based on handcrafted feature engineering with conventional learning and deep learning approach for facial age prediction from 1999 to 2020. The state of art shows that initially researchers focused on age group prediction rather than exact age estimation (regression). Evaluation criteria are based on accuracy for facial age group classification which is shown in percentage on a scale of 0 to 100. Mean absolute error (MAE) is used for measuring the performance of age regression which is defined as the mean of absolute errors for a given population. The MAE has mathematically represented in (1). Here T represents the total size of population (test set); P_i is the predicted age and G_i is the ground truth of ith image of the test set. The error is defined as the absolute difference between predicted age and ground truth. The accuracy is used for age group classification and defined as a percentage of correct age classification over the evaluated population.

$$MAE = \frac{1}{T} \sum_{i=1}^{T} |P_i - G_i| \tag{1}$$

3.2.1 Conventional learning based facial age estimation

The earliest model is given by [56] based on conventional learning for facial age classification of three categories baby, young and senior adult. They used statistical classification on

Table 3 Comparative analysis of performance with different handcrafted features with conventional learning and deep learning approach for facial age prediction techniques: MAE, is measured for AGE Regression(R) while recognition error or accuracy is evaluated for age group Classification(C)

	Authors	Dataset	Feature Extraction	R/C	MAE Accuracy
Conventional learning	Kwon et al. [56]	Private	Statistics Method	C	– – –
	Ramesha et al. [51]	Private	Global,Grid	C	90%
	Pirozmand [79]	FG-NET	Gabor-PCA+LDA	C	90%
	Chikkala [17]	FGNET	WFPDP-GLCM	C	96.5%
		MORPH			97.5%
	Hong et al. [46]	MORPH-II	Bisection Search tree(BST)	R	3.64
	Guo et al. [40]	YGA	BIF,age manifold	R	3.91
		FG-NET			4.77
	Guo et al. [39]	YGA	BIF+MFA(Marginal Fisher Analysis)	R	2.63
	Guo et al. [37]	MORPH	Kernel Partial Least Squares (KPLS)	R	4.18
	Chang et al.[12]	MORPH-II	BIF Scattering Transform	R	3.74
	Hsu et al. [47]	MORPH	CBIF(Component Bioinspired feature)	R	3.21
		FG-NET			3.38
	Suo et al. [89]	FG-NET	AND-OR Graph	R	4.68
	Geng et al. [35]	FG-NET	AAM	R	6.77
	Chao et al. [14]	FG-NET	AAM	R	4.4
	Geng et al. [33]	FG-NET	IIS-LLD	R	5.77
	Fu et al. [31]	UIUC-IFP	Discriminative Aging Manifold	R	3.0
	Luu et al. [70]	FG-NET	Contourlet appearance	R	4.12
	Thukral et al. [91]	FG-NET	2Dshape Grassmann manifold	R	6.2
Deep learning	Wang et al. [96]	FG-NET	CNN	R	4.26
		MORPH-II			4.77
	Niu et al. [75]	MORPH-II	CNN	R	3.42
	Rothe et al. [81]	LAP	CNN	R	5.007
	Rothe et al. [83]	MORPH-II	DEX	R	3.25
		MORPH-II	DEX (fine		2.68
		FG-NET	tune IMDB-WIKI)		3.09
	Chen et al. [16]	MORPH	Ranking CNN	R	2.96
	Pan et al. [77]	FG-NET	CNN	R	2.68
		MORPH-II			2.16
	Zhang et al. [104]	FG-NET	AL-RoR-34	R	2.39
		MORPH			2.36
	Liu et al. [66]	MORPH	LSDML-ResNet 101	R	3.08

Table 3 (continued)

Authors	Dataset	Feature Extraction	R/C	MAE Accuracy
Taheri et al. [90]	FG-NET	DAG-VGG16	R	3.08
	MORPH			2.81
	FG-NET	DAG-GoogLeNet	R	3.05
	MORPH			2.87
Li et al. [62]	FG-NET	BridgeNet	R	2.56
	MORPH			2.38
Agbo et al. [5]	FG-NET	Lightweight CNN	R	3.05
	MORPH			2.31
Liu et al. [67]	FG-NET	MA-SFV2	R	3.81
	MORPH			2.38
Wang et al. [97]	FG-NET	CSC+STD	R	4.01
	MORPH	Pooling		3.66
Liao [64]	FG-NET	Deep SRC+HSVR	R	4.65
	MORPH			3.64

the basis of skin wrinkle analysis. Sobel edge detection is used for wrinkle pattern detection on facial skin and ratios of the distance between facial parts including wrinkles is taken as a feature. This method has high complexity on a small dataset. An extension of the wrinkle extraction approach, different edge extraction methods were used as Canny Edge detector for grid global features on face [51], Gabor filter edge extraction [79], Gabor Wavelet [46] at different scale and orientation. [79] extracted wrinkles by using 12 different Gabor wavelets and developed the Gabor-PCA-LDA technique for age group classification into classes of baby, child and adult with an error of 6.07. [46] extended the work based on Gabor wavelet for facial features. They reduced the dimension of extracted features with PCA, classified the age group using LDA and achieved an error of 4.715 on MORPH-II dataset.

Chikkala et al. [17] divided the age into six categories and introduced wavelet based four-pixel diamond pattern gray level co-occurrence matrix model. They achie-ved 97.5%, 96.5% age classification accuracy on MORPH-II and FG-NET dataset respectively. Clas-sification accuracy with WFPDP-GLCM feature extraction [17] is higher as compared to Gabor-PCA feature extraction [79] on FG-NET dataset because it asses relation of inner(connected component) and outer(not connected component) diamond corner pixel of the third-order four pixels of wavelet image. This reduced the dimension of the features, hence the computational cost is also reduced.

