International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Oct-2022)

https://dx.doi.org/10.12785/ijcds/120194

Automatic Age Estimation of Persons with Dark Skin Tone Using Deep Learning Approach

V. Osekhonmen Abhulimen[1] and O. Erastus Ogunti[2]

[1]*Department of Electrical and Electronics Engineering,Federal University of Technology Akure, Akure, Nigeria*
[2]*Department of Computer Engineering, Federal University of Technology, Akure, Akure, Nigeria*

Received 22 Jan. 2021, Revised 15 Jul. 2022, Accepted 23 Jul. 2022, Published 31 Oct. 2022

Abstract: Age estimation has become very important for many organizations and human endeavour. Traditional machine learning methods have previously been deployed in previous times by many researchers to solve this problem automatically. However, the use of deep learning methods in recent times has shown superior performance in artificial intelligence tasks. This study employs the deep learning method gleaned from the ResNet50 convolutional neural network (CNN) to solve this problem; but on persons with dark skin tone, because pre-existing automatic age estimation models were trained on datasets with a limited population of persons with darker skin tones, as recent studies have shown that the aging features of dark skin tone persons cannot be learnt from persons with light skin pigmentations (white skin tone persons). A combination of persons with dark skin tones from the UTKFace, APPA-REAL and BlackFaces datasets was used to train the CNN. At the end of the experiment, the proposed approach attained a mean absolute error of 5.21 years on the validation set, and showed good performance on age estimation of dark skin tone persons.

Keywords: Deep Learning, ResNet50, Convolutional Neural Network (CNN), UTKFaces, APPA-REAL, BlackFaces, dark skin

1. INTRODUCTION

Deep learning is regarded as a subfield of machine learning that has shown superior performance in a variety of artificial intelligence (AI) tasks [1]. It mimics the workings of the neurons in the human brain in processing information and making patterns in decision making. This is made possible by its ability to unsheathe high-level, complex abstractions as data representations through a stratified learning process, more quickly than traditional machine learning methods. Hence, using a deep learning approach will help to learn the important features by itself, rather than manually selecting the pertinent features [2]. The superiority of its performance over traditional machine learning methods has led to a surge in recent times in the investigation of its possible application in the automatic estimation of ages, which has become a very crucial part of many organizations and human endeavour.

Automatic age estimation has become very important in solving many real-life problems ranging from security control and surveillance, electronic customer relationship management (ECRM), age verification in sports and job recruitment, biometric recognition system, health care, among many others. Nevertheless, certain issues in automatic age estimation are still ambiguous as age estimation of unrefined real-life faces are yet to meet the optimum prerequisites of real-life applications. This is largely due to the application of traditional machine learning methods proposed by several studies for age estimation over the past years. These traditional methods were reliant on the dissimilarities in dimensions of facial descriptors and facial features and therefore lacks the capability to tackle the diverse levels of variations observed in these challenging unrefined imaging conditions [3]. However, in recent times, deep learning method based on CNNs have shown encouraging performance in the automatic estimation of ages using facial images, and this has led to a plethora of techniques based on CNNs to tackle this problem [4].

However, due to the lack of open and cross-racial facial aging datasets, many recent automatic age estimation CNNs were trained mostly on images of persons with light skin pigmentations, having Western European facial features and descriptors, and therefore fall short of persons with dark skin tones. Studies have shown that the more unique aging features of dark skin persons, make them unlearnable from other skin tones [5]. In this study, a method to automatically and more accurately estimate the ages of persons with dark skin tones, by using a dataset consisting solely of dark skin persons to train the ResNet50 [6] CNN is proposed.

2. REVIEW OF RELATED WORKS

The possible application of automatic facial age estimation in solving real-life problems has led to a surge in the study area. Although, deep learning methods based on CNNs is showing superior performance, the use of traditional machine learning methods previously used formed the bedrock that led to the implementation of deep learning approach based on CNNs.

