International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

The Influence of Activation Functions in Deep Learning Models Using Transfer Learning for Facial Age Prediction

Gilbert George*[1], Steve Adeshina[2], Moussa Mahamat Boukar [3]

Submitted: 16/09/2023 **Revised**: 18/11/2023 **Accepted**: 30/11/2023

Abstract: Due to its superior performance, the convolutional neural network (CNN) has been extensively applied in image recognition. A facial age prediction technique based on the CNN model is suggested in this research, due to its many uses in the sports industry, access control and age verification systems. The activation function is at the center of the CNN model's complicated hierarchical structure since it has the nonlinear properties that give the deep neural network its accurate artificial intelligence. The ReLu function is one of the best common activation functions, however, it has flaws. It is likely to manifest as the phenomena of neuronal necrosis since the derivative of the ReLu function is always zero when the input value is negative. The impact of the activation function in the CNN model is also examined in this research to address the aforementioned issue. We look at both linear and nonlinear activation functions for facial age prediction using three facial datasets, namely UTK Facial, IMDB-WIKI and CASIA African Facial datasets. In facial age prediction tasks built on the Keras framework, we have investigated Linear and ReLu activation functions in the output layer on four CNN architectures namely VGG16, VGG19, ResNet50 and MobileNet. we compared each model using the two activation functions in the last output dense layer. The experimental results show that MobileNet using ReLu Performed the best with a Mean Absolute Error of 1.03 on the CASIA African Facial Dataset and VGG16 with a Mean Absolute Error of 1.75 using the linear activation function on the UTK Facial dataset, we also demonstrate that the convolutional neural network based on the modified activation function outperforms most state-of-the-art activation models in terms of performance.

Keywords: *Activation Function, facial age detection, Transfer Learning, Regression, Deep learning, Convolutional Neural Networks (CNNs).*

1. Introduction

The task of automatically estimating age has gained popularity due to the many industries it may be applied to, particularly social media and e-commerce. Currently, various e-commerce websites may provide product recommendations to their users based on their historical preferences; with seeming age, more practical recommendations are likely. Facial age prediction is a challenging task due to the inherent complexity and variations in facial appearances caused by factors such as genetics, lifestyle, and environmental factors. Convolutional neural networks (CNNs) have shown remarkable success in various computer vision tasks, including facial age prediction. This paper focuses on utilizing transfer learning with VGG16[1], [2], ResNet50, MobileNet and VGG19[2], which are deep CNN architectures widely adopted for image recognition tasks, to improve the accuracy of facial age prediction and examine the effect of two activation functions (Rectified Linear Unit(Relu) and Linear (identity)) in improving performance. The accuracy of a neural network's predictions depends on the number of layers it employs, but more significantly, on the kind of activation function [3]it employs. No manual details the minimum or maximum number of layers that should be used for neural networks to achieve better results and accuracy, but as a general guideline, it is recommended to utilize at least two layers.

The kind of activation function that should be employed is also not mentioned in any of the literature. Studies and research have shown that utilizing a single or several hidden layers in a neural network lowers the prediction error. In the actual world, errors have non-linear properties that affect the neural networks' capacity to understand false data. Therefore, in a neural network, non-linear activation functions are preferred over linear activation functions. The kind of activation function a neural network uses determines how accurate its predictions are. Non-linear activation functions are the most widely utilized activation functions like ReLu [3]. If an activation function is not defined, a neural network behaves exactly like a linear regression model, with the anticipated output being equal to the input. The same holds if a linear activation function is applied, in which case the output is the same as the input given plus some error. The network can only adapt to linear changes in the input if a linear activation function is utilized since its boundary is linear. Age can be expressed as an integer or a floating-point number, but it also has some coherence, making it possible to compare a person's facial features across a few age groups (for instance, age 20-22) [4]–[8]. Hence, we can look at the age problem as either regression

[1] *Department of Computer Science, Baze University, Abuja, Nigeria*
ORCID ID: 0000-0003-3080-5646
[2] *Department of Engineering, Nile University, Abuja, Nigeria.*
ORCID ID: 0000-0003-0405-5589
[3] *Department of Computer Science, Nile University, Abuja, Nigeria.*
ORCID ID: 0000-0001-8287-6257
** Corresponding Author Email: gilbertgeorge007@gmail.com*

where a specific age is predicted or classification where different images are grouped into different age groups. We focus on age as a regression problem in the current study. We demonstrate this by using the suggested technique on three benchmark datasets, mainly UTK facial image, IMDB-WIKI and Casia African datasets [9]. The ability of Artificial Neural Networks to adapt their behaviour to the changing properties of the system is its most alluring feature. The performance of Artificial Neural Networks has been studied and improved over the past few decades by several researchers and scientists by optimizing training techniques, hyperparameter tuning, learning parameters, or network structures, but activation functions have received less**.** attention. In this paper we investigate the use of linear and non-linear activation functions for facial age predictions, using transfer learning on famous Deep learning architectures.

