Computational Engineering and Technology Innovations

www.ceti.reapress.com

Comput. Eng. Technol. Innov.Vol. 1, No. 2 (2024) 86–97.

Paper Type: Original Article

Human Age Prediction Based on Facial Patterns Using CNN

Ramin Safa[1,*], Nikola Petrovic[2], Victoria Nozick[3], Lorenzo Cevallos-Torres[4]

[1] Department of Computer Engineering, Ayandegan Institute of Higher Education, Tonekabon, Iran; safa@aihe.ac.ir.
[2] Department of Mechanical Engineering, University of Nis, Serbia-rad po izboru; nikolapetroviclogistika@gmail.com.
[3] Department of Operations and Information Management, Aston Business School, Aston University, B4 7ET Birmingham, United Kingdom; victorianozick79@gmail.com.
[4] Brigham Young University–Idaho, USA; cev18002@byui.edu.

Citation:

Received: 15 March 2024 *Revised: 12 May 2024* *Accepted: 15 June 2024*	Essienu, I., &, E. (2024). Human age prediction based on facial patterns using CNN. *Computational engineering and technology innovations*, *1*(2), 86-97.

Abstract

Deep learning has established its dominance and proved to be a powerful tool for increased accuracy and the ability to handle and process big data. One such popular Deep Neural Networks (DNN) is Convolutional Neural Network (CNN), a class of deep NN, applied to analyze and predict visual imagery. In this paper age estimation problem is adressed. Advancement in technology has increased the expectation of consumers. Face detection and age prediction are the two top technological trends on social media platforms. The trained model attempts to identify an image and then predict its age using deep learning models. We have used CNN algorithm for the final model after the effort was made in selecting the most efficient algorithms among RNN and GAN. The model is trained using different age classifiers and results obtained showed improvement in the performance of age estimation.

Keywords: Convolutional Neural Network, Deep learning, Machine learning, Age estimation.

1 | Introduction

In this modern world of advanced technology, social media is used by almost all people. We can see people often upload many images on social media and this is in trend. Many new features come with every update and we can see that they have a variety of improvements in them. We have witnessed the dawn of facial detection frameworks. The applications which have face detection make it easy in identifying the person. The feature "tag" present in Facebook and Instagram helps in identifying the person in just a click. Remarkable developments in the field of image processing and computer vision have made face detection and identification tasks a lot easier. Our goal is to predict the age of the identified face in a given image.

Corresponding Author: safa@aihe.ac.ir
10.48314/ceti.v1i2.30

Age classification requires certain classifiers and also training needs to be given to the model. This is an inherently challenging problem compared to other tasks present in computer vision. For instance, in Facebook, if we upload an image with college boys enjoying picnics, it might caption it with "fun with each other" but if it could predict the age, then it may better caption it with "fun with college friends" based on the classifier output. This might result in a more sophisticated feature in the social media applications.

Age prediction using speech (or vocal tract features) and text are challenging since these features are loosely correlated with age. Accuracy is still the main problem with age prediction in order to convert it into an efficient commercial product. Almost all the age prediction techniques reported in the literature use the dimensions of facial features which are not efficient. However, machine learning could not make full use of a huge number of images and datasets from different sources for improving accuracy of classification performance [1].

In this article, an effort has been made to reduce the difference between auto-face recognition capabilities and those of age prediction methods by incorporating different face recognition techniques. The face recognition techniques which have been established previously have shown the tremendous progress can be made using Convolutional Neural Networks (CNN). Hence, the CNN architecture has been utilized for face detection and age prediction by considering the limited availability of data set containing face images with accurate age.

Deep learning models are constructed through Neural Network (NN) architecture. Spatial features of the images are reserved in such an architecture since it consists of densely connected kernels. CNNs are part of Artificial Neural Networks (ANN) exclusively for image analysis problems. CNN models deployed over the last decade are showing outstanding results when applied for different machine learning problems.

CNN is a deep learning algorithm, which facilitates the identification of different objects from an input image and differentiate those from one another. CNN requires very less pre-processing compared to other classification algorithms. As the features are hand-made in primitive methods, the CNN can learn these features with adequate training.

The use of age prediction systems is extensively increasing in recent times and it includes application in the areas like computer vision. Computer vision applications include human computer interaction algorithms and techniques. Age prediction models have proven to be of great value in the scenarios such as detection of fraud in alcohol purchasing since there is an age limit for it. Similarly, detection of fraud in driving automobiles and numerous such applications like airport security, hotel, hospital security, electronic customer service, etc. The limitation of human capability of predicting age accurately is solved by using such models. Also, the problem of classification of people in category based on age like children, teenagers, senior citizens etc. can be solved using age prediction.