Guo et al. [40] introduced biologically inspired feature extraction using Gabor filters with different orientations and scales. Standard deviation and MAX operation are performed with different scales and orientations using Gabor kernel. For achieving better performance, they empirically optimize C units through standard deviation and MAX operation for each kernel. The dimension of biologically inspired features is reduced through PCA. Support Vector Machine (SVM) is used for classification of age groups on FGNET dataset and achieved classification error as 4.77. BIF improves age estimation accuracy but produces huge dimensions feature vector. Later Guo et al. [39] combined manifold learning with bio-logically inspired features and achieved better accuracy for age prediction. The dimensions

of BIF are reduced by using manifold learning, but it is sensitive to image misalignment due to translation, rotation and scale. They studied gender effects on age identification and applied different classifiers to identify gender and age estimation. For further improvement Guo and Mu [37] developed Kernel Partial Least Squares (KPLS) regression for age identification. KPLS has a flexible output vector and multiple labels in the same output vector to overcome classification difficulties. KPLS reduced dimension and performed single step age learning. Scattering transform is a generalization of BIF, and is used for representing feature vectors of facial images which are reactive to large deformations and insensitive to small displacement and translation. Co-occurrence coefficients generated by the scattering transform are used to characterize texture [12]. In 2017, Hsu et al. [47] used Component Bio-Inspired Feature (CBIF) to perform regression using SVR and classification using SVM and achieved 3.38 and 3.21 MAE on FG-NET and MORPH dataset respectively which is the best-reported result using BIF features on FGNET and MORPH dataset. Suo et al. [89] introduced hierarchical partitioning model by analyzing it as a Markov process, in which age progression partition is done by an AND-OR graph.

First Active Appearance Models (AAMs) introduced by [19]; used for computing geometric and texture variation features on the face. Firstly, points are set on face followed by Procrustes (statistical shape) analysis [19]. Then, the variation is evaluated using eigenvalue analysis to form texture shape correlations for appearance model [19]. At first Lanitis et al. [57] proposed AAM for facial age prediction, later Geng et al. [35] adopted it to generate ageing pattern vector based on Aging Pattern Subspace (AGES) algorithm, in which facial age is determined by projecting into the subspace of the face image. Aging pattern is a sequence of individual ageing using facial images. Specific age is determined through the location in ageing pattern. Extension of this model is done by Chang and Chen [13] as KAGES (Kernel AGing pattErn Subspace). They considered learning of nonlinear subspaces for human age prediction. Age estimation performance is improved by these methods, however, it is difficult to find sequence or longitudinal images of ageing individual faces [13]. Chao et al. [14] proposed a method to extract features through AAM and support vector regression for age estimation. This method obtained good results but with added computational complexity.

Geng et al. [33] suggested multi-label distribution for the facial age estimation in which each facial image can be used to train the model for adjacent ages and chronological age using improved iterative scaling learning from label distribution (IIS-LLD) and conditional probability neural network(CPNN) method. A hybrid combination of BIF, Gabor, AAM, LBP with dynamic deep sparse coding features is proposed to achieve robust results of facial age estimation [64]. Sparse coding method represents features with locality constraints.

Fu and Huang [31] proposed manifold learning for age estimation in which age features are learned on various subjects at a different age. Manifold methods use Orthogonal LPP (OLPP) and other techniques as Neighborhood Preserving Projections(NPP), Principle Component Analysis(PCA), Locality Preserving Projections (LPP) to convert features into low dimensions for each facial age. After that linear regression is performed for age estimation. Use of OLPP made manifold model flexible and achieved MAE value of 3.0. Disadvantages of manifold models are their sensitivity to image misalignment and large size training data.

Contourlet Appearance Model(CAM) feature extraction [70] achieves better accuracy as compared to 2D shape Grassmann manifold features [91] on FGNET dataset. CAM reconstructs unseen textures more accurately by decoupling nonsub-sampled contourlet transform and facial landmark fitting.

3.2.2 Deep learning based facial age estimation

Deep learning based methods require a huge amount of data and good computing infrastructures like GPUs for regularization and training. In 2015, Wang et al. [96] used CNN with 3 Conv, 2 Pool, 1 Fully Connected layers to represent features and linear support vector regression (SVR) for age estimation and achieved MAE of 4.77 and 4.26 on MORPH and FGNET dataset respectively. Niu et al. [75] introduced ordinal ranking CNN with 3 ConvNet, 3 Norm and 2 Pool layers. In this, there is a chain of basic CNN's respective to each age group for training and cumulative results are presented for age prediction. It decreased prediction error compared to softmax with 3.27 MAE on MORPH dataset, 3.34 MAE on Asian face age dataset. Rothe et al. [81] introduced special CNN, Deep Expectation algorithm (DEX) based on VGG-16 architecture that was pretrained on ImageNet. The results achieved MAE as 5.007. In 2018, Rothe et al. [83] optimized the MAE using regularization and fine-tuning on IMDB-WIKI and achieved MAE of 2.68 years on MORPH dataset. Chen et al. [16] also proposed a Ranking CNN with 3 Conv layers and sub-sampling layers, and 3 fully connected layers. Ranking- CNN outperforms the existing methods for age estimation where the MAE is 2.69 years on MORPH dataset. Pan et al. [77] proposed CNN with softmax loss and mean variance loss. The model was pre-trained on IMDB-WIKI and outperforms on MORPH and FGNET dataset with MAE 2.16 and 2.68 respectively. Liu et al. [66] used label sensitive deep metric learning(LSDML) model to predict age which was based on the fact that age labels of human are chronologically correlated. they used ResNet-101 architecture and achieved 3.08 MAE on MORPH dataset.

