Deep Learning and Genetic Algorithms Approach for Age Estimation Based on Facial Images

Idowu T. Aruleba and Yanxia Sun*

Department of Electrical and Electronic Engineering Science,
University of Johannesburg, Johannesburg 2006, South Africa
Email: ysun@uj.ac.za
*Corresponding author

Manuscript received September 13, 2023; revised May 29, 2024; accepted July 4, 2024; published 18 November 2024

***Abstract*—Age estimation of an individual facial image has become a fascinating research topic due to its wide range of applications in real-world scenarios. In literature, significant research has been done using various techniques and approaches; these studies gave a good outcome, making this area of research a state-of-the-art area for research and giving space for more enhanced accuracy. This study aims to improve age estimation using facial biometric features by applying deep learning and transfer learning techniques. By doing this, the research aims to solve the problem of inaccurate age estimation based on facial images.**

This study proposed using an improved Genetic Algorithm coupled with a Convolutional Neural network (CNN) model (EfficientNet-B0) to estimate age on the Adience benchmark dataset. This study applied a Genetic algorithm for the selection of hyperparameters to help achieve an optimal result. The EfficientNet-B0 + Genetic Algorithm (GA) model's estimation accuracy yielded a good accuracy of 86.5%, which shows an improvement compared to work in the literature that used other models.

***Keywords*—age estimation, feature extraction, deep learning, machine learning, neural networks**

I. Introduction

Numerous researchers have broadly researched age estimation in the Computer Vision field study. Age estimation can determine one's age by taking advantage of different biometric features [1, 2]. Age estimation has been approached using different determinants such as Gait [3] and Sclera [4]. This research seeks to extract features based on biometrics from an individual's face, as the human face remains the most assertive and informative trait for a biometric recognition system [5]. These biometric features include skin texture, wrinkles, skin colour, hair colour, and freckles. Previous studies have shown promising results when applied in a real-world scenario, such as security alerts, digital signage, access control and monitoring, and enhanced human-computer Interaction [6–9]. These impactful applications and challenges in getting a more accurate solution have brought about numerous active research in this field. Some challenges have been identified which must be investigated to get a more accurate result. Firstly, regarding the facial ageing of a person in different age groups (i.e., face progression and regression), no dataset has the facial image of a person in a different age group [10]. Secondly, a problem has been gathering huge datasets to train machine learning models.

Furthermore, facial images in the currently available datasets are not highly definition regarding occlusions, lightening, orientation, poses, and facial expression [6, 7]. These different factors have been obstacles to facial images being trained on models. A suitable approach to solving this problem is using a traditionally handcrafted machine learning algorithm or extracting relevant facial features vital in ageing information [7]. However, this handcrafted method requires much human effort and engineering and is time-consuming. The introduction of deep learning algorithms has aided in automatically learning features in an ordered manner at different phases. This powerful feature has made it robust compared to the traditional machine learning algorithm and solved the problem of handling big data [6, 11]. Convolutional Neural Network (CNN) is one of the deep learning-based techniques that can build an end-to-end model (i.e., from feature extraction to classification/regression process), which has also been extended to the age estimation task. However, training CNN from scratch could be difficult because it has numerous trainable parameters.

Transfer learning in this framework is simply the transfer of knowledge from a pre-trained model that has already been learned from a related task to a new task [12]. The advent of machine learning led to the introduction of pre-trained models such as AlexNet [13], Residual Neural Network (ResNet) [14], GoogleNet [15], SqueezeNet [16], Visual Geometry Group Network (VGG-Net) [17], The Extreme Inception (XceptionNet) [18] and EfficientNet [19]. These models have already been applied to many tasks and have proven to be efficient, save time and computational power, and limit overfitting, especially when the dataset is not huge. However, these models improve through discreet and proper fine-tuning to obtain a significant result. The proposed work further examines a variant of the EfficientNet ConvNet model. In addition, to explore and introduce meta-heuristic optimisation techniques to accurately fine-tune the hyperparameters to train the ConvNet. The technique this study will apply is the Genetic algorithm, which was selected because of its robustness in selecting efficient parameters. The proposed work will be compared with existing literature to fathom if the approach achieves a state-of-the-art result.

The novelty of this study is in the comprehensive integration of various biometric features, including skin texture, wrinkles, skin colour, hair colour, and freckles in age estimation to achieve a more precise age prediction. Furthermore, the study leveraged CNN and transfer learning as innovative approaches to improve the efficiency and generalisation of age estimation and save time and computational resources.

