Musik in Bayern
ISSN: 0937-583x Volume 89, Issue 8 (Aug -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*256*

Accurate Age and Gender Prediction Using Hybrid CNN-LSTM Model from Face Images

Jayabharathi P[1], Rohini[2]

[1]Research Scholar, School of Computing sciences VISTAS Chennai, India,

[1] Department of Computer Application, A.M. Jain College,Chennai

[2]Professor, School of Computing Sciences VISTAS Chennai

[1]jayabharathi.p@gmail.com
[2]*indiarrohini16@gmail.com*

To Cite this Article

Jayabharathi P, Rohini. "Accurate Age and Gender Prediction Using Hybrid CNN-LSTM Model from Face Images**".*Musik in Bayern*, *89*(7), 163–169.https://doi.org/10.15463/gfbm-mib-2024-256**

Article Info

Received: 02-04-2024 Revised: 25-04-2024 Accepted: 5-07-2024 Published: 03-08-2024

Abstract— Many real-world applications benefit greatly from the categorization of ages and genders. Commercial and real-world applications cannot yet accurately age and gender predict real-life faces. An extremely difficult task is determining an individual's age and gender only by looking at their face. This is because there are so many different ways that facial images may vary within a person's age and classification categories. Face photos may be used to estimate age and gender using a hybrid Convolution Neural Network (CNN) and Long Sort-Term Memory (LSTM) model described here. Data cleaning and augmentation techniques have been used to preprocess the dataset. A CNN-LSTM model is utilized to identify gender and age from human facial images. In order to prevent overfitting, a Dropout Layer (DL) is added in the mid of the feature extraction and the LSTM learning blocks. Face images from IMDB and the Adience dataset, a standard dataset for face images, were utilized in this study. We test the proposed CNN-LSTM model against DCNN, SVM, and K-Means clustering algorithms to see how well it performs. The parameters of accuracy, precision, recall, and F1-Score are used to assess the results. Results demonstrate that the proposed CNN-LSTM outperforms the previous algorithms with accuracy around 98%, precision around 96%; recall around 95%; F1-scores around 75%; and accuracy around 98%.

Keywords— Convolutional Neural Network, Long Short Term Memory, Artificial intelligence, Deep Learning.

I. INTRODUCTION

As technology advances, it is becoming simpler and convenient to determine a person's age and gender using programs that combine pattern recognition and picture processing. There are limitations that prevent us from seeing the right age among the photographs, so determining someone's age isn't as simple as it would seem. [1]. Biometrics, vending machines, HCI, and video surveillance, as well as the entertainment and beauty sectors could all profit from utilizing age and gender classification in their systems.[1]. However, there are still a number of unsolved concerns when it comes to age and gender categorization. Commercial and real-world applications cannot yet reliably age and gender predict real-life faces. There have been a slew of approaches to the classification problem during the last several years. Many of these systems are made by hand and do not accurately predict the age

DOI https://doi.org/10.15463/gfbm-mib-2024-*256*

and gender of images taken in the wild. These demanding unconstrained imaging situations necessitate methods that don't rely on the inherent variability of human facial features and descriptors. These have been masterfully crafted by hand.[2].

Moreover, the lack of big and balanced datasets with precise labels and the non-linear connection between age/gender and face images add to this difficulty. For this assignment, most of the datasets available have significant age or gender imbalances, with a large proportion of persons falling between the 20-75 age bracket. [3]

In recent years, the advancement of AI models for facial recognition and classification has received considerable attention and has played a significant role in finding answers to complicated real-life challenges. Artificial Intelligence (AI) encompasses the Gender and Age Prediction (DL) implementation to faces. Gender and age may be inferred from a picture in this approach [4]. An extremely difficult task is determining an individual's age and gender only by looking at their face. This is because there are so many different ways that facial images may vary within a person's age and gender classification categories. Age and gender prediction has been the subject of a number of studies in recent years. Earlier studies relied heavily on manually derived facial picture attributes, which were then fed into a classifier. Deep learning models have had great success in computer vision issues in the previous decade, which has led to a change in recent efforts on age and gender prediction [5]. There has been a substantial amount of work done on estimating age based on a facial image using actual or biological age estimation [6].

II. RELATED WORKS

Salma Fayaz Bhat et al [7] used CNNs to extract information that improved the accuracy of gender prediction significantly. CNN coupled with deep learning approaches have been used to reach cutting-edge performance. Extensive trials on the IMDB-WIKI dataset, the biggest publicly accessible compendium of facial images with Gender labels, were used to determine the image-based Gender estimation. An algorithm has been presented by Rajendr et al. [8]. An improved CNN is implemented in the proposed algorithm, making it more accurate than current systems. Brown hair, smiling, and wearing spectacles are some of the most prominent traits used in the algorithm's attempt to predict class labels. In order to generate the category class label from an input image, neural networks are preferable to machine learning techniques, since they are more efficient at this task.