2. Literature Review

2.1. Artificial Neural Network

An Artificial Neural Network (ANN) is a computational network that mimics the human brain. It has three main layers. input, hidden, and output [10]. The input layer interacts with the environment, the hidden layer processes the inputs obtained by its previous layer, and the output layer is considered the network's output. The output layer is created following the demands of the application. By adding bias to the weighted sum and deciding whether to fire a neuron or not, the activation function determines whether a neuron should fire. Depending on the type of activities, a variety of activation functions, including Rectified Linear Unit (ReLU), SoftMax, sigmoid, etc., can be used. Logistic loss, cross-entropy loss (log loss), categorical cross-entropy loss, exponential loss, square loss, hinge loss, generalized smooth hinge loss, etc. The instability between the predicted value and the actual label is measured by loss functions. By altering the model's parameters in response to the loss function's output, optimizers connect the loss function and model parameters and this in turn can improve accuracy [11], [12].

2.2. Convolutional Neural Network

This is a type of feed-forward neural network. Feature learning and classification are its two components. Convolutional and pooling layers make up the first [13] section, and the fully linked layer makes up the second. A convolutional layer's neurons' primary task is to search for particular features. Spatial size is decreased (just in width and height; not in depth) by the pooling layer. The features are integrated to form a model using the fully connected layers.

2.3. Activation Functions and their needs

When given some data as input, neural networks which are networks of numerous layers of neurons made up of nodes can classify and predict certain types of data. An input layer, one or more hidden layers, and an output layer are present [3], [14]. Every layer has nodes, and each node has a weight that is taken into account while processing data from one layer to the next.

The output signal of a neural network without an activation function would purely be a straightforward linear function or a polynomial of degree one. Although linear equations are straightforward and quick to solve, their complexity is constrained, and they cannot learn and recognize intricate mappings from data [14]. In most cases, a neural network without activation functions performs as a linear regression model with low performance and power. A neural network should be able to model complex types of data, such as photos, videos, audio, speech, and text, in addition to learning and computing a linear function. Hence, why we employ activation functions and artificial neural network techniques like Deep Learning to interpret challenging, highly dimensional, and nonlinear datasets. These techniques require models with multiple hidden layers and intricate architectures for knowledge extraction, which is once more our main objective.

Non-linear functions [3], [14] are those that have more than one degree and exhibit curvature when plotted. A neural network must be able to learn, represent, and analyse any data as well as any arbitrary complicated function that links inputs to outputs. Because they can calculate and learn any given function, neural networks are frequently referred to as universal function approximators. In neural networks, any process that can be imagined can be represented as a functional computation.

To make the network dynamic, add the ability for it to extract intricate and difficult information from data, and express nonlinear convoluted random functional mappings between input and output, we must apply an activation function [3]. Therefore, we can accomplish non-linear mappings from inputs to outputs by introducing non-linearity with the aid of non-linear activation functions to the network. To use the backpropagation optimization strategy to compute errors or losses concerning weights and ultimately optimize weights using Gradient Descend or any other optimization technique to reduce errors, an activation function must be differentiable.

2.4. Categories of Activation Functions for Regression

The most crucial building blocks of a neural network are the net inputs, which are transformed into output results known as unit activation by applying a function called the activation function, threshold function, or transfer function, which is a scalar-to-scalar transformation. Squash functions include permitting a neuron's output at a constrained amplitude and across a constrained range [3]. A squashing function reduces the output signal's amplitude to a finite value. The threshold-based classifier must be taken into account when using an activation function because it determines whether the value of the linear transformation must activate the neuron or not. Alternatively, we can say that a neuron is activated if the input to the activation function is greater than a threshold value, or if it is deactivated otherwise [14]. In that instance, the output is not used as the next layer's input. The most used activation functions for regression are.

A. Linear

The input has a direct proportional relationship with the linear activation function. Because the binary step function lacks an x component, its primary flaw is that it has zero gradients. The application of a linear function can get rid of that. It can be characterized as.

$$F(x) = ix \quad (1)$$

Where *i* is the constant which the user can set.