By comparing its proposed approach with existing methods in the literature, the study aims to establish whether its methodology achieves state-of-the-art results. This comparative analysis evaluates and validates the proposed solution's effectiveness and competitiveness. The research findings can also be valuable in real-world scenarios such as security alerts, digital signage, access control, and human-computer interaction.

The Age Estimation using the techniques applied in this study can be applied in several practical applications. Firstly, age estimators can be integrated into facial recognition and security systems to enhance the identification of underage individuals who may attempt to access age-restricted areas or content. Furthermore, the technology can be applied in content creation, marketing, and advertising. Companies can use age estimation to target a specific group of people for their content or products based on age. Businesses can use the data from age estimation to gather valuable insights about the age distribution of their customers, which may help them to refine further or develop their strategies or products.

Section II describes work related to our approach. Section III explains the materials and methods utilised in this research, including EfficientNet variant B0 and genetic algorithm. The results of the experiments and discussion are presented in Section IV, while Section V concludes the paper.

II. Literature Review

Several studies have discussed and presented methods for age estimation [20–22]. For example, Thaneeshan *et al.* [23] proposed an efficient CNN model to estimate age and gender using the multi-label classification approach. They also modelled a visualisation technique to analyse their classification results and identify landmark facial regions. To evaluate that model, they trained and tested their model using the Adience dataset. The model achieved an accuracy of 57.60% on age estimation. In 2020, Gyawali *et al.* [24] used facial images for age identification by removing unnecessary features other than faces from the image dataset using Multi-task Cascaded Convolutional Neural Network (MTCNN). They further employ crop techniques and augmentation to enhance their approach performance. They built the model using VGG-Face with the help of the transfer learning technique; they evaluated their model using the Adience benchmark dataset.

Taheri and Toygar [25] proposed a new architecture of Deep Neural Network for age estimation named Direct Acyclic Graph CNN (DAG-CNN). DAG-CNN exploits multi-stage feature fusion. The architecture was built using two instances, firstly by adding multi-scale output connections to the fundamental backbone of two pre-trained models (VGG-16 and GoogleNet). Secondly, DAG-CNN fuses age estimation's extraction and classification phases into a single learning process, utilises multi-scale features, and automatically executes score-level fusion for numerous classifiers. The study achieved a lower Mean Absolute Error (MAE) on MORPH-2 and FGNet datasets than in previous studies. Deng *et al.* [26] reported that most previous works only learn a single age feature while neglecting the gender and ethnicity features. This approach greatly impacts the age pattern. They proposed a compact multi-feature learning and fusion age estimation method. Using a regression-ranking age estimator, they examined the age estimation task to convert fusion features into exact age.

To perform face detection and age estimation subsequently in real-time, Castellano *et al.* [7] proposed a fully automated system using "You Only Look Once" YOLOv5 for face detection and EfficientNet for classification. They trained their work on a robust novel MIVIA Age dataset. In contrast, a study by Aruleba and Viriri [6] utilised the fine-tuned CNN-based EfficientNet architecture (B0 – B6) to classify ages. This application was the first of its kind, as reported in their work. They used the FaceNet model to perform face detection, facial landmark detection, and alignment preprocessing. In their experimentation, the EfficientNet-B4 variant performed best on both datasets used. The EfficientNet-B4 gave a state-of-the-art result showing that the model could perform better if more experimentations, such as advancement in the preprocessing of the images and learning facial representation, were implemented. Dagher and Barbara [27] designed a stratified model consisting of pre-trained models with two classes to generate high age estimation accuracy. They coupled GoogleNet with an optimum age gap that can organise the face images in the age group. However, these methods could not handle unconstrained imaging conditions with variations in poses, expression, illumination, etc. Due to the lack of vast and reliable annotated datasets for training deep neural networks, Greco *et al.* [28] proposed an effective training method of CNNs for age estimation based on Knowledge distillation. Knowledge distillation is a method in which a small CNN model is trained efficiently with reduced resources such as memory and processing time [29]. The method consists of extracting the class probability vector called the teacher, and the vectors were later adopted to train the smaller model called the student. The model was trained using LFW+, LAP 2016, and Adience datasets. Their model performed speedily up to 15 times the average training speed.

Meanwhile, limitations have hindered efficient age estimation, lacking vast and reliable annotated datasets for training deep neural networks and pre-trained models. Therefore, this study intends to fill this gap using transfer learning by employing a pre-trained model to investigate further these limitations and how to improve them.