Hence, in this article CNN Architecture has been used for the age prediction. The CNN architecture accuracy depends on the depth of the network and the depth can be increased by considering the large amount of input samples. This is the main drawback of the CNN architecture. Since, good results are achieved through the CNN model, certain architecture-tuning methods have been performed like image augmentation, tuning epochs, handling underfitting and overfitting to overcome the above-mentioned drawback.

2 | Literature Survey

Initial attempt to predict human age from facial images dates back to over two decades. Due to non-availability of large datasets, most of the approaches were based on single local features. A texture feature-based age prediction methodology reported in [1]. An approach based on Biologically Inspired Features (BIF) and statistical age classifier proposed by Guo et al. [2]. Use of Gabor features and fuzzy-linear discriminant analysis based approach is reported in [3]. An attempt to age prediction based on Local Binary Patterns (LBP) was proposed by Gunay et al. [4].

With the remarkable developments in the field of Graphics Processing Units (GPUs), deep learning models have gained much popularity in classification and decision making tasks. Most of the recent deep learning based approaches to age prediction tasks are inspired from the CNN architectures [6]. A Deep Neural Network (DNN) architecture for age gender and skin color detection is reported in [7]. An attempt to fusion features based age prediction was proposed by Huerta et al. [8].

Liao et al. [9] proposed an approach for age estimation based on combination of CNN and divide-and-rule strategy with Mean Absolute Error (MAE) of 3.48 per age prediction. A deep learning classifier based age and gender prediction using unfiltered images is reported in [10] with accuracy from 83%-93% on OIUAdience dataset. A light weight CNN and data augmentation based age prediction is proposed by Liu et al. [11] with MAE of 2.68-3.81 on MORPH2 and FG-NET datasets. Attentional Convolutional Network (ACN) based age prediction methodology was proposed by Abdolrashidi et al. [12] with an accuracy of 91% on UTK-FACE dataset. More recently a multi feature and multi-fusion network based approach was reported in [13] with MAE of 2.47-2.67 on standard datasets.

3 | Working Methodology

The working methodology along with CNN model and its implementation and training of the model is explained in this section.

3.1 | The CNN Model

The architecture of the NN is having many categories. To understand the model let us take the example of LeNet5 architecture for the analysis and also understanding purpose. The LeNet5 architecture is a structure of total seven layers, except to the input layer, and each of other have the training parameters, and also each layer in the architecture contains numerous feature maps, through which input features are extracted through a convolution kernel. And each feature contains multiple neurons.

As represented in the Figure above, the input image size of (32×32) at the input layer is considered. The input layer consists of a convolutional layer (C1) and it consists of 6 convolutional kernels of size 5×5. After processing this convolutional layer, there are 28×28×6 neurons and hence the number of parameters that can be trained will be ((5×5) +1×6).

The generic convolution of an image can be represented by the equation given below.

$$\mathrm{conv}(I, K)x, y = \sum_{i=1}^{nH} \sum_{j=1}^{nW} \sum_{k=1}^{nC} Ki, j, klx + i - 1, y + j - 1, k.$$

The down sampled layer(S2) is the next layer in the architecture after C1 as represented in the *Fig. 1*. The input to this down sampled layer is the output from C1 layer. The size of this input is 28×28. The process of sampling includes addition of four numbers to the original image and then it is multiplied by a training feature and a training offset is added to the output through sigmoid function. The number of neurons are 14×14×6 in the S2 layer.

The step of down sampling the image's key features by summing up the information. This is performed through every channel and hence it only affects the dimensions height and width (n_H, n_W) and keeps number of channel (n_C) intact. For a given image, we pass a filter, with no parameters to train that follows a certain pattern and then apply a function on only selected features. We have:

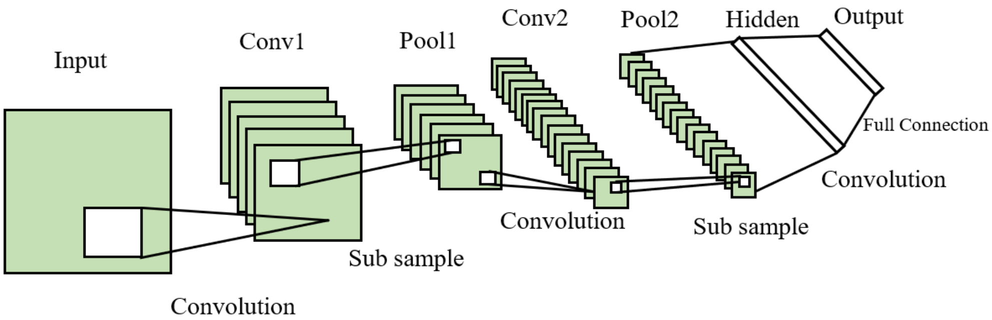


Fig. 1. CNN implementation.

$$\dim\big(\mathrm{conv}(I,K)\big) = (n_H + 2p - f, n_W + 2p - f); s = 0,$$

Where n_C is the number of channels and n_H, n_W are the dimensions. In some cases, it is of advantage to shrink the number of channels without changing dimensions (1×1convolution). However, the filter parameters are decided after back propagation.