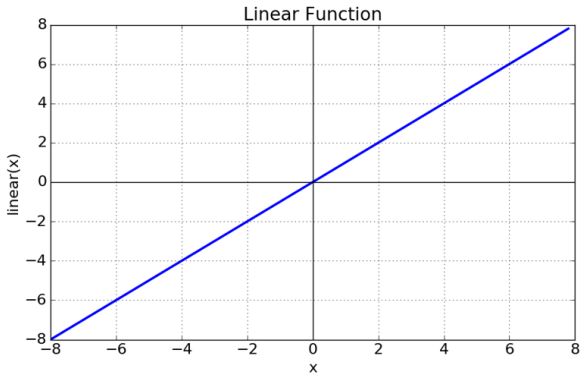

Fig.1. shows the Linear activation function graph.

Using a linear function has limited advantages because the neural network's error would not be reduced. After all, the gradient would be the same for each iteration. Additionally, the network would not be able to extract intricate patterns from the data. Therefore, linear functions are appropriate for straightforward tasks and situations where interpretability is not necessary.

B. ReLu Activation Function

The non-linear activation function known as ReLU, or corrected liner unit, is frequently employed in neural networks. The advantage of utilizing the ReLU function is that not every neuron is triggered simultaneously. This suggests that a neuron would not stop functioning until its linear transformation's output is zero. Mathematically, it can be expressed as

$$F(x) = \max(0, x) \quad (2)$$

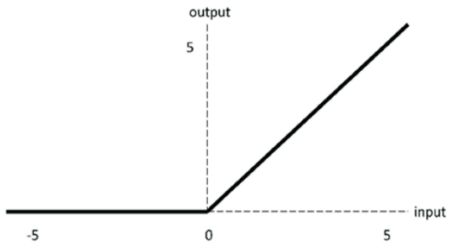

Fig2. shows the ReLu activation function graph

ReLU is more effective than other functions since only a portion of the neurons are engaged at once rather than all of them simultaneously. In rare circumstances, the gradient value is zero, which prevents the neural network training's backpropagation step from updating the weights and biases.

2.5. Related Works

wang et al [3] identified a problem with relu that will manifest as the phenomena of neuronal necrosis since the derivative of the ReLu function is always zero when the input value is negative. The impact of the activation functions (sigmoid, tanh, ReLu, leaky ReLus and softplus–ReLu) in the CNN model was examined in their research to address the issue. A novel piecewise activation function was also suggested by the CNN model's activation function design approach. In facial expression recognition tasks built on the Keras framework, five commonly used activation functions which are the sigmoid, tanh, ReLu, leaky ReLu, and softplus-ReLu have been investigated and compared. The experimental findings on the JAFFE and FER2013 facial expression databases demonstrate that the convolutional neural network based on the modified activation
function LS-ReLu is effective.

$$f(x) = \begin{cases} \frac{x}{1+|x|} & x \leqslant 0 \\ \max(0, x) & k \geqslant x > 0 \\ \log(a*x+1) + |\log(a*k+1) - k| & x > k \end{cases}$$

Fig3. shows the new activation function proposed

From the experiments carried out their new activation function had an accuracy of 99.91% while sigmoid had an accuracy of 96%, Tanh has an accuracy of 94%, ReLu had 95% and softplus-relu had 57% .

Anami & Saganal [14] used three deep learning models which are ResNet, MobileNet, and EfficientNet to classify indoor scenes and how activation functions affect categorization accuracy. In the work, three activation functions (tanh, ReLU, and sigmoid) are used. The authors divided the MIT-67 indoor dataset into scenes with and without people and the classification accuracy is tested. The approach is innovative in that it divides the dataset into two groups—one with humans and one without based on the geographical layout and segregation. Efficient Net has performed well among the three pre-trained models using the sigmoid activation function this is no doubt surprising as sigmoid is relatively better for binary classifications.

Su, suggested that [15] By adding more convolution kernels, a better can feature extraction method can be achieved, to help the model converge more quickly, the data collection can be suitably expanded. The experiment's impact of various activation functions is compared, and the images are categorized using the SoftMax approach. The classification of images of normal and diseased thyroid tissue is then done using the improved model. The model produces better categorization results, with an average classification accuracy of 96.6%.

In Li et al [16] the primary goal of the research is to develop a straightforward convolutional neural network classification model for three popular datasets. Comparative experiments are carried out by altering experimental settings (such as activation function, pooling method, output size, etc.), and the impact of various factors on classification and recognition accuracy is examined using various data sets. According to experimental findings, maximal pooling and the relu activation function are more suited for categorizing image data sets.