III. Materials And Methods

A. Datasets

This study uses two different datasets, UTKFace and Adience. These two datasets were combined to produce a single large dataset. The datasets are described as follows.

1) Adience

The Adience dataset is a benchmark of facial photos gathered with the intention of the outcome being as true as possible to the challenges of real-world imaging conditions. The dataset is made up of facial images that are captured by the cameras from mobile devices such as smartphones or tablets and includes varying images such as occlusions, poor resolutions, expressions, and also out-of-plane pose variations. It has 26,580 images taken from a total of 2,284 subjects. The dataset is already grouped and labelled into eight classes (0–2, 4–6, 8–13, 15–20, 25–32, 38–43, 48–53, 60+). The dataset was taken in the wild with diverse faces possessing variations in poses, lightning, appearance, and more. Fig. 1 shows the age-group distribution from the Adience dataset.

The Adience dataset is more comprehensive since it is an unconstrained benchmark for combined age and gender estimate, unlike other datasets (like Morph II) where the face photos are captured in a controlled environment. Testing this benchmark can more accurately reflect the performance because our goal is to enhance the precision and accuracy of age estimation.

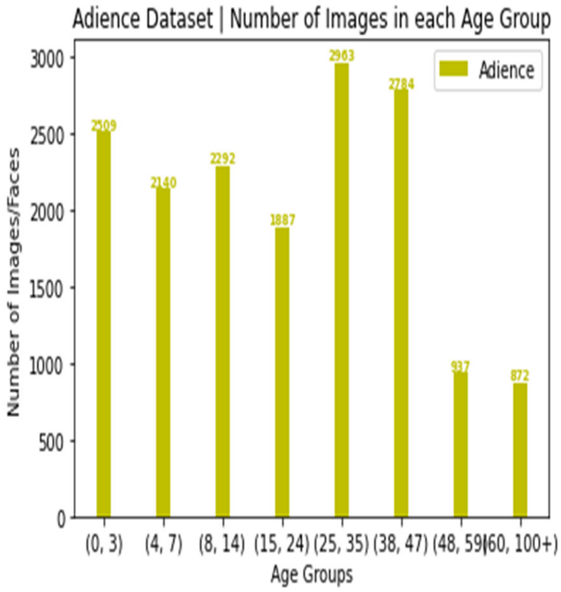

Fig. 1. Adience dataset distribution.

2) *UTKFace*

UTKFace has over 20,000 images with age, gender, and ethnicity annotation. Like Adience, UTKFace also has huge variations in appearance, poses, occlusion, resolution, etc. and can be used for several purposes such as age estimation, landmark localisation, face detection and age regression or progression. The images in the UTKFace dataset are labelled based on gender, age, and ethnicity. Fig. 2 shows the age distributions in the UTKFace dataset.

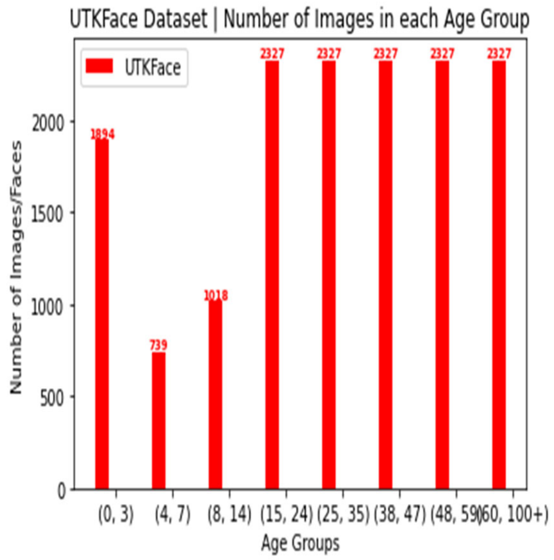

Fig. 2. UTKFace dataset distribution.

3) *UTKFace + Adience*

The two datasets were combined using the Adience default age classes. For UTKFace ages that did not fall into a class, we moved them to the nearest class. For instance, ages like two years moved to the 0–2 class, three years to the 0–2 class, 13 years to the 8–12 class, 29 years to 25–32years and so on. Fig. 3 shows the distribution of age after the combination.