Kwon and da Vitoria Lobo [7], proposed a method that was centered on the use of geometrics of wrinkle patterns, in order to classify facial images. These images were classified into groups of three, using six distance ratios to compute and discriminate children from adults, Snakelets were used to dredge up curves and to also extract wrinkle patterns present on certain areas of the skin. In their experiment, a small-scale dataset consisting of just 47 images was used. Lanitis et al. [8] deployed the use of Active Appearance Model (AAM) for aging simulations based on facial appearance, by using a linear model to learn shape features as well as intensity from facial images using a given set of indicators and a regression-based method. Guo et al. [9], applied Bio-Inspired Features (BIF) to estimate the actual age of a person, and it yielded meaningful results. Guo et al. [10] proposed the LARR (Locally Adjusted Robust Regressor) for automatic age estimation making use of Support Vector Regressor (SVR); while Choi et al. [11] proposed an upgraded stratified classifier, which was gleaned on Support Vector Machines (SVM) as well as SVR for age estimation of facial images.

Recent studies on age estimation using deep learning include: Yi et al. [12], which introduced a multi-scale end-to-end CNN based method to predict a person's age from pixels of their facial images directly. Their method showed good performance on the MORPHII dataset.

Wang et al. [13] proposed an approach based on a CNN with 6 layers, which were used to extract features of facial aging from the face, while SVM and SVR were combined and deployed to learn the aging patterns of the face. Partial Least Squares (PLS) as well as Canonical Correlation Analysis (CCA) were deployed as algorithms for regression. Their method was evaluated on the MORPHII and FGNET datasets, and showed good performance.

Rothe et al. [14] introduced the Deep Expectation (DEX) approach in order to predict a person's age group from a single facial image, before predicting the actual age of the person. The DEX technique was capable of predicting the apparent age without using explicit facial features. Their approach was evaluated on the FGNET, MORPHII and CACD datasets after fine-tuning on the IMBD-WIKI dataset, and it performed well.

Anand et al. [15] deployed several pre-trained deep CNNs in order to extract facial aging features from a single face image. To minimize the dimensions of feature space, they deployed the use of a feature-level-fusion, before estimating the age of the facial image by making use of a feed-forward neural network (FFNN). Their approach was assessed on the Adience and AmIFace (a private) datasets.

Taheri and Toygar [16] proposed a Directed Acyclic Graph Convolutional Neural Network, which they also referred to as DAG-CNN, and it uses the GoogLeNet and VGG16 CNNs as its backbone architecture, and learns distinctive features from several layers of the CNN models. A score level fusion approach is deployed in order to add the features learnt together. The approach was validated on the MORPHII and FGNET datasets, and attained good results. Fariza and Arifin [17] proposed a residual network classification approach known as ResNeXt-50(34x4d) to solve the problem of automatic age estimation, and their approach was evaluated on the UTKFace dataset, and showed excellent performance.

Agbo-Ajala and Viriri [3], proposed a two-level end-to-end CNN for classifying unrefined real-life face images into their respective age as well as gender. The method includes a feature extraction as well as a classification stage with an algorithm that preprocesses the input images. The IMBD-WIKI dataset is used to pre-train the CNN, and then it is fine-tuned on the MORPH2 dataset. In evaluating the performance of their method, the OIU-Adience dataset was used.

Huynh and Nguyen [18] proposed a method by improving upon the Megaage-Asian dataset to tackle the problem of automatic age estimation, as well as gender for Asian Subjects. It involves an image augmentation stage by implementing random erasing and mixup processes, and the images are then used to train the Wide ResNet CNN model for age estimation of Asian subjects.

Abhulimen et al. [19] proposed a deep learning approach based on a fine-tunned ResNet50 CNN by merging images from three datasets (UTKFace, APPA-REAL and Black-Faces) into one, which can be used for the estimation of ages persons across several races.

This study specifically addresses the problem of automatic age estimation of dark skin tone persons, which the reviewed literatures above have failed to address.

3. METHODOLOGY

The following subsections describes the methodologies used for the implementation of this study.

