

A Survey on Deep Learning Face Age Estimation Model: Method and Ethnicity

Hadi A. Dahlan

Faculty of Information Science and Technology
National University of Malaysia
43600 UKM Bangi, Selangor, Malaysia

Abstract—Face age estimation is a type of study in computer vision and pattern recognition. Designing an age estimation or classification model requires data as training samples for the machine to learn. Deep learning method has improved estimation accuracy and the number of deep learning age estimation models developed. Furthermore, numerous datasets availability is making the method an increasingly attractive approach. However, face age databases mostly have limited ethnic subjects, only one or two ethnicities and may result in ethnic bias during age estimation, thus impeding progress in understanding face age estimation. This paper reviewed available face age databases, deep learning age estimation models, and discussed issues related to ethnicity when estimating age. The review revealed changes in deep learning architectural designs from 2015 to 2020, frequently used face databases, and the number of different ethnicities considered. Although model performance has improved, the widespread use of specific few multi-races databases, such as the MORPH and FG-NET databases, suggests that most age estimation studies are biased against non-Caucasians/non-white subjects. Two primary reasons for face age research's failure to further discover and understand ethnic traits effects on a person's facial aging process: lack of multi-race databases and ethnic traits exclusion. Additionally, this study presented a framework for accounting ethnic in face age estimation research and several suggestions on collecting and expanding multi-race databases. The given framework and suggestions are also applicable for other secondary factors (e.g. gender) that affect face age progression and may help further improve future face age estimation research.

Keywords—Deep learning; face age estimation; face database; ethnicity bias

I. INTRODUCTION

Facial aging is a complex biological process. Most researchers in the computer vision and the pattern recognition fields have already found multiple ways to extract information from the face for age estimation/classification. However, not all information extracted can help the system learn. When the system learned from only a specific ethnic sample, it may not estimate/classify the age of other ethnic subjects correctly, even after the face age estimation system improved.

Earlier face aging models combined extractors and classifiers to extract specific aging features and accurately classify the facial image into its correct age. The downside of this approach is that the data needed for learning are usually structured and quantitatively limited; too little or too much data could lead to models learning incorrect patterns, resulting in

inaccurate age classification. Meanwhile, deep learning is another approach that could help algorithms improve the computer's ability to discover common facial aging traits (e.g. aging wrinkles) within vast amounts of data and classify the facial image into its correct age. However, face age databases mostly have limited ethnic subjects, only one or two ethnicities and may result in ethnic bias during age estimation, thus impeding progress in understanding face age estimation.

In this study, the review on face age estimation/classification/distribution examined problems regarding:

- 1) What face databases are frequently used in the age estimation study, and how many different ethnics are in those databases?
- 2) What deep learning technique is used in facial aging research? How did the technique change through time? And do they account for different ethnicities in their studies?
- 3) What are the most used deep learning network architecture and what are their strengths and weaknesses?
- 4) How to obtain more face images of people of different ethnicities in the time of restrictions (e.g. due to quarantine)?

Accordingly, this study surveyed the available face age databases, the most used database in this type of research, and the deep learning techniques used for the face age estimation (or distribution; or classification) model design. More than 50 papers (2015-2020) that used the deep learning method for face age studies were reviewed in this study. The aim of this paper is to survey the different deep learning face age estimation methods and how they account for different ethnicities. By understanding the different deep learning face age estimation methods and the problem related to ethnic bias in their face age estimation, we can discover significant racial traits that could help distinguish unique aging patterns used to solve racial face age estimation problems in real-life applications. Moreover, a framework for studying CNN face age estimation while considering the ethnicities of the subjects is included in this paper to help guide future face age estimation studies that use either the deep learning approach or the standard machine learning approach.

The remainder of this paper is structured as follows: Section 2 mention several related works regarding deep learning and early face age estimation; Section 3 explains the human facial aging and differences in process between several races; Section 4 surveys the face age image databases that can be used for facial age estimation studies and shows the quantities of

each race in each database (if any); Section 5 explains the face age estimation model and reviews the different deep learning techniques proposed between 2015-2020 as well as the databases used. The importance of ethnic traits in age estimation is also highlighted; Section 6 discusses the relevant open issues regarding ethnic characteristics; Section 7 discusses several possible solutions to solve the problems, Section 8 presents the conclusions and Section 9 mentions the future directions.

II. RELATED WORK

The deep learning model has two primary processes: 1) training and 2) inferring. The training phase is the process of labelling large quantities of data (i.e. identifying and memorising the data matching characteristics). Meanwhile, in the inferring phase, the deep learning model decides on the label for the new data using the knowledge gained from the earlier training phase. Manual feature extraction on the data is unnecessary because the model's neural network architecture can learn the feature directly from the data, eliminating the need for data labelling. This learning feature is advantageous when working on large quantities of unstructured data (multiple formats like text and pictures). Recently, deep learning, such as convolutional neural network (CNN), has become well-known in the image processing and pattern recognition fields for its capability to 'learn' from a large number of images and perform specific tasks accurately. The deep learning method can fit the parameters of multi-layered networks of nodes to the vast amounts of data before extrapolating outputs from new inputs. Knowing the commonly used network designs in face age estimation studies and their strengths and weaknesses would be interesting enough.

Recently, face age estimation studies using the deep learning approach to estimate a person's age based on aging features, such as the facial skull shape and aging wrinkle, have increased. These aging features are a person's regular facial aging changes that occur through the years. Nevertheless, considering ethnicity in age estimation can pose a different problem since each ethnicity/race has been confirmed to have a different rate of facial aging [1, 2, 3, 4]. For example, a 20-year-old White subject would look older than a 20-year-old Asian because of their facial bones and skin structures differences [2]. For the CNN model to learn correctly, many datasets containing multiple races with equal ratios are needed.

Although many face databases are available for age estimation, most are racially biased and have just only one or two significant ethnicities. Unbalanced ethnic samples can create problems as age estimation models depend solely on these databases. A bias might occur, for example, when estimating the age of an Asian subject if the majority of ethnicities available in a database are Caucasians/White due to the differences in facial structure and rate of skin aging [1, 2]. In most previous face age estimation/classification/ distribution studies, all sample databases were used for training and testing while utilising different deep learning methods that match their research aim(s) and main objective(s). However, ethnic traits are usually ignored, resulting in very few analyses of racial traits' effects on the face age estimation process. A few reasons for this exclusion: researchers mainly consider racial traits as

age-invariant features, difficulties in capturing a person's face aging progression in a controlled/uncontrolled environment, and capturing >100 face images of different ethnic people in equal quantities can be time-consuming and costly. Nonetheless, it is undeniable that the facial aging process differs between races; therefore, ethnicities should be considered in future research when experimenting with the next CNN age estimation model. Moreover, analyses on the ethnic age difference can contribute to a better understanding of human facial aging.

III. HUMAN FACIAL AGING – ETHNICITIES

Face features and expression are fundamental ways of human communication. Many studies have observed the facial appearance and examined ways to apply the knowledge to real-world applications. One of these studies is face age estimation, which is research on estimating a person's age based on facial appearance observations. Over the years, multiple facial traits help determine a person's age, including the shape of the face, skin texture, skin features, and skin colour contrast [5, 6]. The two predetermined features are as follows: 1) face shape change, particularly the cranium bones that grow with time. This process predominately occurs during childhood to adulthood transition; 2) development of wrinkles or face texture as facial muscle weakens due to decreased elasticity. This process occurs during the transition from adulthood to the senior stage [7, 8].



Fig. 1. Different Ethnic Facial Aging Features for Four Women Aged Over 60 Years Old. from Left to Right: Caucasian, East Asian, Latino/Hispanic, and African (All Images were Taken from [13]).



Fig. 2. Facial Feature and Aging Difference for Adult Caucasian (Top Left) and Asian (Bottom left), while the Baby's Face for the Caucasian is on the Top Right and Bottom right for the Asian (Images were taken from [2], Except for the Caucasian Baby, from [14]).

Internal and external forces act upon the outer and inner skin as a person age, causing some level of damage and changing the skin's appearance. As demonstrated in [9, 10], the older skin was perceived to have a different colour contrast and luminosity than the younger skin. Healthy young skin, which is plumper and emits radiant colour, has a smooth and uniformly fine texture that reflects light evenly. Meanwhile, aged skin tends to be rough and dry with more wrinkles, freckles, and age spots and emits dull colour [11, 12]. However, ethnicities can affect these aging rates because of differences in skull structure and skin type [1] (see Fig. 1). For instance, the skin of a Caucasian subject will gradually have more aging wrinkles when compared to an Asian subject as the age increases from 20 to 39 years old. This phenomenon is due to the different skull and skin structures of each ethnic. Caucasians have a significant angular face, while Asians tend to be broader and less angular, similar to a baby's broad face [2] (see Fig. 2). Due to this broader facial structure, soft-tissue loss in Asians is seen and felt to a lesser extent. Another example is between the Caucasians and the African-Americans' skin. Black skin's epidermis contains a thicker stratum corneum with more active fibroblasts than the Caucasians, making them less affected by photo aging [3, 4]. Although black skin does not tend to get fine lines like white skin, it does get folded when getting older. Such information should be considered to design a more accurate age estimation model which can specify proper age estimation/classification knowledge when dealing with specific ethnic subjects.

IV. FACE AGE DATABASE

Designing face age estimation models require many samples for training and testing. Several studies collected face samples and then made them available to the public so that others might use them in their research. Furthermore, the shared database may serve as a benchmark against which other models can be compared and improved. Table I shows the face databases with age information or labels (henceforth, called Face Age Database). Only two databases captured face images in a controlled environment (MORPH and FACES). In contrast, the rest captured the face image in either a partially controlled or uncontrolled environment. Meanwhile, the FG-NET database has the most undersized samples and subjects, while the IMDB+WIKI database offers the most samples and subjects.