Zhang et al. [104] used Long Short Term Memory(LSTM) based method with Residual Network of Residual Network (RoR) models and constructed an AL-RoR model of a 34 layer network to extract features for age estimation. Firstly, they pre-trained the RoR model on ImageNet dataset then fine-tuned it on IMDB-WIKI-101 dataset. After that, RoR is fine-tuned on target age dataset for global feature extraction and the LSTM unit is used for local feature extraction. Lastly, the local and global features are combined to classify age groups on Adience dataset and achieved 66.82% accuracy. Age regression is achieved by DEX algorithm on FG-NET and MORPH dataset with MAE as 2.39 and 2.36 years respectively. Taheri and Toygar [90] introduced Directed Acyclic Graph Convolutional Neural Network (DAG-CNN) using VGG-16 and GoogLeNet architecture for facial age estimation. DAG-VGG16 architecture achieves MAE 2.81, 3.08 on MORPH-II dataset, FG-NET dataset respectively. While DAG-GoogLeNet gives 2.87, 3.05 MAE on MORPH-II and FG-NET dataset respectively.

Li et al. [62] developed a model called BridgeNet, based on CNN, for age prediction. This model consists of two modules; gating networks and local regressors. In local regressors heterogeneous data is tackled by partitioning the data space into many overlapping sub-spaces. While gating networks selected a bridge-tree structure that learns continuity-aware weights used by the local regressors. These two modules can unitedly be learned in an end-to-end way. Experimental results on the MORPH II, FG-NET datasets achieved 2.38 and 2.52 MAE respectively and proved this model to be effective and outperforms the state-of-the-art methods. Agbo-Ajala and Viriri [5] designed a lightweight CNN model with low training time, for apparent and real and age estimation. The model merges adaptive image augmentation and image pre-processing algorithm. The MAE achieved on FG-NET and MORPH II datasets are 3.05 and 2.01 respectively. Further, Liu et al. [67] proposed mixed attention mechanism (MA-SFV2) based lightweight CNN (ShuffeNetV2) model; Mixed Attention-ShufeNetV2. In this model impact of noise vectors (environmental

information unrelated to face) is reduced by pre-processing images and network overfitting is reduced by data augmentation methods like sharpening, filtering and histogram enhancement etc. The model transforms the output layer, combining regression, classification and distributed learning age estimation methods. The experimental results on MORPH-II and FG-NET datasets achieved 2.68 and 3.81 MAE respectively and proved model applicability in real-life situations, especially in mobile terminals

To estimate age Wang et al. [97] proposed convolutional sparse coding to extract unsupervised learned features of ageing then STD pooling is applied on extracted feature map for better capturing of ageing signs. To find selective features in reduced dimension space manifold learning is used. They got 3.66, 4.01 MAE on MORPH-II and FGNET dataset respectively. Liao [64] used deep sparse representation coding (SRC) for feature extraction and hierarchical support vector regression (HSVR) for age estimation. Extracted features contain age group information and have the advantage of designing hierarchical age estimation method. Hence they achieved the lowest MAE 4.65 on FGNET and 3.64 on MORPH-II dataset respectively as compare to other methods like Gabor, BIF, AAM and LBP + Gabor.

- Analysis of conventional learning v/s deep learning for facial age estimation: In Table 3, different techniques are mentioned which are evaluated on different datasets. MAE on FG-NET dataset is reduced to 3.38 by conventional learning approach with handcrafted features as CBIF while deep learning based technique has improved up to 3.05, 2.39, by DAG-GoogLeNet, DEX respectively.

4 Study and analysis on multi attribute facial recognition(gender recognition and age estimation)

Table 4 shows the different state-of-the-art models for multi-attribute recognition from face (gender and age prediction) using a single model. The analysis of related work is divided into sub-parts as following.

4.1 Conventional learning-based multi-attribute facial recognition

Initially, handcrafted feature engineering is used for joint estimation of age and gender recognition using a single model. In this, Eidinger et al. [26] proposed the localization of facial features with alignment which was based on localization uncertainty estimation. They used dropouts with SVM for gender and age estimation in the wild. The approach achieved 77.8% and 45.1% accuracy for gender and age group recognition respectively on uncontrolled and most challenging Adience dataset. Guo and Mu [38] used the concept of BIF features with multi-attribute recognition (gender, age, race) and concluded that joint feature model achieves better results compared to individual feature models. Han et al. [44] used BIF feature extraction and a hierarchical classifier for gender and age prediction and got better results than human on MORPH dataset but this approach is not suitable for unconstrained datasets. For gender attribute, [38] achieved better accuracy compared to [44] while the results for age are vice-versa using BIF features on MORPH dataset.

4.2 Deep learning based multi-attribute facial recognition

Recent studies show that CNN is the most used architecture for gender and age estimation, as a CNN model can learn a compact and discriminating feature representation when the

Table 4 Comparison of different multi attribute (age and gender) based facial recognition models. The table mentions the state-of-the-art methods including name of author, year, dataset, method used and performance achieved in term of accuracy and MAE

Authors	Methods	Dataset	Accuracy/MAE
Eidinger et al. [26]	LBP	Adience	Age group: 45.1%
	four-patch	(13,000; 3,300)	Gender: 77.8%
	LBP SVM	Images of Groups	Age group:66.6%
	Dropout	(3,500; 1,050)	Gender:88.6%
Guo and Mu [38]	BIF	MORPH II	Gender: 98.5%
			Age:3.92
			Age group:70.0%
Han et al. [44]	BIF	MORPH II	Age:77.4%
	SVM		Gender:97.6%
Yi et al. [100]	Multi-Scale	MORPH II	Age:3.63
	CNN	(10,530, 44,602)	Gender:98.0%
Levi et al. [60]	CNN(3Conv	Adience	Gender:86.8%
	,2FC)One		Age group:50.7%
	CNN/attribute		
Uricar et al. [92]	VGG-16	ChaLearn LAP	Gender: 89.2%
	One SVM/		Age:0.24 error
	attribute		
Wang et al. [94]	AlexNet +	MORPH II	Gender: 98.0%
	DMTL		*Age* : 85.3%
Li et al. [61]	Tree CNN	MORPH II	Gender:98.4%
	features		Age:3.61 MAE
Shin et al. [86]	CNN,SVM	Faces of Asian	Gender:75.91%
		Non Asian	Age:54.98%
		Celebrity	
Zhang et al. [103]	CNN,RoR	Adience	Gender: 93.24%
			Age:66.74%
Liao et al. [65]	LDNN	LFW	Gender:77.79%
			Age:39.50%
Han et al. [43]	DMTL	LFW+	Gender:96.7%
	based on		Age:75.0% MAE:4.5
	modified	MORPH II	Gender:98.3%
	AlexNet		Age:85.3%,MAE:3.0
Duan et al. [25]	CNN+ELM	MORPH	Gender: 87.3%
			Age MAE:3.44
Das et al. [22]	MTCNN	UTKFace	Gender: 98.23%
			Age:70.1%
Lee et al. [58]	LMTCNN	Adience	Gender: 85.16%
			Age:70.78%
Debgupta et al. [24]	ResNet	APPA-REAL	Gender: 96.26%
			Age MAE:1.65