In [17] Kothari et al. provided a comparison analysis of a custom CNN model to a pre-trained model. The article employs two models, the first of which is a pre-trained MobileNet model and the second of which is a customized

CNN model. The outcomes of each model that was used are compared using prediction accuracy and MAE. MAE is used as a performance metric for age whereas accuracy is used for gender and ethnicity. The UTKFace dataset is used to analyze the models.

In this paper [7] the authors introduce a two-part age estimate system to close the age estimation gap. A specially created gender classifier that separates males and females is the first element. The second module uses two different models for age estimation. Model B is trained exclusively on photographs of men, while Model A is trained exclusively on images of women. The system takes an input image, determines the gender of the face, and then sends the image to the appropriate model depending on the gender label that is predicted. The age estimation models were updated to match the specifics of each case from the Visual Geometry Group (VGG16) networks. Individually, the models produce accuracy levels of more than 85%, and the system achieves an 80% total accuracy. To validate the performance on unseen data, the proposed system is trained and evaluated on the UTK-Face dataset and cross-validated on the FG-NET dataset.

In [18] Akhand et al. calculate age from frontal, semi-frontal, and profile photos. They also examined several deep learning models (ResNet18, ResNet34, ResNet50, Inceptionv3 and Dense Net) and they obtained a mean absolute error of 2.64 using ResNet34.

3. Methodology

3.1. Datasets

For facial age prediction, a dataset comprising facial images with corresponding age labels is required for that we selected three datasets for this experiment, which are UTK faces, CASIA African Facial, and IDMB_WIKI datasets which are openly available from the following link.

I. https.//www.kaggle.com/datasets/abhikjha/utk-face-cropped
II. https.//www.kaggle.com/datasets/abhikjha/imdb-wiki-faces-dataset
III. http.//www.idealtest.org/#/datasetDetail/24

3.1.1. UTK Facial Dataset

One of the most used datasets created for age determination from facial photos is the UTK image benchmark. The dataset is one of the largest-scale facial dataset with a wide age range (from 0 to 116 years old), it has over 20,000 facial photos with annotations for age, gender, and ethnicity. There is a wide range of poses, facial expressions, lighting, occlusion and resolution. Several tasks, including face detection, age estimation, age progression and regression, and landmark localization could be performed using this dataset. We used pictures from ages 5-30, each consisting of 450 images which made a total of 11400 images.

Fig4. image showing example of facial images in the UTK dataset

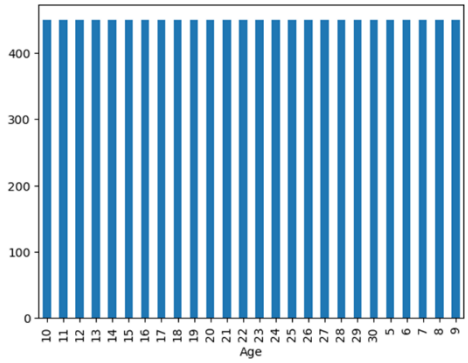

Fig 5. image showing the distribution of from ages 5-30 from the UTK Facial dataset used for the experiments

Table 1. Age distribution from the UTK Facial dataset used for the experiments.

Age	*Amount*
5	450
6	450
7	450
8	450
9	450
10	450
11	450
12	450
13	450
14	450
15	450
16	450
17	450
18	450
19	450
20	450
21	450
22	450
23	450
24	450
25	450
26	450
27	450
28	450
29	450
30	450

3.1.2. IMDB-WIKI

This is the largest publicly accessible dataset of face photos with gender and age labels. We used a total of 3659 images from the
Wiki dataset for our experiments.

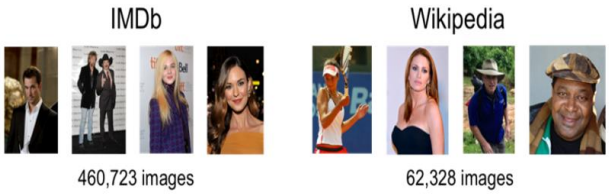

Fig6. images showing example of facial images in the IMDB-WIKI dataset

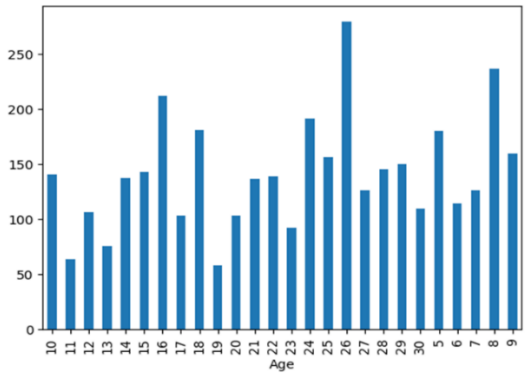

Fig7. image showing the distribution of from ages 5-30 from the WIKI Facial dataset used for the experiments

Table 2. The distribution of the ages between 5-30 from the WIKI dataset used for the experiment.

