2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC) | 978-1-7281-9183-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICAECC50550.2020.9339503

A Lightweight Deep Convolutional Neural Network Model for Real-Time Age and Gender Prediction

Md. Nahidul Islam Opu*, Tanha Kabir Koly†, Annesha Das‡ and Ashim Dey§
Department of Computer Science & Engineering
Chittagong University of Engineering and Technology
Chittagong-4349, Bangladesh
Email: *u1604073@student.cuet.ac.bd, †u1604079@student.cuet.ac.bd, ‡annesha@cuet.ac.bd, §ashim@cuet.ac.bd

***Abstract*—Recognition of age and gender has become a significant part of the biometric system, protection, and treatment. It is widely used for people to access age-related content. It is used by social media in the distribution of layered advertising and promotions to expand its scope. Application of face detection has grown to a great extent that we should upgrade it using various methods to achieve more accurate results. In this paper, we have developed a lightweight deep Convolution neural network model for real-time age and gender prediction. For making the training dataset more diverse, Wiki, UTKFace, and Adience datasets have been merged into one containing 18728 images. Using this vast mixed dataset, we have achieved accuracy of 48.59% and 80.76% for age and gender respectively. Further, the model is tested in real-time. Different experimental investigations on the prepared dataset show that with most recent approaches, our model provides competitive prediction accuracy.**

***Keywords*—Convolutional Neural Network, Deep Learning, Facial Images, Age and Gender Prediction, Real-time recognition system.**

I. Introduction

Human age and gender are considered as important biometric trait for human identification. Age and gender prediction refers to the process of recognizing a person's face in the picture and identifying if a person is male or female and predicting age. These two attributes play a vital role in our social life. Recognition of face attributes in real-time is a very promising research topic. Recent research suggests that the aging characteristics deeply learned from huge data contribute to a substantial improvement in facial image-based age evaluation efficiency. For a growing number of applications, automatic age and gender detection have become important, especially after the rise of social networks and social media. In the real world, there are many technologies available that are related to age estimation and gender prediction, including product sales, biometrics, cosmetology, forensics, entertainment, etc. [1]. Age prediction plays an significant role in crime investigation also as it helps to find the actual criminal based on the person's age.

However, the performance of achievable processes on real world raw images is still not up to the standard when it comes to the realistic method of face recognition. Nevertheless, there is still a significant lack of performance of established techniques on real world images, particularly when compared to the enormous performance leaps for the associated face recognition task. While research on age estimation extends over decades, the research of apparent age evaluation or age as interpreted from a face picture by other humans is a recent effort. It should be noted that age detection from a single picture is not an easy task to accomplish because the perceived age depends on several factors and same-aged people look very different in different parts of the world. Successful age and gender prediction from facial picture captured under real-world conditions will lead to improving the results of identification. The applications of age and gender classification systems have been growing fast in recent years due to its improved technology such as haar cascade classifiers, deep multi-task learning and OpenCV etc. [2]–[4]. Recently deep neural networks have become popular for numerous applications to improve accuracy. A deep learning age and gender classification approach is proposed in this paper, taking into account significant constraints of the mobile application.

In this work, we implemented a deep learning Convolutional Neural Network (CNN) solution to age and gender prediction from a single face image combining three datasets with age and gender labels. This paper shows that with the exercise of profound CNN as presented here, it is possible to achieve remarkable performance. The main objectives of our work are:

- Build a lightweight CNN model.
- Train the CNN model using a large combined dataset.
- Estimate age and predict gender from facial image in real time.

Our multi-task learning system allows optimal features to be shared and learned in order to enhance recognition efficiency for both tasks. The CNN architecture that our model employs is designed specifically for age and gender estimation to increase the time efficiency and reduce the model size while retaining recognition output consistency.

The rest of the paper is summarized as follows. Related literature and studies are presented in Section II. The proposed methodology is described in Section III. In Section IV, results and performance analysis is presented. Finally, the paper is concluded in Section V.

II. Literature Review

There is several notable research done in the area of age and gender prediction using facial images. The early methods for this sector were focused mainly on extraction and calculation of the facial features.

In [5], a deep neural network was used that is computationally inexpensive and provides good accuracy on many competitive datasets. A deep CNN network was proposed in [6] that was composed of locally connected hidden layers. The dataset CAS-PEAL and FEI have been used for training.