After sampling is done through S2, the next layer is a convolutional layer which consists of 16 kernels and all the features of this convolutional layer is same as C1. This layer is named as C3 as depicted in the *Fig. 1*.

The size of the output layer is 10×10. The features of the S2 layer are connected with all the features that are in the C3 layer. The combination of features from this level of layer is dissimilar from the features from previous one.

The S4 layer and S2 layer are similar which has a sample size of 16. The convolutional kernel size remains 5×5. The image size changes to (5-5+1) in C5 layer. The fully connected layer is the last layer of the architecture.

3.2 | Convolution Model Construction

To understand the CNN architecture further and how the model is constructed let us consider the *Fig. 2*.

According to the typical CNN architecture shown in *Fig. 2*, the input image after undergoing some pre-processing techniques is passed to the CNN architecture. The next stages are convolution and down sampling. We can also add other layers like normalization stage, timing check stage for more accuracy. The final stage is the output. This process is repeated until we get the desired output without over-fitting.

In the proposed model, fifty-four peoples face were collected. By the use of image augmentation, each image is increased by 10 number and hence the total sample size is 540.

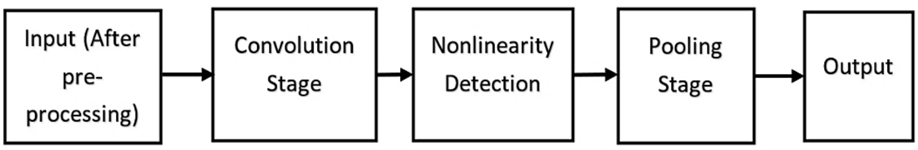


Fig. 2. Typical CNN architecture.