A. Data Collection

The first stage in the implementation of the study methodology in this study is the collection of facial images of persons with dark skin. Conventionally, most studies refer to the MORPH2, FGNET, UTKFace, IMDB-WIKI and APPA-REAL datasets when they tackle the problem of automatic facial age estimation, since they are among the most popular public datasets used. However, they only contain a small percentage population of persons with dark skin tone. Therefore, models trained on those datasets underperformed as expected, when presented with dark skin persons, because dark skin persons have some unique aging characteristics that cannot be learnt from other races [5]. In achieving a dataset of solely dark skin persons, we extracted images of persons with dark skin tones from the cross-racial UTKFace and APPA-REAL datasets.

The UTKFace [20] dataset contains 23,704 facial images of both male and female with ages ranging from 1-116 years, with each face image annotated with a specific age, gender and racial group. A total of 5,026 images of persons with dark skin tone were extracted from the UTKFace dataset.

While the APPA-REAL [21] dataset contains 7,591 facial images of male and female subjects ranging from 1-95 years, a total of 440 images of persons with dark skin tone was extracted from it by manual inspection; as the images contained in the dataset were not annotated with their specific racial group.

Deep learning models and algorithms are only as good as the data from which they are built [22]; hence, they tend to show better performance on larger-scale datasets. So as to further enhance the performance of the deep learning model and also augment the number of images already collected from both datasets, we created a small-scale dataset solely of dark skin persons. This was done by collecting facial images of black persons alongside their respective age (at the time the images were taken), from social media platforms such as Facebook, Instagram, Twitter and Google images, and also by taking real time photographs of persons within our immediate environment. We referred to this dataset as "BlackFaces". At the time of this study, BlackFaces consist of 3,000 facial images of equal number of male and female between the ages of 1-92 years.

B. Data Pre-processing

In pre-processing the data, the images from the APPA-REAL dataset and BlackFaces dataset were re-annotated with the same format as the UTKFace dataset so as to ensure easy readability of the data by the CNN model. Also, all the images were resized to 224x224 pixels after which they were collected in a single folder as a single dataset.

We implemented data augmentation processes such as flipping and rotating, on the dataset so as to further improve the performance of the CNN model, since our training dataset was not sufficiently large, as it contains a total of 8,466 images.

A sample of facial images contained in the facial aging dataset we used for training the deep learning model is shown in Figure 1.

C. Dataset Split

In order to train the CNN, the facial aging dataset was first split by using a stratified cross validation (CV) approach which has been defined by equation 1. 80% of the images in the dataset was used for training, while 20% was used for testing. 10% of the training set was set aside for validation.

Figure 1. Sample of dark skin persons in the facial aging dataset

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L(yi, \hat{F}^{-\kappa(i)}(x_i)) \tag{1}$$

where L represents the loss function, and $\hat{F}$ represents the ResNET50 model to be trained. $\kappa(i)$ means that the model $\hat{F}$ is trained without the training patterns in the same split-up of the dataset as in pattern i.

D. ResNet50 CNN

For the task of automatic age estimation, we deployed the ResNet50 CNN [6]. It is a variant of the residual network architecture. It was initially pre-trained for image classification task on the ImageNet dataset. It is made up of 48 convolution layers, alongside 1 maxpool and 1 average pool layer. The residual network architecture makes use of skip connections which permits an input x to skip one or several layers without having to pass through some layer weights. This is illustrated in Figure 2.
The desired base mapping, may be denoted as;

$$H(x) \tag{2}$$

and it fits the stacked nonlinear layers mapping given by;

$$F(x) := H(x) - x \tag{3}$$

hence the default mapping is rearranged to become;

$$H(x) := F(x) + x \tag{4}$$

This helps the ResNet50 architecture overcome the problem of vanishing/exploding gradient observed when a CNN is stacked up with more layers, which usually leads to accuracy getting saturated and degrading rapidly leading to an increase in the training error. The inner architecture of the ResNet50 model shows that it is made up of 5 phases, with each phase having a convolution as well as an identity block. Each convolution block and identity block also have 3 convolutional layers. This is depicted with the image of