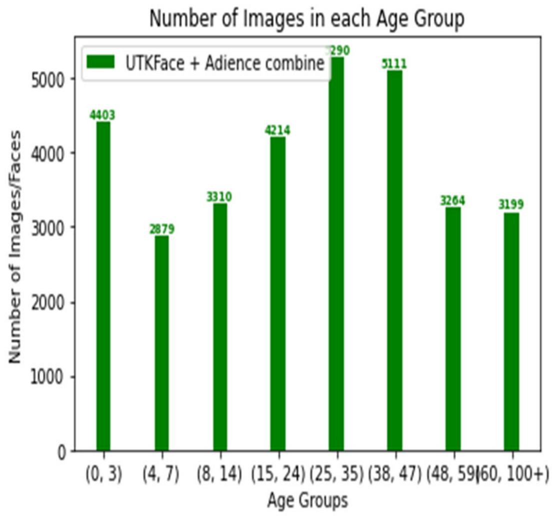

Fig. 3. UTKFace + Adience dataset distribution.

B. *EfficientNet B0 and Architecture*

In this research, we used the EfficientNet B0, a variant of the EfficientNet covNet developed from the neural network. It was developed by applying numerous objectives of neural architecture search that optimises both floating-point operations and accuracy. The B0 variant outperformed the ResNet-50 and DenseNet-169 when trained on imageNet with fewer parameters of 5.3 million compared to the 26 million and 14 million parameters, respectively. EfficientNet models were scaled from the B0 variants as the baseline model; it uses various compound coefficients, and all scaled variants have consistently reduced parameters and FLOPS by order of magnitude as compared to other ConvNets, order of magnitude such as up to 8.4x parameter reduction and up to 16x FLOPS reduction.

C. *Genetic Algorithm (GA)*

To optimise the model and achieve optimal results, we must employ a robust algorithm such as Evolutionary Algorithms (EA) that searches, solves, optimises, and achieves an optimal solution. This EA can be subdivided into different variations and operations with their strength; these sub-divisions are Genetic Algorithm (GA), Evolution Strategies, Differential Evolution, Genetic Programming, and Estimation of Distribution Algorithm. This research would employ the Genetic Algorithm to help cover a wide search space, including the Convolution layers, the fully connected layer, dropout rate, filter size, and pooling size.

A Genetic Algorithm is a population-based stochastic algorithm whereby every solution correlate to a Chromosome, and each parameter represents a gene. GA has three major operators: selection, crossover, and mutation [30]. Compared to random local search, GA has proven more robust as it focuses on the optimum solution instead of trying random solutions [31]. The genetic algorithm uses method sets summarised into major steps: initial population, selection, crossover, and mutation [32]. The GA first populates randomly using Gaussian random distribution to expand the diversity. This population incorporates multiple solutions representing an individual's chromosome; these chromosomes, in turn, have a set of variables that activate the genes. The initial population circulate the solutions across the search space as evenly as possible to be more robust in finding a promising region and expanding the diversity. Secondly, the GA uses the roulette wheel to assign probabilities to individuals and select them to form the next generational proportional to their fitness value. Other selection operators that have been used in literature are Rank selection [31], Boltzmann selection [33], Truncation selection [34], Local selection [35], Fuzzy selection [36], and Proportional selection [37].

After the selection phase, the crossover phase occurs, where the generation of new offspring is produced that will later be added to the new generation population. The crossover application can be made using different methods like the single-point crossover, multi-point crossover [38], uniform crossover [39], shuffle crossover, and order-based crossover [40]. The single-point approach is the most commonly used, which we will also explore in this research. The mutation is another vital phase that follows the selection, and it is used to expand the diversity in the generation and bypass impulsive convergence. Different mutation methods are swap mutation, random resetting, scramble mutation, and inversion mutation [41]. When set to a low mutation, the random mutation is more efficient than other approaches to avoid randomness in the search space [42]. After the new generation has been successfully generated, the generations will be passed through several iterations following the network's latest update. The successful completion of the iteration produces the optimal or close-to-optimal solution for the optimization; Fig. 4 shows the GA framework. GA is mathematically represented by.

$$GA = \{P(0), N, g, s, l, p, f, t\} \tag{1}$$

$P(0) = (x_1(0), x_2(0), \ldots, x_N(0)) \in I^N$, denotes the initial population size; g denotes the initial genetic operators; s denotes the reduction strategy; l denotes the length of string (chromosome); f denotes the fitness function$[f : 1 \rightarrow R^+]$; and t denotes a terminal law$[t : 1^N \rightarrow \{0,1\}]$.