Table 4   (continued)

Authors	Methods	Dataset	Accuracy/MAE
Debgupta et al. [24]	ResNet	APPA-REAL	Gender: 96.26%
			Age MAE:1.65
Gurnani et al. [41]	Sailency Map	Adience	Gender:62.11%
	+AlexNet		Age:91.8%
Yoo et al. [102]	CMT + LE	MORPH-II	Gender: 99.28
			Age MAE:2.89
Agbo et al. [4]	Deep CNN	OIU-Adience	Gender:96.2%
			Age:93.8%
Khan et al. [53]	MCFP-	Adience	Gender:93.6%
	DCNNs		Age:69.4%
		LFW	Gender:94.1%
		FERET	Gender:100%

training data size is huge. Yi et al. [100] proposed Multi-Scale CNN which performs better for gender and age estimation simultaneously as compared to biologically inspired features and achieved lower error on MORPH dataset. Levi and Hassner [60] used CNN with 3 conv layer and 2 fully connected layers on unconstrained Adience benchmark for age and gender classification. It shows that CNN improves the performance of gender and age recognition compared to handcrafted features of LBP [26] on unconstrained data sets having small resolution face images. Uricar et al. [92] used CNN VGG-16 architecture to learn deep features with individual SVM classifiers for each attribute. They used the pre-trained ImageNet network which was fine-tuned on ChaLearn 2015 LAP dataset and used structured output SVM (SO-SVM) for prediction of gender and apparent age.

Wang et al. [94] used deep multi-task learning to learn homogeneous and heterogeneous attributes and outperformed on MORPH dataset. Li et al. [61] used tree-like CNN for non-rigid variations in appearance of the face which performs better than AlexNet. The MAE of age regression is reduced to 3.61 and accuracy of gender recognition increases to 98.4% on MORPH dataset which is better than multi-scale CNN of [100]. Shin et al. [86] proposed a CNN-SVM based gender and age estimation model with ethnicity difference on Asian and Non-Asian celebrities face images. Zhang et al. [103] proposed CNN with residual networks of residual networks RoR to classify gender and age in wild using Adience dataset and achieved 93.24% and 66.74% accuracy of gender recognition and age group classification respectively. The method outperforms handcrafted features of LBP [26] as well as CNN method of [60]. Liao et al. [65] proposed a local deep neural network in which facial region is covered by using 9 overlapping patches for each image for reducing training time. They performed tests on Adience and LFW databases and achieved 1% lower recognition rate compared to the original algorithm that used 100 patches per image.

Han et al. [43] has improved the DMTL of [94] using modified AlexNET and achieved an accuracy of 98.3% for gender recognition, MAE 3.0 for age regression on MORPH dataset. Improved DMTL results are better-compared to [65] on LFW dataset. Duan et al. [25] introduced a hybrid approach based on CNN and ELM (Extreme Machine Learning) for joint prediction of gender and age from face images. The problem of over-fitting is resolved by ELM without tuning the biases and weights. They achieved MAE of 3.44

for age estimation and accuracy as 87.3% for gender classification on morph dataset. Das et al. [22] introduced multitask-CNN (MTCNN) model to recognize joint attributes (age and gender) by minimizing the inter-class bias. This model used a combined dynamic loss for age and gender attributes. They achieved 98.23% and 70.1% accuracy for gender and age classification respectively on UTK face dataset. UTK contains juvenile faces also which have limited features on the face which makes age and gender recognition harder. Lee et al. [58] introduced Lightweight multi-task CNN (LMTCNN) with 2 convolutional layers (separate depth-wise), 1 common convolutional layer and 2 fully connected layers for joint gender and age classification. In this, the inference time was reduced for achieving better FPS. Accuracy achieved was 85.16% and 70.78% for gender and age recognition respectively on Adience dataset which under-performed for gender recognition and outperformed for age estimation compared to [60]. Recently Debgupta et al. [24] used wider ResNet to solve the problem of vanishing gradients in deep learning for multi-attribute recognition. They achieved 96.26% accuracy for gender recognition and MAE of 1.65 for age regression on APPA-REAL dataset.

Saliency features are equivalent to the human visual system. Gurnani et al. [41] proposed Multi-level Network (ML-Net) to evaluate the saliency map for detection of the face and subsequently AlexNET model of CNN is applied to classify facial attributes. They achieved an accuracy of 91.8%, 62.11% for age attribute and gender attribute with saliency map features respectively on Adience dataset while same architecture without saliency map features achieved accuracy 83.4%, 52.2% for age and gender recognition respectively.

Multitask learning is used to enhance age estimation by making use of auxiliary tasks, like gender recognition, which is linked with the primary task. In classic multitask learning, it is difficult to describe the relationship in primary and auxiliary tasks; how the auxiliary tasks improve the model for the primary objective is ambiguous. Yoo et al. [102] developed a conditional multitask (CMT) deep learning model in which an age variable is architecturally factorized into gender conditioned age probabilities in DCNN. Another critic limitation for the training of age estimation models is that accurate training labels with discrete age values are insufficient. To increase the number of accurate training labels they developed a label expansion (LE) mechanism. For verifying the generality of the model, intensive experiments are performed on FG-NET and MORPH-II datasets MAE for age estimation was 3.43 and 2.89 respectively.