Age	*Amount*
5	180
6	114
7	126
8	236
9	159
10	140
11	63
12	106
13	75
14	137
15	143
16	212
17	103
18	181
19	58
20	103
21	136
22	139
23	92
24	191
25	156

3.1.3. CASIA African facial Dataset

The images in the database were taken in Kaduna, Nigeria, which is an African country. Approximately 1150 individuals took part in the practice of capture. The dataset has different ethnic Nigeria tribes, for this experiment, we made use of the Hausa ethnic group. The images in the database comprise a total of 38,546 images from 1,183 subjects. We made use of a total of 10921 facial images distributed to ages 10-30, table 2 shows the age distribution of the dataset.

Fig8. images showing example of facial images in the CASIA African dataset.

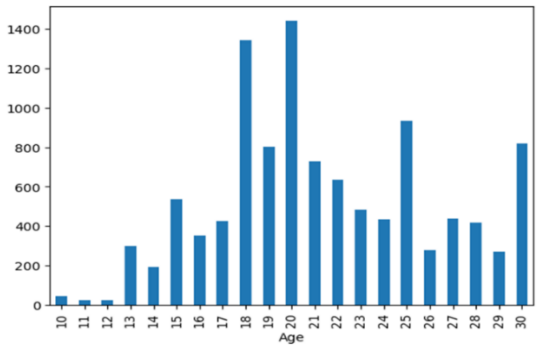

Fig9. image showing the distribution of from ages 10-30 from the CASIA facial dataset used for the experiments.

Table 3. The distribution of the ages between 10-30 from the CASIA Facial dataset used for the experiment.

Age	*Amount*
10	46
11	23
12	23
13	297
14	193
15	538
16	352
17	425
18	1242
19	804
20	1441
21	729
22	636
23	483
24	435
25	932
26	280
27	483
28	417
29	269
30	818

3.2. Training

The input image was rescaled to 120 X 120 pixels and we used the optimization function adam, and a dropout of the fully connected layer of 0.2 which was to regularize the network while training. We downloaded the various deep learning models from the TensorFlow keras library, and then we froze the convolutional layers and added our fully connected layers for the prediction to take place, we used three dense layers which had 50, 20 and 1 outputs hence for prediction. We then specified the activation function at the last dense layer which would be either ReLu or linear for the experiments.

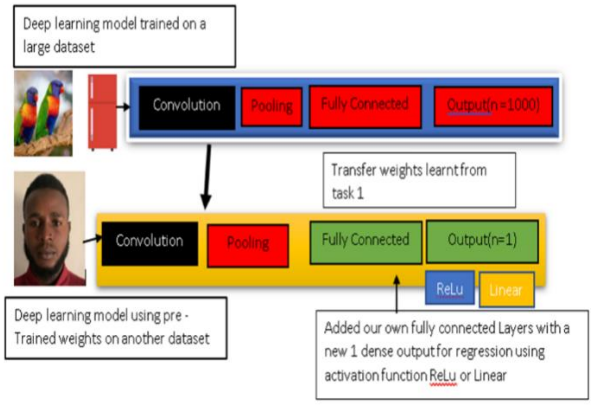

Fig10. Image showing the methodology for the experiments using transfer learning

We also used the early stopping function to stop training once there are no improvements in accuracy, the RGB image is passed to the convolution area for feature extraction (pre-trained) and then passed to the fully connected layer which is then passed to the output layer where the probability of the output layer is calculated. We used a total of 11700 images which had 450 images in each age label i.e. 5-30 for the UTK dataset, we used a total of 10921 images from the CASIA dataset and lastly, we used 3659 images from the WIKI dataset. we then used the Python function" train_test_split()" to split the X images and labels into training and test set, using the ration of 70.30 for facial datasets.