Zukang Liao et al. [7] proposed 9 overlapping patches per photo instead of a hundred patches to cover the whole region. Accuracy was achieved by the nine-patched method was 95.072% on the Labeled Face in the Wild (LFW) dataset and 78.63% on the Adience dataset for gender prediction. For age classification, accuracy was 40.25%. Rothe et al. [8] approached a deep learning method for age prediction from a single face picture without using facial landmarks. Their main contributions are the IMDB-WIKI dataset, regression formula by deep classification, and achieving accuracy 64.0%. In [9], a hybrid architecture was introduced that consists of a CNN and an extreme learning machine (ELM). Using datasets MORPH-ll and Adience benchmark, achieved accuracy for age and gender prediction respectively is around 52% and 88% on average.

For age and gender prediction, computationally efficient CNN model was designed for mobile platform in [10]. In [4], authors developed an android app for age and gender using LBP (Local Binary Patterns) classifier and LBPH (Local Binary Pattern Histogram) model. A lightweight CNN was implemented for mobile application in [11] using the Adience dataset and gained accuracy using LMTCNN-2-1 for age top-1 is 44.26% and gender top-1 is 85.16%. Also a smartphone-based implementation was done using ocular images for gender prediction in [12]. VISOB dataset was used and gained accuracy was 78% and 86.2% from iPhone and oppo images.

Gabor filter as input in CNN and Adience dataset were used in [13] and achieved accuracy for age and gender is respectively 61.3% and 88.9%. A video-based implementation was done by using Dempster-Shafer theory to generate classifiers using different datasets such as IMFDB, Kinect, EmotiW 2018 and IJB-A in [14]. They increased the accuracy of the age and gender detection by 2-5% and 5-10% correspondingly. A research was carried out using feed-forward propagation neural networks at a finer level with 3-sigma control limits in [15]. By using the JAFFE dataset, they gained accuracy of 95% for age and gender detection.

Using the (LFW) dataset and Adience dataset, another local deep neural network was introduced in [16]. Accuracy of 96.02% and 80.64% for gender prediction was obtained from these two datasets and accuracy of 44.36% for age prediction was achieved from the proposed model. Another research was done by two learning methods -single task learning (STL) and deep multi-task learning (DMTL) in [3]. Gained accuracy for CNN+STL and CNN+DMTL is 80.11% and 91.34% for gender prediction. For age prediction, gained mean absolute error (MAE) is 4.01. and 4.00 respectively. Insha Rafique et al. [2] proposed a deep CNN for eight age groups and two gender groups. They gained 79% accuracy using deep CNN model and Haar Cascade Classifier. Recently in [17], a research was done using CNN with 'Google' and 'IMFDB' datasets and gained approximately 80% accuracy. Another research was done with the IMDB-WIKI dataset and achieved accuracy for gender prediction is approximately 96.50% and age prediction is 85% for their own dataset [1].

In recent years, other researchers developed a Deep CNN architecture for gender detection with standard accuracy and low computational cost. They compared their architecture with other popular CNN with common datasets namely IMDB-WIKI, LFW, and Adience dataset [18]. In [19] they implemented a lightweight model for age and gender estimation. They have used multi-task learning for an embedded system using multiple datasets such as MORPH-II, FGNET, and MegaAge-Asian datasets. In [20], they have used the same dataset MORPH-II and FG-NET and developed an age estimation process based on a lightweight CNN and Data Augmentation. They proposed a mixed attention method by combining regression and classification formulas. In [21], a system was developed to detect gender, age and emotion by using CNN. They have used a haar cascade classifier and 10000 images to train their model. A deep learning classifier was developed to predict age and gender in [22] from unfiltered images. Their CNN architecture consists of two parts that extract features and classify. They trained their network on IMDB-WIKI, MORPH-II and OIC- Adience datasets and improved accuracy for age and gender prediction.

Studying and summarizing the above researches, there are some drawbacks of their implementations such as CNN architecture, big sized CNN model, high computing cost, dataset processing, and so on.

Analyzing the previous works, we can state two prime factors contributing to the success of age and gender prediction using CNN which are:

- Changes in architecture.
- Dataset preparation.