Agbo-Ajala and Viriri [4] proposed a model for gender and age group classification from unfiltered real-life face images. The model contains image pre-processing that prepares and processes the input images and a CNN that does feature extraction and the classification. The network is pre-trained on an IMDB-WIKI dataset and fine-tuned on MORPH-II dataset. The accuracy achieved for gender and age group classification was 96.2 and 93.8 respectively on OIU-Adience dataset. Khan et al. [53] developed a face parsing model MCFP-DCNNs using multi-class face segmentation (MCFS) and deep convolutional neural networks (DCNNs) for gender and age classification. Face image is divided into seven parts (eyes, hair, eyebrows, nose, skin, mouth and back). Model is trained via a DCNNs model by extracting information from various facial parts and probabilistic classification is used to generate Probability maps for seven facial classes. For feature extraction from the corresponding probability maps another DCNNs model is used for gender and age recognition. A series of experiments are performed to investigate which face parts help in gender and age classification. Experiments on Adience dataset got 93.6%, 69.4% accuracy for gender and age recognition respectively.

5 Analysis and discussion

The analysis of different studies is discussed with deep analysis of their pros and cons.

From the above study we conclude that unconstrained dataset are needed for training and validation of models for real time applications of gender and age prediction. For these uses cases, the LFW [48], IMDB-WIKI [82], LAP [27] and Adience [26] datasets are available. IMDB is the largest dataset while LFW and Adience are the most challenging dataset. MORPH is the most commonly used data set in literature which is captured in controlled environment with some real life challenges.

- Facial growth effects the FAR: teenagers have stressing of soft tissues as initial sign of ageing. Adult ageing is affected by morphological changes in wrinkles, skin textures and facial lines on the forehead with different shape as horizontal and vertical. The size of face grows with age and it is shown that performance of facial attribute recognition (gender and age) degrades on child's faces compared to adult faces [98].
- Uncontrolled environment or image captured in real life includes various challenges of emotions, obstacles, occlusions, scale, illumination, camera focus, the orientation of camera etc., and all these affect the performance of facial attribute recognition [68].

MTL V/S STL based approach for facial attribute recognition: In the literature, mostly single attribute learning is used for facial attribute prediction (age, gender). The comparison of the state-of-the-art methods using STL based approach as shown in Table 2 shows that the accuracy of facial gender recognition is optimized up to 99.28% using deep learning on LFW database and 79.3% on Adience dataset using conventional learning. It is derived from the Table 3 that the MAE of age prediction has been reduced up to 2.16 years on the MORPH II database and 2.39 years on FGNET dataset using deep learning methods based on STL approach. The MTL based approach, DMTL with modified Alexnet has achieved 98.3% and 85.3% accuracy on gender and age classification respectively; with an MAE of 3.0 years on MORPH II dataset [43]. The same model has also achieved an accuracy of 96.7% and 75.0% with an MAE of 4.5 years on LFW dataset.

These are different models of deep learning approach for gender and age prediction so they will take more compounded memory and time during inference for gender and age estimation for a given face. The compounded time (inference time for gender by a gender model + inference time for age by age model) is needed for image acquisition, feature engineering, classification or regression in STL based approach. While MTL based approach takes single (same) inference time during image acquisition and feature engineering but compounded time for classification or regression. This makes MTL based approach faster compared to STL based approach. Further, the MTL based approach needs to save weight matrix and inference graph in memory which makes it better for saving memory compared to STL based approach. The faster technique and memory saving features of MTL makes it more suitable for deployment in edge detection, where the challenges are limited computation power and memory.

It is concluded that deep learning outperforms compared to handcrafted engineering but a huge amount of computation resources and data is needed for training and regularization.

The main challenges in age and gender estimation are that the facial appearance changing rates are different at different ageing stages. Changes in a child or young faces are faster compared to old faces so gender recognition in small children is very difficult because male and female both look alike. Similarly, age estimation causes more error in older faces [34].

The age variation characteristics are:

1. Aging process is very slow and irreversible so it is uncontrollable.
2. Obtaining sufficient amount of training data is extremely laborious for age estimation.
3. Aging patterns are different for every person which are affected by various external factors including weather conditions, health, living style and genetic structure etc.
4. Prediction of age after the knowledge of gender is an easy task and the accuracy also increases as the age classification depends upon the gender of a person also.
5. The attributes of male and female are entirely different for determining age.
6. The result of age classification in man is individually affected by different attributes like colour and texture of skin, similarly in women these attributes also adversely affect the age.
7. Face features like eyes are more useful for gender prediction while eyes and mouth are mostly used for age prediction [65].

6 Conclusions

Multi attribute heterogeneous prediction is a problem of gender classification and age regression in a single network. It is more challenging compared to single attribute prediction problem but has vast use-cases. It takes lesser time in feature extraction and lesser memory consumption due to lower size of weights model. The multi-attribute prediction models include attribute heterogeneity and attribute correlation both in a single network and allows category-specific feature learning for heterogeneous attributes and shared feature learning for all attributes. The review outcomes are that deep learning (Alexnet) based approaches outperforms compared to handcrafted feature engineering, however, they need huge resources for computation and data regularization. There are still some challenges in real-time facial gender classification and age estimation which include face localization in wild, feature detection of juvenile age group or children, blur and de-focused faces, expression of the person, occlusions and ethnicity (race). In the current pandemic of Coronavirus, human persons are using face masks and hence we have the extraordinary challenges of face localization and hidden face feature extraction for occluded faces which makes it very difficult to detect. Future work is needed on these issues to make better and robust facial attribute recognition to handle real-life situations.