3.2.1. Transfer Learning Workflow

The transfer learning workflow involves the following steps.

a. Preparing the dataset. Splitting the dataset into training, validation, and testing sets.

b. Pre-training. Utilizing the pre-trained VGG16 or VGG19 model, which is trained on a large-scale image classification task such as ImageNet, to extract features from facial images.

c. Fine-tuning. Modifying the last few layers of the pre-trained model or adding new fully connected layers to adapt the model for age prediction. These layers are initialized randomly and trained on the specific age prediction task as seen in fig6.

d. Training and Evaluation. Training the modified model using the training set and evaluating its performance on the validation set. Iterative optimization techniques such as gradient descent and backpropagation are applied.

e. Testing. Assessing the final model's accuracy by evaluating it on the testing set. Metrics such as mean absolute error (MAE) and mean squared error (MSE) can be used to quantify the prediction accuracy. In this work, we utilize the MAE as our evaluation metric.

$$MAE = \frac{\sum_0^N |\bar{y} - y|}{N} \quad (3)$$

In a regression model, R-Squared (also known as R^2 or the coefficient of determination) is a statistical metric that quantifies how much of the variance in the dependent variable can be accounted for by the independent variable. R-squared, or the "goodness of fit," measures how well the data match the regression model. A greater r-squared shows that the model can explain more variability. R-Squared ranges from 0-1.

$$R2 = \frac{SSr}{SSt} \quad (4)$$

where SSt represents the total sum of squares and SSr represents the residual sum of squares. We would be using both MAE and R2 to determine how well the models performed. The suggested age prediction system using deep CNNs and transfer learning is shown in Fig 6. One of the deep learning models would be attached to the base model with fully connected layers and an output. in Fig.6, a deep CNN model is initially imported with its pre-trained weights. The architectures were developed for the ImageNet object classification problem, so each deep CNN has 1000 units in its last layer. Hence, we remove its fully connected layers and added our custom fully connected layer with one output for age prediction using an activation function of either ReLu or Linear.

3.2.2. Experimental Setup

Using Jupyter Notebook, we put the suggested strategy into practice. Intel 12 gen I-core 7 with 10 cores has been used for training the networks. For regression, the learning rate was set to 0.001 and the momentum to 0.9. We used mean absolute error (MAE) for regression in the performance metric. Human age exhibits some coherence, and facial shape can vary slightly but not enough for the human eye to notice. For instance, the age patterns of people who are 21 and 23 years old may be similar, and in some instances, it may be difficult for a human to tell them apart. On the other hand, for regression, we minimized MAE.

4. Experimental Results

We show the experimental results, using Linear and ReLu for the activation function of the last output layer of our transfer learning model, mainly VGG16, VGG19, MobileNet and ResNet-50, to investigate the mean absolute error of the regression process. These popular activation functions are frequently used in neural network topologies, especially in regression processes. By current best practices, we computed the gradient and updated the weight matrix using Adam Optimizer. We trained with a constant velocity of 0.9 throughout.

Table 4. The different deep learning models performance

Model	*Activation Function*	*MAE*	*R^2*	*Dataset*
VGG16	Relu	1.45	0.80	Casia African Dataset
VGG16	Linear	1.21	0.85	Casia African Dataset
VGG19	Relu	1.32	0.83	Casia African Dataset
VGG19	Linear	1.34	0.84	Casia African Dataset
RESNET 50	Relu	3.13	0.21	Casia African Dataset
RESNET 50	Linear	3.19	0.28	Casia African Dataset
MobileNet	Relu	**1.03**	0.89	Casia African Dataset
MobileNet	Linear	1.10	0.88	Casia African Dataset
VGG16	Relu	4.23	0.52	IMDB-Wiki Dataset
VGG16	Linear	4.19	0.52	IMDB-Wiki Dataset
VGG19	Relu	4.33	0.50	IMDB-Wiki Dataset
VGG19	Linear	4.23	0.51	IMDB-Wiki Dataset
RESNET 50	Relu	18.03	-5.04	IMDB-Wiki Dataset
RESNET 50	Linear	5.67	0.19	IMDB-Wiki Dataset
MobileNet	Relu	**3.96**	0.56	IMDB-Wiki Dataset
MobileNet	Linear	4.01	0.54	IMDB-Wiki Dataset
VGG16	Relu	2.38	0.77	UTK Facial Dataset
VGG16	Linear	**1.75**	0.85	UTK Facial Dataset
VGG19	Relu	17.48	-5.47	UTK Facial Dataset
VGG19	Linear	2.22	0.80	UTK Facial Dataset
RESNET 50	Relu	4.80	0.39	UTK Facial Dataset
RESNET 50	Linear	4.60	0.38	UTK Facial Dataset
MobileNet	Relu	2.02	0.81	UTK Facial Dataset
MobileNet	Linear	2.01	0.80	UTK Facial Dataset

The results obtained from the experiment show that MobileNet Transfer Learning Model outperforms the other models in two datasets with a MAE of 1.03 using the ReLu activation function in the CASIA African Dataset and 3.96 using the ReLu for IMDB-WIKI dataset. VGG16 using linear activation function performed best for UTK facial dataset with a MAE of 1.75. From the results obtained so far, we see, the importance of the activation function has been seen, with each model's performance using a different activation function.