Statistical pattern recognition has been suggested by Ishita Verma et al [9]. The suggested technique uses a Convolutional Neural Network (ConvNet / CNN), a Deep Learning algorithm, to extract features. For example, CNN learns the weight and biases associated with various parts of an image, allowing it to discern between them. Compared to other classification techniques, ConvNet needs much less preprocessing. Despite the fact that the filters are constructed by manually, ConvNets may be trained to identify and use these features. Gender and age group of facial images may be accurately predicted using a deep learning framework developed by Amirali Abdolrashidi et al. [10]. This model is able to concentrate on the most significant and relevant regions of the face thanks to the use of an attention mechanism. This helps it produce more accurate predictions. With the use of a multi-task learning approach, they were able to substantially improve the accuracy of their age prediction model by including gender predictions into its feature embedding.

Vikas Sheoran et al [11] have proposed two techniques to for age and gender categorization using a unique CNN architecture and transfer learning-based pre-trained models. We were capable of defeating overfitting thanks in large part to these pre-trained models. When evaluated on real-world images, it was observed that their models generalized quite well with minimal overfitting. Using the synergy of two classifiers, Mingxing Duan and colleagues [12] have developed a hybrid structure that incorporates a Convolutional Neural Network (CNN) and an Extreme Learning Machine (ELM). Their strengths are utilized in a hybrid architecture. ELM classifies the intermediate results after CNN has extracted the features from the input pictures. To test their hybrid structure, they have used two well-known datasets, MORPH-II and Adience Benchmark.

Page Layout

Your paper must use a page size corresponding to A4 which is 210mm (8.27") wide and 297mm (11.69") long. The margins must be set as follows:

- Top = 19mm (0.75")
- Bottom = 43mm (1.69")
- Left = Right = 14.32mm (0.56")

Your paper must be in two column format with a space of 4.22mm (0.17") between columns.

III. PROPOSED SOLUTION

In this paper, a hybrid CNN-LSTM model is developed for age and gender estimation from face images. Initially the dataset has been preprocessed by applying data cleaning and data augmentation operations. Then CNN-LSTM model is applied for predicting the gender and age from human face images. In order to avoid overfitting problem, a DL is included in the mid of the feature extraction and LSTM learning blocks.

IV. DATA CLEANING AND BALANCING

DOI https://doi.org/10.15463/gfbm-mib-2024-*256*

To overcome the overfitting problem of the network, the first and foremost step is the data adequacy for training the deployed network. In the case of the age estimation task, we observed that this dataset consists of a huge number of images for people aged between compared to children and elderly people. In this situation, if we train the model with this class sparsity, it is impossible to obtain best outputs for the class which is imbalanced in real-time. This data sparsity problem is avoided by taking the following steps:

Randomly choose the number of samples from the class which has sufficient observations so that the comparative ratio among the class will be retained;

- Manually filter out the wrongly annotated samples from each class
- The data from the classes with lesser samples are enhanced from another benchmark dataset (Adience dataset)
- Perform necessary offline data augmentation functions to make the class balanced. The performed data augment operations, such as right flipping, rotation, scaling and adding noise.

During training, we perform online data augmentation by rescaling the input image into 256X256 pixels and taking a center crop of 224X224 pixels from the 256X256 size image and pass it on for the training.

V. HYBRID CNN-LSTM MODEL

The CNN characteristics are extracted using three 1D convolutional layers (CLs). Between the two CL, layers of MaxPooling (MP) and Rectified Linear Unit (ReLU) were put together. There are several techniques to learn low-level properties in an input, but one of the most effective is to use CLs. Features in the input are tracked using CLs, which create a feature map. This has a restriction. Little changes in an input feature's location will produce a new map of that feature. While an activation function improves the model's capacity to learn complex structures, a pooling layer is often applied after a CL to help mitigate the feature map's inherent invariance. When building our model, we introduced a layer called MaxPooling, which is a two-fold downsampling scheme, which minimizes the total computational load.