References

1. (2017) Japanese smokers to face age test. http://news.bbc.co.uk/2/hi/asia-pacific/7395910.stm
2. Abbas H, Hicks Y, Marshall D, Zhurov AI, Richmond S (2018) A 3d morphometric perspective for facial gender analysis and classification using geodesic path curvature features. Computational Visual Media 4(1):17–32
3. Afifi M, Abdelhamed A (2019) Afif4: Deep gender classification based on adaboost-based fusion of isolated facial features and foggy faces. J Vis Commun Image Represent 62:77–86
4. Agbo-Ajala O, Viriri S (2020) Deeply learned classifiers for age and gender predictions of unfiltered faces. The Scientific World Journal 2020
5. Agbo-Ajala O, Viriri S (2020) A lightweight convolutional neural network for real and apparent age estimation in unconstrained face images. IEEE Access 8:162800–162808
6. Alamri T, Hussain M, Aboalsamh H, Muhammad G, Bebis G, Mirza AM (2013) Category specific face recognition based on gender. In: 2013 international conference on information science and applications (ICISA). IEEE, pp 1–4

7. Alexandre LA (2010) Gender recognition: A multiscale decision fusion approach. Pattern Recognition Letters 31(11):1422–1427
8. Antipov G, Berrani S-A, Dugelay J-L (2016) Minimalistic cnn-based ensemble model for gender prediction from face images. Pattern Recognition Letters 70:59–65
9. Aslam A, Hussain B, Cetin AE, Umar AI, Ansari R (2018) Gender classification based on isolated facial features and foggy faces using jointly trained deep convolutional neural network. Journal of Electronic Imaging 27(5):053023
10. Baluja S, Rowley HA (2007) Boosting sex identification performance. Int J Comput Vis 71(1):111–119
11. Bissoon T, Viriri S (2013) Gender classification using face recognition. In: 2013 international conference on adaptive science and technology. IEEE, pp 1–4
12. Chang K-Y, Chen C-S (2015) A learning framework for age rank estimation based on face images with scattering transform. IEEE Trans Image Process 24(3):785–798
13. Chang K-Y, Chen C-S, Hung Y-P (2011) Ordinal hyperplanes ranker with cost sensitivities for age estimation. In: CVPR 2011. IEEE, pp 585–592
14. Chao W-L, Liu J-Z, Ding J-J (2013) Facial age estimation based on label-sensitive learning and age-oriented regression. Pattern Recogn 46(3):628–641
15. Chen B-C, Chen C-S, Hsu WH (2015) Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. IEEE Transactions on Multimedia 17(6):804–815
16. Chen S, Zhang C, Dong M, Le J, Rao M (2017) Using ranking-cnn for age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5183–5192
17. Chikkala R, Edara S, Bhima P (2019) Human facial image age group classification based on third order four pixel pattern (tofp) of wavelet image. Int Arab J Inf Technol 16(1):30–40
18. Cirne MVM, Pedrini H (2017) Gender recognition from face images using a geometric descriptor. In: 2017 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 2006–2011
19. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23(6):681–685
20. Cottrell GW, Metcalfe J (1991) Empath: Face, emotion, and gender recognition using holons. In: Advances in neural information processing systems, pp 564–571
21. D Amelio A, Cuculo V, Bursic S (2019) Gender recognition in the wild with small sample size-a dictionary learning approach. In: International symposium on formal methods. Springer, pp 162–169
22. Das A, Dantcheva A, Bremond F (2018) Mitigating bias in gender, age and ethnicity classification: a multi-task convolution neural network approach. In: Proceedings of the European conference on computer vision (ECCV), pp 0–0
23. Deb D, Nain N, Jain AK (2018) Longitudinal study of child face recognition. In: 2018 international conference on biometrics (ICB). IEEE, pp 225–232
24. Debgupta R, Chaudhuri BB, Tripathy BK (2020) A wide resnet-based approach for age and gender estimation in face images. In: International conference on innovative computing and communications. Springer, pp 517–530
25. Duan M, Li K, Yang C, Li K (2018) A hybrid deep learning cnn–elm for age and gender classification. Neurocomputing 275:448–461
26. Eidinger E, Enbar R, Hassner T (2014) Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security 9(12):2170–2179
27. Escalera S, Fabian J, Pardo P, Baró X, Gonzalez J, Escalante HJ, Misevic D, Steiner U, Guyon I (2015) Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results. In: Proceedings of the IEEE international conference on computer vision workshops, pp 1–9
28. Fu Y, Hospedales TM, Xiang T, Yao Y, Gong S (2014) Interestingness prediction by robust learning to rank. In: ECCV
29. Fu Y, Guo G, Huang TS (2010) Age synthesis and estimation via faces: A survey. IEEE Trans Pattern Anal Mach Intell 32(11):1955–1976
30. Fu Y, Huang TS (2008) Human age estimation with regression on discriminative aging manifold. IEEE Transactions on Multimedia 10(4):578–584
31. Fu Y, Huang TS (2008) Human age estimation with regression on discriminative aging manifold. IEEE Transactions on Multimedia 10(4):578–584
32. Gallagher AC, Chen T (2009) Understanding images of groups of people. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 256–263
33. Geng X, Yin C, Zhou Z-H (2013) Facial age estimation by learning from label distributions. IEEE Trans Pattern Anal Mach Intell 35(10):2401–2412
34. Geng X, Zhou Z-H, Smith-Miles K (2007) Automatic age estimation based on facial aging patterns. IEEE Trans Pattern Anal Mach Intell 29(12):2234–2240

35. Geng X, Zhou Z-H, Zhang Y, Li G, Dai H (2006) Learning from facial aging patterns for automatic age estimation. In: Proceedings of the 14th ACM international conference on Multimedia, pp 307–316

36. Guo G (2012) Human age estimation and sex classification. In: Video analytics for business intelligence. Springer, pp 101–131