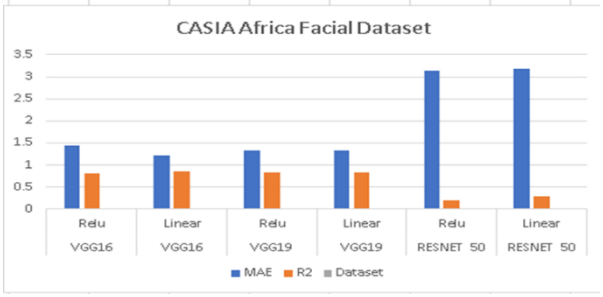

Fig11. Image showing the graphical illustration of the different model's performance for the CASIA Africa Dataset.

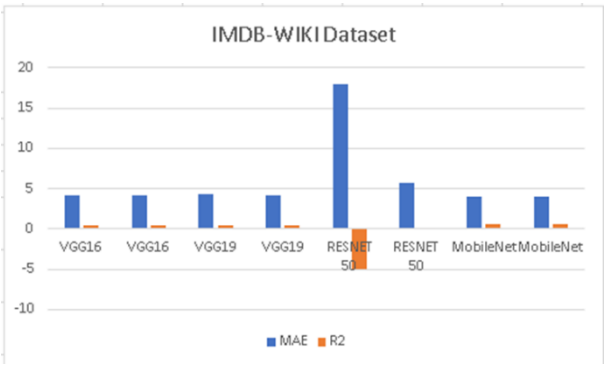

Fig12. Image showing the graphical illustration of the different model's performance for the IMDB-WIKI Dataset.

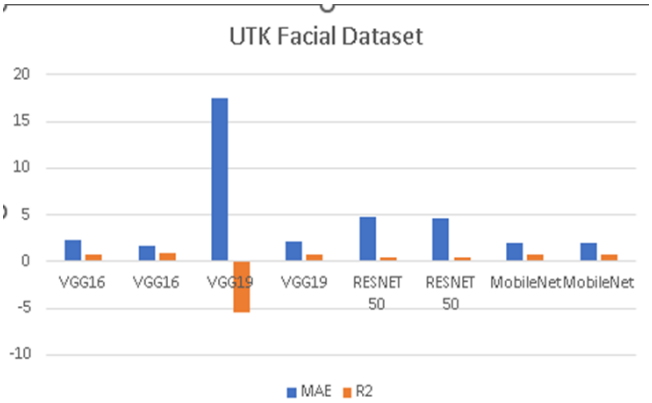

Fig13. Image showing the graphical illustration of the different model's performance for the UTK Facial Dataset.

5. Conclusion

Prediction of Facial ages is important in areas such as sports, election verification systems, medical sciences and forensics. We have used the concepts of transfer learning in a task to understand the context of a person's biological age. Using ResNet50, MobileNet, VGG16 and VGG19 CNN models we investigated the performance of these models using two activation functions, ReLU and Linear. We conducted separate experiments on facial age prediction using three different farcical datasets with people from different ethnicity using the UTK Facial, Casia Africa and IMDB-Wiki datasets. MobileNet with ReLu activation

function, has produced a superior result with a mean absolute error of 1.03. Using testing, the number of epochs is set at 20. The investigation is innovative because it was not conducted by prior researchers regarding facial age prediction, and the comparison with the leading-edge publication has shown this.

6. Conflicts of interest

The authors declare no conflicts of interest.

References.

[1] "VGGNet-16 Architecture. A Complete Guide | Kaggle." https.//www.kaggle.com/code/blurredmachine/vggnet-16-architecture-a-complete-guide (accessed May 07, 2023).

[2] S. Mascarenhas, M. A.-2021 I. C. on, and undefined 2021, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," *ieeexplore.ieee.org*, Accessed. May 07, 2023. [Online]. Available. https.//ieeexplore.ieee.org/abstract/document/9687944/

[3] Y. Wang, Y. Li, Y. Song, and X. Rong, "The influence of the activation function in a convolution neural network model of facial expression recognition," *Applied Sciences (Switzerland)*, vol. 10, no. 5, 2020, doi. 10.3390/app10051897.