III. Methodology

The main focus of this work is to develop a CNN model and training it using a large combined dataset. Total methodology can be divided into two main parts:

- Training Phase
- Real-time Testing Phase

A. Training Phase

The steps of training are shown in Fig. 1. The first two steps are, building a CNN and processing the dataset. After that, using the processed dataset, the CNN model is trained. And the last step is to save the model for future use.

This is because training a deep learning model takes a lot of time and resources. So it is not feasible to train each time before predicting. After successful training, the model with the weights can be saved in memory. Whenever a prediction is required, the model is loaded from memory and used for prediction. This method also enables low-end devices, such as IoT devices, smartphones to perform complex tasks as

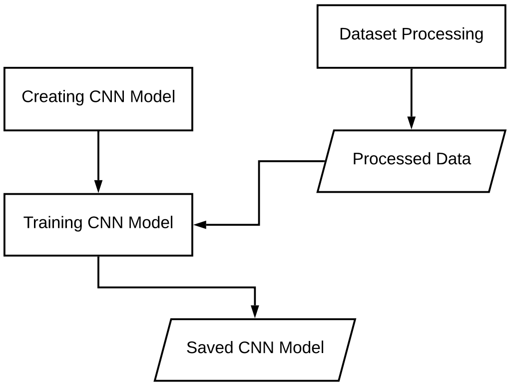

Fig. 1. Training Phase

prediction or classification, as training is not possible in these low-end devices.

We have learnt two main ideas given below from the deep learning research which are applied in our work:

- The more balanced and diverse the datasets are, the more the network understands to generalize and the more resilient it becomes to overfitting.
- The deeper the neural networks are, the greater the ability to model extremely non-linear shifts.

Dataset preparation and model creation and train are described here:

1) Preparing Dataset: For this study, we have selected three publicly available facial datasets which are, Wiki [8], UTKFace [23] and Adience Dataset [24]. Preprocessing steps are different for each of them. For the wiki dataset, the photos which have only one face are selected. Then the age is calculated as mentioned in (1).

$$\text{age} = \text{date of photo taken} - \text{date of birth} \qquad (1)$$

For the UTKFace and Adience dataset, much processing is not needed as these datasets are well organized. We have merged these three datasets into one.

Fig. 2. Dataset Example

The total age range is divided into 8 age groups. These groups are: 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60-100 as it has been done in Adience dataset. The total dataset is then down-sampled according to age so that each age group has an equal number of images which results in a total of 18728 images. Fig. 2 shows some sample images from the dataset. The dataset distribution based on age and gender is shown in Fig. 3. Finally, the dataset is divided into three parts,

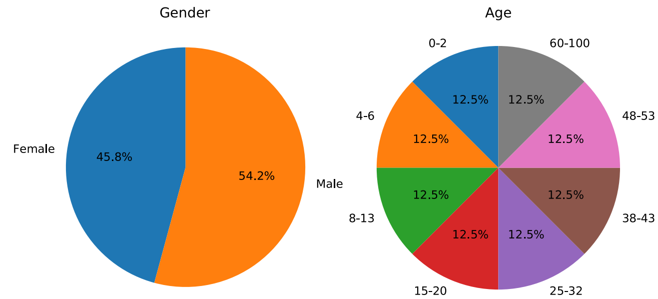

Fig. 3. Dataset Distribution

with an 8:1:1 ratio of training, validation and testing.

2) Creating and Training CNN Model: The CNN is a deep neural network architecture used in computer vision, such as image recognition. Computer vision and image recognition are not new concepts rather old ones. The architecture of CNN as shown in Fig. 4, has two parts which are feature extraction and then classification. The feature extraction part contains several convolution layers and pooling layers. This convolution process extracts features from the input images. The output of this process is called a feature map.

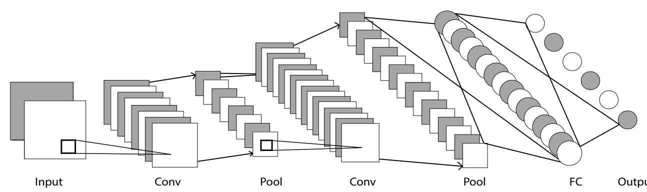

Fig. 4. A Simple CNN Architecture

Pooling layer is also called subsampling or down sampling. This layer reduces the feature map by retaining only the most important information such as taking the average or maximum value. The classification unit generates output according to input data. This unit has fully connected layers. It indicates that all the neurons of the previous layer are connected with all the neurons in the next layer. Typically it uses "Softmax" operation.