37. Guo G, Mu G (2011) Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In: CVPR 2011. IEEE, pp 657–664

38. Guo G, Mu G (2014) A framework for joint estimation of age, gender and ethnicity on a large database. Image Vis Comput 32(10):761–770

39. Guo G, Mu G, Fu Y, Dyer C, Huang T (2009) A study on automatic age estimation using a large database. In: 2009 IEEE 12th international conference on computer vision. IEEE, pp 1986–1991

40. Guo G, Mu G, Fu Y, Huang TS (2009) Human age estimation using bio-inspired features. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 112–119

41. Gurnani A, Shah K, Gajjar V, Mavani V, Khandhediya Y (2019) Saf-bage: Salient approach for facial soft-biometric classification-age, gender, and facial expression. In: 2019 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 839–847

42. Gutta S, Huang JRJ, Jonathon P, Wechsler H (2000) Mixture of experts for classification of gender, ethnic origin, and pose of human faces. IEEE Transactions on Neural Networks 11(4):948–960

43. Han H, Jain AK, Wang F, Shan S, Chen X (2017) Heterogeneous face attribute estimation: A deep multi-task learning approach. IEEE Trans Pattern Anal Mach Intell 40(11):2597–2609

44. Han H, Otto C, Liu X, Jain AK (2014) Demographic estimation from face images: Human vs. machine performance. IEEE Trans Pattern Anal Mach Intell 37(6):1148–1161

45. Hassner T, Harel S, Paz E, Enbar R (2015) Effective face frontalization in unconstrained images. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4295–4304

46. Hong L-J, Wen D, Fang C, Ding X-Q (2013) Face age estimation by using bisection search tree. In: 2013 international conference on machine learning and cybernetics, vol 1. IEEE, pp 370–374

47. Hsu G-SJ, Cheng Y-T, Ng CC, Yap MH (2017) Component biologically inspired features with moving segmentation for age estimation. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 540–547

48. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst

49. Jain A, Huang J (2004) Integrating independent components and linear discriminant analysis for gender classification. In: Sixth IEEE international conference on automatic face and gesture recognition, 2004. Proceedings. IEEE, pp 159–163

50. Jia S, Lansdall-Welfare T, Cristianini N (2016) Gender classification by deep learning on millions of weakly labelled images. In: 2016 IEEE 16th international conference on data mining workshops (ICDMW). IEEE, pp 462–467

51. Ramesha K, Raja KB, Venugopal KR, Patnaik LM (2010) Feature extraction based face recognition, gender and age classification

52. Khan A, Majid A, Mirza AM (2005) Combination and optimization of classifiers in gender classification using genetic programming. International Journal of Knowledge-based and Intelligent Engineering Systems 9(1):1–11

53. Khan K, Attique M, Khan RU, Syed I, Chung T-S (2020) A multi-task framework for facial attributes classification through end-to-end face parsing and deep convolutional neural networks. Sensors 20(2):328

54. Kim H-C, Kim D, Ghahramani Z, Bang SY (2006) Appearance-based gender classification with gaussian processes. Pattern Recogn Lett 27(6):618–626

55. Kumar N, Berg AC, Belhumeur PN, Nayar SK (2009) Attribute and simile classifiers for face verification. In: 2009 IEEE 12th international conference on computer vision. IEEE, pp 365–372

56. Kwon YH, da Vitoria Lobo N (1999) Age classification from facial images. Computer Vision and Image Understanding 74(1):1–21

57. Lanitis A, Taylor CJ, Cootes TF (2002) Toward automatic simulation of aging effects on face images. IEEE Trans Pattern Anal Mach Intell 24(4):442–455

58. Lee J-H, Chan Y-M, Chen T-Y, Chen C-S (2018) Joint estimation of age and gender from unconstrained face images using lightweight multi-task cnn for mobile applications. In: 2018 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, pp 162–165

59. Leng X, Wang Y (2008) Improving generalization for gender classification. In: 2008 15th IEEE international conference on image processing. IEEE, pp 1656–1659

60. Levi G, Hassner T (2015) Age and gender classification using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 34–42

61. Li S, Xing J, Niu Z, Shan S, Yan S (2015) Shape driven kernel adaptation in convolutional neural network for robust facial traits recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 222–230

62. Li W, Lu J, Feng J, Xu C, Zhou J, Tian Q (2019) Bridgenet: A continuity-aware probabilistic network for age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1145–1154

63. Li Z, Zhou X, Huang TS (2009) Spatial gaussian mixture model for gender recognition. In: 2009 16th IEEE international conference on image processing (ICIP). IEEE, pp 45–48

64. Liao H (2019) Facial age feature extraction based on deep sparse representation. Multimedia Tools and Applications 78(2):2181–2197

65. Liao Z, Petridis S, Pantic M (2017) Local deep neural networks for age and gender classification. arXiv preprint arXiv:1703.08497

66. Liu H, Lu J, Feng J, Zhou J (2017) Label-sensitive deep metric learning for facial age estimation. IEEE Transactions on Information Forensics and Security 13(2):292–305

67. Liu X, Zou Y, Kuang H, Ma X (2020) Face image age estimation based on data augmentation and lightweight convolutional neural network. Symmetry 12(1):146

68. Liu Z, Luo P, Wang X, Tang X (2015) Deep learning face attributes in the wild. In: Proceedings of the IEEE international conference on computer vision, pp 3730–3738

69. Lu L, Shi P (2009) A novel fusion-based method for expression-invariant gender classification. In: 2009 IEEE international conference on acoustics, speech and signal processing. IEEE, pp 1065–1068

70. Luu K, Seshadri K, Savvides M, Bui TD, Suen CY (2011) Contourlet appearance model for facial age estimation. In: 2011 international joint conference on biometrics (IJCB). IEEE, pp 1–8

71. Mäkinen E, Raisamo R (2008) An experimental comparison of gender classification methods. Pattern Recogn Lett 29(10):1544–1556