[4] J. Prajapati, A. Patel, and P. Raninga, "Facial Age Group Classification," *IOSR Journal of Electronics and Communication Engineering*, vol. 9, no. 1, pp. 33–39, 2014, doi. 10.9790/2834-09123339.

[5] K. R, "DEEP LEARNING FOR AGE GROUP CLASSIFICATION SYSTEM," *INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES*, vol. 4, no. 2, pp. 16–22, Dec. 2018, doi. 10.29284/IJASIS.4.2.2018.16-22.

[6] M. F. Mustapha, N. M. Mohamad, G. Osman, and S. H. A. Hamid, "Age group classification using Convolutional Neural Network (CNN)," in *Journal of Physics. Conference Series*, 2021. doi. 10.1088/1742-6596/2084/1/012028.

[7] V. Raman, K. Elkarazle, and P. Then, "Gender-specific Facial Age Group Classification Using Deep Learning," *Intelligent Automation and Soft Computing*, vol. 34, no. 1, 2022, doi. 10.32604/iasc.2022.025608.

[8] A. Tunc, S. Tasdemir, M. Koklu, and A. C. Cinar, "Age group and gender classification using convolutional neural networks with a fuzzy logic-based filter method for noise reduction," *Journal of Intelligent and Fuzzy Systems*, vol. 42, no. 1, 2022, doi. 10.3233/JIFS-2191206.

[9] J. Muhammad, Y. Wang, C. Wang, K. Zhang, Z. Sun, and S. Member, "CASIA-Face-Africa. A Large-scale African Face Image Database", Accessed. Jul. 04, 2023. [Online]. Available. http.//www.cripacsir.cn/dataset/

[10] M. S. Islam, I. Hussain, M. M. Rahman, S. J. Park, and M. A. Hossain, "Explainable Artificial Intelligence Model for Stroke Prediction Using EEG Signal," *Sensors (Basel)*, vol. 22, no. 24, Dec. 2022, doi. 10.3390/s22249859.

[11] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G. Z. Yang, "XAI—Explainable artificial intelligence," *Sci Robot*, vol. 4, no. 37, Dec. 2019, doi. 10.1126/SCIROBOTICS.AAY7120.

[12] K. Mohammed and G. George, "IDENTIFICATION AND MITIGATION OF BIAS USING EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) FOR BRAIN STROKE PREDICTION," *Open Journal of Physical Science (ISSN. 2734-2123)*, vol. 4, no. 1, pp. 19–33, Apr. 2023, doi. 10.52417/OJPS.V4I1.457.

[13] V. Sheoran, S. Joshi, and T. R. Bhayani, "Age and Gender Prediction Using Deep CNNs and Transfer Learning," in *Communications in Computer and Information Science*, 2021. doi. 10.1007/978-981-16-1092-9_25.

[14] B. S. Anami and C. V. Sagarnal, "Influence of Different Activation Functions on Deep Learning Models in Indoor Scene Images Classification," *Pattern Recognition and Image Analysis*, vol. 32, no. 1, 2022, doi. 10.1134/S1054661821040039.

[15] L. Su, "Intelligent Strategy by Deep Learning for Thyroid Case Images Classification," in *Proceedings - 2020 13th International Conference on Intelligent Computation Technology and Automation, ICICTA 2020*, 2020. doi. 10.1109/ICICTA51737.2020.00076.

[16] J. Li, C. Nanchang, and K. Song, "Research on image classification based on deep learning," in *Proceedings - 20th IEEE/ACIS International Summer Conference on Computer and Information Science, ICIS 2021-Summer*, 2021. doi. 10.1109/ICIS51600.2021.9516872.

[17] S. Kothari, S. Deshmukh, and S. Mehta, "Comparison of Age, Gender and Ethnicity Prediction Using Traditional CNN and Transfer Learning," in *2022 13th International Conference on Computing Communication and Networking Technologies, ICCCNT 2022*, 2022. doi. 10.1109/ICCCNT54827.2022.9984552.

[18] M. A. H. Akhand, Md. Ijaj Sayim, S. Roy, and N. Siddique, "Human Age Prediction from Facial Image Using Transfer Learning in Deep Convolutional Neural Networks," pp. 217–229, 2020, doi. 10.1007/978-981-15-3607-6_17.