Table I reveals that most of the subjects in the databases are Caucasian/White, whereas Table II provides the ethnic count. Correspondingly, the ethnic percentage is shown in Fig. 3, which reveals very few databases with non-Caucasians/non-White ethnic (White = 80%; Black = 3%; Asian = 8%; and Others = 9%). This gap creates an imbalance in the databases when ethnicity is considered to estimate the age of non-Caucasian/non-White races. Moreover, not all the databases have ethnic information (e.g. IMDB+WIKI, FERET, and Webface). The lack of ethnic labels can make it difficult for face age model researchers to divide samples into their appropriate ethnicity, eventually treated as one of their research limitations.

TABLE I. SUMMARY OF FACE DATABASES WITH AVAILABLE INFORMATION

Year	Database	Samples	Environment		Age Range	Ethnic
			C	UC		
1998	FERET [21]	14,126 samples; 1,199 subjects	√	√	Not mentioned (real age)	Not mentioned
2002	FG-NET [22]	1,002 samples; 82 subjects		√	0-69 (real age)	All White/Caucasian
2004	LIFESPAN [23]	1, 142 samples; 575 subjects	√	√	18-93 (age group)	African-American:89; Caucasian:435; Others:52
2005	FRGC [24]	44,278 samples; 568 subjects	√	√	16-77 (real age)	White:386; Asian:125; Others:57
2006	MORPH [25]	55,134 samples; 13,618 subjects	√		16-77 (real age)	White-Black ratios 4:1; Others-very small
2008	YGA [26]	8,000 samples; 1,600 subjects		√	0-93 (real age)	Not mentioned
2009	GROUPS [27]	28,231 samples; 28,231 subjects		√	0-66+ (age group)	Not mentioned
2010	FACES [28]	2, 052 samples; 171 subjects	√		19-80 (age group)	All White/Caucasian
2012	Webface [29]	59, 930 samples		√	1-80 (real age)	Not mentioned
2014	Adience [30]	26,580 images; 2,284 subjects		√	0-60 (age group)	Not mentioned
2014	CACD [31]	160,000 samples; 2,000 subjects		√	16-62 (real age)	Not mentioned
2015	Chalearn 2015 [32]	4, 699 samples		√	Not mentioned (real age)	Not mentioned
2016	Chalearn 2016 [33]	7, 591 samples		√	Not mentioned (real age)	Not mentioned
2017	AgeDB [34]	16,516 samples; 570 subjects		√	1-101 (real age)	Not mentioned
2018	IMDB+WIKI [35]	523,051 samples; 20,284+		√	0-100 (real age)	Not mentioned
2007	Iranian face [15]	3,600 samples; 616 subjects		√	2-85 (real age)	All Iranian
2013	IMFDB [16]	34,512 samples; 100 subjects		√	Not mentioned (age group)	All Indian
2016	AFAD [17]	164,432 samples		√	15-40 (real age)	All Asian
2017	APPA-REAL [36]	7,591 samples; 7,000+ subjects		√	0-95 (real age)	Caucasian: 6,686; Asian: 674; Afro-American: 231

(C – captured or collected in a controlled environment; UC – captured or collected in an uncontrolled environment)

TABLE II. ETHNIC COUNT (BASED ON TABLE I)

Ethnic Count	Subjects (Approx. ≈)
White	7760
Black	320
Asian	799
Others	825

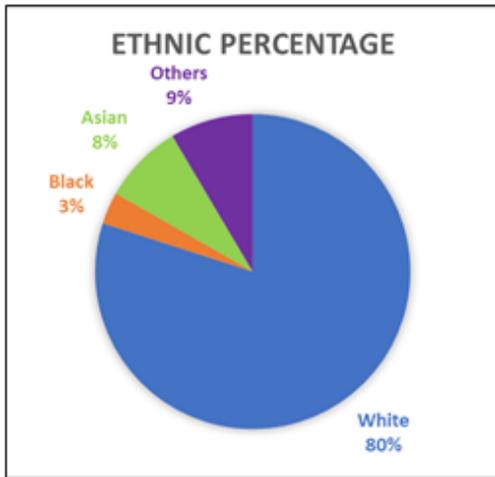


Fig. 3. Ethnic Percentage from All Databases (Based on known Data from Table I).

Some studies, however, have collected particular ethnic subjects with age information, such as Bastanfard A. et al. (Iranian face database) [15]; Setty S. et al. (Indian Movie Face database - IMFDB) [16]; and Niu Z. et al. (Asian Face Age Dataset - AFAD) [17]. Furthermore, there are ethnic-specific databases that can be used for face age research (see Table I coloured in grey). However, no studies have used these databases for deep learning age estimation research in the past six years; these databases are either not considered benchmark databases or less known by the face age estimation community.

V. FACE AGE DATABASE ESTIMATION MODEL

In one of the earliest face age model studies, Kwon and Lobo [18] classified age into three categories: infant, adult, and senior, and used simple feature extraction and machine learning for face age classification. Subsequently, computer science and pattern recognition researchers introduced various age classification/estimation methods [19, 20]. Earlier machine learning methods typically included one (or more) feature extractor and one (or more) age classifier (or estimator). The feature extractors can be holistic (e.g. whole facial shape), local (e.g. aging wrinkle), or both. The selection of feature extractors is influenced by the database used, with most of the sample quantity used by these methods being less than that of the deep learning approach.

Previous machine learning approaches can produce precise estimation (or classification) using just one or two databases, but are confined to those databases and could give an erroneous estimation if a wild sample is used for testing instead. Moreover, it is difficult for most machine learning approaches

to analyse unstructured data; they require additional tasks to divide the problem and later recombine the results to form a conclusion, which takes time and resources. Nevertheless, the deep learning method's known capability and strength have shifted the face aging system approach.

A. Deep Learning Approach

The rise of deep learning in image processing and machine learning has also impacted face age estimation. Better age estimation performance is strictly associated with the depth of the used network in the deep learning method, and it has become the generalist network adopted for feature extraction, including deep architectures that require a considerable amount of image samples, such as AlexNet, VGG-Net, VGG-Face, GoogLeNet, and Residual Networks (ResNet) [37]. VGG-Net has been reported to be one of the most effective deep learning architectures for age estimation. Notwithstanding, new studies continue to propose deep architecture designs for improving model accuracy when processing a sample of subjects' faces captured in an uncontrolled environment.

Deep learning face age research can be classified into three types: 1) classification age (CA) - classify the face age with several classes equal to the number of the considered age groups; 2) estimation age (EA) - estimate age using a regressor; and 3) distribution age (DA) - a modified CA strategy obtained by substituting the one-hot encoding vector with a statistical distribution centred on the estimated age [37]. Furthermore, the deep learning approach is much more accurate than other older machine learning methods at estimating age from sample images captured in the wild (uncontrolled environment). Nevertheless, if the subject's ethnicity in the dataset is not considered, the ethnic bias will persist.

B. Deep Learning Model Method and Ethnicity Bias in Database

When searching for papers on face age research, this study focused on research that used the deep learning method from 2015-2020. Deep learning has the potential to revolutionise computer science and machine learning. Furthermore, data biases are becoming more important with the rise of more powerful machine learning, which deep learning takes advantage of when dealing with large amounts of data. The search was conducted using a variety of web search engines, including Google Scholar and Web of Science.

Table III displays the search results, which include the following information: publisher, year of publication, network architecture, domain area, selected databases, and ethnicity consideration. From 2015-2016, the most commonly used network architectures were well-known general architectures such as GoogleNet, VGG-Net, and DCNN (or Deep-CNN). As the year progressed, an increasing number of studies began to design the architectural network or modify the general CNN network architecture to fit their research objectives. As a result, the network design became more complex to produce a more accurate novel model (e.g. by combining multiple CNN networks to create a hybrid network). In the research domain area, there have been 33 EA studies, 20 CA studies, and only 8 DA studies. However, there is no significant preference between the research domain and databases used in the studies.

Therefore, it can be inferred that most of these databases can be used in all deep learning areas: EA, CA, and DA.

TABLE III. DEEP LEARNING FACE AGE RESEARCH AREA AND AGE DATABASES USED FOR TRAINING AND TESTING (FROM 2015-2020)