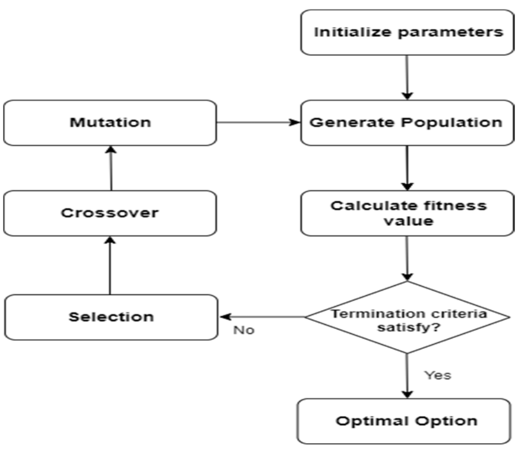

Fig. 4. Genetic algorithm framework.

D. Proposed Approach

Cropped faces were downloaded and resized to 224 × 224 (EfficientNet B0 original image size) to implement the age classification model, so no extra data-cleaning method was required. This research focused on two experiments: training our model using the combined dataset and training using the Adience dataset. Both were tested using 25% of the dataset to report the validation accuracy. Training and testing were not done on UTKFace as it is most suitable for a regression problem. To balance these datasets, this study used dataset mean value to cap and set the limits. To reproduce the results across both experiments, the training and validation of images were set the same throughout the experiment. This study integrates EfficientNet convolutional neural network with the GA-based method. Adience and UTKFace were merged in the first experiment, as explained earlier in section III(3). A custom data generator class is created to correctly feed the data into the neural network later and construct a data processing pipeline simultaneously.

Firstly, this study processed and created batches of images so images could be chosen easily according to the batch they belong to. For each image opened, augmentation is applied at the training step and age is extracted. This step creates a formed batch with images and their corresponding ages. Secondly, the pre-trained EfficientNet model proposed in this study is initiated using the transfer learning approach. To implement the EfficientNet-B0, layers are added to the top layer starting with the Batch Normalisation, GlobalAveragePooling2D, which reduces overfitting by performing sub-sampling (sub-sampling in this study aided to further scale down the total number of parameters in a way that it puts cluster of neurons together as a categorical neuron), Dropout layers which process the information being fed from the model and a dense layer (256, 128, 8) with softmax activation function. The softmax activation function is mathematically defined as in Eq. (2).

$$Softmax\ (x)_i = \frac{e^{x_i}}{\sum_{k=1}^{N} e^{x_k}} \tag{2}$$

where x = the vector of outputs from the neural networks
e = a constant of 2.718
ith = the entry in the softmax output vector, and softmax (x) is the predicted probability of the test input related to class i.

This study carried out two experiments. After setting the pre-trained model, this study employed a GA-based model for the parameter selections in compiling the model. The genetic algorithm was defined, and each chromosome considers the hyperparameters to be evaluated in the network. The parameters considered were learning rate, optimisation function, and batch size. The algorithm was first explored on a small scale of 4 × 4, the population and generation. The list of possible values for each chromosome are for learning rate, we selected 0.001, 0.0001, 0.00001; for optimisation function, we selected adamax, adadelta, adam, adagrad, sgd, rmsprop; for the batch size, we selected 32, 64, 128; and for the activation function, we selected elu, relu, tanh. We set the generations at four, the threshold at 0.6, and the number of populations at four. After training for best fit, GA selected

0.00001, adam, 64, and relu for learning rate, optimisation function, batch size and activation function, respectively.

IV. Results And Discussion

This section presents the experimental results. All the experiments were carried out using Google Colab with a GPU (Tesla K80) runtime type of 15GB. The frameworks used in the experimentation, evaluation, and testing of the efficiency of the age estimation model were state-of-the-art, such as OpenCV and Tensorflow libraries. We extended the performance and capability of these frameworks into the classification task.

The performance evaluation metrics used in this study include accuracy, precision, recall, and F1-score.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F\ measure = \frac{2 \times precision \times recall}{precision + recall}$$

where TP is True Positives
FP is False Positives
TN is True Negatives and
FN is False Negatives

In the first experiment, we used the combined Adience and UTKFace, balanced and the mean for each age class capped at 5291, giving us 33,503 image samples. These images were fed into our fine-tuned EfficientNet Model and trained using 40 epochs. The outcome on the combined dataset only gave us 69% validation accuracy, and the result overfits; checking from the classification report, some of the age group accuracies were low due to uneven group range. The result of the research was higher than Thaneeshan *et al.* [23], where the authors recorded an accuracy of 57.60% on age estimation. Custom age class is not possible on Adience as it is already a group. We had to put UTKFace in the same classes. So, the problem is traced to the age distribution in both datasets.