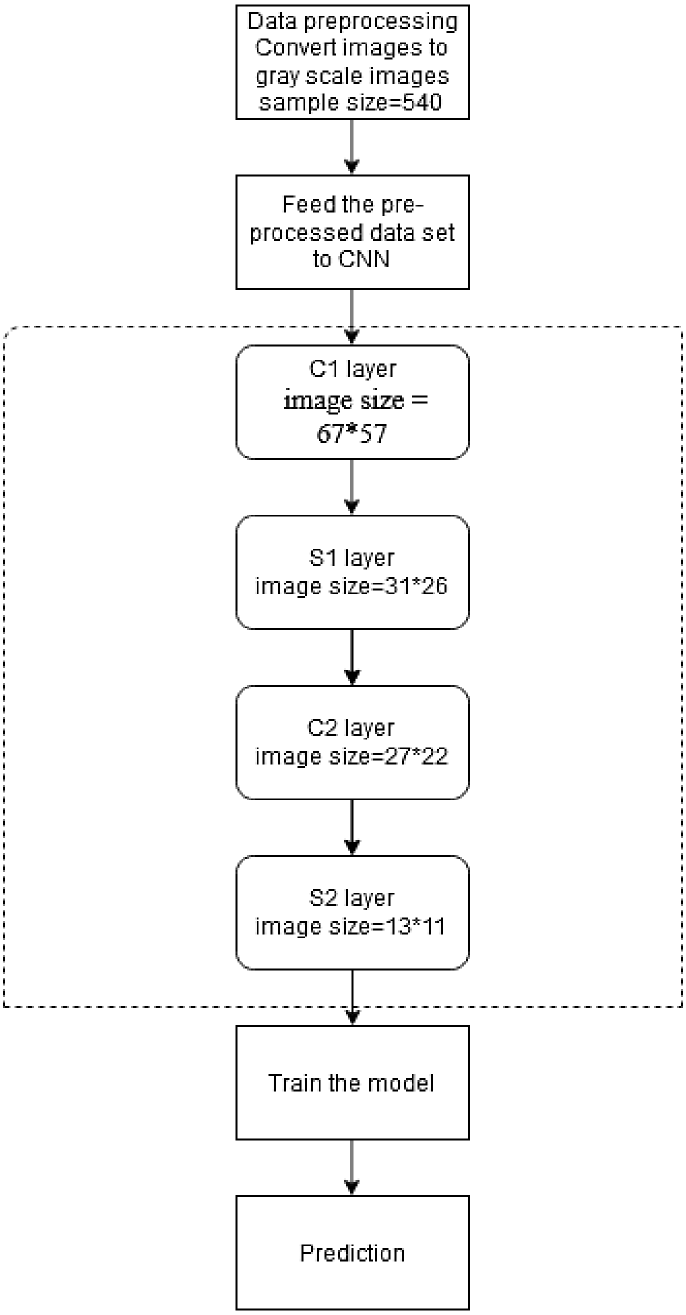


Fig. 3. Flowchart of implementation of the model.

Each image is of the size (67×57=3819). The convolution kernel size is 5×5. The images are converted into grayscale images by pre-processing techniques before feeding to CNN. By the design of this architecture, the image size becomes (67-5+1)×(57-5+1) = (63,53). On this image, the down sampling is performed. The size of the image after down sampling is (31×26). The input to the next convolution stage is the output of the previous stage. This procedure of convolution plus sampling is carried out at every layer till we get an image of size (13×11). The flowchart in *Fig. 3* represents the working of CNN model for our dataset along with size of images at each stage.

4 | Results

4.1 | Dataset Generation

Collecting Social media platforms will require privacy permission and it might turn to security hazard and its time consuming. Hence, the set of images is taken from available datasets from Kaggle and photographs of people we know from family and friends. The dataset we selected has 540 images. While using machine learning algorithms on small datasets, over fitting is the key problem [14] and this problem is addressed by using CNN that have huge parameters. Hence, overfitting problem needs a special attention.

The approach we have used in this paper performs prediction of age perfectly. Our CNN model consist of two convolutional layers that are fully connected with small number of neurons. Small network design is employed in order to take less risk of overfitting. Prediction of age on dataset needs to differentiate between 10 different age classes.

4.2 | Performance Parameters

In order to improve the performance of the convolutional network, we intended to stress upon the following steps.

4.2.1 | Tuning parameters

Parameters like epochs, learning rates can be fine-tuned to improve the performance. After doing experimentations on no of epochs to be used, we can observe that training loss and accuracy cannot be further increased.

4.2.2 | Image data augmentation

CNN learns the features from the data automatically only when huge amount of training data is available. Image data augmentation is used in the cases where training data available is not huge. We used image augmentation parameters like zoom, shear, rotation etc. to increase the existing data samples by 3 times to 4 times.

4.2.3 | To handle overfitting and underfitting problem

Underfitting occurring due to less features may be solved through increasing the depth of the network. Overfitting can be encountered by training larger datasets and cross validation.

4.2.4 | Use of grayscale images instead of color images

Although using RGB images will bring out more features, it also leads to overfitting of the model. Hence, we have used grayscale images. We can convert RGB images to gray scale using average method or weighted method. The Average method calculates the average value of R, G, and B into grayscale value:

$$\text{Grayscale} = (R + G + B) / 3.$$

In the weighted method, the wavelengths of red, green and blue is taken and grayscale value is calculated according to the equation given below:

$$\text{Grayscale} = 0.299R + 0.587G + 0.114B.$$

The age distribution of the collected dataset is shown in *Fig. 5*. The performance metrics that we made use of in this project are cross entropy (a loss function for the CNN) and accuracy (MAE). However, this is a multiclass prediction in which the similarity between age group 21-30 and 31-40 is more than between age group 71-80. Hence, the confusion matrix is also used.

Fig. 4. Image augmentation example.

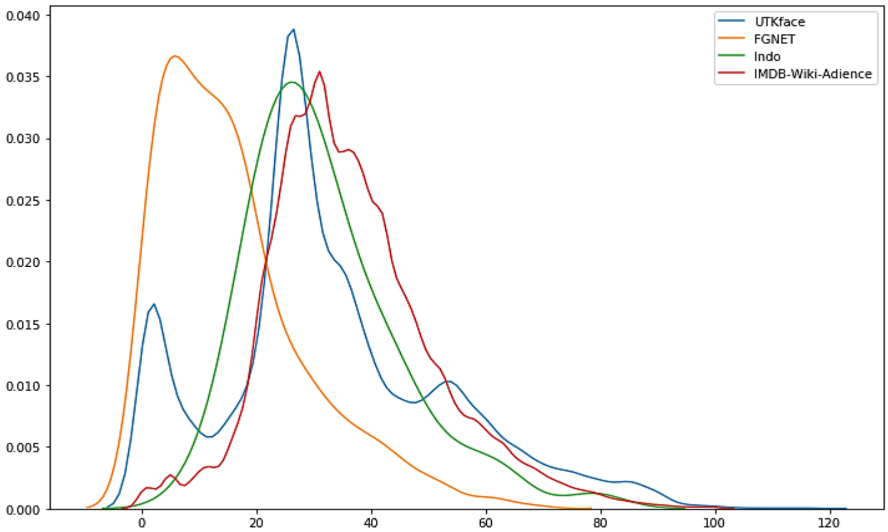


Fig. 5. Age distribution in the dataset.