Ref.	Publisher	Year	Network Architecture	Domain Area			Database Model										Account Ethnic
				CA	DA	EA	M	C15	I	A	F	C16	W	G	P	O	
[38]	IEEE	2015	GoogLeNet		√	√	√	√					√			√	No
[39]	IEEE	2015	VGG-Net	√				√	√								No
[40]	IEEE	2015	GoogLeNet			√	√	√		√	√		√			√	No
[41]	IEEE	2015	VGG-Net & GoogleNet			√		√									No
[42]	IEEE	2015	VGG-Net & Novel arch.		√		√	√		√	√				√		No
[43]	IEEE	2015	DLA			√	√				√						No
[44]	IEEE	2015	Tree kernel adaptive CNN			√	√					√					No
[45]	Elsevier	2015	LeNet			√	√								√		No
[46]	IEEE	2015	Novel arch.	√						√							No
[47]	IEEE	2015	DCNN-H-3NNR			√	√	√		√							No
[48]	IEEE	2016	DCNN			√		√		√	√			√			No
[49]	IEEE	2016	VGG-Net			√						√					No
[50]	IEEE	2016	Novel arch.	√			√									√	No
[51]	IEEE	2016	VGG-Net			√	√										No
[52]	IEEE	2016	Compact-CNN			√	√								√	√	Yes
[53]	Elsevier	2016	DCNN	√										√			No
[54]	IEEE	2016	GilNet; AlexNet; VGG-Net	√						√							No
[55]	IEEE	2016	VGG-Net	√	√			√	√							√	No
[56]	IEEE	2016	DADL		√			√*	√			√**					No
[57]	IEEE	2016	VGG-Net	√				√	√			√					No
[58]	IEEE	2016	VGG-Net	√					√			√					No
[59]	IEEE	2016	DCNN	√						√						√	No
[60]	IEEE	2016	Novel arch.		√		√				√					√	No
[61]	IEEE	2017	AGEn & MO-CNN	√			√	√	√		√	√				√	No
[62]	IEEE	2017	Multitask CNN			√	√	√	√	√	√						No
[63]	IEEE	2017	ODFL & ODL	√			√	√			√						No
[64]	Elsevier	2017	GA-DFL	√			√	√			√						No
[65]	Elsevier	2017	VGG-Net CNN+LDAE		√		√		√		√	√			√		No
[66]	Elsevier	2017	Novel arch.	√	√	√	√						√				Yes
[67]	Elsevier	2017	D2C			√	√						√				No
[68]	PMLR	2017	R-SAAFc2			√		√*	√	√	√			√			No
[69]	IEEE	2017	Deep-ROR	√					√	√							No
[70]	IEEE	2017	DMTL			√	√									√	Yes
[71]	IEEE	2017	M-LSDDL	√			√	√	√	√	√		√	√			No
[72]	IEEE	2017	DMTL			√	√	√	√							√	Yes
[35]	Springer	2018	VGG-Net	√			√	√	√	√	√					√	No
[73]	IEEE	2018	VGG-Net-GPR			√	√	√	√							√	Yes
[74]	IEEE	2018	ELM			√	√		√	√		√					Yes
[75]	ALM-DL	2018	ScatNet	√												√	No
[76]	IEEE	2018	CMT-deep network			√	√				√		√				No
[77]	Elsevier	2018	DAG-CNNs			√	√		√		√						No
[78]	Springer	2019	CNN+triplet ranking	√		√	√			√						√	No
[79]	Elsevier	2019	DeepAge			√	√			√							No
[80]	IEEE	2019	SADAL			√	√			√							No
[81]	IEEE	2019	Novel Arch.			√		√									No
[82]	IEEE	2019	Multitasks-AlexNet			√			√							√	No
[83]	IEEE	2019	ODFL & ODL			√	√	√			√					√	No
[84]	IEEE	2020	SADAL & VDAL			√	√	√			√						No
[85]	Elsevier	2020	LRN		√	√	√	√								√	Yes
[86]	IEEE	2020	CR-MT	√		√	√						√			√	No
[87]	SYMMETRY	2020	MA-SFV2	√		√	√				√						No
[88]	IEEE	2020	DOEL-groups			√	√	√	√		√					√	No
[89]	IEEE	2020	MSFCL			√	√	√		√	√					√	No

(CA – Classification Age; DA – Distribution Age; EA – Estimation Age; M – MORPH; C15 - ChaLearn2015; I – IMDB+WKI; A – Adience; F – FG-NET; C16 - ChaLearn2016; W – Webface; G – GROUP; P – Private DB; O – Others; *√* - variation of ChaLearn2015; **√** - variation of chalearn2016).

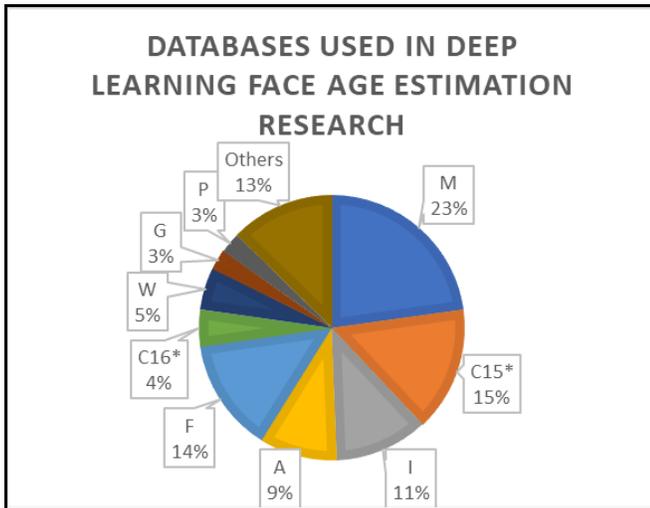


Fig. 4. Breakdown of Databases used in Deep Learning Face Age Estimation Research (based on Table III). ‘*’ Means that there are Multiple Versions of the Databases Included in the Count (M – MORPH; C15 - ChaLearn2015; I – IMDB+WIKI; A – Adience; F – FG-NET; C16 - ChaLearn2016; W – Webface; G – GROUP; P – Private DB)

Meanwhile, Fig. 4 depicts the most commonly used databases for face age research (derived from Table III), indicating that MORPH [25] is the most commonly used database because it has the highest sample count (55,134 samples) captured in a controlled environment. Because of this advantage, the MORPH database is the best benchmark database for comparing the CNN model performance with other models since it can lessen the influence of unwanted factors that may affect the overall estimation results. The MORPH database, on the other hand, has an unbalanced ratio of races in its dataset (refer to Table I), which can lead to ethnic bias when estimating age.

The second most used face age database is ChaLearn2015 [32], which was explicitly developed for the ICCV 2015 ChaLearn Looking at People Apparent Age Estimation Challenge [32]. This challenge event was a competition to build the best appearance age estimation model, and most of the authors of the research surveyed in this study competed in it. ChaLearn2016 [33], the fifth most used database, is the second/expanded version of the ChaLearn2015 database. The lack of ethnicity records for subjects in both ChaLearn databases makes analysing the effect of ethnicity on a model's overall performance difficult, even though both databases have a diverse set of races.

The FG-NET [22] database comes in third place, with images captured in uncontrolled real-life conditions that are not equally distributed across age groups and has the lowest

samples (1,002 samples) compared to other databases. FG-NET has been used in face age research since around 2005 [5], making it one of the most well-known databases used primarily for comparing model performance in the face age research community. Despite this, the majority of its subjects are Caucasians/Whites. When used in the CNN model, a small dataset should be fine-tuned or pre-trained with another database with large sample size, such as IMDB+WIKI. The IMDB+WIKI [35] dataset contains images with one or more people in them, as well as annotations for researchers' reference when there are multiple people in one image. However, there is no proper explanation for which annotation refers to which person in the image of multiple people. Therefore, studies primarily use this database for pre-training deep networks due to its large sample size. Because of the lack of annotation, no model performance results for IMDB+WIKI are shown in Table IV, which reveals the model performance on studies based on their selected databases. Although the IMDB+WIKI database samples contain multiple ethnicities, no annotation for a subject's ethnicity is available.

Adience [30], a database for gender and age group classification, comes in fourth place, with subjects drawn from real-world conditions. Its sources are mostly Flickr albums uploaded from smartphone devices. This database was made available to the general public under the Creative Commons (CC) licence. Meanwhile, in fifth place is Webface [29], a database collected for the experiments of a PhD thesis, and in sixth place is GROUP [27], a collection of images of people captured in a group (hence the name) that includes age group and gender information. Nevertheless, none of these three databases has a record of the subject's ethnicity.

Although some studies used/included their own database, these private databases [42, 52, 55, 65] contain no information about the subject's ethnicity. Furthermore, some of them were only used to fine-tune network models [55], [65]. CACD [31], LIFESPAN [23], LFW [90], FACES [28], FRGC [24], AFAD [17], and FERET [21] are the remaining databases used in the age model. Only a few of these were used in the face age deep learning research (categorised as 'Others' in Table III).

Meanwhile, some of the studies used a different database for pre-training their models (e.g. face detection in images) than the one used for age estimation, such as the CelebFaces Attributes (CelebA) [91] and ImageNet databases [92]. The CelebA database was built using the CelebFaces [91] face verification database with face attribute annotations. ImageNet, on the other hand, is a database for object classification and detection. Both databases lack age information and were primarily used for pre-training/fine-tuning the network model in these studies [39, 52, 57, 69, 70, 74].

TABLE IV. SUMMARY OF BEST CNN MODEL PERFORMANCE ON SELECTED DATABASES (FROM 2015-2020)

Ref.	Year	Best CNN Model Performance on Selected Databases <i>Database_used(performance_measurement)</i>	Account Ethnic
[38]	2015	C15(MAE = 3.33); C15-testset(e-error = 0.27)	No
[39]	2015	C15-validset(MAE = 3.22); C15-testset(e-error = 0.260)	No
[40]	2015	C15-validset(e-error = 0.309); C15-testset(e-error = 0.290)	No
[41]	2015	C15-validset(MAE = 3.29; e-error = 0.285); C15-testset(e-error = 0.287)	No
[42]	2015	C15-validset(e-error = 0.338); C15-testset(e-error = 0.306)	No