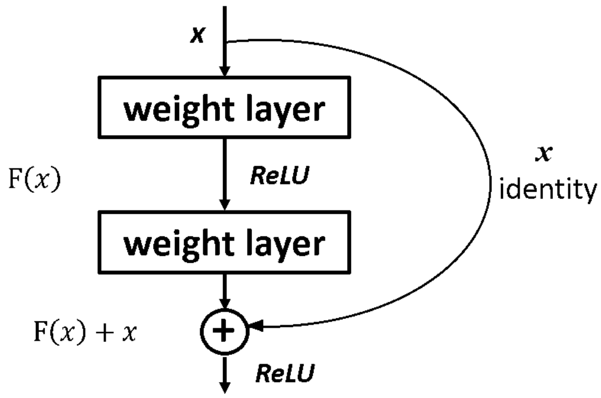

Figure 2. Building block of ResNet50 (identity mapping) [6]

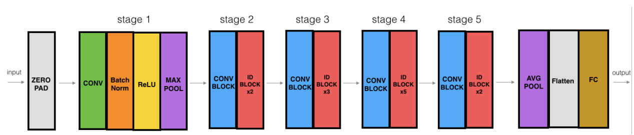

Figure 3. Architecture of ResNet50 CNN [23]

Figure 3 and further illustrated with Table 1.

ResNet50 CNN models are easier to train; more indulgent of hyperparameters, as well as regularization and initial learning rate; and have better generalization capabilities, because of the principle of skip connections and identity mapping it utilizes.

E. Evaluation Metrics

In evaluating the performance of the ResNet50 CNN in the age estimation of dark skin persons, the mean absolute error (*MAE*) was used. It is computed with the expression

TABLE I: ResNet50 CNN Architecture

Layer name	Output size	RestNet50
conv1	112×112	7×7, 64 stride 2
conv2	56×56	3×3 max pool, stride 2
		$\begin{pmatrix} 1\times1 & 64 \\ 3\times3 & 64 \\ 1\times1 & 256 \end{pmatrix} \times 3$
conv3	28×28	$\begin{pmatrix} 1\times1 & 128 \\ 3\times3 & 128 \\ 1\times1 & 512 \end{pmatrix} \times 4$
conv4	14×14	$\begin{pmatrix} 1\times1 & 256 \\ 3\times3 & 256 \\ 1\times1 & 1024 \end{pmatrix} \times 6$
conv5	7×7	$\begin{pmatrix} 1\times1 & 512 \\ 3\times3 & 512 \\ 1\times1 & 2048 \end{pmatrix} \times 3$
	1×1	*average pool, 1000-id fc, softmax*
FLOPs		3.8×10^9

TABLE 2: Hyperparameters used for training the RESNET50 CNN in this study

ResNet50 CNN hyperparameters used for training	
Image size	224x224
Colour mode	RGB
Batch size	32
Random state	42
Optimizer	Stochastic Gradient Descent (SGD)
Learning rate	$1 \times 10e - 04$
Evaluation metric	MAE
Epoch	12
SoftMax layer	116
Frozen layers	100%

of equation (5).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y} - E_i(O)| \tag{5}$$

here n denotes the sum of facial images deployed for testing in the experiment; $\hat{y}$ represents the actual age of a person i; and $E_i(O)$ is the predicted or estimated age of the person. The lower the value of *MAE*, the better the system performs.

F. Estimation Error

The error in the age estimation of a person is obtained with equation (6) as;

$$Error = E_age - A_age \tag{6}$$

where; E_age = Estimated age, and A_age = Actual age

G. Experimental Hyperparameters

A summary of the hyperparameters used to train the ResNet50 CNN model in this study is shown in Table 2.

H. System's Flowchart

The flowchart for the automatic age estimation system developed in this study is shown in Figure 4.

I. Computational Complexity

The bulk of this study was implemented on Python programming language, making use of the ktrain library [24], on a 12GB memory NVIDIA Telsa K80 GPU.

4. RESULTS/DISCUSSIONS

The training accuracy or model loss curve of the trained ResNet50 model is depicted in figure 5.