Table 1. adience+utkface dataset classification report

	Precision	Recall	F1-score	Support
(0, 3)	0.90	0.93	0.91	1095
(4, 7)	0.78	0.71	0.75	727
(8, 14)	0.77	0.79	0.78	834
(15, 24)	0.63	0.69	0.66	1035
(25, 35)	0.62	0.59	0.61	1283
(38, 47)	0.56	0.58	0.57	1266
(48, 59)	0.55	0.55	0.55	840
(60, +)	0.78	0.71	0.74	838
Accuracy			0.69	7918

For the second experiment, we used Adience alone, as reported in the literature, for good performance. We used 16,384 images after balancing by setting the mean cap limit at 2,327. We fed the processed images into our fine-tuned model using the same epochs as the first experiment. The model was validated using 25% of the dataset (4096 images). After training, we had a result of 86.5%, as shown in Figure 5, which shows improved performance compared to results in the literature. We further compare our proposed model to the existing model, as shown in Table 2, to ascertain the efficiency of EfficientNet. Table 1 shows the adience+utkface dataset classification report. Our outcome can be seen to have outperformed related works with a significant difference. In contrast to Dagher and Barbara [27], the results of this model contained unconstrained imaging conditions with variations in poses, expression, illumination, etc. Furthermore, if we investigate the trainable parameters of all related works compared, EfficientNet B0 has the lowest parameters with 5.3 million. Figure 6 shows the confusion matrix of the outcome.

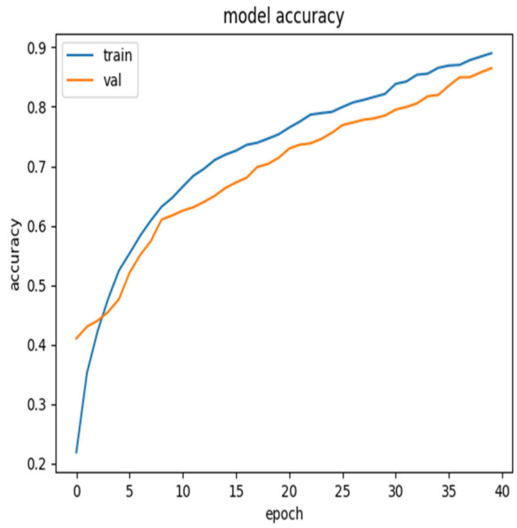

Fig. 5. EfficientNet-B0 + GA accuracy on Adience dataset.

Table 2. Comparison with existing approaches that employed a pre-trained model and evaluated

References	Architecture	Accuracy (%)
[43]	Deep ROR	67.3
[25]	VGG-Face	70.96
[23]	CNN	57.6
[44]	Lightweight Deep CNN	60.03
[6]	EfficientNet-B4	81.1
Proposed	EfficientNet-B0 + GA on Adience_UTKFace	69
Proposed	EfficientNet-B0 + GA on Adience	86.5

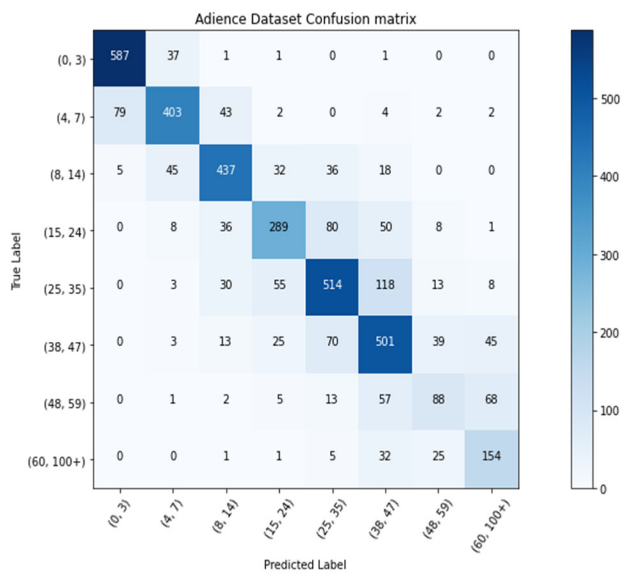

Fig. 6. Confusion matrix for EfficientNet-B0 + GA on Adience dataset.