[43]	2015	M(MAE = 4.77); F(MAE = 4.26)	No
[44]	2015	M(MAE = 3.61); W(MAE = 7.27)	No
[45]	2015	M(MAE = 3.88); FRGC(MAE = 3.31)	No
[46]	2015	A(AEM = 50.7±5.1, AEO = 84.7±2.2)	No
[47]	2015	C15-validset(e-error = 0.359); C15-testset(e-error = 0.373)	No
[48]	2016	A(AEM = 52.88±6%, AEO = 88.45±2.2); C15-validset(e-error = 0.297)	No
[49]	2016	C16-validset(MAE = 3.85, e-error = 0.330); C16-testset(e-error = 0.370)	No
[50]	2016	M(MAE = 3.27); AFAD(MAE = 3.34)	No
[51]	2016	M(MAE = 3.45)	No
[52]	2016	M(MAE = 3.23); P(acc. = 88%)	Yes
[53]	2016	G(AEM = 56%, AEO = 92%)	No
[54]	2016	A(mean AEM = 57.9%)	No
[55]	2016	C15-validset(e-error = 0.261); C15-testset(e-error = 0.241)	No
[56]	2016	C15-validset(MAE = 1.76, e-error = 0.134); C15-testset(e-error = 0.321)	No
[57]	2016	C16-validset(e-error = 0.240); C16-testset(e-error = 0.336)	No
[58]	2016	C16-testset(e-error = 0.367)	No
[59]	2016	A(acc. = 42%); FERET(acc. = 86.4%)	No
[60]	2016	M(MAE = 2.78); F(MAE = 2.80)	No
[61]	2017	M(MAE = 2.52); F(MAE = 2.96); CACD (ave. MAE = 4.68); C15-validset(MAE = 3.21, e-error = 0.28); C15-testset(MAE = 2.94, e-error = 0.264); C16-testset(MAE = 3.82, e-error = 0.310)	No
[62]	2017	F(MAE = 2.00); C15-validset(e-error = 0.293)	No
[63]	2017	M(MAE = 2.92); F(MAE = 3.71); C15-validset(MAE = 3.95, e-error = 0.312)	No
[64]	2017	M(MAE = 3.25); F(MAE = 3.93); C15-validset(MAE = 4.21, e-error = 0.369)	No
[65]	2017	F(MAE = 2.84); M(MAE = 2.35); P(MAE = 4.33); C16 (e-error = 0.241)	No
[66]	2017	M(ave. MAE = 2.96); W(ave. MAE = 5.75)	Yes
[67]	2017	M(ave. MAE = 3.06); W(ave. MAE = 6.104)	No
[68]	2017	F(MAE = 3.01 MAE); A(AEM = 67.3, AEO = 97.4)	No
[69]	2017	A(AEM = 67.34 ± 3.56%, AEO = 97.51 ± 0.67%)	No
[70]	2017	M(acc. = 85.30 ± 0.6%)	Yes
[71]	2017	M(MAE = 2.89); F(MAE = 3.31); A(AEM = 60.20±5.3%, AEO = 93.70± 2.3%); C15-validset(e-error = 0.315)	No
[72]	2017	M(MAE = 3.00); LFW(MAE = 4.50)	Yes
[35]	2018	M(MAE = 2.68); F(MAE = 3.09); CACD(MAE = 4.79); A(AEM = 64.00±4.2%, AEO = 96.60±0.9%)	No
[73]	2018	M(MAE = 2.93); CACD (MAE = 5.22); C15-validset(MAE = 3.30, e-error = 0.290)	Yes
[74]	2018	M(MAE = 2.61); A(AEM = 66.49 ± 5.08%); C16-validset(MAE = 3.67, e-error = 0.325); C16-testset(e-error = 0.368)	Yes
[75]	2018	LIFESPAN(MAE = 4.01); FACES (MAE = 5.95)	No
[76]	2018	M(MAE = 2.89); F(MAE = 3.43)	No
[77]	2018	M(MAE = 2.81); F(MAE = 3.05)	No
[78]	2019	M(MAE = 2.87); A(AEM = 63.10 ± 1%, AEO = 96.7 ± 0.4%)	No
[79]	2019	M(MAE = 2.87); F(MAE = 7.08)	No
[80]	2019	M(MAE = 2.75); F(MAE = 3.67)	No
[81]	2019	C15-testset(MAE = 6.031, e-error = 0.441)	No
[82]	2019	Wiki(MAE = 5.47); UTKFace(MAE = 9.54); AgeDB(MAE = 10.01)	No
[83]	2019	M(MAE = 2.92); F(MAE = 3.71); APPARENT-AGE(MAE = 3.95)	No
[84]	2020	M(MAE = 2.57); F(MAE = 2.98); C15(MAE = 3.58, e-error = 0.285)	No
[85]	2020	M(MAE = 1.90); C15-validset(MAE = 3.05, e-error = 0.274); MegaAge-Asian(CA(7) = 91.64)	Yes
[86]	2020	M(ave. MAE = 2.36); CACD(MAE = 4.48); Webface(MAE = 5.67)	No
[87]	2020	M(MAE = 2.68); F(MAE = 3.81)	No
[88]	2020	M(MAE = 2.75), F(MAE = 3.44); AgeDB(MAE = 5.69); C15-validset(MAE = 2.93, e-error = 0.258); C15-testset(MAE = 2.71, e-error = 0.247)	No
[89]	2020	M(MAE = 2.73); F(MAE = 2.71); A(AEM = 65.3%, AEO = 96.3%); MEGAAGE-ASIAN(MAE = 2.81, CA(3)(62.89%), CA(5)(82.46%))	No

M – MORPH; C15 – ChaLearn2015; A – Adience; F – FG-NET; C16 – ChaLearn2016; W – Webface; G – GROUPS. Note that other databases that are not shown in this table but shown in Table III (e.g. I – IMDB+WKI) were used for pre-training by the studies.

C. Deep Learning Technique Strengths and Weaknesses

A review of the different deep learning architectural networks used in previous studies revealed several techniques that are frequently used for face age estimation. Table V summarises the network architectures frequently used in the age estimation studies reviewed, as well as their strengths and weaknesses. As previously stated, the main goal of deep learning face age estimation is to find the best method for learning the face aging features from a large sample of data and then use the information to distinguish the different ages of test subjects. Each study's architecture was chosen based on its

research aim and objectives, such as the problem(s) to solve that can help improve face age classification/estimation/distribution. The problems include face detection, landmark localisation, optimisation, regression, classification, feature extraction, residual learning part, sampling technique, layer size (depth and width), discriminative distance, learning speed, training and/or testing process and others. This study identified several known network architectures that were frequently used in comparison to the others [93]. Among these network architectures are the following:

TABLE V. SUMMARY OF NETWORK ARCHITECTURES MOSTLY USED BY AGE ESTIMATION STUDIES IN THIS SURVEY

Architecture	Background (referred from [93])	Learning Methodology	Strength	Weakness	Author(s) that Used the Architecture
LeNet	<ul style="list-style-type: none"> - Invented in 1998 by Yann Lecun. - First popular CNN architecture. 	Spatial exploitation	<ul style="list-style-type: none"> - Small and simple design. - A good introduction to neural networks for beginners. 	<ul style="list-style-type: none"> - Problem to detect all aging features. Require extensive training. - Speed and accuracy are outperformed by newer network architecture. 	[45]
AlexNet	<ul style="list-style-type: none"> - Introduced in 2012 at the ImageNet Large Scale Visual Recognition Challenge. - Uses ReLu, dropout and overlap pooling. - First major CNN model that used GPUs for training. 	Spatial exploitation	<ul style="list-style-type: none"> - Using GPUs for training leads to faster training of models. - ReLu helps lessen the loss of features and improve model training speed. 	<ul style="list-style-type: none"> - Authors require to find design solutions on how to compete with other newer network architectures that are more accurate and faster. 	[54, 82]
VGG-Net	<ul style="list-style-type: none"> - Visual geometric group (VGG) was introduced in 2014. - It groups multiple convolution layers with smaller kernel sizes. 	Spatial exploitation	<ul style="list-style-type: none"> - Homogenous topology. - Smaller kernels. - Good architecture for benchmarking face age estimation - Pre-trained networks for VGG-Net are freely available. 	<ul style="list-style-type: none"> - Computationally expensive as more layer increases. - Face age estimation studies need to consider the vanishing gradient problem to improve the estimation performance. 	[35, 39, 42, 49, 51, 54, 55, 57, 58, 65, 73]
GoogleNet	<ul style="list-style-type: none"> - Researchers at Google introduced GoogleNet in 2014. - Introduced block concept. - Split transform and merge idea. - In a single layer, multiple types of 'feature extractors' are present to help the network perform better. 	Spatial exploitation	<ul style="list-style-type: none"> - Trains faster than VGG-Net. - Smaller pre-trained size than VGG-Net. - Training network has many options to solve tasks. 	<ul style="list-style-type: none"> - Heterogeneous topology design require face age estimation studies to make thorough customisation - from module to module. 	[38, 41, 43]
ResNet	<ul style="list-style-type: none"> - Introduced in 2015. - Residual learning. - Identity mapping-based skip connections. 	Depth + multi-path	<ul style="list-style-type: none"> - Capable of skipping learned feature(s), reducing training time and improve accuracy. - Solve the vanishing gradient problem faced by VGG-Net. - Possible to train very deep networks and generalise well. 	<ul style="list-style-type: none"> - Computationally expensive as more layer increases. 	[69, 71]
Novel Arch.	<p>Most designs were expanded/modified/or built from scratch based on the previously available architectures (e.g. AlexNet, VGG-Net, etc.)</p>	-	<ul style="list-style-type: none"> - Specialise in learning face representation for different ages. - Improving several parts of the network based on the study's aim and objectives. 	<ul style="list-style-type: none"> - Cater to a very specific problem(s). - Time-consuming when building from scratch. 	[42, 43, 44, 46, 47, 48, 50, 52, 53, 54, 56, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89]

1) *LeNet*: Yann Lecun invented the LeNet architecture in 1998 to perform optical character recognition (OCR), and its design is smaller and simpler than the rest of the network architectures. For beginners, this network is a good way to learn neural networks and can be used for face age estimation studies, such as in [45]. However, due to its simple design, the network requires additional improvement that the designer must build from scratch if used for face age estimation. It is also outclassed by newer models in terms of speed and accuracy when used as is, with no modifications.

2) *AlexNet*: Alex Krizhevsky introduced the AlexNet architecture in 2012, and it was the first major CNN model to use graphics processing units (GPUs) for training, which aided in training speed. Meanwhile, ReLU, dropout and overlap pooling were used to reduce feature loss and improve training speed. This architecture design was used in [54], [82] for face age classification and estimation, respectively. Their accuracy performance, however, was inferior to that of the model that used the LeNet network design [45] (see Table IV). This implies that, even though AlexNet is a newer network than LeNet, proper modification, structuring, and organisation of the AlexNet network are still required to achieve the best face age estimation (or classification) performance.

3) *VGG-Net*: Introduced in 2014, the VGG model improves training accuracy by improving its depth structure. The addition of more layers with smaller kernels increases nonlinearity, which is good for deep learning. This study discovered that VGG-Net is the most commonly used network model among the many available (11 papers). One of the possible explanations is that the VGG pre-trained networks are freely available online. Although it is the best architecture for benchmarking on the face age estimation task, the performance obtained by studies that used this model is not the best, but it is also not the worst. This could be due to the vanishing gradient problem, one of the main challenges faced when using VGG-Net, which occurs when the number of layers exceeds 20, causing the model to fail to converge to the minimum error percentage. When this happens, the learning rate slows to the point where no changes are made to the model's weights. Furthermore, using VGG-Net can be time-consuming because the training process can exceed a week, especially if it was built from scratch. As a result, when using the VGG-Net network for face age estimation, users must address the vanishing gradient problem as well as the training time.

4) *GoogleNet*: A class of architecture designed by Google researchers that won ImageNet 2014. Instead of a sequential architecture design, GoogleNet opted for a split transform and merge design, in which a single layer can have multiple types of "feature extractors". In addition, GoogleNet has a smaller pre-trained size and trains faster than VGG-Net [93]. One drawback of GoogleNet is that almost every module must be customised. As a result, when designing a face age estimation using GoogleNet, users must customise from module to

module. This study discovered that only [38, 40, 41] used this network architecture.

5) *ResNet*: ResNet was introduced in 2015 and provides residual learning to help solve the vanishing gradient problem (from the VGG-Net architecture). Furthermore, ResNet can have a deeper network (more layers) than VGG-Net while avoiding performance degradation. ResNet is a concept in which if a feature has already been learned, it can be skipped and focus can be given to newer features, thereby improving training time and accuracy. On the other hand, the ResNet structure design is primarily concerned with how deep the structure should be. If ResNet is chosen for face age estimation, the designer must consider how the network should be structured to learn multiple aging features. Adding more layers is one of the common ideas. However, this could result in a longer learning time for the model (it can take several weeks); therefore, the designer must also account for this. This study discovered that only a few face age estimation studies used Resnet architecture/concept in their design [69, 71].

6) *New Arch*: Is a network architecture created by expanding previous architectures, modifying them, or building the network from scratch. These architectures were created specifically to find the best network approach for learning how to best estimate age. For example, a facial image with a specific age can be affected by facial variations caused by external factors, such as lighting, which can lead to a neighbouring age category being predicted as the final bias. The study in [80] attempted to address this problem by proposing a network composed of a generator that could generate discriminative hard-examples (taken from extracted features done by a deep CNN) to complement the training space for robust feature learning and a discriminator that could determine the authenticity of the generated sample using a pre-trained age ranker [80]. This approach offers designers the 'freedom' to create the best solution to a given problem. The designs can be based on available networks and further modified to their preferences, rather than being limited to the original design architecture. This study found that most of the previous studies, particularly those conducted in 2020, tend to propose their own architectural network design. However, one major drawback of this design approach is that the designer may take a long time to modify/create networks when compared to using available networks.

D. Model Performance Evaluation

Multiple protocols and performance calculations were used in the studies to evaluate model performance. Table IV shows the performance of the CNN models used in the studies on the databases that they were tested on. The evaluation protocol is a method for studies to determine the optimal number of training and testing datasets for their chosen databases. Meanwhile, the performance calculation allows studies to compare the estimation/classification/ distribution accuracy of their own model to that of others. Because of the numerous ways for designing protocols and performance calculations, problems

arise when performance is compared on the same database but using different evaluation methods, resulting in a unanimous 'agreement' from most of the studies that specific performance calculation(s) should be used for comparison's sake for a specific database. Among the performance calculations used to evaluate the accuracy of the face age deep learning model are:

1) *Mean absolute error (MAE)*: a widely used performance evaluation for age estimation studies that measures the error between the predicted and actual ages. MORPH, FG-NET, ChaLearn2015, and ChaLearn2016 are examples of databases that used this evaluation method. The model performance improves as the MAE value decreases.

2) *E-Error*: is the performance calculation used in apparent age estimation. This evaluation metric was used to compare the performance of studies that used ChaLearn2015 [32] and ChaLearn2016 [33] datasets. The lower the e-Error, the better the performance.

3) *Accuracy of an exact match (AEM)*: a method of calculating accuracy that calculates the percentage of correctly estimated/classified age per the total number of test images used. This type of evaluation metric was used by the Adience database. The higher the AEM value, the better the performance. Some studies went so far as to include the standard deviation value in their evaluation.

4) *Accuracy error of one age category (AEO)*: Is another type of evaluation metric used on the Adience database, in which errors of one age group are also included as correct age classifications. The higher the AEO value obtained, the better the overall model performance.

5) *Cumulative score (CS)*: is defined as the percentage of images with an error of no more than a certain number of years. The evaluation is usually shown as a curve on a graph (which is not depicted in this paper), with the x-axis representing the error level in years and the y-axis representing the cumulative score (in percentage value). This type of evaluation performance was sometimes combined with the MAE evaluation method in studies that used MORPH, FG-NET, and other earlier year databases. Meanwhile, studies that used the MegaAge-Asian database present some of their results in terms of $CA(\theta)$, where θ is the allowable age error corresponding to the cumulative accuracy, which several of them are shown in Table IV.

Because the studies reviewed from 2015-2020 (see Table IV) used different databases, analysing and comparing their performance progress was difficult. Therefore, only the most frequently used databases were chosen and averaged to create a line chart depicting the performance progress of face age research from 2015-2020. Fig. 5 illustrates the average yearly performance for two different databases: MORPH and FG-NET. As shown in Fig. 5, the MAE values for the MORPH database decreased from 2015-2020, but not for FG-NET. The chart may imply that models applied to the MORPH database improved over six years, whereas FG-NET did not. Table VI shows the average MAE and its standard deviation for each

year; the improvement might be valid for MORPH since most of the standard deviation obtained is low (< 0.3). However, the implication for FG-NET may be invalid because only a few studies used this database in 2015-2016. Most of the standard deviation for 2017-2020 is high (> 0.4), meaning that the MAE results obtained by the different studies are too wide apart. Among other databases, the performance of the MOPRH database appears to be the best. The samples captured in a controlled environment help the models to better identify aging features because unwanted factors are absent (e.g occlusion). Meanwhile, the low quantity (1,002 images) and low quality (old images captured in an uncontrolled environment) samples of FG-NET might hinder the CNN model learning process in the studies. Nonetheless, some studies were able to obtain low MAE values using the FG-NET database: [62] MAE = 2.00 and [89] MAE = 2.71.

Regarding publishers, from 2015-2020 (see Fig. 6), IEEE is the publisher with the highest reviewed papers in this study. Elsevier is in second place, and Springer is in third. The bar chart in Fig. 6 shows that the number of published papers increased in 2016, but then declined until 2018, and then remained relatively low until 2020. The figure seems to imply that the deep learning approach is becoming less attractive to the face age research community, but this is most likely not the case. When a more robust, advanced, and practical deep learning technique becomes available, a resurgence may occur.

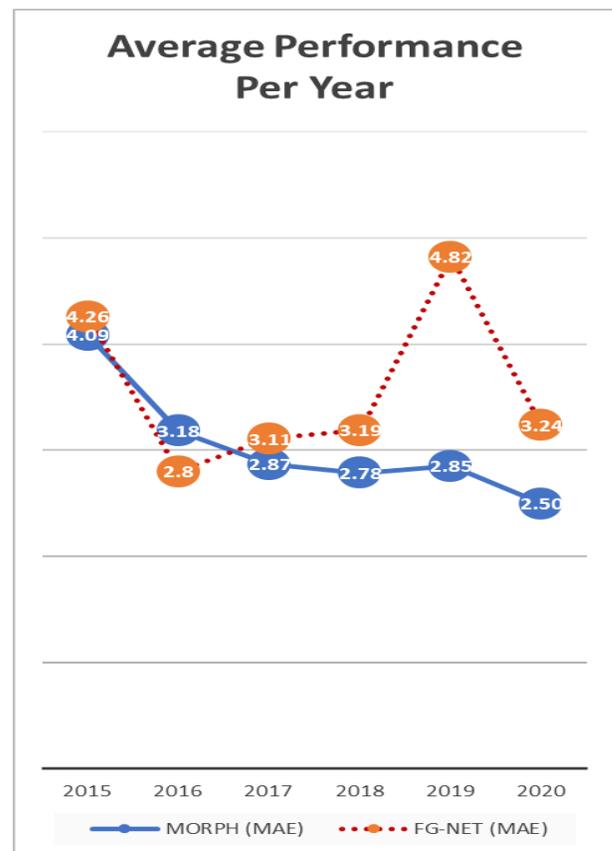


Fig. 5. Line Chart Shows Face Age Research's Performance Progress from 2015-2020 for MORPH and FG-NET (based on Table VI).

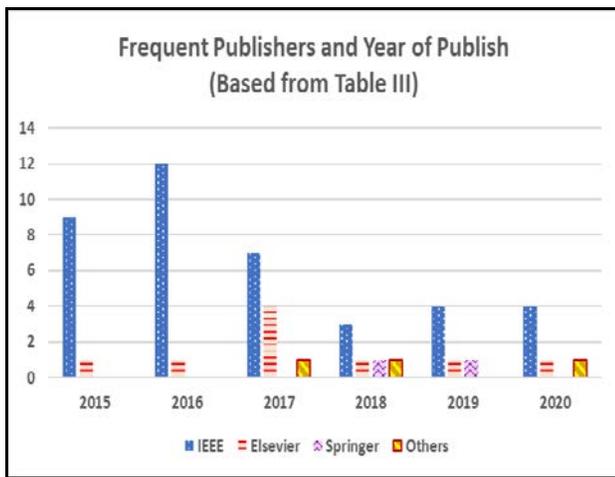


Fig. 6. Publishers and Frequency of Publication of Deep Learning Face Age Estimation Papers, Including Publication Year (based on Table III).

TABLE VI. PERFORMANCE PROGRESS OF FACE AGE RESEARCH FROM 2015-2020 FOR MORPH AND FG-NET DATABASES

Year	MORPH	FG-NET
	MAE (std. deviation)	MAE (std. deviation)
2015	4.09 (0.61)	4.26 (-)
2016	3.18 (0.28)	2.8 (-)
2017	2.87 (0.29)	3.11 (0.63)
2018	2.78 (0.14)	3.19 (0.21)
2019	2.85 (0.07)	4.82 (1.96)
2020	2.50 (0.33)	3.24 (0.49)

VI. OPEN ISSUES

As mentioned in the facial aging section, different ethnic subjects age differently, which means that a 20-year-old White subject would look older than a 20-year-old Asian. More research into the effects of ethnicity on face age estimation is needed. However, most studies focus only on primary aging features, such as face shape and aging wrinkles, and ignore secondary ones, such as racial facing aging traits. There are two possible reasons why studies did not take ethnic traits into account. The first is the perception that secondary aging features are non-essential for better model performance. A few CNN face age estimation studies have disproved the perception that race is unimportant. The second possible reason is that the lack of race variety in most databases causes researchers to overlook racial traits as one of the aging estimation problems in the first place.

Among papers reviewed, only seven considered ethnicity traits in the face age estimation experiment [52, 66, 70, 72, 73, 74, 85]. Studies in [52, 66, 70] considered ethnicity in the model learning performance and discovered that it does improve age estimation. However, these studies did not investigate the influence of racial traits on effectiveness in facial age estimation. Meanwhile, the study in [73] inferred that performance would improve if both gender and race information were included. Another study [74] discovered that gender and race could easily affect its age estimation model. Combining all of the age, gender, and race features further improved the age estimation performance of the model. Lastly, [85] explored the impact of ethnicity and gender on age

estimation, stating that having more samples for a specific ethnic can increase age estimation accuracy. These studies suggest that the CNN model can be further improved if the overall framework takes ethnicity into account first. When a large number of samples are available, CNN models perform better at discovering significant aging features, as this also improves the learning of racial aging traits.

Meanwhile, other studies on face age estimation that used machine learning rather than CNN have demonstrated the importance of ethnic aging traits. Ricanek et al. [94], for example, investigated the ethnicity of the subject and introduced the least angle regression (LAR) method, which was conducted on three databases: MORPH, FG-NET, and PAL, with five races included (African-American, Asian, Caucasian, Hispanic, and Indian). In another study, Akinyemi and Onifade [95] improved the performance of their model by incorporating ethnic parameters for African and Caucasian people into the GroupWise age ranking model. The FG-NET and FAGE databases were used in their experiment. FAGE is a locally collected dataset of 238 images of 209 black (African) individuals aged 0 to 41 years.

Shin et al. [96] presented an age estimation system that considered ethnic differences for Asians and Non-Asians using CNN and support vector machine (SVM). The proposed age estimation system outperformed the standard system when trained on an ethnicity-biased database. The study relied on LFW [90] and its samples, with Asians in the datasets consisting of Korean, Japanese, and Chinese web celebrities [96]. Several other studies, however, were unable to investigate their approach further due to a lack of multiple races in their datasets [94, 95, 96]. Hence, the importance of having more race variety in databases is demonstrated.

Table III shows that the databases in the deep learning age estimation model mostly favour Caucasian subjects. The race variety in the databases is imbalanced; in most databases, the Caucasian/White subjects are always the majority, while other races are either underrepresented or missing. Moreover, some of the databases with large samples and multiple races have no information on the subjects' ethnicity. There are only a few ethnic-specific databases, such as AFAD, IMFDB, and Iranian Face available. It would be beneficial to have more multiple ethnic databases with large samples and races that are evenly distributed.

VII. DISCUSSION AND SUGGESTION

The first problem to address is the negative perception that ethnicity is not a critical aging factor. Researchers should be informed more about the importance of ethnic traits in the aging face; thus, this paper aimed to raise awareness on this to others. Moreover, face age estimation research should be expanded; more researchers should consider the secondary aging traits when building CNN face age estimation models. The research scope should not be limited to primary aging features (face shape and aging wrinkles) but also expanded to secondary features that can help distinguish unique aging traits that occur only in specific races. One suggestion is to create a framework for organising different racial samples in a database before being used for a CNN model. The steps of the framework are as follow:

1) First, decide on the number of races to be included in the study and then collect as many samples as possible for each race while ensuring the samples are similar in quantity. This may require the creation of multiple databases with various ethnicities (e.g. using MORPH [25], IMFDB [16], Iranian face [15], and MEGAAGE-ASIAN [89] databases together). Because CNN would be the model approach, having a large sample size would not be an issue for the CNN learning process - it is required. The study must also decide whether to use all samples or specific ones based on the research aim and objective(s).

2) Next, apply the necessary image processing to the sample images, such as face detection, face landmark, and face alignment.

3) Each database's estimation performance is evaluated using an evaluation protocol. Multiple ethnic subjects from chosen databases are mixed and segregated into specific training and test sets when accounting for ethnicity in age estimation. The ethnic effect analysis requires two protocols: one that considers ethnicity and one that does not. The first protocol requires different ethnic subjects within these sets to be divided equally in quantity. Training and testing, for example, take up 80% and 20% of the total samples, respectively. When the samples are made up of two races (e.g. Caucasian and Asian), half of the training samples should be Caucasian and the other half should be Asian. Similarly with the test samples - half is Caucasian, and another half is Asian. The second protocol is similar to the first, except that the different ethnic subjects are split randomly rather than equally.

4) Afterwards, run the samples into the CNN model and analyse the result in terms of the ethnicity's effect on the overall age estimation. Search for any significant finding regarding the ethnicities traits that can be exploited in future age estimation studies.

Fig. 7 shows the proposed framework for studying CNN face age estimation while considering the ethnicities of the subjects. This framework can guide future face age estimation studies that use either the deep learning approach or the standard machine learning approach.

The review of the papers revealed that most studies did not consider using other ethnic-specific databases (e.g. Iranian [15], Indian [16]), even though these databases are available for use (see Table I). Benchmark databases like MORPH and FG-NET are more preferred because it is safer since these databases are frequently used and have long been used for comparison; thus, making it easy to perform comparative analysis. Nevertheless, using only the same benchmark databases and ignoring other available ethnic-specific databases can pose a risk, which will hinder the face age estimation research's progress in understanding the overall ethnic factor in facial aging process. Suppose various databases are continuously and increasingly used throughout the years. There will be enough results to allow meaningful comparison between studies, resulting in new benchmark databases that can be used and compared in the future.

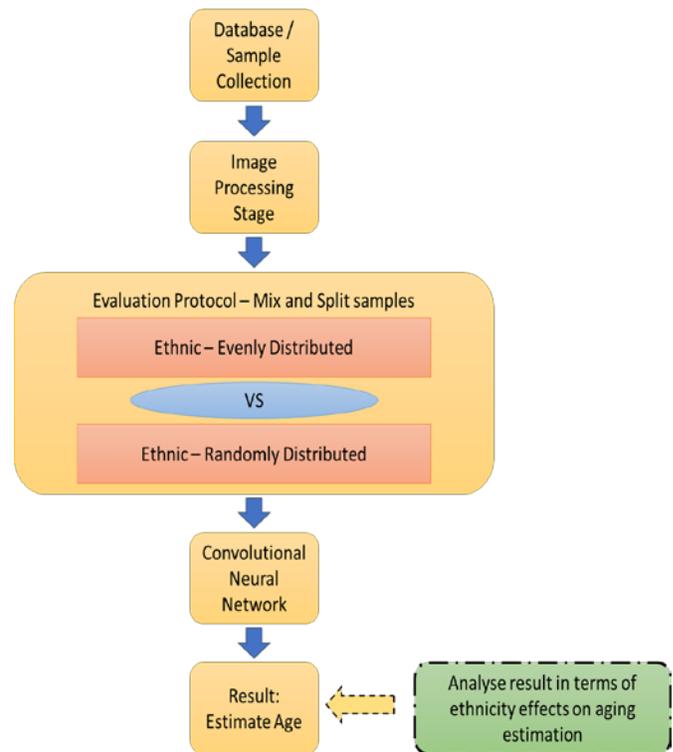


Fig. 7. Framework for Organising Different Racial Samples for CNN Face Age Estimation Study - Subject's Ethnicity Accounted for.

Although many public databases are available for face age estimation studies, very few are non-Caucasian/non-White databases. Accordingly, two suggestions could enable the collection of more ethnic-specific samples; either for private or public use:

1) Organise an ethnic-biased age estimation contest and develop an ethnic-specific dataset like how it was done for the ICCV 2015 ChaLearn [32] challenge dataset. This approach can help increase ethnic-biased age estimation studies from contestants and ethnic-specific database usage (e.g. AFAD, IMFDB, and Iranian database). These databases may become benchmark databases themselves later on.

2) In dangerous times, such as the current COVID-19 pandemic, most work and communication are now done online. Governments, businesses, educational institutions, medical institutions, and others are now using communication platforms for videoconferencing, online meetings, workspace chat, online classes, and even file sharing. One of the communication platform's primary functions is video streaming, which can accommodate up to nine people (or more) concurrently. This video streaming function allows researchers to organise a video conference for a group of volunteers to collect ethnic-specific samples for face age studies by capturing volunteers' face images during video streaming. Researchers must first decide whether to collect samples in a controlled/uncontrolled environment, for example, by requesting volunteers to standardise their background colours (use one colour) and stand still while researchers prepare to capture their faces (controlled

environment). Researchers must also decide whether to capture a single face image or multiple faces at once. However, the size and quality of the faces in the video may differ between users. Therefore, this should be taken into account when trying to use this approach to collect samples from volunteers, which can be co-workers or students (if the researcher is also an educator). Moreover, additional information about volunteers, such as their age and ethnicity, can be directly requested and recorded for research purposes. Microsoft Teams, Zoom, and Google Meet are some of the communication platforms that are available for use. Fig. 8 shows an example of captured face images using Microsoft Teams (single face or multiple faces).

When collecting samples, likely, some people would not be willing to help or give any personal information. Therefore, proper planning on target subjects selection before collecting their face images is required.



Fig. 8. Samples of Face Images - Captured using Microsoft Teams (Images taken from [97]).

It would be interesting to develop the suggested model framework with different ethnic races for face age recognition. Significant racial traits might be discovered, which can further distinguish the aging processes between different ethnic people. This discovery could further improve the understanding of racial aging traits, particularly concerning the face and the development of a model that can learn and identify those traits. Additionally, using the suggested sample collection method to collect and capture own samples may help ease the collection process. Aside from face age studies, the collected face images/samples can also be used for other facial image studies, such as emotion recognition and ethnic recognition. These suggestions, however, are beyond the scope of this study and will be considered in future research.

VIII. CONCLUSION

The analysis in this paper focused on ethnic consideration in the dataset used for the last six years for accurate age estimation using the deep learning approach. This paper specifically analysed 53 papers on deep learning face age estimation, model performance, selected databases, and whether or not any face ethnicity traits analysis was performed when estimating age. This paper also highlighted 19 database papers that promote the use of publicly available databases for age estimation research, as well as information on multiple database ethnicities. Although the deep learning approach improves face age estimation over time, it can be further enhanced by understanding how ethnicity affects face age estimation and designing an evaluation protocol that takes the subjects' ethnic traits into account. Moreover, a sizeable multi-racial database is needed for the investigation of aging in different ethnic groups. Therefore, it is crucial to collect the necessary information to create an extensive database with well-distributed age and ethnic labels. Suggestions for capturing samples were also provided to help researchers in increasing their ethnic-specific samples for private or public use.

IX. FUTURE DIRECTION

Making the collected ethnic-specific samples public and sharing them via web image collection sites can increase interest in conducting more ethnicity-based face age estimation research. More robust deep learning face age estimation models can be developed by performing more such studies, sample collection, and analyses in the future. Future research could also discover significant racial traits that could help distinguish unique aging patterns used to solve racial face age estimation problems in real-life applications. Proper planning and key considerations must be made when collecting samples, such as ensuring personal data privacy or a subject's consent. Additionally, it would be good to reiterate the benefit of having more samples for studies beyond facial age recognition.

ACKNOWLEDGMENT

The authors are grateful to the Faculty of Information Science and Technology, The National University of Malaysia, for supporting and contributing to this study; under the grant code GGPM-2019-038.

REFERENCES

- [1] N. A. Vashi, M. B. D. C. Maymone, & R. V. Kundu, "Aging differences in ethnic skin," *The Journal of clinical and aesthetic dermatology*, vol. 9, no. 1, pp. 31, 2016.
- [2] Y. Shirakabe, Y. Suzuki, & S. M. Lam, "A new paradigm for the aging Asian face," *Aesthetic plastic surgery*, vol. 27, no. 5, pp. 397-402, 2003.
- [3] A. E. Brissett, & M. C. Naylor, "The aging african-american face," *Facial Plastic Surgery*, vol. 26, no. 2, pp. 154-163, 2010.
- [4] M. O. Harris, "The aging face in patients of color: Minimally invasive surgical facial rejuvenation—A targeted approach," *Dermatologic therapy*, vol. 17, no. 2, pp. 206-211, 2004.
- [5] M. G. Rhodes, "Age estimation of faces: A review," *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, vol. 23, no. 1, pp. 1-12, 2009.
- [6] M. S. Zimble, M. S. Kokoska, & J. R. Thomas, "Anatomy and pathophysiology of facial aging," *Facial plastic surgery clinics of North America*, vol. 9, no. 2, pp. 179-87, 2001.
- [7] N. Ramanathan, R. Chellappa, & S. Biswas, "Age progression in human faces: A survey," *Journal of Visual Languages and Computing*, vol. 15, 3349-3361, 2009.
- [8] G. Y., Guo, & T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 11, pp. 1955-1976, 2010.
- [9] R. Russell, et al. "Facial contrast is a cue for perceiving health from the face," in *Journal of Experimental Psychology: Human Perception and Performance*, vol. 42, no. 9, pp. 1354, 2016.
- [10] C. Trojahn, G. Dobos, A. Lichterfeld, U. Blume-Peytavi, & J. Kottner, "Characterizing facial skin aging in humans: disentangling extrinsic from intrinsic biological phenomena," in *BioMed research international*, 2015.
- [11] M. S. Zimble, M. S. Kokoska, & J. R. Thomas, "Anatomy and pathophysiology of facial aging," *Facial plastic surgery clinics of North America*, vol. 9, no. 2, pp. 179-87, 2001.
- [12] T. Igarashi, K. Nishino, & S. K. Nayar, "The appearance of human skin: A survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 1, pp. 1-95, 2007.
- [13] N. A. Vashi, M. B. D. C. Maymone, & R. V. Kundu, "Aging differences in ethnic skin," *The Journal of clinical and aesthetic dermatology*, vol. 9, no. 1, pp. 31, 2016.
- [14] Stock photo caucasian baby. 2021. Crello. [online] Available at: <<https://crello.com/unlimited/stock-photos/166742418/stock-photo-caucasian-baby/>> [Accessed 22 July 2021].
- [15] A. Bastanfard, M. A. Nik, & M. M. Dehshibi, "Iranian face database with age, pose and expression," in *Machine Vision*, pp. 50-55, 2007.
- [16] S. Setty, et al., "Indian movie face database: a benchmark for face recognition under wide variations," in *2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pp. 1-5, December, 2013.
- [17] Z. Niu, M. Zhou, L. Wang, X. Gao, & G. Hua, "Ordinal regression with multiple output cnn for age estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4920-4928, 2016.
- [18] Y. H. Kwon & N. da Vitoria Lobo. "Age classification from facial images," *Computer vision and image understanding*, vol. 74, no. 1, pp. 1-21, 1999.
- [19] R. Angulu, J. R. Tapamo, & A. O. Adewumi. "Age estimation via face images: a survey," *EURASIP Journal on Image and Video Processing*, 2018, no. 1, pp. 42, 2018.
- [20] O. F. Osman, & M. H. Yap. "Computational intelligence in automatic face age estimation: A survey," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 3, pp. 271-285, 2018.
- [21] P. J. Phillips, H. Wechsler, J. Huang, & P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and vision computing*, vol. 16, no. 5, pp. 295-306, 1998.
- [22] A. Lanitis, C. J. Taylor, & T. F. Cootes, "Toward automatic simulation of aging effects on face images," in *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 4, pp. 442-455, 2002.
- [23] M. Minear, & D. C. Park, "A lifespan database of adult facial stimuli," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 4, pp. 630-633, 2004.
- [24] P. J. Phillips, et al., "Overview of the face recognition grand challenge," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, Vol. 1, pp. 947-954, June, 2005.
- [25] K. Ricanek, & T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *IEEE 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 341-345, April, 2006.
- [26] Y. Fu & T. S. Huang, "Human age estimation with regression on discriminative aging manifold," in *IEEE Transactions on Multimedia*, vol. 10 no. 4, pp. 578-584, 2008.
- [27] A. C. Gallagher, & T. Chen, "Understanding images of groups of people," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 256-263, June, 2009.
- [28] N. C. Ebner, M. Riediger, & U. Lindenberger, "FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behavior research methods*, vol. 42, no. 1, pp. 351-362, 2010.
- [29] S. Zheng, "Visual image recognition system with object-level image representation," *Doctoral dissertation*, 2012.
- [30] E. Eiding, R. Enbar, & T. Hassner, "Age and gender estimation of unfiltered faces," in *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170-2179, 2014.
- [31] B. C. Chen, C. S. Chen, & W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *European conference on computer vision*, Springer, Cham, pp. 768-783, September, 2014.
- [32] S. Escalera, et al., "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, 2015.
- [33] S. Escalera, et al., "Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1-8, 2016.
- [34] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, & S. Zafeiriou, "Agedb: the first manually collected, in-the-wild age database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 51-59, 2017.
- [35] R. Rothe, R. Timofte, & L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," in *International Journal of Computer Vision*, vol. 126, no. 2-4, pp. 144-157, 2018.
- [36] E. Agustsson, et al., "Apparent and real age estimation in still images with deep residual regressors on appereal database," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 87-94, May, 2017.
- [37] V. Carletti, A. Greco, G. Percannella, & M. Vento, "Age from faces in the deep learning revolution," in *IEEE transactions on pattern analysis and machine intelligence*, 2019.
- [38] X. Liu, et al. "Agenet: Deeply learned regressor and classifier for robust apparent age estimation," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 16-24, 2015.
- [39] R. Rothe, R. Timofte, & L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 10-15, 2015.
- [40] Y. Zhu, Y. Li, G. Mu, & G. Guo, "A study on apparent age estimation," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 25-31, 2015.
- [41] Z. Kuang, C. Huang, & W. Zhang, "Deeply learned rich coding for cross-dataset facial age estimation," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 96-101, 2015.
- [42] X. Yang, et al. "Deep label distribution learning for apparent age estimation," in *Proceedings of the IEEE international conference on computer vision workshops*, pp. 102-108, 2015.

- [43] X. Wang, R. Guo, & C. Kambhampettu, "Deeply-learned feature for age estimation," in 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 534-541, January, 2015.
- [44] S. Li, J. Xing, Z. Niu, S. Shan, & S. Yan, "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, 2015.
- [45] I. Huerta, C. Fernández, C. Segura, J. Hernando, & A. Prati, "A deep analysis on age estimation," Pattern Recognition Letters, vol. 68, pp. 239-249, 2015.
- [46] G. Levi, & T. Hassner, "Age and gender classification using convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 34-42, 2015.
- [47] R. Ranjan, S. Zhou, J. Cheng Chen, A. Kumar, A. Alavi, V. M. Patel, & R. Chellappa, "Unconstrained age estimation with deep convolutional neural networks," in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 109-117, 2015.
- [48] J. C. Chen, A. Kumar, R. Ranjan, V. M. Patel, A. Alavi, & R. Chellappa, "A cascaded convolutional neural network for age estimation of unconstrained faces," in 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1-8, September, 2016.
- [49] F. Gurpinar, H. Kaya, H. Dibeklioglu, & A. Salah, "Kernel ELM and CNN based facial age estimation," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 80-86, 2016.
- [50] Z. Niu, M. Zhou, L. Wang, X. Gao, & G. Hua, "Ordinal regression with multiple output cnn for age estimation," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4920-4928, 2016.
- [51] R. Rothe, R. Timofte, & L. Van Gool, "Some like it hot-visual guidance for preference prediction," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5553-5561, 2016.
- [52] Y. Yang, F. Chen, X. Chen, Y. Dai, Z. Chen, J. Ji, & T. Zhao, "Video system for human attribute analysis using compact convolutional neural network," in 2016 IEEE International Conference on Image Processing (ICIP), pp. 584-588, September, 2016.
- [53] Y. Dong, Y. Liu, & S. Lian, "Automatic age estimation based on deep learning algorithm," Neurocomputing, vol. 187, pp. 4-10, 2016.
- [54] G. Ozbulak, Y. Aytar, & H. K. Ekenel, "How transferable are CNN-based features for age and gender classification?" in 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1-6, September, 2016.
- [55] G. Antipov, M. Baccouche, S. A. Berrani, & J. L. Dugelay, "Apparent age estimation from face images combining general and children-specialized deep learning models," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 96-104, 2016.
- [56] Z. Huo, X. Yang, C. Xing, Y. Zhou, P. Hou, J. Lv, & X. Geng, "Deep age distribution learning for apparent age estimation," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 17-24, 2016.
- [57] M. Uricár, R. Timofte, R. Rothe, J. Matas, & L. Van Gool, "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, 2016.
- [58] R. Can Malli, M. Aygun, & H. Kemal Ekenel, "Apparent age estimation using ensemble of deep learning models," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 9-16, 2016.
- [59] B. Hebda, & T. Kryjak, "A compact deep convolutional neural network architecture for video based age and gender estimation," in 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 787-790, September, 2016.
- [60] Z. Hu, Y. Wen, J. Wang, M. Wang, R. Hong, & S. Yan, "Facial age estimation with age difference," in IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3087-3097, 2016.
- [61] Z. Tan, J. Wan, Z. Lei, R. Zhi, G. Guo, & S. Z. Li, "Efficient group-n encoding and decoding for facial age estimation," in IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 11, pp. 2610-2623, 2017.
- [62] R. Ranjan, S. Sankaranarayanan, C. D. Castillo, & R. Chellappa, "An all-in-one convolutional neural network for face analysis," in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 17-24, May, 2017.
- [63] H. Liu, J. Lu, J. Feng, & J. Zhou, "Ordinal deep learning for facial age estimation," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 2, pp. 486-501, 2017.
- [64] H. Liu, J. Lu, J. Feng, & J. Zhou, "Group-aware deep feature learning for facial age estimation," in Pattern Recognition, vol. 66, pp. 82-94, 2017.
- [65] G. Antipov, M. Baccouche, S. A. Berrani, & J. L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," in Pattern Recognition, vol. 72, pp. 15-26, 2017.
- [66] J. Xing, K. Li, W. Hu, C. Yuan, & H. Ling, "Diagnosing deep learning models for high accuracy age estimation from a single image," in Pattern Recognition, vol. 66, pp. 106-116, 2017.
- [67] K. Li, J. Xing, W. Hu, & S. J. Maybank, "D2C: Deep cumulatively and comparatively learning for human age estimation," in Pattern Recognition, vol. 66, pp. 95-105, 2017.
- [68] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, & J. H. Saltz, "Convnets with smooth adaptive activation functions for regression," in Proceedings of machine learning research, vol. 54, pp. 430, 2017.
- [69] K. Zhang, et al., "Age group and gender estimation in the wild with deep ror architecture," IEEE Access, vol. 5, pp. 22492-22503, 2017.
- [70] F. Wang, H. Han, S. Shan, & X. Chen, "Deep multitask learning for joint prediction of heterogeneous face attributes," in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 173-179, May, 2017.
- [71] H. Liu, J. Lu, J. Feng, & J. Zhou, "Label-sensitive deep metric learning for facial age estimation," IEEE Transactions on Information Forensics and Security, vol. 13, no. 2, pp. 292-305, 2017.
- [72] H. Han, A. K. Jain, F. Wang, S. Shan, & X. Chen, "Heterogeneous face attribute estimation: A deep multitask learning approach," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 11, pp. 2597-2609, 2017.
- [73] J. Wan, Z. Tan, Z. Lei, G. Guo, & S. Z. Li, "Auxiliary demographic information assisted age estimation with cascaded structure," IEEE transactions on cybernetics, vol. 48, no. 9, pp. 2531-2541, 2018.
- [74] M. Duan, K. Li, & K. Li, "An Ensemble CNN2ELM for Age Estimation," IEEE Transactions On Information Forensics And Security, vol. 13, no. 3, 2018.
- [75] H. F. Yang, B. Y. Lin, K. Y. Chang, & C. S. Chen, "Joint estimation of age and expression by combining scattering and convolutional networks," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 14, no. 1, pp. 9, 2018.
- [76] B. Yoo, Y. Kwak, Y. Kim, C. Choi, & J. Kim, "Deep facial age estimation using conditional multitask learning with weak label expansion," IEEE Signal Processing Letters, vol. 25, no. 6, pp. 808-812, 2018.
- [77] S. Taheri, & Ö. Toygar, "On the use of DAG-CNN architecture for age estimation with multi-stage features fusion," Neurocomputing, vol. 329, pp. 300-310, 2019.
- [78] W. Im, S. Hong, S. E. Yoon, & H. S. Yang, "Scale-Varying Triplet Ranking with Classification Loss for Facial Age Estimation," in Computer Vision – ACCV 2018. Lecture Notes in Computer Science, vol. 11365. Springer, Cham, 2018.
- [79] O. Sendik, and Y. Keller. "DeepAge: Deep Learning of face-based age estimation," in Signal Processing: Image Communication, vol. 78, pp. 368-375, 2019.
- [80] S. Penghui, L. Hao, W. Xin, Y. Zhenhua, & S. Wu. "Similarity-aware deep adversarial learning for facial age estimation," in 2019 IEEE International Conference on Multimedia and Expo (ICME), pp. 260-265, July, 2019.
- [81] C. Miron, V. Manta, R. Timofte, A. Pasarica, & R. I. Ciucu, "Efficient convolutional neural network for apparent age prediction," in 2019 IEEE

- 15th International Conference on Intelligent Computer Communication and Processing (ICCP), pp. 259-262, September, 2019.
- [82] N. Savov, M. Ngo, S. Karaoglu, H. Dibeklioglu, & T. Gevers, "Pose and Expression Robust Age Estimation via 3D Face Reconstruction from a Single Image," in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 0-0, 2019.
- [83] H. Liu, J. Lu, J. Feng, & J. Zhou. "Ordinal Deep Learning for Facial Age Estimation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 2, pp. 486-501, 2019.
- [84] H. Liu, P. Sun, J. Zhang, S. Wu, Z. Yu, & X. Sun. "Similarity-Aware and Variational Deep Adversarial Learning for Robust Facial Age Estimation," IEEE Transactions on Multimedia, 2020.
- [85] P. Li, Y. Hu, X. Wu, R. He, & Z. Sun. "Deep label refinement for age estimation," Pattern Recognition, vol. 100, p. 107178, 2020.
- [86] N. Liu, F. Zhang, & F. Duan. "Facial Age Estimation Using a Multitask Network Combining Classification and Regression," IEEE Access, vol. 8, pp. 92441-92451, 2020.
- [87] X. Liu, Y. Zou, H. Kuang, & X. Ma. "Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network," Symmetry, vol. 12, no. 1, pp. 146, 2020.
- [88] J. C. Xie, & C. M. Pun. "Deep and Ordinal Ensemble Learning for Human Age Estimation From Facial Images," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2361-2374, 2020.
- [89] M. Xia, X. Zhang, L. Weng, & Y. Xu. "Multi-Stage Feature Constraints Learning for Age Estimation," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2417-2428, 2020.
- [90] G. B. Huang, M. Mattar, T. Berg, & E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, October, 2008.
- [91] Z. Liu, P. Luo, X. Wang, & X. Tang, "Deep learning face attributes in the wild," in Proceedings of the IEEE international conference on computer vision, pp. 3730-3738, 2015.
- [92] O. Russakovsky, et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, vol. 115, no. 3, pp. 211-252, 2015.
- [93] A. Khan, A. Sohail, U. Zahoora, & A. S. Qureshi. "A survey of the recent architectures of deep convolutional neural networks," Artificial Intelligence Review, vol. 53, no. 8, 5455-5516, 2020.
- [94] K. Ricanek, Y. Wang, C. Chen, & S. J. Simmons, "Generalized multi-ethnic face age-estimation," in 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, pp. 1-6, September, 2009.
- [95] J. D. Akinyemi, & O. F. Onifade, "An ethnic-specific age group ranking approach to facial age estimation using raw pixel features," in 2016 IEEE Symposium on Technologies for Homeland Security (HST), pp. 1-6, May, 2016.
- [96] M. Shin, J. H. Seo, & D. S. Kwon, "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), pp. 567-572, 2017.
- [97] Microsoft.com. 2021. Online Meeting Software, Video Conferencing | Microsoft Teams. [online] Available at: <<https://www.microsoft.com/en-my/microsoft-teams/online-meetings>> [Accessed 3 May 2021].