It is stated earlier that a deeper model gives a better result. But also the deeper model takes a lot of time to train and consumes large memory. Moreover, a deeper model is not good to integrate into mobile devices. So, we have tried to make trade-off between these.

The model that is used consists of a basic building block as shown in Fig. 5. 32 filters are used in each layer. Different sizes of kernels: 3X3, 5X5, 3X1, 1X3, 1X5, 5X1, 1X1 are applied in different layers. Max pooling is used for down-sampling without reducing dimension. By connecting three of these blocks, the CNN model is built. The last layers are fully

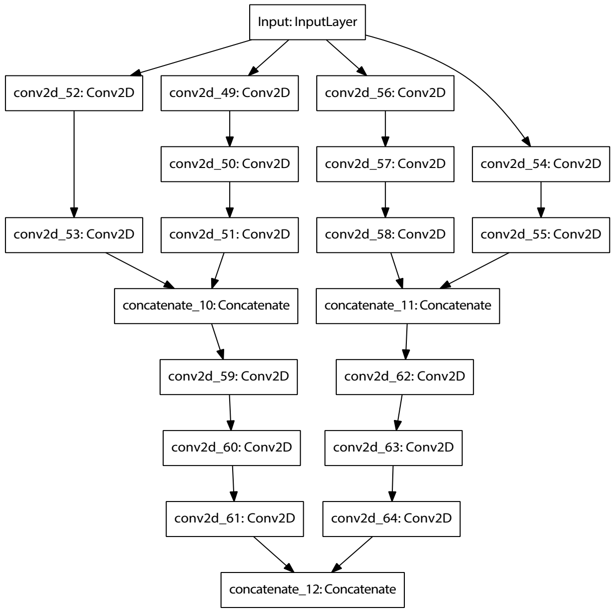

Fig. 5. Basic Building Block of Proposed CNN Architecture

connected layers. The output layer for gender prediction has two output units or classes, 1 for male and 0 for female. The output layer for age prediction has 8 output units or classes.

B. *Real-time Testing Phase*

In the real-time testing phase, after the completion of training, the model is used for predicting captured images via a camera and displaying the result continuously on the screen. This involves several steps as shown in Fig. 6.

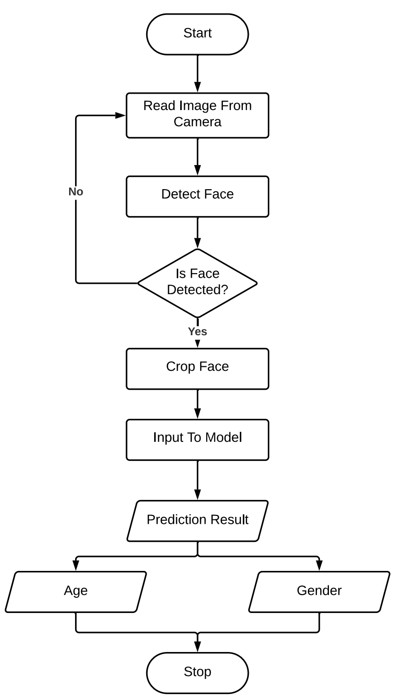

Fig. 6. Real-time Testing Phase

First, images are read from the camera. The face is detected from the images using OpenCV and haar cascade classifiers. OpenCV is a library of programming functions specially targeted at computer vision in real-time. And haar-cascade classifiers can be explained as a trained machine learning model that can detect a face. The face is cropped from the main image and sent to the model for prediction. Then, predicted age and gender are shown as output.

IV. Experimental Results & Analysis

Experiments were conducted on grayscale images for 100 epoch with a batch size of 32. Images have been resized to 64x64 to reduce training time.

A. *Performance Analysis*

One of the primary goals of this paper is to create a lightweight model. By the word lightweight model we mean a model that has a small number of parameters, and a small model size that is suitable for mobile integration. The total number of parameters of the proposed model is 210,050. And the final model size is 2.60MB which is very lightweight to execute in any mobile platform. Where some well-known models like VGG, Reset has size more than 200MB [25].
After 100 epochs, 81.35% and 51.59% training accuracy have been achieved for gender and age respectively. Test accuracy for gender is 80.76% and for age is 48.59% on our combined dataset. Fig. 7 shows the model accuracy and loss for age and gender separately.