The study also has some limitations. Firstly, the study did not conduct a data-cleaning exercise when processing the images. However, real-world data often contains noise, outliers, and inconsistencies that can affect model training. Therefore, the absence of data cleaning steps may leave the model susceptible to noise in the dataset. Also, the approach did not include testing on the UTKFace dataset because it is considered more suitable for age regression rather than classification. However, excluding this dataset may limit the evaluation of the model's performance across different age estimation tasks. Lastly, the validation accuracy was reported using 25% of the dataset. Although this is common, choosing the validation set's size can influence the reported accuracy. Smaller validation sets may lead to higher variance in reported results.

V. Conclusion

This study presented herein focuses on age estimation based on facial images employing a pre-trained model (EfficientNet-B0) and Genetic algorithm, which are adept in handling facial features and parameter selection. EfficientNet was fine-tuned using the transfer learning technique and it proved to be efficient in handling features and, with the help of the GA, selecting optimal parameters. The model gave a good accuracy of 86.5% on the Adience dataset. This study shows that applying GA could help get optimal parameters and reduce the cost of training power. The EfficientNet-B0 shows that good accuracy can be achieved even with lesser training parameters, saving computational resources. We will consider applying this approach to UTKFace as a regression task for future work.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Idowu Aruleba conducted the research, analysed the data, and wrote the paper. Professor Sun proofread the paper, made corrections and gained more insight into the final draft. All authors had approved the final version.

References

[1] A. Lanitis, "Facial age estimation," *Scholarpedia*, vol. 5, no. 1, p. 9701, 2010.

[2] V. Carletti, A. Greco, G. Percannella and M. Vento, "Age from faces in the deep learning revolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2113–2132, 1 Sept, 2020. doi: 10.1109/TPAMI.2019.2910522.

[3] T. U. Islam, L. K. Awasthi and U. Garg, "Gender and age estimation from gait: A review," in *International Conference on Innovative Computing and Communications*, Springer, Singapore, 2021, pp. 947–962.

[4] S. Das, I. De Ghosh and A. Chattopadhyay, "Deep age estimation using sclera images in multiple environment," in *Applied Information Processing Systems*, Springer, Singapore, 2022, pp. 93–102.

[5] P. Punyani, R. Gupta and A. Kumar, "Neural networks for facial age estimation: a survey on recent advances," *Artif. Intell. Rev.* vol. 53, pp. 3299–3347, 2020. https://doi.org/10.1007/s10462-019-09765-w

[6] I. Aruleba, and S. Viriri, "Deep learning for age estimation using EfficientNet," in *International Work-Conference on Artificial Neural Networks*, Springer, Cham, pp. 407–419, June 2021.

[7] G. Castellano, B. D. Carolis, N. Marvulli, M. Sciancalepore, and G. Vessio, "Real-time age estimation from facial images using yolo and efficientnet," in *International Conference on Computer Analysis of Images and Patterns*, Springer, Cham, pp. 275–284, September 2021.

[8] O. Agbo-Ajala, and S. Viriri, "Age estimation of real-time faces using convolutional neural network," in *International Conference on Computational Collective Intelligence*, Springer, Cham, pp. 316–327, September 2019.

[9] S. Han, "Age estimation from face images based on deep learning," in *2020 International Conference on Computing and Data Science (CDS)*, IEEE, pp. 288–292, August 2020.

[10] D. V. Atkale, M. M. Pawar, S. C. Deshpande, and D. M. Yadav, "Multi-scale feature fusion model followed by residual network for generation of face aging and de-aging," *Signal, Image and Video Processing*, pp. 1–9, 2021.

[11] N. Sharma, R. Sharma, and N. Jindal, "Machine learning and deep learning applications-a vision," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 24–28, 2021.

[12] L. Torrey, and J. Shavlik, "Transfer learning. In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques," *IGI global*, pp. 242–264, 2010.

[13] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.* vol. 25, pp. 1097–1105, 2012.

[14] K. He, G. Gkioxari, P. Doll'ar, R. Girshick, "Mask R-CNN". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, 2017.

[15] C. Szegedy, *et al*. "Going deeper with convolutions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.

[16] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 0.5 MB model size," arXiv preprint, arXiv:1602.07360, Nov. 2016.

[17] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint, arXiv:1409.1556, 2014

[18] F. Chollet, "Xception: deep learning with depthwise separable convolutions," In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, 2017.

[19] M. Tan, Q. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," In: *International Conference on Machine Learning*, *PMLR*, pp. 6105–6114, May 2019.

[20] B. Zhang, and Y. Bao, "Age Estimation of Faces in Videos Using Head Pose Estimation and Convolutional Neural Networks," *Sensors*, vol. 22, no. 11, p. 4171, 2022.