From the graph of Figure 5, it can be interpreted that the model performed optimumly on the training and validation set, with a decreasing validation loss, till the 12th epoch, attaining an *MAE* of 5.21 years on the validation set. In order to avert bias in the CNN model selection, we also carried out the same evaluation on the test set, and an *MAE* of 5.03 was obtained. This result shows that the system misses a person's actual age by a margin of approximately

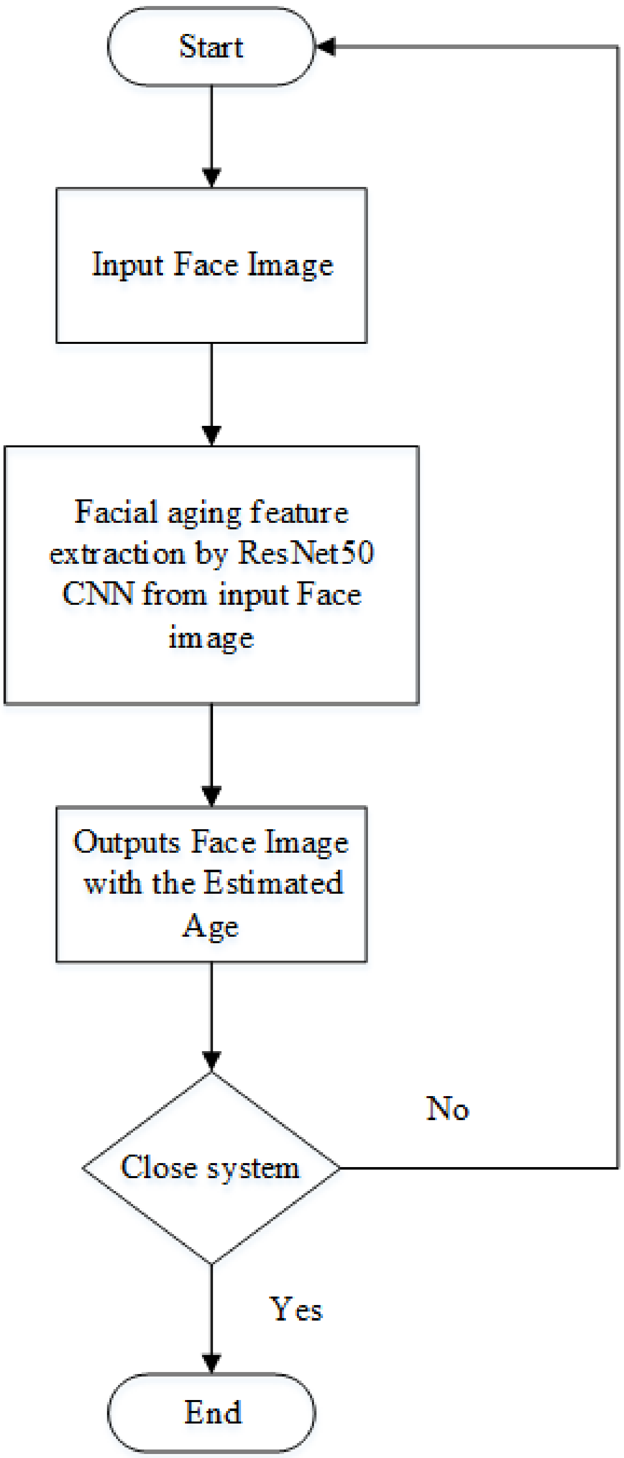

Figure 4. The age estimation system's flowchart

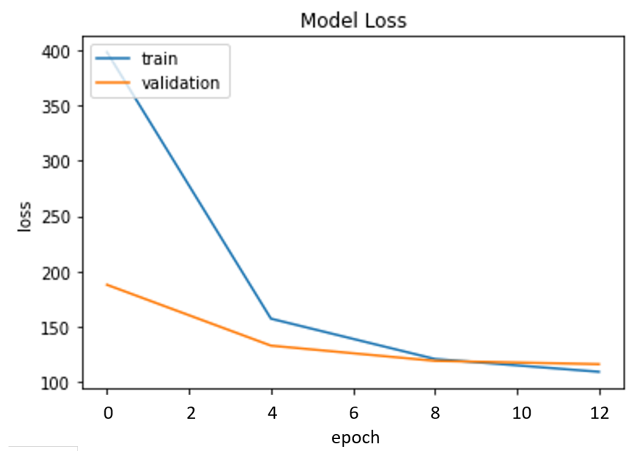

Figure 5. Model loss curve for the trained ResNet50

±5 years on average. This hypothetically indicates that aside the early childhood phase of a person - when the face is most likely to change in a dramatic fashion, the face of a dark skin person is likely not to change much in a period of 5 years.