4.3 | Statistical Analysis

The distribution of training dataset, test data set and the confusion matrix between various parameters of the dataset are represented in the diagrams below:

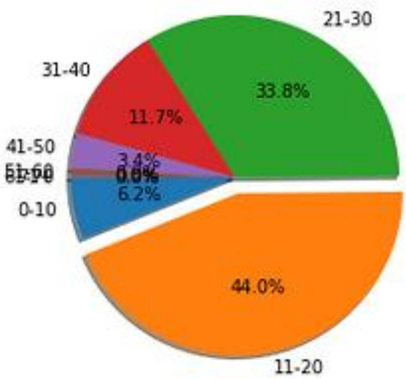


Fig. 6. Train dataset.

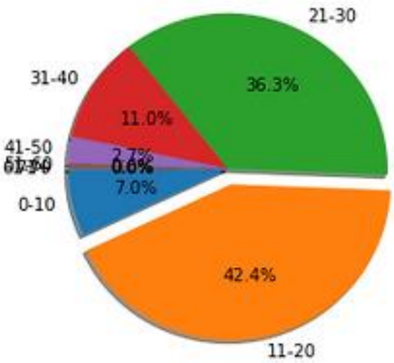


Fig. 7. Test dataset.

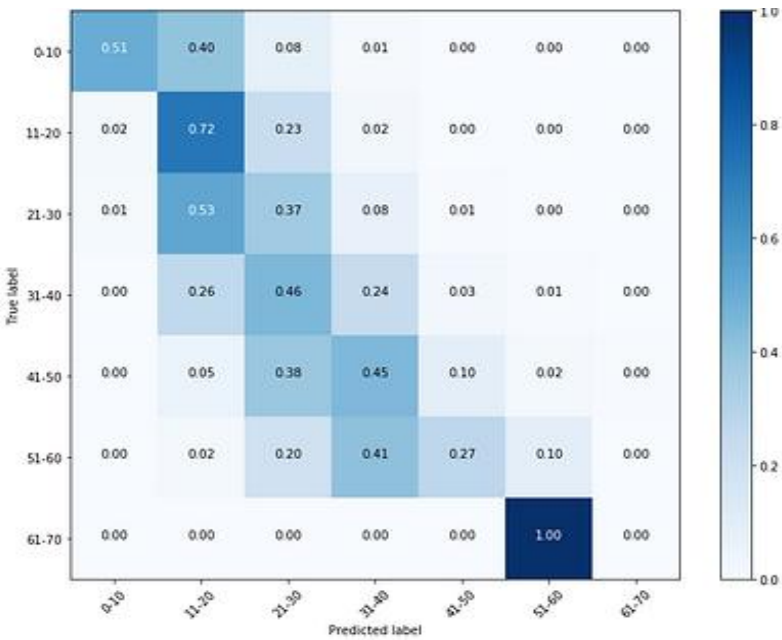


Fig. 8. Confusion matrix (normalized).