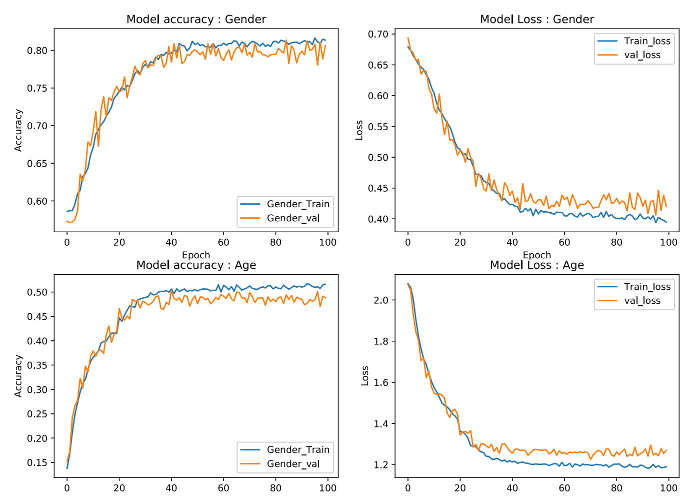

Fig. 7. Accuracy vs Epoch and Loss vs Epoch

In Table I, results are compared with several research works. The model is tested on the complete Adience dataset to compare with other research work. Accuracy of 46.71% for age and 81.38% for gender are obtained. The model is then trained only on the Adience dataset for the same purpose.

This is done after splitting the dataset into train, validation, and test (8:1:1). The training accuracy of 86.68% for gender and 64.22% for age are achieved and testing accuracy is 60.03% for age and 85.77% for gender. Our model is also

TABLE I
COMPARISON AMONG SEVERAL STATE OF THE ART

Research Work	Dataset Used	Test Accuracy	
		Age	Gender
[5]	Adience	61.3%	91%
[7]	Adience + LFW	40.25%	78.63%
[9]	Adience + MORPH-II	52.3%	88.2%
[13]	Adience	61.3%	88.9%
[11]	Adience	44.26%	85.16%
[16]	Adience + LFW	44.36%	80.64%
Proposed Model	Adience	60.03%	85.77%
Proposed Model	Adience + Wiki + UTKFace	48.59%	80.76%

tested, which has been trained using the Adience dataset, on both Wiki and UTKFace dataset. It shows moderate accuracy of 79% and 72% on gender respectively. But it performs very poorly for age, showing less than 32% accuracy.

B. Real-time Testing

The model that trained on a mixed dataset was tested in real-time data. The image was processed as described in Section III-B. We got promising output in the real time test. In real-time prediction run-time of the model was average 0.0654 ± 0.0056 seconds per prediction for 100 predictions on the Windows platform. Fig. 8 shows the real time test.

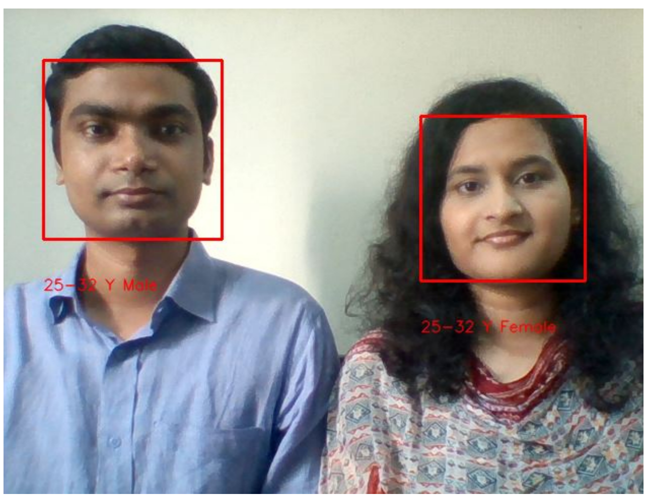

Fig. 8. Real-time Prediction

V. CONCLUSION & FUTURE WORKS

In this work, we developed a lightweight CNN model which is ideal to integrate in mobile devices. And we have achieved this without compromising too much accuracy. The model achieved accuracy of 48.59% for age and 80.76% for gender using a large combined dataset. Comparing with other state of the art works, it is clear that the model built on the mixed dataset performs well on unknown data and shows good results on the real-time test. We plan add more datasets from different sources and increase accuracy for age. We also want to develop a smartphone application that can predict gender and age in real-time using the proposed model. And our other idea is to upgrade the model for special cases such as faces with a mask.