72. Mansanet J, Albiol A, Paredes R (2016) Local deep neural networks for gender recognition. Pattern Recogn Lett 70:80–86

73. Moeini H, Mozaffari S (2017) Gender dictionary learning for gender classification. J Vis Commun Image Represent 42:1–13

74. Moghaddam B, Yang M-H (2002) Learning gender with support faces. IEEE Trans Pattern Anal Mach Intell 24(5):707–711

75. Niu Z, Zhou M, Wang L, Gao X, Hua G (2016) Ordinal regression with multiple output cnn for age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4920–4928

76. Osman OF, Yap MH (2018) Computational intelligence in automatic face age estimation: A survey. IEEE Transactions on Emerging Topics in Computational Intelligence 3(3):271–285

77. Pan H, Han H, Shan S, Chen X (2018) Mean-variance loss for deep age estimation from a face. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5285–5294

78. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The feret evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 22(10):1090–1104

79. Pirozmand P, Amiri MF, Kashanchi F, Layne NY (2011) Age estimation, a gabor pca-lda approach. J Math Comput Sci 2(2):233–240

80. Ricanek K, Tesafaye T (2006) Morph: A longitudinal image database of normal adult age-progression. In: 7th international conference on automatic face and gesture recognition (FGR06). IEEE, pp 341–345

81. Rothe R, Timofte R, Van Gool L (2015) Dex: Deep expectation of apparent age from a single image. In: Proceedings of the IEEE international conference on computer vision workshops, pp 10–15

82. Rothe R, Timofte R, Van Gool L (2018) Deep expectation of real and apparent age from a single image without facial landmarks. Int J Comput Vis 126(2-4):144–157

83. Rothe R, Timofte R, Van Gool L (2018) Deep expectation of real and apparent age from a single image without facial landmarks. Int J Comput Vis 126(2-4):144–157

84. Scherbaum K, Sunkel M, Seidel H-P, Blanz V (2007) Prediction of individual non-linear aging trajectories of faces. Computer Graphics Forum 26(3):285–294

85. Shakhnarovich G, Viola PA, Moghaddam B (2002) A unified learning framework for real time face detection and classification. In: Proceedings of Fifth IEEE international conference on automatic face gesture recognition. IEEE, pp 16–23

86. Shin M, Seo J-H, Kwon D-S (2017) Face image-based age and gender estimation with consideration of ethnic difference. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 567–572

87. Simanjuntak F, Azzopardi G (2019) Fusion of cnn-and cosfire-based features with application to gender recognition from face images. In: Science and information conference. Springer, pp 444–458

88. Somanath G, Rohith MV, Kambhamettu C (2011) Vadana: A dense dataset for facial image analysis. In: 2011 IEEE international conference on computer vision workshops (ICCV Workshops). IEEE, pp 2175–2182

89. Suo J, Zhu S-C, Shan S, Chen X (2009) A compositional and dynamic model for face aging. IEEE Trans Pattern Anal Mach Intell 32(3):385–401

90. Taheri S, Toygar O (2019) On the use of dag-cnn architecture for age estimation with multi-stage features fusion. Neurocomputing 329:300–310

91. Thukral P, Mitra K, Chellappa R (2012) A hierarchical approach for human age estimation. In: 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 1529–1532

92. Uricar M, Timofte R, Rothe R, Matas J, Van Gool L (2016) Structured output svm prediction of apparent age, gender and smile from deep features. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 25–33

93. van de Wolfshaar J, Karaaba MF, Wiering MA (2015) Deep convolutional neural networks and support vector machines for gender recognition. In: 2015 IEEE symposium series on computational intelligence. IEEE, pp 188–195

94. Wang F, Han H, Shan S, Chen X (2017) Deep multi-task learning for joint prediction of heterogeneous face attributes. In: 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017). IEEE, pp 173–179

95. Wang J-G, Li J, Lee CY, Yau W-Y (2010) Dense sift and gabor descriptors-based face representation with applications to gender recognition. In: 2010 11th international conference on control automation robotics & vision. IEEE, pp 1860–1864

96. Wang X, Guo R, Kambhamettu C (2015) Deeply-learned feature for age estimation. In: 2015 IEEE winter conference on applications of computer vision. IEEE, pp 534–541

97. Wang X, Li R, Zhou Y, Kambhamettu C (2017) A study of convolutional sparse feature learning for human age estimate. In: 2017 12th IEEE international conference on automatic face and gesture recognition (FG 2017). IEEE, pp 566–572

98. Wilkinson CM, Ferguson E (2016) Juvenile age estimation from facial images. Science & Justice

99. Yang Z, Ai H (2007) Demographic classification with local binary patterns. In: International conference on biometrics. Springer, pp 464–473

100. Yi D, Lei Z, Li SZ (2014) Age estimation by multi-scale convolutional network. In: Asian conference on computer vision. Springer, pp 144–158

101. Yildirim ME, Ince OF, Salman YB, Song JK, Park JS, Yoon BW (2016) Gender recognition using hog with maximized inter-class difference. In: VISIGRAPP (3: VISAPP), pp 108–111

102. Yoo B, Kwak Y, Kim Y, Choi C, Kim J (2018) Deep facial age estimation using conditional multitask learning with weak label expansion. IEEE Signal Process Lett 25(6):808–812

103. Zhang K, Gao C, Guo L, Sun M, Yuan X, Han TX, Zhao Z, Li B (2017) Age group and gender estimation in the wild with deep ror architecture. IEEE Access 5:22492–22503

104. Zhang K, Liu N, Yuan X, Guo X, Gao C, Zhao Z, Ma Z (2019) Fine-grained age estimation in the wild with attention lstm networks. IEEE Transactions on Circuits and Systems for Video Technology

105. Zhang Z, Song Y, Qi H (2017) Age progression/regression by conditional adversarial autoencoder. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5810–5818