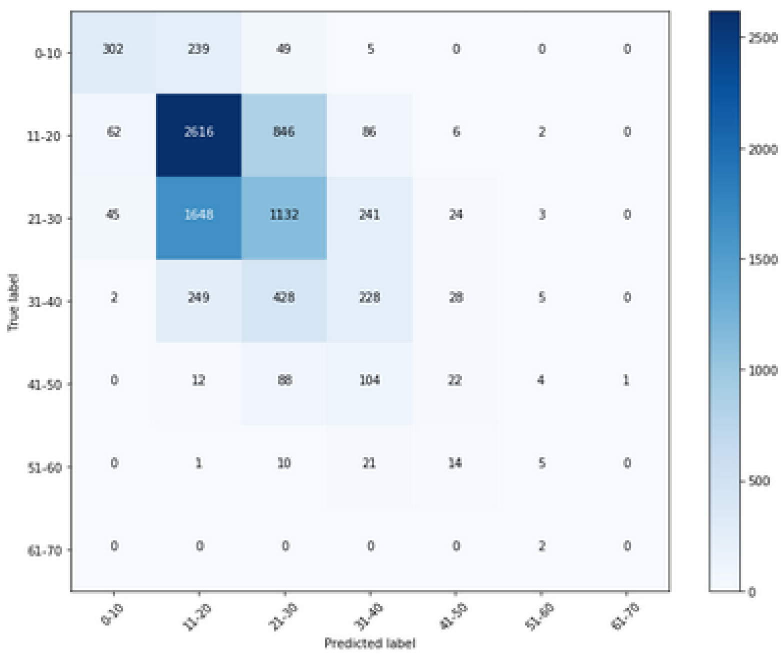


Fig. 9. Confusion matrix (prediction by model).

The details of the basic CNN architecture used is given in *Table 1*. The results of age prediction model after preprocessing of the data (conversion to grayscale images), tuning the parameters (epoch values) are shown in *Fig. 11*. The left side five images from the output are numbered as 1-5 in the above table given.

The right side five images from the output are numbered as 6-10 in the above table given. With the above CNN architecture using deep learning we obtained validation accuracy of 80% as shown in *Fig. 12*.

Table 1. CNN architecture details.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 198, 198, 8)	80
conv2d_1 (Conv2D)	(None, 196, 196, 16)	1168
conv2d_2 (Conv2D)	(None, 194, 194, 32)	4640
max_pooling2d	(None, 97, 97, 32)	0
conv2d_3 (Conv2D)	(None, 95, 95, 32)	9248
max_pooling2d_1	(None, 47, 47, 32)	0
conv2d_4 (Conv2D)	(None, 45, 45, 64)	18496
max_pooling2d_2	(None, 22, 22, 64)	0
global_avg_pooling2d	(None, 64)	0
dense (Dense)	(None, 20)	1300
dense_1 (Dense)	(None, 11)	231
Total Params: 35.163.		
Trainable Params: 35.163.		
Non-trainable: 0.		

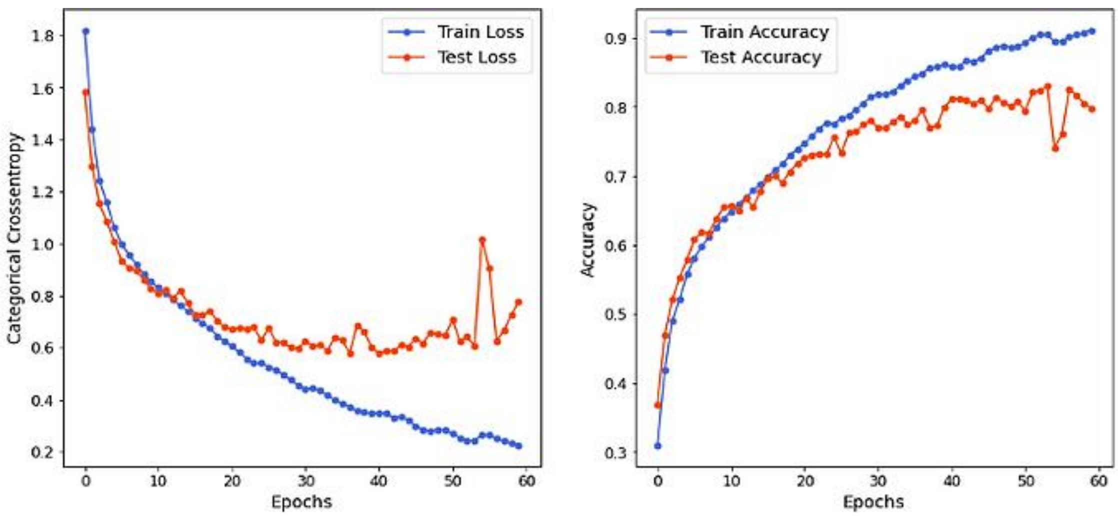


Fig. 10. Line plots to show train and test loss due to epochs.