However, the age estimation from facial images is generally quite difficult to predict, this is because aging patterns are personalized for each individual. The aging pattern of a person is hugely dependent on his or her genetic structure, as well as a combination of several extrinsic factors like, health, lifestyle and environmental conditions [25].

Figure 6 shows a sample of images where the trained ResNet50 model made some good estimation. The actual age of each person and the error in estimation are also shown. Figure 7 shows a sample of images with top losses in age estimation.

From the sample images of Figure 6, it can be observed that the ResNet50 model gave a good performance in estimating the ages of each dark skin person, and not exceeding the benchmark *MAE* of $\approx$ 5 obtained in this study. The reverse is the case in the age estimation of the images of Figure 7. The poor estimation is as a result of some extreme conditions of variability observed on the images, such as occlusion, filtered faces, makeup, extreme facial expressions, blurring and light direction.

5. CONCLUSION

In this study, a deep learning approach based on ResNet50 CNN, for improved accuracy in the automatic estimation of the ages of dark skin tone persons using their facial images has been proposed. A facial aging dataset developed from the merging of dark skin persons from the UTKFace, APPA-REAL and BlackFaces datasets was used to train the model, after applying some image augmentation techniques. The experimental results shows that the approach obtains promising outcomes attaining an *MAE* of 5.21 years on the validation set. We plan to investigate the