REFERENCES

[1] B. Agrawal and M. Dixit, "Age estimation and gender prediction using convolutional neural network," in *International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology*. Springer, 2019, pp. 163–175.

[2] I. Rafique, A. Hamid, S. Naseer, M. Asad, M. Awais, and T. Yasir, "Age and gender prediction using deep convolutional neural networks," in *2019 International Conference on Innovative Computing (ICIC)*. IEEE, 2019, pp. 1–6.

[3] D. S. Al-Azzawi, "Human age and gender prediction using deep multi-task convolutional neural network," *Journal of Southwest Jiaotong University*, vol. 54, no. 4, 2019.

[4] A. Salihbašić and T. Orehovački, "Development of android application for gender, age and face recognition using opencv," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2019, pp. 1635–1640.

[5] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood, "Dager: Deep age, gender and emotion recognition using convolutional neural network," *arXiv preprint arXiv:1702.04280*, 2017.

[6] K. Zeeshan, H. Kaleem, R. Malik, and S. Khalid, "Deepgender : real-time gender classification using deep learning for smartphones," *Journal of Real-Time Image Processing*, 2017.

[7] Z. Liao, S. Petridis, and M. Pantic, "Local deep neural networks for age and gender classification," *arXiv preprint arXiv:1703.08497*, 2017.

[8] R. Rothe, R. Timofte, and L. V. Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision*, vol. 126, no. 2-4, pp. 144–157, 2018.

[9] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning cnn–elm for age and gender classification," *Neurocomputing*, vol. 275, pp. 448–461, 2018.

[10] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6848–6856.

[11] J.-H. Lee, Y.-M. Chan, T.-Y. Chen, and C.-S. Chen, "Joint estimation of age and gender from unconstrained face images using lightweight multi-task cnn for mobile applications," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2018, pp. 162–165.

[12] A. Rattani, N. Reddy, and R. Derakhshani, "Convolutional neural networks for gender prediction from smartphone-based ocular images," *IET Biometrics*, vol. 7, no. 5, pp. 423–430, 2018.

[13] S. Hosseini, S. H. Lee, H. J. Kwon, H. I. Koo, and N. I. Cho, "Age and gender classification using wide convolutional neural network and gabor filter," in *2018 International Workshop on Advanced Image Technology (IWAIT)*. IEEE, 2018, pp. 1–3.

[14] A. Kharchevnikova and A. V. Savchenko, "Neural networks in video-based age and gender recognition on mobile platforms," *Optical Memory and Neural Networks*, vol. 27, no. 4, pp. 246–259, 2018.

[15] M. Dileep and A. Danti, "Human age and gender prediction based on neural networks and three sigma control limits," *Applied Artificial Intelligence*, vol. 32, no. 3, pp. 281–292, 2018.

[16] Y. Zhang and T. Xu, "Landmark-guided local deep neural networks for age and gender classification," *Journal of Sensors*, vol. 2018, 2018.

[17] K. Jain, M. Chawla, A. Gadhwal, R. Jain, and P. Nagrath, "Age and gender prediction using convolutional neural network," in *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*. Springer, 2020, pp. 247–259.

[18] A. Greco, A. Saggese, M. Vento, and V. Vigilante, "A convolutional neural network for gender recognition optimizing the accuracy/speed tradeoff," *IEEE Access*, vol. 8, pp. 130 771–130 781, 2020.

[19] H.-T. Q. Bao and S.-T. Chung, "A light-weight gender/age estimation model based on multi-taking deep learning for an embedded system," in *Proceedings of the Korea Information Processing Society Conference*. Korea Information Processing Society, 2020, pp. 483–486.

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

[21] S. Manasa, J. S. Abraham, A. Sharma, and K. Himapoornashree, "Age, gender and emotion detection using cnn," *International Journal of Advanced Research in Computer Science*, vol. 11, no. Special Issue 1, p. 68, 2020.

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

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

[24] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[25] J. Fu and Y. Rui, "Advances in deep learning approaches for image tagging," *APSIPA Transactions on Signal and Information Processing*, vol. 6, 2017.