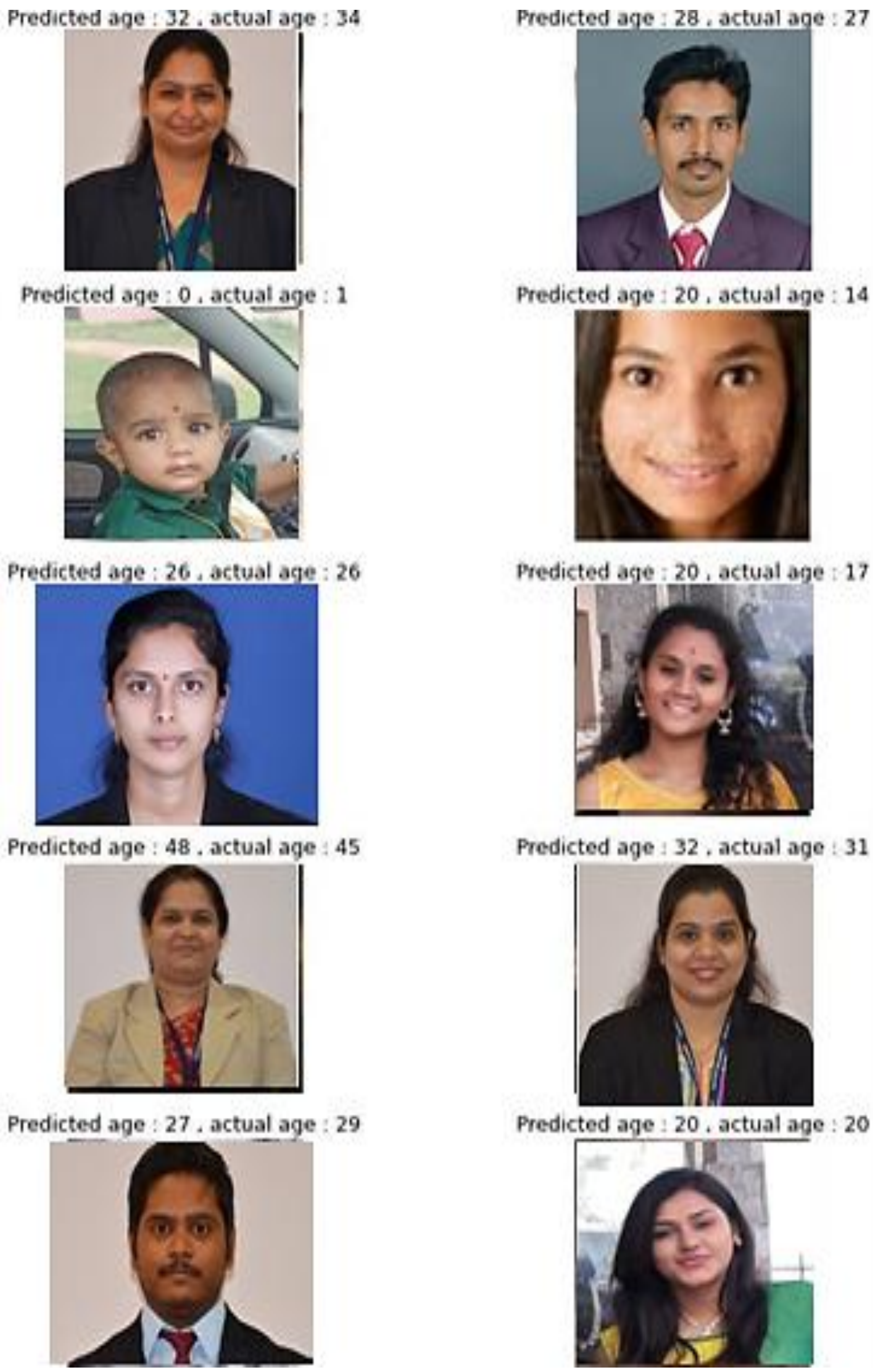
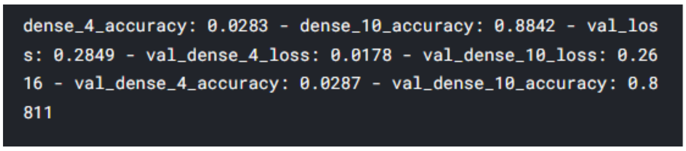


Fig. 11. Results of age prediction model.



```
dense_4_accuracy: 0.0283 - dense_10_accuracy: 0.8842 - val_los
s: 0.2849 - val_dense_4_loss: 0.0178 - val_dense_10_loss: 0.26
16 - val_dense_4_accuracy: 0.0287 - val_dense_10_accuracy: 0.8
811
```

Fig. 12. Accuracy of the model.

The model is evaluated based MAE as given in *Table 2*.

Table 2. Tabular representation of age prediction with mean error rate.

Image No.	Predicted Age	Actual Age	Mean Error Rate in Percentage
1	32	34	5.88
2	0	1	0
3	26	26	0
4	48	45	6.66
5	27	29	6.89
6	28	27	3.70
7	20	14	42.85
8	20	17	17.64
9	32	31	3.22
10	20	20	0

5 | Conclusion

In this paper we use state of the art deep learning architecture and show that by combining the age, calculations, there is an improvement in the overall accuracy of the output. With more resources and time, the parameters used would have been fine-tuned and more modified architectures could have been used. Therefore, we think that the accuracy of the model is good and is better than the existing models, but can be improved further by using more input data and a better network architecture or algorithm.