[21] D. Kwasny, D. Hemmerling, "Gender and age estimation methods based on speech using deep neural networks," *Sensors*, vol. 21, p. 4785, 2021. https://doi.org/10.3390/s21144785

[22] H. Sun, H. Pan, H. Han and S. Shan, "Deep Conditional Distribution Learning for Age Estimation," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4679–4690, 2021. doi: 10.1109/TIFS.2021.3114066.

[23] R. Thaneeshan, K. Thanikasalam, amd A. Pinidiyaarachchi, "Gender and age estimation from facial images using deep learning," In *2022 7th International Conference on Information Technology Research (ICITR)*, IEEE, pp. 1–6, Dec. 2022.

[24] D. Gyawali, P. Pokharel, A. Chauhan and S. C. Shakya, "Age range estimation using MTCNN and VGG-Face model," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, pp. 1–6, 2020. doi: 10.1109/ICCCNT49239.2020.9225443.

[25] 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

[26] Y. Deng, S. Teng, L. Fei, W. Zhang, and I. Rida, "A multifeature learning and fusion network for facial age estimation," *Sensors*, vol. 21, no. 13, p. 4597, 2021

[27] I. Dagher, and D. Barbara, "Facial age estimation using pre-trained CNN and transfer learning," *Multimedia Tools and Applications*, pp. 1–12, 2021.

[28] A. Greco, A. Saggese, M. Vento, andd V. Vigilante, "Effective training of convolutional neural networks for age estimation based on knowledge distillation," *Neural Computing and Applications,* pp. 1–16, 2021

[29] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint, arXiv:1503.02531, 2015

[30] S. Mirjalili, "Evolutionary algorithms and neural networks," *Studies in computational intelligence*, p. 780, 2019.

[31] H. M. Balaha, M. Saif, A. Tamer, and E. H. Abdelhay, "Hybrid deep learning and genetic algorithms approach (HMB-DLGAHA) for the early ultra-sound diagnoses of breast cancer," *Neural Computing and Applications*, pp. 1–25, 2022

[32] D. Whitley, "A genetic algorithm tutorial," *Statistics and Computing*, vol. 4, no. 2, pp. 65–85, 1994.

[33] D. E. Goldberg, "A note on Boltzmann tournament selection for genetic algorithms and population-oriented simulated annealing," *Complex Systems*, vol. 4, pp. 445–460, 1990.

[34] T. Blickle, and L. Thiele, "A comparison of selection schemes used in evolutionary algorithms," *Evolutionary Computation*, vol. 4, no. 4, pp. 361–394, 1996.

[35] R. J. Collins, and D. R. Jefferson, "Selection in massively parallel genetic algorithms," *University of California (Los Angeles), Computer Science Department*, pp. 249–256, 1991.

[36] H. Ishibuchi, and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets and Systems*, vol. 141, no. 1, pp. 59–88, 2004.

[37] J. J. Grefenstette, and J. E. Baker, "How genetic algorithms work: a critical look at implicit parallelism," *The proceedings of the third. International conference on genetic algorithms*, 1989.

[38] L. J. Eshelman, R. A. Caruana, and J. D. Schaffer, "Biases in the crossover landscape," In *Proceedings of the Third International Conference on Genetic Algorithms*, 1989.

[39] E. Semenkin, and M. Semenkina, "Self-configuring genetic algorithm with modified uniform crossover operator," In *International Conference in Swarm Intelligence*, Springer, Berlin, Heidelberg, pp. 414–421, June 2012.

[40] L. Davis, "Applying adaptive algorithms to epistatic domains," In *IJCAI*, vol. 85, pp. 162–164, August 1985.

[41] T. V. Mathew, "Genetic algorithm," *Report submitted at IIT Bombay*, 2012.

[42] J. Genlin, "Survey on genetic algorithm," *Computer Applications and Software*, vol. 2, no. 1, pp. 69–73, 2004.

[43] K. Zhang, "Age group and gender estimation in the wild with deep RoR architec-ture," *IEEE Access*, vol. 5, pp. 22492–22503, 2017.

[44] M. N. Islam Opu, T. K. Koly, A. Das and A. Dey, "A Lightweight Deep Convolutional Neural Network Model for Real-Time Age and Gender Prediction," *2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, Bengaluru, India, 2020, pp. 1–6. doi: 10.1109/ICAECC50550.2020.9339503.