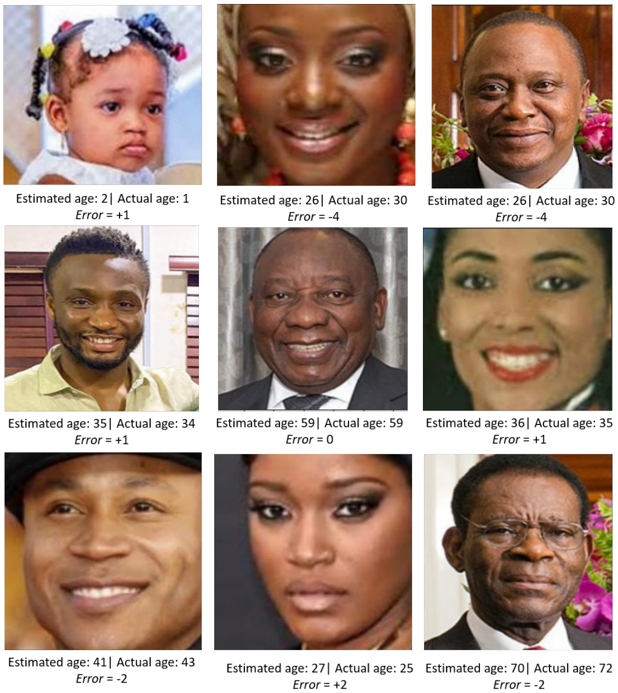

Figure 6. A sample of some accurately estimated facial images

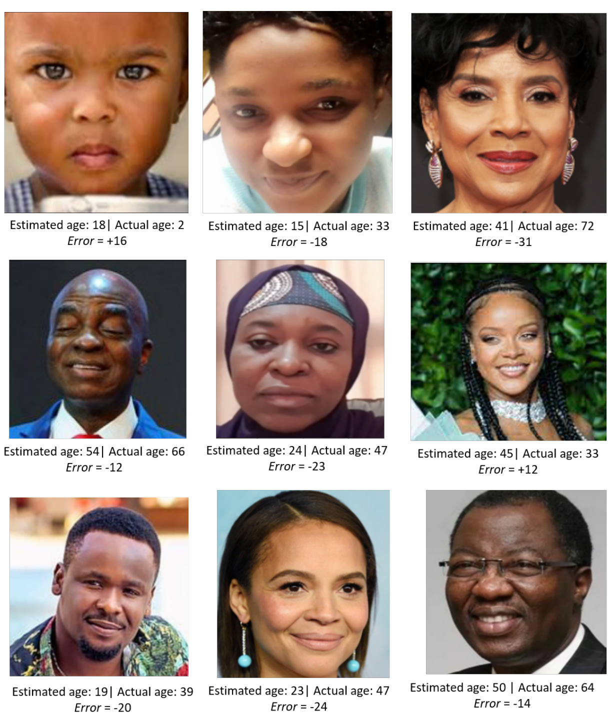

Figure 7. A sample of some estimated images with top errors

CNN model on larger-scale database of dark skin persons, in order to increase the performance and minimize the loss in automatic age estimation of persons with dark skin tones.

References

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[2] H. D. Wehle, "Machine learning, deep learning and ai: what's the difference," *Data Scientist Innovation Day*, pp. 2–5, 2017.

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

[4] O. Abhulimen and E. Ogunti, "Facial age estimation using deep learning: A review," *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, vol. 8, pp. 13 927–13 946, 2021.

[5] 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*, vol. 196, p. 102961, 2020.

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

[7] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Computer vision and image understanding*, vol. 74, no. 1, pp. 1–21, 1999.

[8] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on pattern Analysis and machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002.

[9] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 112–119.

[10] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.

[11] 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 recognition*, vol. 44, no. 6, pp. 1262–1281, 2011.

[12] D. Yi, Z. Lei, and S. Z. Li, "Age estimation by multi-scale convolutional network," in *Asian conference on computer vision*. Springer, 2014, pp. 144–158.

[13] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2015, pp. 534–541.

[14] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE international conference on computer vision workshops*, 2015, pp. 10–15.

[15] A. Anand, R. D. Labati, A. Genovese, E. Munoz, V. Piuri, and F. Scotti, "Age estimation based on face images and pre-trained

convolutional neural networks," in *2017 IEEE symposium series on computational intelligence (SSCI)*. IEEE, 2017, pp. 1–7.

[16] S. Taheri and Ö. Toygar, "On the use of dag-cnn architecture for age estimation with multi-stage features fusion," *Neurocomputing*, vol. 329, pp. 300–310, 2019.

[17] A. Fariza, A. Z. Arifin *et al.*, "Age estimation system using deep residual network classification method," in *2019 International Electronics Symposium (IES)*. IEEE, 2019, pp. 607–611.

[18] H. T. Huynh and H. Nguyen, "Joint age estimation and gender classification of asian faces using wide resnet," *SN Computer Science*, vol. 1, no. 5, pp. 1–9, 2020.

[19] O. Abhulimen, A. Fadamiro, and E. Ogunti, "Cross-racial automatic facial age estimation using deep learning," *International Journal of Emerging Trends in Engineering Research*, vol. 9, no. 9, pp. 6437–6440, 2021.

[20] S. Y. Zhang, Zhifei and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[21] S. Escalera, X. Baró, H. J. Escalante, and I. Guyon, "Chalearn looking at people: A review of events and resources," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 1594–1601.

[22] G. Motroc, "A machine learning model is only as good as the data it is fed." [Online]. Available: https://jaxenter.com/apache-spark-machine-learning-interview-143122.html

[23] P. Dwivedi, "Understanding and coding a resnet in keras." [Online]. Available: https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33

[24] A. S. Maiya, "ktrain: A low-code library for augmented machine learning," *arXiv preprint arXiv:2004.10703*, 2020.

[25] K. Sveikata, I. Balciuniene, J. Tutkuviene *et al.*, "Factors influencing face aging. literature review," *Stomatologija*, vol. 13, no. 4, pp. 113–116, 2011.

V. Osekhonmen Abhulimen is presently pursuing a Masters degree in the Department of Electrical and Electronics Engineering, Federal University of Technology Akure, Nigeria. His study focus includes artificial intelligence, deep learning, image processing and data analytics.

O. Erastus Ogunti is a Professor in the Department of Computer Engineering, Federal University of Technology Akure, Nigeria. He has over 25 years of teaching and study experience. His study interests include fuzzy systems and control, software engineering, computer networks, IoT, AI, image processing and machine learning.