There will be different paths in future building off of this work to include using age estimation for face recognition, to improve the experience with images on social media, and a lot more. Making use of additional training data that will become readily available in future to carry out age estimation and it will allow successful techniques from other types of regressions with huge dataset that can be used.

Acknowledgments

The authors wish to extend their thanks to their affiliated institutions for backing this research. Special appreciation goes to the students and staff who contributed to data collection and provided technical assistance throughout the model training and evaluation processes.

Author Contributions

Ramin Safa: Conceptualization, Methodology, Model Design, Writing–Original Draft.

Nikola Petrovic: Data Curation, Formal Analysis, Writing–Review & Editing.

Victoria Nozick: Supervision, Validation, Project Administration.

Lorenzo Cevallos-Torres: Software Implementation, Visualization, Resources.

All authors have reviewed and approved the final manuscript.

Funding

This research did not receive any external funding.

Data Availability

The datasets utilized and/or analyzed in this study can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest

The authors indicate that there are no conflicts of interest.

References

[1] Fu, Y., Guo, G., & Huang, T. S. (2010). Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11), 1955–1976. https://doi.org/10.1109/TPAMI.2010.36

[2] Lanitis, A., Taylor, C. J., & Cootes, T. F. (2002). Toward automatic simulation of aging effects on face images. *IEEE transactions on pattern analysis and machine intelligence*, 24(4), 442–455. https://doi.org/10.1109/34.993553

[3] Guo, G., Mu, G., Fu, Y., & Huang, T. S. (2009). Human age estimation using bio-inspired features. *2009 IEEE conference on computer vision and pattern recognition, CVPR 2009* (pp. 112–119). IEEE. https://doi.org/10.1109/CVPRW.2009.5206681

[4] Gao, F., & Ai, H. (2009). Face age classification on consumer images with gabor feature and fuzzy lda method. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (pp. 132–141). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-01793-3_14

[5] Günay, A., & NabIyev, V. V. (2008). Automatic age classification with LBP. *2008 23rd international symposium on computer and information sciences, ISCIS 2008* (pp. 1–4). IEEE. https://doi.org/10.1109/ISCIS.2008.4717926

[6] Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 34–42). IEEE. https://B2n.ir/ud1451

[7] Abdulnabi, A. H., Wang, G., Lu, J., & Jia, K. (2015). Multi-task CNN model for attribute prediction. *IEEE transactions on multimedia*, 17(11), 1949–1959. https://doi.org/10.1109/TMM.2015.2477680

[8] Huerta, I., Fernández, C., Segura, C., Hernando, J., & Prati, A. (2015). A deep analysis on age estimation. *Pattern recognition letters*, 68, 239–249. https://doi.org/10.1016/j.patrec.2015.06.006

[9] Liao, H., Yan, Y., Dai, W., & Fan, P. (2018). Age estimation of face images based on CNN and divide-and-rule strategy. *Mathematical problems in engineering*, 2018(1), 1712686. https://doi.org/10.1155/2018/1712686

[10] Agbo-Ajala, O., & Viriri, S. (2020). Deeply learned classifiers for age and gender predictions of unfiltered faces. *Scientific world journal*, 2020(1), 1289408. https://doi.org/10.1155/2020/1289408

[11] Liu, X., Zou, Y., Kuang, H., & Ma, X. (2020). Face image age estimation based on data augmentation and lightweight convolutional neural network. *Symmetry*, 12(1), 146. https://doi.org/10.3390/SYM12010146

[12] Abdolrashidi, A., Minaei, M., Azimi, E., & Minaee, S. (2020). Age and gender prediction from face images using attentional convolutional network. *ArXiv Preprint ArXiv:2010.03791*. https://arxiv.org/abs/2010.03791

[13] Deng, Y., Teng, S., Fei, L., Zhang, W., & Rida, I. (2021). A multifeature learning and fusion network for facial age estimation. *Sensors (basel, switzerland)*, 21(13), 4597. https://doi.org/10.3390/s21134597

[14] Chen, S., Zhang, C., Dong, M., Le, J., & Rao, M. (2017). Using ranking-cnn for age estimation. *Proceedings-30th IEEE conference on computer vision and pattern recognition, CVPR 2017* (pp. 5183-5192). IEEE. https://doi.org/10.1109/CVPR.2017.86