The dropout layer is a novel technique to eliminate the problem of overfitting in the creation of any deep learning model. This layer randomly selects and disables the neurons throughout training. For the purpose of preventing overfitting, For the final output, the dropout layer is connected to the sequence learning block's output, and the fully connected layer is connected to that dropout layer as well. When building a CNN model, it is typical to use a coarse-to-fine approach. Because of the large number of tunable factors, this structure adds a greater level of computational complexity. Using the pyramid design, which was also used to develop our CNN feature extraction block, we started with a large number of kernels and decreases the amount by a constant as we progressed upwards from the pyramid. For the first CL, we choose a kernel size of 48, which is then lowered to 32 and 16 for the 2nd and 3rd CLs, respectively. Overfitting is prevented, and the number of trainable parameters is reduced, using this structure.

The sequence learning block was applied to 3 layers of LSTN network, each layer consisting of 20 neurons. Return sequence is assigned as true for layer 1 and 2 of LSTM and false for the layer 3. As a result, the final LSTM layer will only produce one hidden state for the network. Before the fully connected layer, we used a dropout layer to prevent over-fitting and other problems. 20 neurons make up the entire layer. According to the number of lookaheads, the output layer's number of neurons vary from one to six (up to 3 h ahead load forecasting).

As a result, we used a loss function known as the mean absolute error (MAE) (MAE). The training procedure begins with the loading of both training and validation data. The validation loss is calculated and checked at the end of each period to see if it has decreased. If the validation loss is reducing and the epochs are increasing, it is stored with the updated weights. While this is true, the learning rate is slowed and the number of epochs is increased if the validation loss is not minimized for 10 successive epochs. The training is completed when the epoch is equal to 150. In order to avoid overfitting, the test data is loaded with a last-optimum model for prediction and classification.

VI. CNN ARCHITECTURE

In all of our tests, we've used the network design we've presented to classify participants by age and gender. With only 3 CLs and 2 Fully Connected Layers (FCLs), this network has a minimal number of neurons on each layer to begin with. When compared to the massive designs, this is a small thing. As a result of both our intention to decrease the danger of overfitting as well as the nature of our challenges, we chose a smaller network design. Age classification on the Adience set has eight classes and gender classification only two. It's a far cry from facial recognition, which requires tens of thousands of identify classes to train. The network immediately processes all three colour channels. 256X256 images are rescaled to 227X227 before being fed into the network. The next 3 CLs are then defined as follows:

1. A ReLU, the maximum value of 3 X 3 regions with two-pixel strides, and a local response normalising layer follow 96 filters of size 3X7X7 pixels.
2. The 96 X 28 X 28 output from the preceding layer is processed by 256 filters of size 96 X 5 X 5 pixels in the second CL. ReLU, a max pooling layer, and a local response normalising layer are then utilised with the same hyper parameters.

DOI https://doi.org/10.15463/gfbm-mib-2024-*256*

3. Before applying ReLU and a max pooling layer, 384 filters of 256 X 3 X 3 pixel size are applied to the 256 X 14 X 14 blob in the final CL layer. Following that, the FCLs are defined as follows:
4. A first FCL with 512 neurons follows the ReLU and dropout layers and gets the output of the 3 CL.
5. Another layer with 512 neurons that receives the 512-dimensional output of the previous layer and is completely connected, followed by a ReLU and a dropout layer.
6. An entirely new layer has been added for the last two classes in terms of gender and age.
7. The output of the soft-max layer allocates a probability to each class once all layers have been applied. For a prediction, the most likely class is chosen for the given test image.

Explanation for the Architecture
Step1: Take the Data as input
Step2: Apply Convolution algorithm on the collected data
Step3: Apply ReLU then send this into Pooling
Step4: After Pooling norm1 LRN will be applied then Convolution 2 algorithm will be applied
Step5: Then ReLU2 will be continued and entered in to pooling 2 operation
Step6: Then this process will continue untill FC6 will be attained.
Step7: Based on FC6 dropout and inner product will be calculated.
Step8: After FC8 probability of SOFTMAX will be found and the final Probability will be calculated

VII. TESTING AND TRAINING

Initialization. The weights in each layer are generated randomly using a Gaussian distribution with a standard deviation of only 0.01. Networks are not started with pre-trained models; instead, they are trained from scratch utilising just images and labels provided by their benchmark. When compared to facial-recognition-based CNN implementations, which require images from hundreds of thousands to train the model, this is a stark contrast.

Using sparse binary vectors corresponding to the ground truth classes, the target values for training are represented. Gender and age classification tasks use label vectors with the number of classes in the number of training images. For each training image, the goal, label vector contains a 1 in the index of ground truth and 0s elsewhere.

Network training. Overfitting may be minimised in two ways in addition to using a lean network architecture. Dropout education is used in the first place (i.e. randomly setting the output value of network neurons to zero). A dropout ratio of 0.5 exists in one of the network's two dropout layers (a 50 percent probability of changing the output value of a neuron to zero). