Facial Age Estimation Using Deep Learning: A Review

Osekhonmen V. Abhulimen
Elizade University

Facial Age Estimation Using Deep Learning: A Review

Osekhonmen Abhulimen
Department of Electrical and Electronics Engineering
Federal University of Technology, FUTA,
Akure, Nigeria.
bulioses@gmail.com

Erastus Ogunti
Department of Electrical and Electronics Engineering
Federal University of Technology, FUTA,
Akure, Nigeria.
oguntig@gmail.com

***Abstract*— In recent times, there have been a surge in the number of researchers, adopting ways and methods in which Artificial Intelligence (AI) can be achieved even with better accuracy than humans and previous handcrafted methods by using deep learning due to its superior performance. Facial age estimation is quite a difficult one and not until recent years that quality results are being obtained by using Deep Learning; and hence it is becoming a crucial system for biometric security framework and has a host of other applications in the real-world. The use of deeply learned Convolutional Neural Networks (CNNs) algorithms has aided in enhancing the accuracy of automatic facial age estimation. In this paper, an in-depth review of research efforts on facial age estimation using deep learning is carried out. Details on other handcrafted techniques adopted by Researchers till date for automatic facial age estimation are also included in order to give an overall idea to interested Researchers. CNNs and their models are reviewed; data sets used, major contributions, performance evaluation metrics, their results and future areas of research in this field as well as challenges are reported.**

Keywords*— *Automatic Facial Age Estimation (AFAE); Convolutional Neuraal Networks (CNNs); Deep Learning; Feature Extraction.

I. INTRODUCTION

Deep learning, a subset of Machine learning, has been extensively used in recent times for Artificial intelligence (AI). Artificial intelligence (AI) which is the capability of machines to imitate intelligence in human behavior is attained by examining how the human brain think and how they learn, decide and work in the process of solving a specific problem. The outcomes of this study is what is been used as a basis of developing and designing intelligent systems that would be used for facial age estimation. It is also applied in speech recognition, understanding natural language, image recognition, etc.

Face remains one of the most prominent and informative traits used for biometric recognition systems. The facial appearance of a person is affected mostly by increase in age, and the effects of facial aging are believed to be majorly credited to the movement of bones, growth and skin related deformations associated with the introduction of wrinkles and reduction of muscle strength [2, 70]. Bone growth usually occur during childhood, while at adulthood, the most profound age-related changes are associated with texture changes observed mostly on the face. The observation of aging-related features on the face enables us to predict the age of others just by mere staring at their faces. Consequently, systems to estimate age, have been developed, but the process of ageing is not quite linear, hence it is highly complicated and it differs for different persons [61].

Applications of Automatic Facial Age Estimation (AFAE) systems can be seen in;

(a) HCI – (Human–Computer Interaction),
(b) web content filtering and surveillance,
(c) retrieval of information - where it permits the retrieval of image from a large database of images, by the process of age image query. For example, a website known as “Flickr.com” contains a very large number of facial images, which are uploaded by constant users of varying ages. It can take advantage of automatic facial age estimation application in search of friends and identification [6], and in;
(d) E-CRM (Electronic Customer Relationship Management) systems [61].

Humans are very likely to fail in accurately carrying out age estimation, hence, it is very vital to develop Automatic Facial Age estimation (AFAE) systems that could outperform human performance [61]. Enhancing the ability of machines to recognize and interpret facial images has helped to ameliorate the interaction between humans and machines and has also been very useful in systems to identify faces, genders, ethnicity, predict ages and perceive emotions.

In this paper, we aim at reviewing the most acknowledged researches about facial age estimation technique using deep learning while presenting the challenges found, performance evaluation and popular datasets. Section 2 gives a review of Facial Age Estimation, reviews of Deep learning approach and models and popular datasets used for AFAE are in sections 3 and 4 respectively. Section 5 reviews the performance evaluation measures used for facial age estimation problems. We present in section 6 review on AFAE using handcrafted and deep learning methods. Finally, section 7 presents the conclusion and insight in probable future research.

II. FACIAL AGE ESTIMATION

Aging is referred to as an inevitable, uncontrollable, irreversible and stochastic process that leads to variations in facial texture and shape [93]. Though aging is believed to be a stochastic process with each individual having varying patterns of aging, there are some common similarities and variations that could be modelled to estimate the age of an individual [2]. The formative stage also known as childhood phase and adulthood or aging phase are two different phases in the life of humans that are apparent with respect to facial growth [53]. Aging brings about a notable change in the shape of a face during the formative years and a somewhat large texture variation with small changes in shape in older age groups [28, 63]. Craniofacial growth is the major causes of shape variations in younger age groups. It has been shown by craniofacial studies that the faces of humans transition from circular to oval as they age [68]. These changes gave way to alterations in the position of fiducial landmarks [21]. In the course of craniofacial development, the forehead slopes back and releases space on the cranium. The mouth, ears, nose, and eyes will increase to cover interstitial space already created. The chin becomes more protrusive as the cheeks extend [2]. According to [103], as one increases in age, facial blemishes like freckles, age spots and wrinkles begin to develop. Beneath the skin, the cells that produces melanin are damaged as a result of exposure to UV (ultraviolet) rays from the sun. Age spots and freckles appear due to melanin overproduction. In consequence, light reflecting collagen becomes less-uniformly distributed and also decreases, hence making facial skin tone to become non-uniform. Parts affected majorly by sunlight are the nose-bridge, nose, forehead and upper cheek. As the skin gets older biologically, collagen beneath it is lost [28]. The effect of gravity and loss of collagen make the skin to become leathery, darker, less elastic and thinner. Wrinkles and facial spots then begin to gradually appear. The schema of bones underneath the skin may also begin to depreciate leading to increased variations in skin textures and development of wrinkles. These variations in texture and shape across ages are what is modelled and implemented for the automatic estimation of human ages.

According to [31] facial aging has 3 specific attributes;

(a) Aging is uncontrollable and inevitable. No human can delay, advance or avoid it. The process of aging is quite slow, but irreversible.

(b) Aging patterns are personalized, i.e., humans age differently. The pattern of aging of an individual is largely reliant on his or her genetic makeup and also various extrinsic factors like lifestyle, environmental conditions and health.

(c) Facial variations caused by aging as well as achieved aging patterns are not permanent.

According to [67], till date it is impossible to accurately predict age due to its reliance on accurate features, hence posing a huge challenge to automatic facial age estimation systems. Also, the availability of adequate databases that contains all age as well as gender annotation for carrying out rigorous research is limited. Factors like camouflage due to beard, glasses and moustache amplify the struggles faced by AFAE systems. Diet, skin infections, skin texture, cosmetic makeup, face shape and skin texture changes during adulthood and childhood also affect AFAE systems. Facial expressions like laughing, frowning, smiling and crying may bring about wrinkle lines on some parts of the face and hence can distort the accuracy of facial age estimation. Real life environmental problems of occlusion, pose, lightning, race or ethnicity and colour mode also affect the task of age estimation. Gender is also another factor that affects facial age estimation. Some other crucial factors include; living environment, lifestyle, health issues, genes and occupation. These factors affecting the process of facial ageing are classified as intrinsic and extrinsic factors. While Intrinsic factors occur inside the human body, extrinsic factors happen outside.

It is important to note that there are three major aspects centred on facial aging researches. They include:

(i) Age Invariant Face Recognition. This helps to identify human faces irrespective of the aging changes that appears on it. Its application is seen in the real world for biometric authentication and for regeneration of drivers' license or passport [6].

(ii) Age Synthesis. This helps in predicting the future appearance of a facial image at a particular age [89]. It can be applied in the identification of victim in police duty, and in the simulation of the aging features of missing persons from old images.

(iii) Facial Age estimation (which we are mostly concerned with) - It is used in estimating the precise age or age-group of a person from their facial image [5].

An efficient automatic facial age estimation system will have a vast range of application in areas such as; age-invariant face recognition system, security control as well as surveillance, commercial and law enforcement areas, face verification across several ages, biometrics, age-based image retrieval, entertainment centres, electronic customer relationship management and human computer interaction.

III. DEEP LEARNING APPROACH AND MODELS

Since the 1990s, significant progress has been made in the field of computer vision [98]. In recent times, traditional face recognition as well as age estimation methods have been dominated by deep learning methods that are based on CNNs. The major advantage of deep learning method with CNNs is that they can be trained with very large dataset in order to learn the best features to use in representing the data. Expeditious progress have been actuated as a result of the increasing affordability of powerful Graphic processing Units (GPUs) [99] and the advancements in convolutional neural network architectures, which is centred on real world applications. Deep CNNs have been predicted to encompass industrial application and future research, and they are presently being implemented by large corporations like Facebook, Microsoft and Google [43].

A CNN also known as CovNet is an artificial neural network that has so far been widely deployed for the analysis of images. They can also be deployed for other classification or data analysis problems as well. Deep learning with CNNs is presently one of the most saliant methods used for deep learning involving image data. While in traditional machine learning approach important features have to be manually extracted, deep learning approach makes use of raw images which serves as input to learn relevant features. A CNN is made up of an input and output layer, and a number of hidden layers in-between. The layers in between are the convolutional, max-pooling, and fully connected layers. The architectures of CNNs may vary in the number and type of layers employed for a particular task. For categorical feedbacks, the network must consist of a classification function and layer, whereas for continuous feedback, the system must consist of a regression layer at the end of a network. The neurons in each layer of a CNN are usually aligned in a 3-dimensional (3D) fashion, and transform a 3D output from a 3D input [74].

The configurations of CNNs comprise of a number of hidden layers. In each of these layers, activation volumes are transformed with the use of differentiable functions. Four major layers exist, which are used in building CNN configurations;

1. **Convolutional Layer (ConV)**: This layer helps to derive an activation map from the input data.
2. **Rectified Linear Unit Layer (ReLU)**: Here, negative values are filtered in order to provide positive values for a faster training time.
3. **Pooling Layer (POOL)**: This layer performs the function of nonlinear down-sampling and helps to reduce the number of parameters for a much simpler output.
4. **Fully Connected Layer (FC)**: This layer helps to compute the class probability scores by producing a vector of N dimensions as output, with N being the number of classes. All the neurons in a CNN are connected to this layer.

A SoftMax layer performs the duty of allotting decimal probabilities to each outputted neuron, and it is often serves as the final layer.

An example of a CNN is shown in Fig. 1.

One major difficulty that usually emanate with training of CNNs is the large number of parameters needed to be learned, which may result to the issue of over-fitting. To overcome this, methods such as dropout, data augmentation and stochastic pooling have been introduced. Additionally, CNNs are often subjugated to pre-training, which is a process used to initialize the network with pre-trained parameters rather than randomly set parameters. Pre-training can impel the learning process of the network and also improve its generalization capability.

In general, CNNs have shown to supersede traditional handcrafted machine learning methods in a variety of pattern recognition and computer vision tasks [9]. Their exceptional performance in combination with the relative ease in training them are the major reasons that explain the rise in their popularity for facial age estimation problems over the last few years

Depending on this basic concept of facial age estimation, Researchers have proposed many deep learning models based on Convolutional Neural Networks [73], to aid in carrying out their research ideas.

State-of-the-art and popular CNN architectures like AlexNet, GoogLeNet, VGG16, ResNet, SqueezeNet, DAG-CNN and Xception are briefly reviewed in the following subsections.

A. AlexNet

This convolutional neural network (CNN), was developed by Alex Krizhevsky in 2012 [43]. It famously won the 2012 ImageNet LSVRC-2012 competition with a top-5 error rate of 15.3%. It consists of 8 layers in total, 5 of which are convolutional layers, which is usually accompanied by Max-Pooling layers; while the last 3 layers are the fully connected (FC) layers. After each convolutional and FC layer, a ReLU activation layer is attached. Also attached at the end of each fully connected layer is the dropout layer, which helps to tackle the problem of overfitting. The architecture for AlexNet is shown in fig. 2.

B. VGGNet

The VGG architecture was designed by Oxford Visual Geometry Groups (VGG) [75]. It behaves just like a conventional CNN with significantly increased layers. Its architectural form is made up of 16 convolutional layers (VGG-16) (which could be up to 19 (VGG-19)). 5 of these layers are hidden layers of max-pooling which uses a ReLU activation function, assigned with the responsibility of providing nonlinearity and spatial pooling. Additional non-linearity is introduced to the decision function by the presence of 1×1 convolution filters. The spatial padding is retained after convolution in some layers. VGG-16 uses three fully-connected layers with similar arrangement and a softmax layer as the output. The architecture for VGG-16 is shown in fig. 3. The total number of parameters used in this large scale CNN architecture is 138, and still it is predominantly used for research purposes, due to its uniformity and simplicity. It achieved a top 5 error of 7.32% in the ILSVR (ImageNet) competition in 2014, which is much reduced in comparison to that of AlexNet.

C. GoogLeNet

GoogLeNet is a 22-layer deep convolutional neural network designed by [9]. It is a variant of the Inception Network, but with deeper and wider layers. It is made up of 9 Inception modules joined together for the purpose of attaining a deeper architecture. A number of the inception modules are usually accompanied by Max-pooling layers, which helps to minimize the volume of the parameter. GoogLeNet makes use of global average pooling instead of fully-connected (FC) layers found in previous CNN architectures; this helps to reduce the weight size. Unlike AlexNet and VGGNet, the size of convolution for each layer is not fixed for GoogLeNet. In GoogLeNet as shown in fig. 4, every inception module consists of a few convolutional kernels, which are of sizes; 1×1, 3×3 and 5×5. The 1×1 convolutional layers are responsible for feature dimensionality reduction whilst moving further into the network. The GoogLeNet architecture performs better than the AlexNet and VGGNet architectures in regards to top 5 error which was minimized to 6.67% [67]. This architecture was the winner at the ILSVRC 2014 image classification challenge. In 2015, [32] proposed a more improved version of Inception (Inception V3), to enhance the performance on “ImageNet” classification accuracy.

D. DAG-CNN

The DAG-CNN known as Directed Acyclic Graph-Convolutional Neural Network was designed by Yang and Ramanan in 2015 [87]. It is built in an end-to-end pattern as shown in fig. 5. Just like traditional CNNs, DAG-CNN is made up of Convolution (conv), Multi-output (ReLU), Normalization (Norm) and Pooling layers, which are joined together and referred to as a chain-structure backbone. The backbone of the DAG-CNN architecture is the ReLU layer which is allied to the average pooling layer, then accompanied by the Normalization layers which is finally sent to a fully connected layer. Summed together and finally sent as input to the SoftMax output function, is a number (*N*) of such output scores from the different layers in order to produce the final predictions [67].

E. ResNet

In order to further enhance the performance of the already existing architectures (AlexNet, VGGNet, GoogLeNet and DAG-CNN), [39] developed the Residual Network (ResNet). It was the winner of the ILSVC 2015 challenge. The ResNet architecture brought about the idea of residual module and identity mappings not found in previous architectures. Like in the GoogLeNet architecture, ResNet uses global average pooling instead of fully-connected layers, which makes the model size to be smaller. ResNets work majorly on the concept of skip connections which allows going deeper into the network architecture with ease. A ResNet module may be designed by making use of identity mapping which will create a pathway between the input layer as well as the output layer and then skipping some of the layers that are in between. Other model types of ResNet examined till date are the ResNet-18, 34, 50, 101 and ResNet-152. Best performance on the top five error was obtained by the ResNet50 model with a value of 4.49%, which outperforms that of Inception and VGG-16. Results obtained for automatic age estimation and other classification tasks using ResNet are found to be more accurate when compared with others. The architecture of ResNet-50 is shown in fig. 6 for a better knowledge of its framework.

F. SqueezeNet

In 2017, [41] introduced a lighter CNN architecture (when compared with other previous architectures) called SqueezeNet that achieved an “AlexNet level” of accuracy on ImageNet, having about 50% less parameters. With the help of model compression techniques, it was possible to compress the model to less than 0.5MB, that is, about 510x smaller than the AlexNet model. Its advantages over other larger CNN architectures include; minimum communication within servers during distributed training, fewer bandwidth required to export a new model directly from the cloud, and ease of deploying on Field Programmable Gate Arrays (FPGAs) as well as other hardware with limited memory. The SqueezeNet model depend on an expansive and smaller phase of solely 1×1 and 3×3 convolutions, because they use fire module which helps in reducing the spatial volume size present in the network with its few amount of filters and the presence of global average pooling instead of fully connected (FC) layers. SqueezeNet architecture is shown in Fig. 7.

G. Xception Network

The Xception architecture (Extreme Inception) was presented by Chollet in 2017 [17]. It is a representation of Inception modules in CNNs as being an intermediary step between regular convolution and the depthwise separable convolution operation, that is, a depthwise convolution followed by a pointwise convolution. Unlike conventional convolution, convolution across all channels does not need to be performed, hence the number of connections are fewer and the model is usually light. The Xception network has 36 layers of convolution, which are assembled into 14 modules, all of which have residual connections, excluding the topmost and bottommost modules.

Summarily, the Xception network cam be described as a linear stack of depthwise separable convolution layers having residual connections. This enables the architecture to be easily defined and modified. Xception outperforms other architectures on the ImageNet, but it is slightly slower than the Inception modules. This architecture is shown in Fig. 8.

IV. DATASETS

For improved accuracy and efficiency of facial age estimation systems, the CNN models depends hugely on the type of dataset or database used for training and testing it. The required dataset should have minimum variability in conditions such as texture, lighting, pose, with an adequate number of facial images of individuals at different ages or age-groups. It should also comprise of different races/ethnicity and gender in order for it to be more expressive [78].

In this section, we review the popular datasets used majorly for automatic facial age estimation purposes.

- **Face and Gesture Recognition Network (FGNET)**

The FGNET age dataset is a facial image database made available to the public. It was collected in 2004 and contains 1002 colour and gray scale facial images of 82 different persons at different ages, which ranges from 0-69 years. However, there are more number of facial images between the ages of 0 and 40 years in the database. Data files which contains the locations of 68 facial landmarks and the age of the subject in each image are also available [45]. The database displays a considerable variability in illumination, resolution, expressions and viewpoint. The variability is because images were gotten by scanning pictures of individuals that were found in personal collections. This made some of the images present in the dataset to have occlusion problems of different forms.

- **Face Recognition Technology (FERET)**

This database was championed by the Department of Defence (DoD) Counterdrug Technology Development Program Office, under the Face Recognition Technology (FERET) program [65], with the aim of developing a large database for automatic face recognition and age estimation systems that may be helpful in assisting intelligence, lwa enforcement and security agencies in carrying out their duties. The database was collected between 1993 and 1996. It is made up of 1,564 sets of facial images for a total of 14,126 gray scale images, which contains 1,199 subjects and a duplicate of 365 sets of facial images. There is over 2 years difference between the first and last images, for some of the images in the dataset, with some subjects been photographed more than once. With this, the Researchers were able to study and examine the changes in the appearance of a subject over a period of one year. The images in the database were taken in a controlled environment which makes the quality of the images better, but with a variation of race, gender and pose.

- **People Action for Learning (PAL Network) Database**

PAL database [54], contains about 1,140 (225 males and 350 females) facial images with age range spanning from 19-93 years, and split into four different age group comprising of: 18–29 (218 images), 30–49 (76 images), 50–69 (123 images) and 70+ (158 images). It was assembled at the University of Michigan in order to aid facial aging researches. The images in the database represent a wide range of ages and have fine resolution, beneficial for the purpose of solving AFAE problems.

- **MORPH Database**

This is a publicly available face database, which was collected at the University of North Carolina suituated in Wilmington by the Face Aging Group [71]. It is of two variations: the MORPH longitudunal database and the Academic MORPH database.

The MORPH longitudunal database is presently the largest longitudinal facial recognition database in the world, with 400,000+ images of about 700,000 subjects and it was recently doubled in size. The longitudinal aspect of the database shows that there are multiple images of a given subject, over a period of time. It is licensed for commercial and developmental purposes only. The images are of 8-bit colour, and their sizes may be different. The database is annotated with labels for age, height, gender, weight, race and coordinates for eye.

The Academic MORPH dataset consists of two different Albums. Albulm 1 contains 1724 facial images of 515 subjects aged between 27 and 68; while Album 2 is made up of 55,134 facial images of about 13,000 subjects, having ages ranging between 16 and 80, with the average number of facial images per subject been four. Conversely, the gender and ethnicity distributions are very uneven in the database, with variations in expression, illumination and resolution.

- **3D Morphable Dataset**

The 3D morphable dataset [11] is a sturdy depiction for human faces in 3-Dimensions. The dataset is made up of 3-Dimensional scans of 200 adult faces (100 males females each) and 238 teenage faces (125 males and 113 females), aged between 8 and 16 years; with a resolution of 320x240 vertices per scan (320x240).

- **FACES Database**

This is a dataset that tells the facial expressions of younger, middle-aged and older males and females, that were conceived between the year 2005 and 2007 [24]. It was collected at the Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany. It contains facial images of about 171 persons having ages ranging from 19-80. Each subject has their faces in two sets of six facial expressions (sadness, neutrality, fear, anger, happiness and disgust), which results in 2,052 facial images in total.

- **CACD**

Cross-Age Celebrity Dataset (CACD) [15] consist of over 160,000 facial images of 2000 celebrities collected from the internet across a period of ten years (2004-2013). The age of subjects in the dataset ranges from 16 to 62 years.

- **OUI (Organization Unique Identifier)-Adience Database**

The OUI-Adience dataset [25] is a publicly known facial age estimation database only made available under the Creative Commons (CC) license. It consists of about 26,580 facial images of 2,284 subjects, cutting across eight age groups (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+) of different individuals. It is useful for experimental works on age, gender and label classification.

- **Waseda Human–Computer Interaction Technology (WIT-DB) database**

This database [82] contains facial images of about 5500 different subjects from Japan – 3000 males and 2500 females. It has about 26,222 total images with a contribution of 1 to 14 images per subject. The dataset is divided into 11 different age groups ranging from 3-85 years. The face images are of normal facial expression with unoccluded frontal views, and they show a number of degrees of variability in their appearance that reflect a 'real-world' image.

- **APPA-REAL Database**

The ChaLearn-Looking at People (LAP) Apparent and Real Age Estimation (APPA-REAL) facial image database [26] was developed in the year 2016. It consists of 7,591 facial images, showing their actual and apparent age labels. Human Observer's labelled the images. Their votes on each image are collected for collating the mean and the standard deviation in order to determine the final age label of the face image. The age range of subjects in the dataset is between 0-95 years, and the images were collected under different conditions.

- **UTK-Face Database**

The UTKFace database [83] is a large face dataset that contains about 24,000 facial images with a very wide age span ranging from 0-116 years. The images are labelled with annotations of age, gender and ethnicity. The dataset can be used for a variety of tasks involving facial age estimation, age progression/regression, face detection and landmark localization. The images cover large variation in pose, facial expression, facial expression, resolution, illumination and occlusion.

- **IMDB (Internet Movie Database)-WIKI Dataset**

The IMBD-WIKI dataset [73] is one of the largest publicly available dataset used for the task of age estimation of persons in the wild, It contains about 524,230 facial images with crawled age information ranging from 0 to 100 years. It consists of about 461,871 facial images of 100,000 very prominent Actors listed on the IMDB website, and 62,359 images of Actors from Wikipedia, using the same criteria for selection. Many low-quality images, like "human comic" images, severe facial mask, multi-person images, sketch images, full body images, blank images, etc; are present in the dataset. This is because they were obtained directly from the websites.

- **Asian Face Age Database (AFAD)**

The Asian Face Database also known as AFAD is a publicly available large dataset which was collected in 2016 [59]. It was proposed for evaluating the performance of age estimation. The dataset contains 164,432 face images of descents from Asia, with their corresponding age and gender labels. It is made up of 63,680 and 100,752 facial images of females and males respectively, with their ages ranging between 15 and 40. The images found in this database were gotten from the "RenRen social network (RSN)", mostly used by Asian students.

- **Iranian Face Database (IFDB)**

IFBD [7] is a small database collected by the Department of Engineering, University of Karaj, between September 2006 – January 2007, for the sole purpose of researches having to deal with age estimation, race detection, facial surgery among others. It contains about 3600 coloured facial images of 616 Iranians, consisting of 487 men and 129 women, having ages ranging between 2 and 85 years. The images in the database were captured in a controlled environment using a digital camera of high-resolution.

- **Lotus Hill Research Institute (LHI) face Database**

The Lotus Hill Research Institute face database [77] consist of 8,000 RGB facial images of Asian subjects, comprising of 4000 males and females with a single image per subject. The age range of subjects in the dataset is between 9-89, with an about 100 images per age. The resolution of the facial images are about 120x160 pixels, with a slight variations in pose and illumination.

- **The AI & R Asian database**

This database [27] was created for the purpose of rendering and synthesis of Asian faces. It is made up of four different databases which are: AI & R V1.0 used for expression, AI & R V2.0 used for aging, AI & R V3.0 used for viewing, and finally AI & R V4.0 used for Illumination. The AI & R V2.0 database that is used for facial age estimation consist of 34 facial images in total, with 17 individuals. The age range is between 22-61 years, The images have a colour depth of 24-bit with a dimension of 640x480 each.

- **AgeDB Database**

The AgeDB database [55] was collected for research purposes in age estimation, face age progression and age invariant face tasks. The face images were colected manually and they exhibit some levels of variations that are expected of images taken under uncontrolled, in-the-wild conditions. These variations in the images appear in expressions, pose and occlusions. The dataset is made up of 16,488 images of policians, actresses, actors and other popular individuals. The images are annotated with labels showing their age, identity and gender attribute. It contains a total of 568 different subjects with their ages spaning between the range of 1 and 101 years.

- **VADANA Database**

The VADANA database was introduced in 2011 [76]. It consists of 2298 facial images of 43 different subjects of faces in-the-wild. Four age age groups are utilized in total, and each image is labelled with a specific group identifier.

- **Kyaw's web-collected Database**

The Kyaw's web-collected database [44] consist of facial images gathered from the internet using the programming interface of the microsoft image search application. The images in the dataset were primed by positioning the eye corner points and cropped to a dimension of 65x75. It contains about 963 different images which are split into four age groups of: 3–13, 23–33, 43–53, and 63–73.

- **Ni's Web-Collected Database**

The Ni's web-collected database [58] is among the largest database used for the purpose of facial age estimation. It is made up of a total of 219,892 facial images of 77,021 subjects obtained from "flickr.com" website and "google image search engine". The age labels of the images in the dataset is between the range of 1-80.

- **Gallagher's Web-Collected Database**

This database [30] was created by collecting facial images from the popular "flickr.com" website. It consist about 28,231 facial images, annotated with gender and age. There are 7 different age groups, ranging from: 0–2, 3–7, 8–12, 13–19, 20–36, 37–65, and 66+ in the database.

- **UIUC-IFP database**

The UIUC-IFP (University of Illinois at Urbana–Champaign-Image formation and Processing group) database [35] is a not-too large dataset not readily available to the public, used for facial age estimation purposes. The dataset consist of 8,000 high-resolution RGB facial images of persons between the ages of 0 and 93 years. It consists of 1,600 different natives from Asia comprising of 800 males and females with each person contributing at least five frontal images each. The images were gathered from an uncontrolled environment with different levels of variation ranging from makeup, facial expression to illumination.

- **Human & Object Interaction Processing (HOIP) database**

The HOIP database [28] consists of 306,600 facial images of about 300 different persons. The age range of persons in the dataset ranges from 15 to 64 years, divided into 10 different age-groups, having an interval of 5 years each.

- **Biometric Engineering Research Center (BERC) Database**

The BERC database [16] was gathered and developed by the "Biometric Engineering Research Center" (BERC), and used for facial aging research purposes. It contains facial images of 390 subjects, having ages ranging between 3 and 83. The images in the dataset have high resolution of about 3648x2736 pixels each, and without variability in illumination and facial expression. The age and gender of subject's images in the dataset are uniformly distributed.

- **Yamaha Gender and Age (YGA) Database**

YGA is a non-publicly available large face dataset, containing 8,000 face images captured outdoors for 1600 subjects having ages ranging between 0 and 93 years [29].

Table 1 shows the summary of the reviewed datasets used for AFAE.

Table 1: Summary of popular datasets used for Facial Age Estimation

S/N	Database	Number of subjects	Database size	Age range
1.	FGNET [45]	82	1002	0–69
2.	FERET [65]	1564	14,126	-
3.	PAL [54]	-	1140	19-93
4.	MORPH (Longitudunal) MORPH(Academic)1 MORPH(Academic)2 [71]	- 515 13,000	700,000 1724 55,134	- 27-68 16-80
5.	3D morphable database [11]	438	–	–
6.	FACES [24]	171	2052	19-80
7.	CACD [15]	2000	160,00	16-62
8.	OUI-Adience [25]	2284	26,580	0-60+
9	WIT [82]	5500	26,222	3–85
10.	ChaLearn LAP [26]	-	7591	0-95
11.	UTKFace [83]	-	24,100	0-116
12.	IMDB-WIKI [73]	100,00	524,230	0-100
13.	AFAD [59]	100,752	164,432	15-40
14.	Iranian Face database (IFDB) [7]	616	3600	2–85
15.	Lotus Hill Research Institute (LHI) database [77]	–	50,000	9–89
16.	AI & R Asian [27]	17	34	22–61
17.	AgeDB [55]	568	16,488	1-101
18.	VADANA [76]	43	2298	-
19.	Kyaw's Web-Collected Database [44]	–	963	3–73
20.	Ni's Web-Collected database [58]	–	219,892	1–80
21.	Gallagher's Web-Collected database 30]	–	28,231	0–66
22.	UIUC-IFP [35]	1600	8000	-
23.	Human and Object Interaction Processing (HOIP) [28]	300	306,600	
24.	BERC database [16]	95	5910	3–83
25.	Yamaha Gender and Age (YGA) [29]	1600	8000	0-93

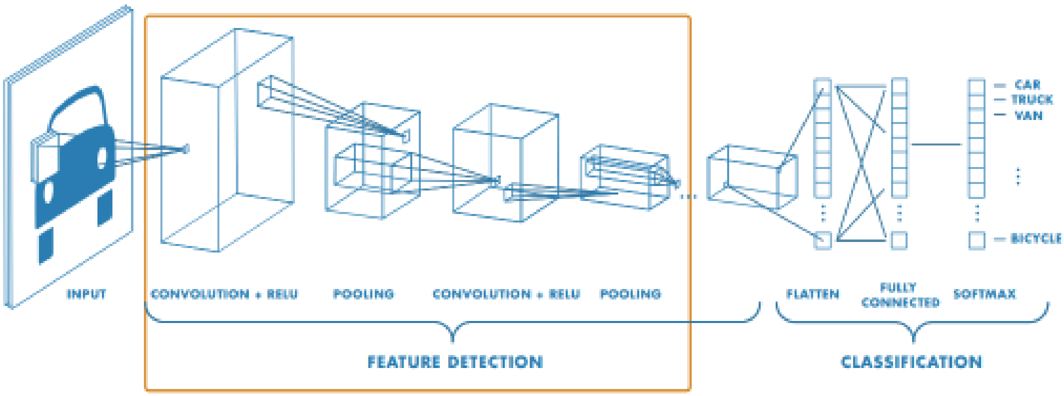

Fig. 1: CNN Architecture [100]

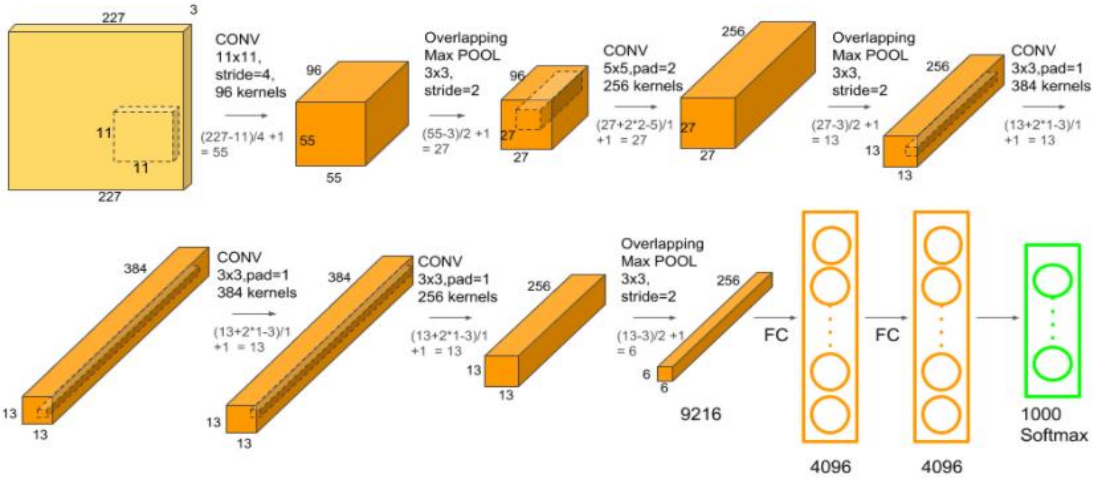

Fig. 2: AlexNet Architecture [57]

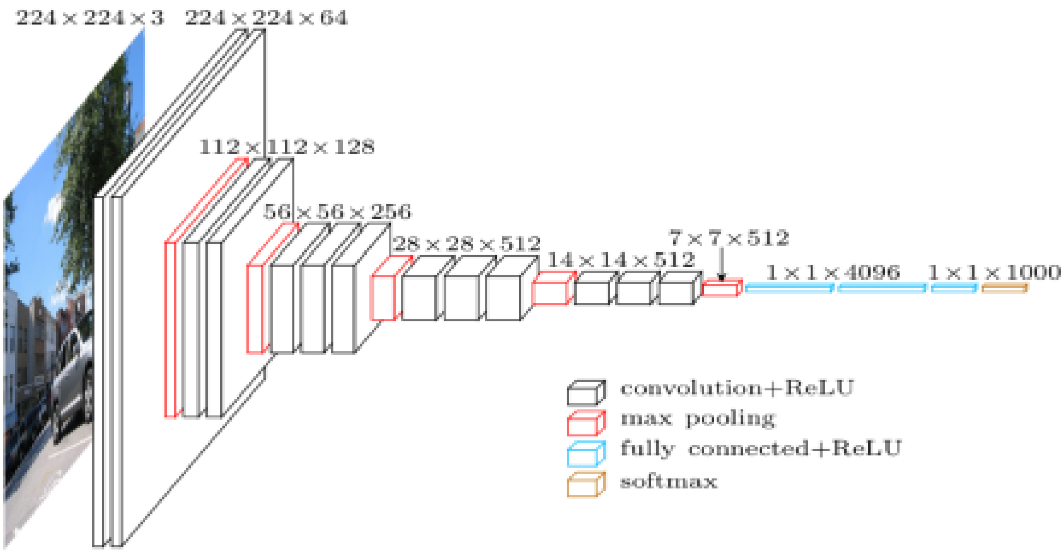

Fig. 3: VGG-16 Architecture [75].

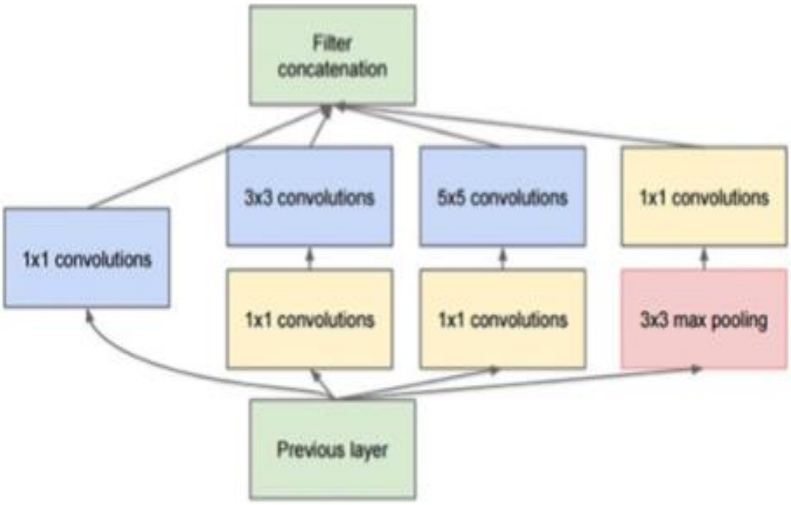

Fig 4: GoogLeNet Architecture [79].

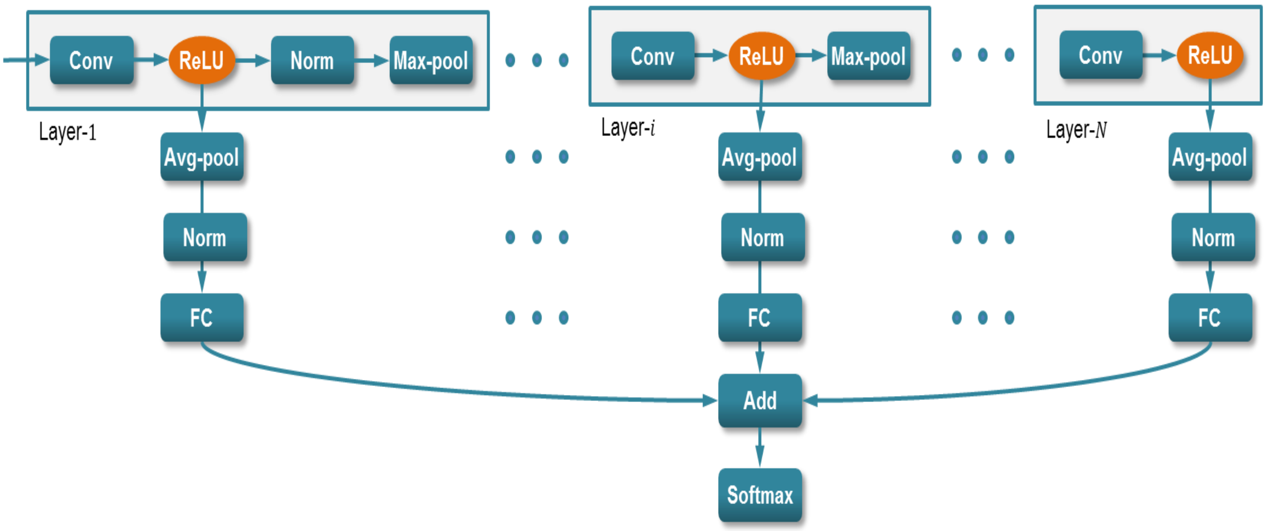

Fig. 5: DAG-CNN Architecture [87].

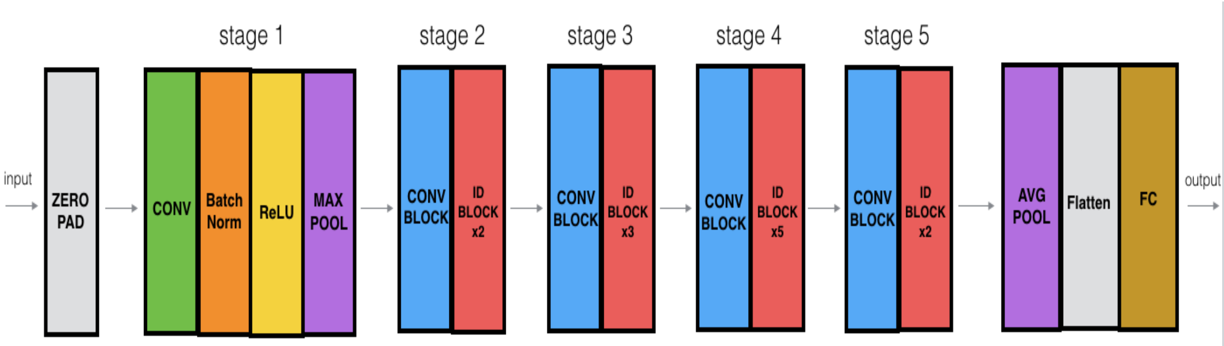

Fig. 6: ResNet50 model [66]

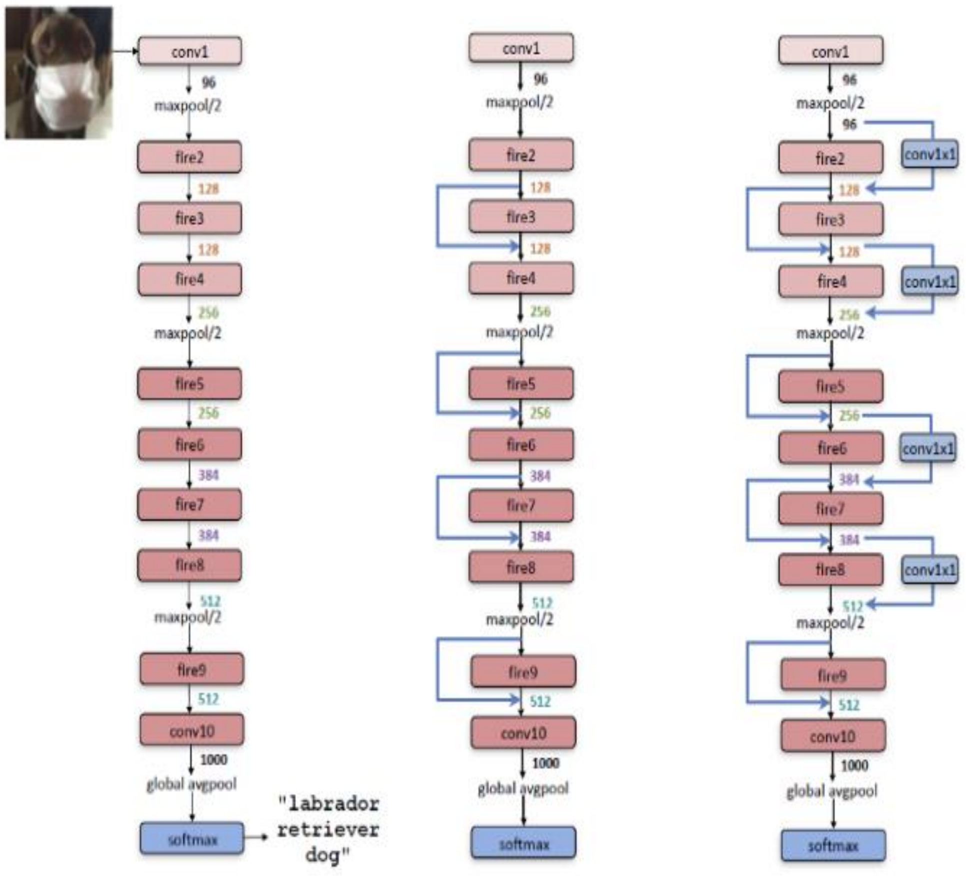

Fig. 7: SqueezeNet (left), SqueezeNet with simple bypass (middle), SqueezeNet with complex bypass (right) [41]

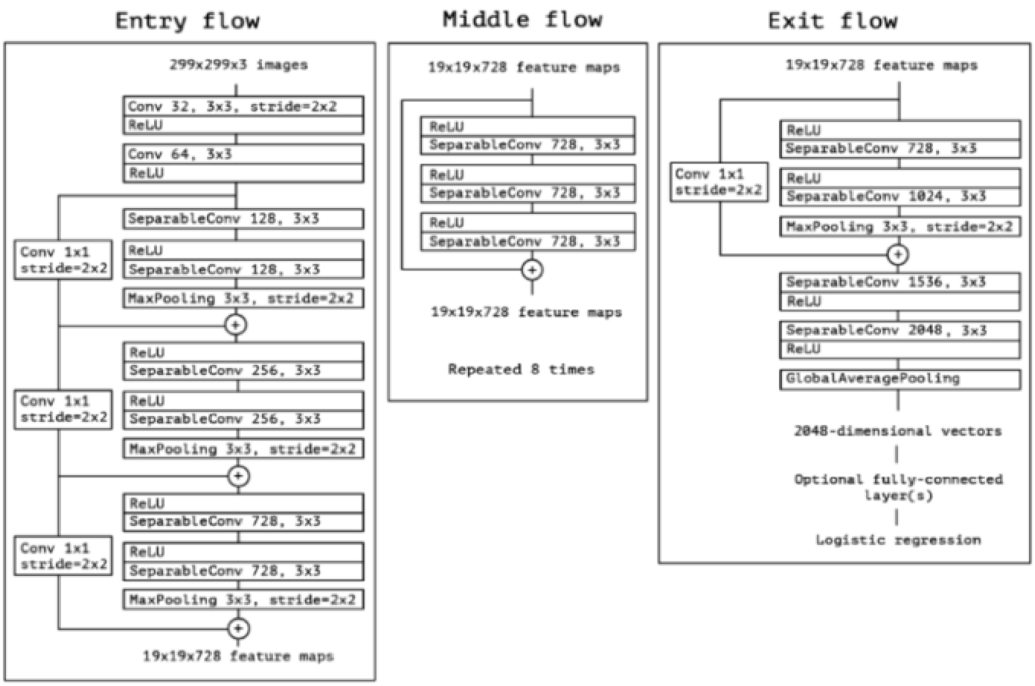

Fig. 8: The Xception architecture [17]

V. PERFORMANCE EVALUATION METRICS

The several metrics mostly employed by Researchers for the performance evaluation of AFAE systems, are:

A. *Mean Absolute Error (MAE)*

This is defined as the sum of errors between the estimated age and the actual or real age, divided by the number of images used for testing. It is the most pupluar performance metrics used for AFAE systems. It is represented by (1).

$$MAE = \frac{1}{n}\sum_{s=1}^{n}|y-\hat{y}| \quad (1)$$

where n is the number of images used for testing in the course of the experiment; y is the real or actual age of a sample s; and $\hat{y}$ is the estimated age of that same sample. The lower the MAE value, the better the perfomance of the system.

B. *Cummulative Score (CS)*

The CS is derived from the MAE, and it gives the proportion of test images having absolute errors lesser than a particular number of years. It is represented by the equation in (2).

$$CS(k) = \frac{n_{e\leq r}}{n} \times 100\% \quad (2)$$

here n refers to the number images used for testing in the experiment; $n_{e\leq k}$ gives the number of test images whose absolute error between the real and estimated age is less than r years.

C. *Normal Score ε*

This takes into consideration the variation of observer's vote given for each image, such as in the LAP dataset [26]. The LAP dataset was labelled by several human observers, while the mean and the standard deviation were computed of labels for each given sample [6]. It's equation is represented by (3).

$$\varepsilon = 1 - e^{-\left(\frac{(i-\mu)^2}{2\alpha^2}\right)} \quad (3)$$

where i is the tested image; μ is the mean of the vote given by the observer for i. The ε score is always between the values of 0 and 1 [22].

VI. ADVANCES ON HANDCRAFTED AND DEEP LEARNING METHODS FOR AUTOMATIC FACIAL AGE ESTIMATION

We give a brief review of the most popular and state-of-the-art or cutting-edge researches on facial age estimation involving traditional handcrafted machine learning methods and more recently deep learning methods in this section.

More recently, deep learning methods have been extensively employed in AFAE systems by Researchers because it has shown superior performance over the other handcrafted machine learning methods like regression [52], classification [95], ordinal ranking [85] and hierarchical age estimation [49].

Extraction of ageing patterns (feature extraction) is the first phase in AFAE. These features have been categorized into: local, global and hybrid [60]. Active appearance model (AAM) [18] as well as Bio-Inspired Features (BIF) created by Gabor filters [37] and Local Binary Patterns (LBP) [33] were also vastly used to learn texture and shape features from the training facial images. The second phase usually involves combining these features descriptors with traditional machine learning classification or regression (K-Nearest Neighbours (KNN), Support Vector Machine, Random Forest, etc.) to estimate the age.

For the first time, in 2009, [36] applied BIF to predict the age of a human and it produced a good performance. The AAM was employed by [46] for facial aging by learning a linear model for shape features and intensity from images and a set of indicators using a regression approach. [37] also investigated the BIF for age estimation, using standard deviation (STD) operator, which helped to reveal local variation and capturing important aging information such as wrinkles. [40] combined a series of local descriptors which combines and utilizes texture as LBP and SURF (Speeded Up Robust Features) [8] and local appearance-based descriptors as Histogram of Oriented Gradient (HOG) [20] under a diversity of setting on the MORPH and FRGC dataset.

In [14], the combination of distance metric learning and dimensionality reduction to investigate the relationships that exist between features of a face and the age annotations, in addition a label sensitive approach and a number of imbalance analysis were introduced during training stage so as to utilize the intrinsic ordinal relationship amid the ages of humans and overcome the problem of data imbalance; a traditional age regression algorithms is then introduced to capture facial aging process for age estimation problems.

For enhanced accuracy of facial age estimation systems, a hybrid approach which combines classification and regression method were used by researchers by taking advantage of the merit from both [28]. The handcrafted ageing patterns are sent as input to the classifiers or regressors for the purpose of age estimation learning. Guo *et al* [35] adopted a Locally Adjusted Robust Regressor known as LARR for learning and estimating the age of a human by using Support Vector Regressor (SVR), and it was tested on the FGNET and UIUC datasets. Choi *et al*. [16] proposed a more improved hierarchical classifier-based Support Vector Machine (SVM) and SVR. Yan *et al*. [85] expatiated on the use of SDP (Semi-Definite Programming) problem to estimate the age of humans as well as pose, which was considered as a non-linear regression problem that can deal with undecided labels expressed as two inequality constraints and non-negative labels expressed as the square norm of a matrix. SDP enhances the generalization and prediction capabilities of the algorithm and also helps to tackle the problem of overfitting. Experiments were done on the FGNET and Yamaha datasets. Zhang *et al*. [94] introduced a Multi-Task Warped Gaussian Process (MTWGP) as a classifier for estimating age, in which age estimation is expressed as a problem involving multi-task regression peculiar to each human (task), while the

MTWGP models similar properties shared by a number of persons (tasks). The algorithm is made up of a learning and a testing procedure. It showed good result when tested on the FGNET and MORPH datasets. Eidinger *et al.* [25] used the droupout-SVM approach for face attribute learning. Zhuang *et al* [96] introduced a patch based Hidden Markov Model (HMM) supervector that is based on Gaussian Mixture Model (GMM) in order to improve facial age estimation accuracy. Each of the images are depicted as sampled patches by capturing the spatial structure of facial images and loosening the presumption of similar face patch dissemination within a facial image; and then the HMM is implemented for unsupervised segmentation of the facial images and each segmented region can be regarded as a hidden state. Euclidian distance is used to make the final decision after calculating the patch HMM super-vector for each facial image.

In recent times, high-level semantic features are developed using deep learning methods based on CNNs in order to attain improved accuracy in AFAE. The multi-level CNNs execute a chain of transformations on the facial image in order to learn a denser transformation representation of the facial image. Further abstract features or characteristics are learnt in the deeper layers which give rise to a more accurate estimation of the class of the ageing patterns. The high-level semantic features extracted by using deep learning methods outperform traditional hand-crafted features [6].

Yang *et al* [88] combined ordinal regression with a convolutional neural network they called ScatNet to extract facial aging features and predict the age of the facial image. They used PCA (principal component analysis) to minimize the feature dimension and lastly 3-layers of fully connected CNN to produce the final decision via category-wise rankers. It was tested on MORPH, Lifespan and FACES datasets, and produced good results.

In [90], Yi *et al*. introduced an end-to-end multi-scale CNN model to estimate human age using their facial images. The input images are cropped into many pixels and inputted to the independent convolutional sub-networks, which are then joined at the FC layer to predict the age. It was tested on the on MORPH2 dataset and obtained an MAE of 3.63 years.

Wang *et al* [84] proposed a novel method to extract facing aging features based on deep learning. A 6-layer CNN was used to extract the features, while Support Vector Regression (SVR) and Support Vector Machines (SVM) were used to learn aging patterns. Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA) were also used as regression algorithms. They didn't put into consideration influence of gender and ethnicity. Experiment was conducted on FGNET and MORPH2 datasets and achieved an MAE of 4.26 and 4.77

In [13] an age ranking based on ordinal regression approach was proposed and was used on the MORPH2 dataset and obtained a MAE of 3.74 years. They failed to put into consideration the changes of neutral facial expressions, which may affect the accuracy of age estimation.

Han *et al* [38] tried to show that a machine could beat human performance in facial age estimation. Their approach yielded a better result than humans on FGNET, MORPH2 and PCSO databases in age estimation, race and gender tasks. Their method used the BIF algorithm to extract the demographic informative features through boosting algorithm from a facial image; and then used the approach of hierarchy where classification phases were accompanied by a regression phase so as to obtain a reliable demographic estimation.

In 2016, Rothe *et al* [72] introduced a method known as DEX (Deep Expectation), to estimate the age group of humans from a single image, before estimating the exact age value. This appeared to be more accurate and faster than when trying to estimate the exact age from a wide age range. The deep age estimation method was able to estimate apparent age without making use of explicit facial features. DEX obtained an MAE of 4.63 and 3.25 years on the FGNET and MORPH2 datasets respectively. When fine-tuned with the IMBD-WIKI (Internet Movie Database-What I Know Is) dataset, it achieved an MAE of 3.09 and 2.68 years on FGNET and MORPH2 respectively.

Liu *et al* [50] introduced an end-to-end large scale deep CNN, they called "AgeNet", for a robust apparent age estimation, while adopting a score fusion of classification and regression models based on Gaussian label and real-value distribution. A general-to specific deep transfer learning approach was also exploited. Their method achieved a cutting-edge performance in the ICCV (International Conference on Computer Vision)-2015 Look for Apparent Age Estimation contest, achieving MAE(s) of 3.3345 and 3.9489 on the MORPH2 and FGNET datasets.

Can Malli *et al* [12] designed an ensemble of deep learning models, where several CNNs were trained for varying group of ages and their age shifted groupings were added together so as to attain a higher accuracy. A fine-tuned VGG16 architecture lay the foundation the CNN model used. The Input image was first detected and five facial landmarks were detected for face alignment and then the faces are cropped to be fed to the CNN. The fully connected layers were finally used in obtaining their SoftMax and final mean scores. Experiment carried out on the IMDB-WIKI and ChaLearn LAP datasets attained a test error rate of 0.367 in the final of the 2016 edition of the ChaLearn LAP contest, though with a lesser error rate on the validation set.

Li *et al*. [47] proposed an end-to-end trainable CNN model that can easily learn discriminative aging features from raw facial images. This they called Deep Cumulatively and Comparatively (D2C) learning for estimating human age. In order to tackle the problem of sample imbalance and learn more efficient facial aging features, two novel layers which are differentiable were introduced; these are the Cumulative Hidden Layer and the Comparative Ranking Layer. One face in various training pairs is introduced in various training pairs in order to maximize the usage of the limited training data. Their model's performance was tested on MORPH2 and WebFace datasets and achieved MAE(s) of 3.06 and 6.04 respectively. The Researchers took into account losses peculiar to D2C architecture, such as cumulative, age and rank losses.

Anand *et al* [3] utilized multiple pre-trained deep CNNs (VGG face CNN and AlexNet) to extract facial aging features from a facial image which serves as input, by applying post-processing strategies in order to enhance the performance of the pre-trained CNN models. Their approach involves extracting aging features from a single input facial image. To decrease the feature space dimension, the proposed method made use of a feature level fusion before finally trying to estimate the age of the face image by making use of a Feed-Forward Neural Network (FFNN). The method achieved and MAE of 3.30, when tested on the Adience benchmark of unfiltered faces for gender and age estimation and AmIFace (a private) datasets.

Zou *et al* [97] introduced an automatic facial age estimation method by using a CNN with 3-convolutional, 3-pooling and 3-fully-connected layers (3C3P3FC). The method involves firstly an image pre-processing stage, where the original image is randomly cropped to 227x227, after the first convolution, it is then resized to 56x56. In order to increase non-linear properties, a ReLU is used as the activation function. Max-pooling is used to perform Down-sampling and then the image size becomes 28x28. The dropout method is used to reduce overfitting. The method achieved excellent result on the MORPH2 dataset attaining an accuracy of 0.7391. Their proposed method was also finetuned on the CaffNet and VGGFace architecture and it attained a peak accuracy of 0.7220 and 0.7756 respectively on the MORPH2 dataset.

Rattani *et al* [69] introduced a novel CNN model that could perform the task of age classification for ocular images from smart-phones. Their model was evaluated on the Adience database for gender and age classification. The CNN consist of 6 layers, with 3 conv. layers and max-pooling layers, 1 flattened layer and 2 FC layers. The model involves 41,416 parameters. The choice of the simpler model was to prevent the problem of overfitting, mostly for the case of small training databases. The accuracy of the results obtained were observed to be comparable to those involving full facial images.

In order to overcome the problem of optical and motion blurring real-time facial images in age estimation, Kang *et al* [42] introduced a deep CNN model based on ResNet152. The model excludes any form of manual intervention while training the feature vector coefficients and weights classifiers, i.e., they are trained automatically. The whole algorithm of their proposed model involves the recognition of face and eye regions by making use of Adaboost detector; redefining and in-plane rotation of the facial image in the region of interest; and then lastly, using the pre-trained ResNet152 model to redefine the face region and estimate the person's age. Their method was experimented on the PAL and MORPH2 dataset and achieved a MAE of 6.0 and 5.78 on the PAL and MORPH2 dataset respectively.

Zaghbani *et al* [92] proposed a novel method to learn aging features and estimate human ages based on Autoencoders in a supervised manner. The proposed work involves two major stages; firstly, AdaBoost algorithm is used to extract and crop the faces from the images, and then a tan inverse formula is exploited for the in-plane rotation of the extracted faces. Secondly, the task of classification centred on Deep Supervised Sparse Autoencoders (DSSAE) is carried out. Experimental results on the FGNET and MORPH dataset were found to be 3.75 and 3.34 respectively, which are much more reduced compared to other cutting-edge methods.

Taheri *et al* [80] combines multistage learned features from a generic feature extractor, a trained CNN model and a selection of age-related handcrafted features. Two approaches were implemented in this method. The first approach involves a feature-level fusion of a variety of handcrafted local feature descriptors of facial components, skin and wrinkle; while the second approach involves score-level fusion of different feature vectors that were learned from multiple layers of a CNN used for age estimation. The method achieved MAE(s) of 3.29 on the FGNET and 3.17 on the MORPH2 datasets respectively.

Duan *et al*. [23] proposed a hybrid structure which includes CNN and Extreme Learning Machine (ELM) which was referred to as CNN2ELM for facial age estimation. While the CNN was implemented for extracting the features from the input images, the ELM was used to classify the intermediate results. Age, race and gender are trained using different target methods. To prevent overfitting, different dropout measures were adopted and details of parameter selection, layer designing as well as the derivation of back-propagation process were provided. The architectural design of the hybrid CNN2ELM is made up of a configuration of convolutional, contrast normalization, max-pooling and ELM layers used for classification. The convolutional layer is responsible for performing convolutions with previous layers and then extract important features from the input feature maps. The Contrast Normalization layer is used for the computation of a variety of feature maps in the same spatial location by making use of subtraction and division normalization operations. The purpose of the Max-pooling layer is to filter out irrelevant information. Lastly, the outputs of the fully connected layers are fed as input to the ELM layer for performing classification task. The processing of CNN2ELM is done as a training and a classification phase. Adience, MORPH and LAP2016 datasets were used to achieve the predicted result which showed better performance than some other prominent deep learning methods.

In [81], Taheri *et al* also made use of the multi-stage features from a variety of layers of two CNN based architectures - GoogLeNet CNN and VGG16 CNN, and introduced a new CNN model for the estimation of human age, which they referred to as Directed Acyclic Graph Convolutional Neural Networks (DAG-CNN). Discriminative features from different layers of the CNN models are learnt automatically by the model and the features learnt are added together by using a Score-level fusion. Two variations of the introduced model were developed, they are; DAG-GoogLeNet which uses GoogLeNet CNN as its backbone architecture; and DAG-VGG16 which uses VGG-16 CNN as its backbone architecture. The performance of architectures are experimented on the MORPH2 and FGNET databases and attained MAE(s) of 2.87 and 3.05 for the DAG-GoogLeNet model; and 2.81 and 3.08 for the DAG-VGG16 model.

Mahjabin *et al*. [101] proposed a facial age estimator based on ResNet50 CNN architecture that implements age estimation majorly as a regression problem. Their methods involve a pre-processing phase which is made up of face detection, image resizing and one hot encoding. In order to improve the number of training data in the dataset, data augmentation approach such as flipping, rotating, zooming, etc were implemented. To minimize loss, the SGD and Adam optimizers were used. The performance of their method was tested on the FGNET dataset and training was done using the APPA-REAL and UTKFace datasets. An MAE of 4.49 and 5.3 on both optimizers were attained.

To mine the continuous relation between age labels effectively, Li *et al*. 48] introduced a CNN based model called BridgeNet to perform the task of facial age estimation. The model consists of local regressors and gating networks that can be both learnt from start to finish. The local regressors tackles various data by splitting the data space into several overlapping subspaces. While the gating networks learn continuity-aware weights by making use of a bridge-tree like structure used by the local regressors. The performance of their method was tested on the MORPH2, FGNET and Chalearn LAP 2015 datasets, and it outperformed other state-of-the-art methods.

So as to overcome the problem of facial images with low-resolution in age estimation tasks, Nam *et al*. [56] proposed a deep CNN-based age estimation model that reconstruct low-resolution facial images as high-resolution facial images by making use of a generative adversarial network (GAN) that preprocesses low-resolution face images before they are been sent as input. For the task of age estimation of the reconstructed face images, the model uses state-of-the-art CNN models such as VGG, DEX and ResNet. The performance of the model is experimented on MORPH, PAL and FGNET datasets, and showed its effectiveness in higher face image resolution reconstruction.

Yousaf, *et al*. [91] introduced a novel approach for solving Age Invariant facial Recognition (AIFR) tasks. It is based on several pretrained CNNs. Viola Jones face detector is first used to detect the test face, and then the necessary age-invariant features are extracted by using a fine-tuned pretrained CNN, which also classifies the test face image as one of the subjects. The weights of the several pretrained CNNs are fine-tuned appropriately by using the back-propagation to learn sturdy and optimum age-invariant features using FGNET database. The performance of the several pretrained CNNs, are compared for the purpose of AIFR. The accuracy of the pretrained CNNs for AIFR were observed to be better than using the default CNNs architectures.

Agbo-Ajala *et al*. [1] developed an end-to-end CNN based model to help in the classification of unconstrained real-life facial images into age and gender. Their approach involves a two-level CNN model which includes feature extraction and classification phase. An image preprocessing algorithm that preprocesses the input facial images is also included. The network is pretrained on the IMBD-WIKI database and fine-tuned on the MORPH2 database. The performance of their model was tested on the OIU-Adience dataset, and it shows that their approach was effective, attaining a state-of-the-art performance in the classification of gender and age group, with an exact accuracy of 16.2% and 3.2% one-off accuracy on age group classification.

Liu *et al*. [51], proposed a lightweight CNN network known as ShuffleNetV2, for facial age estimation on mobile terminals. It is centred on mixed attention mechanism. The Mixed Attention-ShuffleNetV2 (MA-SFV2) model transforms the output layer, by combining classification and regression methods, and highlights vital features by pre-processing the facial images using data augmentation approaches like sharpening, histogram enhancement, filtering, etc. This helps to improve the size of the face image and reduce the problem of overfitting. The experimental results of the method shows its relevance in real-life circumstances, mostly in mobile terminals on the MORPH2 and FGNET datasets.

Recently, Ahmed *et al* [102] proposed a 7-layer CNN model trained for gender classification to also be used for age estimation by transfer learning by fine-tunning the model. The CNN model is used to get information on gender of the face image, then Bayesian optimization which helps to minimize the classification error is applied to the pre-trained CNN. The best network is loaded and evaluated on the test set for age estimation task. FGNET and FERET datasets were used for the evaluation of this method and an MAE of 2.67 and 1.2 were obtained respectively, surpassing other state-of-the-art methods.

A comparison/overview of the various facial age estimation approaches based on hand crafted and deep learning methods reviewed thus far, are highlighted and summarized in Table 2.

Table 2: Overview of researches of facial age estimation using traditional handcrafted machine learning and deep learning methods.

Reference	Used algorithm	Dataset	MAE	CS%	ε
Handcrafted method					
Lanitis *et al* (2002) [46]	AMM and Regression	Private	4.3	-	-
Yan *et al*. (2007) [85]	Semi-Definite Programming (SDP) with non-linear regressor.	FGNET Private YGA:F Private YGA:M	5.78 9.79 10.38	95% 85% 82%	- - -
Zhuang *et* al. (2008) [96]	Image patches + Guasian Mixture Model (GMM)- HMM	Private: F Private: M	6.335 5.397	- -	- -
Guo *et al* (2008) [35]	LARR + SVR	FGNET UIUC-IFP-Y/F UIUC-IFP-Y/M	5/07 5.25 5.30	-	-
Guo *et al* (2009) [36]	BIF	FGNET Private YGA:F Private YGA:M	4.77 3.91 3.47	-	-
Zhang *et al* (2010) [94]	MTWGP	FGNET MORPH	4.83 6.28	-	-
Han *et al* (2013) [37]	BIF with standard deviation (STD) operator	FGNET MORPH2 PCOS	4.6 4.2 5.1	-	-
Chao *et al* (2013) [14]	Label-sensitive relevant component analysis	FGNET	4.4		
Huerta *et al* (2014) [40]	LBP+SURF+HOG+ CCA	MORPH FRGC	4.25 4.17		
Eidinger *et al.* (2014) [25]	LBP + FPLBP + droupout-SVM	Adience GALLAGHER	-	45.1% 66.%	-
Deep learning method					
Yang *et al*. (2013) [88]	CNN – ScaNet with ordinal regression	MORPH2 Lifespan FACES	3.49 5.19 7.04	-	-
Yi *et al*. (2014) [90]	End-to-end CNN with subnetwork per image patch.	MORPH2	3.63	-	-
Han *et al* (2014) [38]	Boosting algorithm + BIF + regression + classification	FGNET MORPH2 PCSO	3.8 3.6 4.1		
Wang *et al*, (2015) [84]	CNN + dimensionality reduction + classifiers (SVR, PLS, CCA)	MORPH2 FGNET	4.77 4.26	-	-
Chang *et al.* (2015) [13]	age ranking based on ordinal regression	MORPH2	3.74	-	-
Liu *et al*. (2015) [50]	AgeNet	MORPH2 FGNET	3.3345 3.9489		0.2872 0.3360
Rothe *et al*. (2016) [72]	DEX – VGG16	MORPH2 FGNET	2.68 3.09		
Can Malli *et al*. (2016) [12]	Fine-tuned CNN based on VGG16	IMDB-WIKI Chalearn LAP			0.3668
Li *et al*. (2017) [47]	D2C for age estimation	MORPH2 Webface	3.16 6.12	- -	- -
Anand *et al*. (2017) [2]	Dimensionality reduction + FFNNs based on fine-tuned VGGNet+AlexNet	Adience AmIFace IMBD-WIKI	3.30	-	-
Zou *et al*. (2017) [97]	3C3P3FC CaffeNet-finetune VGG-finetune	MORPH2 MORPH2 MORPH2	-	74% 72% 77%	-

References	Used Algorithm	Dataset	MAE	CS%	ε
Rattani *et al*. (2017) [69]	CNN for ocular images from smart phones.	Adience	-	46.97%	
Kang *et al*. (2018) [42]	Deep Residual CNN-ResNet152	PAL MORPH	6.0 5.78		
Zaghbani *et al*. (2018) [92]	Autoencoders	FGNET MORPH	3.75 3.34	-	-
Taheri *et al*. (2018) [80]	CNN for fusion based multi-stage age estimation	FGNET MORPH2	3.29 3.17		
Duan *et al*., (2018) [23]	Hybrid architecture: CNN+ELM	Adience MORPH LAP2016	- 2.61 -	66.69% - -	- - 0.3679
Taheri *et al*. (2019) [81]	DAG-CNN	MORPH2 FGNET	DAG-VGG16: 2.81(MORPH2) 3.08 (FGNET) DAG-GoogLeNet: 2.87(MORPH2) 3.05(FGNET)	-	
Mahjabin *et al*. (2019) [101]	ResNet50 CNN	(APPA-REAL + UTKFace) FGNET	Adam – 5.3 SGD – 4.49	67.3% 59.68%	- -
Li *et al*. (2019) [48]	VGGNet-BridgeNet	MORPH FGNET LAP	2.38 2.56 2.98	-	-
Nam *et al*., (2020) [56]	CNN + GAN	PAL MORPH2 FGNET	8.33 9.42 8.56	-	-
Yousaf, *et al*., (2020) [91]	Pretrained AlexNet, VGG16, VGG19, SqueezeNet, GoogLeNet, ResNet18, ResNet50, ResNet101, InceptionV3	FGNET	-	94.69% 91.58% 92.69% 85.48% 94.27% 99.78% 97.77% 98.27% 96.96%	-
Agbo-Ajala *et al*., (2020) [1]	6 layer (4C2FC) fine tunned AlexNet + dropout	OIU-Adience	-	84.8%	-
Liu *et al*. (2020) [51]	MA-ShuffleNetV	MORPH2 FGNET	2.68 3.81	-	-
Ahmed *et al*. (2020) [102]	CNN + Bayesian Optimization	FGNET FERET	2.67 1.2	-	-

VII CONCLUSION AND INSIGHT INTO PROBABLE FUTURE RESEARCH

In this review paper, a survey on recent and state-of-the-art approaches to AFAE using deep learning is carried out. Also a brief review on handcrafted methods used for this task is also carried out. Feature extraction
and learning algorithms are the major tasks involved for any facial age estimation system. Most of the recent researches involved deeply learned CNNs. Best MAE and accuracy are achieved by using deep learning methods based on CNNs, as observed in table 2. Hence CNN replaces the traditional handcrafted features by using supervised or semi-supervised feature extraction. However, the unique performance obtained by the use of deep learning methods can be improved upon. For probable future research, a study on other regression and classification methods that can be used for facial age estimation can be carried out. Gender, ethnicity, facial expressions as well as other demographic features can be experimented on furtherly to test their performance or effect on facial age estimation. Deep learning approach needs a very large dataset with minimum variation in lighting, pose and environmental changes for efficient training and improved accuracy of facial age estimation systems.

REFERENCES

[1] O. Agbo-Ajala and S. Viriri, S. “Deeply learned classifiers for age and gender predictions of unfiltered faces”. *The Scientific World Journal*, *2020*.

A.M. Alberta, K. Ricanek and E. Pattersonb. "A Review of the Literature on the Aging Adult skull and face: implications for forensic science research and applications." *Forensic Sci. Int* 172: 2007, pp 1–9.

[3] A. Anand, R. D. Labati, A. Genovese, E. Muñoz, V. Piuri and F. Scotti, "Age estimation based on face images and pre-trained convolutional neural networks". *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI, USA, 2017, pp. 1-7.

[4] A. Lanitis. 2010. "Facial Age Estimation." *Scholarpedia* 5 (1).

[5] R. Angulu, J.R. Tapamo & A.O. Adewumi. "Age estimation via face images: a survey." *EURASIP Journal on Image and Video Processing*, 2018, pp. 1-35.

[6] M.M. Badr, A.M. Sarhan & R.M. Elbasiony. “Facial Age Estimation using Deep neural networks: A Survey”. In *2019 15th International Computer Engineering Conference (ICENCO)*, 2019, December, (pp. 183-191). IEEE.

[7] A. Bastanfard, M.A Nik and M.M Dehshibi. “Iranian face database with age, pose and expression”. In: Proceedings—International Conference on Machine Vision, ICMV 2007, January, pp 50–55.

[8] H. Bay, T. Tuytelaars and L. Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, May 2006, pp. 404-417. Springer, Berlin, Heidelberg.

[9] Y. Bengio, Y. "Learning deep architectures for AI." *Foundations and Trends in Machine Learning,* 2009, 2 (1): 1–27.

[10] M.M. Sawant and K.M. Bhurchandi. “Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging”. *Artif Intell Rev* **52,** 981–1008 (2019).

[11] V. Blanz and T. Vetter. “A morphable model for the synthesis of 3D faces”. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, July 1999, pp. 187-194.

[12] R. C. Malli, M. Aygun and H. Kemal Ekenel, H. (2016). “Apparent age estimation using ensemble of deep learning models”. *In*: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 9-16.

[13] K. Y. Chang, K. Y and C.S. Chen. “A learning framework for age rank estimation based on face images with scattering transform”. *IEEE Transactions on Image Processing*, *24*(3), 2015, pp. 785-798.

[14] W.L. Chao, J.Z. Liu and J.J. Ding. “Facial age estimation based on label-sensitive learning and age-oriented regression”. *Pattern Recognition*, 2013, *46*(3), 628-641.

[15] B.C. Chen, C.S. Chen and W.H. Hsu. "Cross-age reference coding for age-invariant face recognition and retrieval." *European conference on computer vision.* Springer. 2014, pp. 768–783.

[16] S.E. Choi, Y.J. Lee, S.J. Lee, K.R. Park and J. Kim. “Age estimation using a hierarchical classifier based on global and local facial features”. Pattern Recognit. 2011, vol 44 (6), pp. 1262–1281.

[17] F. Chollet. “Xception: deep learning with depthwise separable convolutions”. In: Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017 January, pp 1800–1807.

[18] T.F. Cootes, G.J. Edwards and C.J. Taylor. “Active appearance models”. *IEEE Transactions on pattern analysis and machine intelligence*, 2001, *23*(6), 681-685.

[19] J.L. Crowley and T. Cootes. "FGNET: Face and Gesture Recognition Working Group,." http://www-prima.inrialpes.fr/FGnet/. 2009.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005, pp. 886-893 vol. 1.

[21] M.M. Dehshibi and A. Bastanfard. "A new algorithm for age recognition from facial images." *Signal Process* vol 90: pp. 2431-2444. 2010.

[22] M. Duan, K. Li and K. Li, "An Ensemble CNN2ELM for Age Estimation," in *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 758-772, March 2018.

[23] M. Duan, K. Li, K, C. Yang and K. Li. “A hybrid deep learning CNN–ELM for age and gender classification”. *Neurocomputing*, *275*(2018), 448-461.

[24] N.C. Ebner, M. Riediger and U. Lindenberger. "FACES - A database of facial expressions in young, middle-aged, and older women and men: Development and validation." *Behav. Res. Methods* 42, 351–362 (2010).

[25] E. Eidinger, R. Enbar and T. Hassner. "Age and gender estimation of unfiltered faces." *IEEE Transactions on Information Forensics and Security.* 2014, pp. 2170-2179.

[26] S. Escalera, X. Baro, H.J. Escalante and I. Guyon. "ChaLearn Looking at People: A Review of Events and Resources." *Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE.* 2017.

[27] Y. Fu and N. Zheng, "M-Face: An Appearance-Based Photorealistic Model for Multiple Facial Attributes Rendering", in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 830-842, July 2006.

[28] Y. Fu, G. Guo and T. S. Huang, "Age Synthesis and Estimation via Faces: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955-1976, Nov. 2010.

[29] Y. Fu, Y. Xu and T.S. Huang. Estimating human age by manifold analysis of face pictures and regression on aging features. In *2007 IEEE International Conference on Multimedia and Expo*. IEEE, July 2007, pp. 1383-1386.

[30] A.C. Gallagher AC and T. Chen. “Understanding images of groups of people”. *In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops* 2009, vol 2009. IEEE, pp 256–263.

[31] X. Geng, Z. Zhou and K. Smith-Miles, "Automatic Age Estimation Based on Facial Aging Patterns," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234-2240, Dec. 2007.

[32] P. Goodman A. Namdeo F. Galatioto M. Bell E. Foster and C. Shield. “Rethinking the inception architecture for computer vision”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818-2826

[33] A. Gunay and V.V. Nabiyev. “Automatic age classification with LBP”. In *2008 23rd International Symposium on Computer and Information Sciences*. October 2008, pp. 1-4. IEEE.

[34] G. Guo, Y. Fu, C. R. Dyer and T. S. Huang, "A Probabilistic Fusion Approach to human age prediction," *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, USA, 2008

[35] G. Guo, Y. Fu, C. R. Dyer and T. S. Huang, "Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression," in *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178-1188, July 2008

[36] G. Guo, G. Mu, Y. Fu & T.S. Huang. “Human age estimation using bio-inspired features”. In *2009 IEEE conference on computer vision and pattern recognition*, June 2009. (pp. 112-119). IEEE.

[37] H. Han, C. Otto and A. K. Jain, "Age estimation from face images: Human vs. machine performance," *2013 International Conference on Biometrics (ICB)*, Madrid, Spain, 2013, pp. 1-8,

[38] H. Han, C. Otto, X. Liu and A. K. Jain, "Demographic Estimation from Face Images: Human vs. Machine Performance," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1148-1161, 1 June 2015.

[39] He K, Zhang X, Ren S, Sun J. (2016). "Deep residual learning for image recognition." *Proceedings of IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, 770–778.

[40] I. Huerta, C. Fernández, and A. Prati. “Facial age estimation through the fusion of texture and local appearance descriptors”. In *European conference on computer vision*, September 2014, pp. 667-681. Springer, Cham.

[41] F.N Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. "Medical image computing and computer - assisted Intervention – MICCAI". *ICLR* 8150 (2017): pp. 1-13.

[42] J.S. Kang, C.S. Kim, Y.W. Lee, Y.W, S.W. Cho and K.R. Park. Age Estimation Robust to Optical and Motion Blurring by Deep Residual CNN. *Symmetry* **2018**, *10*, 108.

[43] A. Krizhevsk, I. Sutskever and G. Hinton G. "ImageNet classification with Deep convolutional neural networks." *Communications of the ACM* 60 , no. 6 (2017), 84-90.

[44] S.P. Kyaw, J. Wang and E.K. Teoh, "Web image mining for facial age estimation," *2013 9th International Conference on Information, Communications & Signal Processing*, Tainan, Taiwan, 2013, pp. 1-5.

[45] A. Lanitis. “Comparative Evaluation of Automatic Age–progression methodologies”. *Eurasip Journal on Applied Signal Processing*, 2008.

[46] A. Lanitis, C. J. Taylor and 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, April 2002.

[47] K. Li, J. Xing, W. Hu, and S.J. Maybank, S. J. “D2C: Deep cumulatively and comparatively learning for human age estimation”. *Pattern Recognition*, vol. *66*, 2017, pp. 95-105.

[48] W. Li, J. Lu, J. Feng, C. Xu,J. Zhou, J and Q. Tian. “Bridgenet: A continuity-aware probabilistic network for age estimation”. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1145-1154.

[49] W. Li, Y. Wang and Z. Zhang, "A hierarchical framework for image-based human age estimation by weighted and OHRanked Sparse Representation-based classification," *2012 5th IAPR International Conference on Biometrics (ICB)*, New Delhi, India, 2012, pp. 19-25

[50] X. Liu, S. Li M. Kan J. Zhang S. Wu, W. Liu, H. Han, S. Shan and X. Chen. “AgeNet:

deeply learned regressor and classifier for robust apparent age estimation". *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, 2015, pp. 16-24

[51] X. Liu, Y. Zou, H. Kuang, H and X. Ma. "Face image age estimation based on data augmentation and lightweight convolutional neural network". *Symmetry* 12, no 1(2020), 146.

[52] K. Luu, K. Ricanek, T. D. Bui and C. Y. Suen, "Age estimation using Active Appearance Models and Support Vector Machine regression," *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, Washington, DC, USA, 2009, pp. 1-5

[53] L.S. Mark, J.B. Pittenger, H. Hines, C. Carello C, R.E. Shaw and Todd J.T. "Wrinkling and head Shape as coordinated sources for age-level information." *Percept Psychophys* 27, 117–124(1980).

[54] M. Minear and D.C. Park. "A lifespan database of adult facial stimuli," Behavior Research Methods." *Instruments, & Computers,* vol. 36, no. 4 (2004), 630–633,.

[55] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, I. and 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.

[56] S. H. Nam, Y. H. Kim, N. Q. Truong, J. Choi and K. R. Park, "Age Estimation by Super Resolution Reconstruction Based on Adversarial Networks," in *IEEE Access*, vol. 8, pp. 17103-17120, 2020.

[57] Neurohive. 2018. "AlexNet – ImageNet Classification with Deep Convolutional Neural Networks." *Popular Networks*, October.

[58] B. Ni, Z. Song and S. Yan S. "Web image mining towards universal age estimator". *In: Proceedings of the 17th ACM international conference on Multimedia*, October 2009, pp. 85–94.

[59] Z. Niu, M. Zhou, L. Wang, X. Gao and G. Hua. "Ordinal regression with multiple output CNN for age estimation". *In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016 December, pp 4920– 4928

[60] O.F. Osman and M.H. Yap. "Computational intelligence in automatic face age estimation: A survey". *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. *3* no. 3(2018), 271-285.

[61] A. Othmani, A.R. Taleb, H. Abdelkawy and A. Hadid. Age estimation from faces using deep learning: A comparative analysis. *Computer Vision and Image Understanding*, *196* (2020), 102961.

[62] G. Panis, A. Lanitis, N. Tsapatsoulis and T.F. Cootes. "Overview of research on facial ageing using the FG-NET ageing database." *IET Biom* 5 (2016), 37–46.

[63] U. Park, Y. Tong and A. K. Jain, "Age-Invariant Face Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947-954, May 2010.

[64] P.J. Phillips, S.Z. Der, P.J. Rauss and O.Z. Der. "FERET (face recognition technology) recognition algorithm development and test Results." *Adelphi, MD: Army Research Laboratory*, 1996.

[65] P.J. Phillips, H. Moon, S.A. Rizv, and P.J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, vol. *22* no. 10 (2000), 1090-1104.

[66] P. Dwivedi. "Understanding and Coding a ResNet in Keras". 2019

[67] P. Punyani, R. Gupta and A. Kumar. "Neural networks for facial age estimation: a survey on recent advances." *Artificial Intelligence Review* (Springer) vol. 53 no. 5 (2020), 3299-3347.

[68] N. Ramanathan and R. Chellappa, "Modeling Age Progression in Young Faces," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, 2006, pp. 387-394.

[69] A. Rattani, N. Reddy and R. Derakhshani, "Convolutional neural network for age classification from smart-phone based ocular images," *2017 IEEE International Joint Conference on Biometrics (IJCB)*, Denver, CO, USA, 2017, pp. 756-761.

[70] M.G. Rhodes. "Age Estimation of Faces: A Review." *Appl. Cognitive Psychology* 23, 2009, pp. 1-12.

[71] K. Ricanek and T. Tesafaye, "MORPH: a longitudinal image database of normal adult age-progression," *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, Southampton, UK, 2006, pp. 341-345.

[72] R. Rothe, R. Timofte, and LV. 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.

[73] T. Sabharwal, R. Gupta, R. Kumar, and S. Jha. "Recognition of surgically altered face images: an empirical analysis on recent advances". *Artificial Intelligence Review*, 52, no. 2 (2019), 1009-1040.

[74] N.E. Sahla. *A Deep Learning Prediction Model for Object Identification.* M.Sc Thesis, 2018.

[75] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv:1409.1556* (2014)

[76] G. Somanath, M. Rohith and C. Kambhamettu, "VADANA: A dense dataset for facial image analysis," *2011 IEEE International Conference on Computer Vision Workshops (ICCV*

Workshops), Barcelona, Spain, 2011, pp. 2175-2182

[77] J. Suo, T. Wu, S. Zhu, S. Shan, X. Chen and W. Gao, "Design sparse features for age estimation using hierarchical face model," *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, Amsterdam, Netherlands, 2008, pp. 1-6

[78] K. Sveikata, I. Balciuniene. and J. Tutkuviene. "Factors influencing face aging. literature review." *Stomatologija* 13, vol. 4 (2011), 113–115.

[79] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelo, D. Erhan, V. Vanhoucke and A. Rabinovich. "Going deeper with convolutions." *Proceedings of IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.

[80] S. Taheri and Ö. Toygar. "Multi-stage age estimation using two level fusions of handcrafted and learned features on facial images". *IET Biometrics* 8, no. 2 (2018), 124-133

[81] S. Taheri and Ö. Toygar. "On the use of DAG-CNN architecture for age estimation with multi-stage features fusion". *Neurocomputing* 329 (2019): 300-10.

[82] K. Ueki, T. Hayashida and T. Kobayashi, "Subspace-based age-group classification using facial images under various lighting conditions," *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, Southampton, UK, 2006, pp. 6 pp.-48.

[83] UTKFace (2017)

[84] X. Wang, R. Guo and C. Kambhamettu, "Deeply-Learned Feature for Age Estimation," *2015 IEEE Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, 2015, pp. 534-541.

[85] S. Yan, H. Wang, X. Tang and T.S. Huang. "Learning auto-structured regressor from uncertain non-negative labels". *In: Proceedings of IEEE conference on computer vision, 2007, pp 1–8*.

[86] S. Yan, H. Wang, T. S. Huang, Q. Yang and X. Tang, "Ranking with Uncertain Labels," *2007 IEEE International Conference on Multimedia and Expo*, Beijing, China, 2007, pp. 96-99.

[87] S. Yang and D. Ramanan. "Multi-scale recognition with DAG-CNNs". In *Proceedings of the IEEE international conference on computer vision*, pp. 1215-1223. 2015.

[88] H.F. Yang, B.Y. Lin, K.Y. Chang and C.S. Chen. "Automatic age estimation from face images via deep ranking". *networks 35* no. 8 (2013), 1872-1886.

[89] M. Yehia. "*Automatic facial age estimation*". Cairo: Ph.D. dissertation, Faculty of Computers & Information, Cairo University, 2011.

[90] Yi, D., Lei, Z., & Li, S. Z. (2014, November). Age estimation by multi-scale convolutional network. In *Asian conference on computer vision* (pp. 144-158). Springer, Cham.

[91] A. Yousaf, M.J. Khan, M. J. Khan, A.M. Siddiqui and K. Khurshid. "A robust and efficient convolutional deep learning framework for age-invariant face recognition. *Expert Systems*, *37*, no. 3 (2020), e12503.

[92] Zaghbani S, Boujneh N, Bouhlel MS. "Age estimation using deep learning". *Comput Electr Eng* 68, pp. 337–347, 2018.

[93] L.A. Zebrowitz. "*Reading Faces: Window to the Soul*". Washington DC: Westview Press, 1997.

[94] Y. Zhang Y and D. Yeung. "Multi-task warped gaussian process for personalized age estimation". *In: IEEE conference on computer vision and pattern recognition (CVPR),* pp. 2622–2629, 2010.

[95] Y. Zheng, H. Yao, Y. Zhang and P. Xu. "Age classification based on back-propagation network. In *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service*, pp. 319-322, August 2013.

[96] X. Zhuang, X. Zhou M. Hasegawa-Johnson and T. Huang. "Face age estimation using patch- based hidden Markov model supervectors". *In: International conference on image and graphics*, 2008, pp 1–4, 2008.

[97] J. Zou, X. Yang, H. Zhang and X. Chen. "Automatic Facial Age Estimate Based on Convolution Neural Network". In *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Springer, Cham, pp. 19-26, August 2017.

[98] X. Lu, X. Duan, X. Mao, Y. Li, and X. Zhang. "Feature extraction and fusion using deep convolutional neural networks for face detection". *Mathematical Problems in Engineering*, *2017*.

[99] R. Ranjan *et al.*, "Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, than Humans," in *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 66-83, January. 2018.

[100] Mathworks – CNN, 2017

[101] S. Mahjabin, M. M. Alam and K. H. Talukder, "Age Estimation From Facial Image Using Convolutional Neural Network (CNN)," *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, Rajshahi, Bangladesh, 2019, pp. 1-4.

[102] M. Ahmed and S. Viriri. "Facial Age Estimation using Transfer Learning and Bayesian Optimization based on Gender Information". *Signal & Image Processing: An International Journal (SIPIJ) Vol*, *11*, 2020.

[103] M.S. Zimbler, M.S. Kokoska and J.R. Thomas." Anatomy and pathophysiology of facial aging". *Facial plastic surgery clinics of North America*, vol. 9 no. 2, pp. 179-87, 2001.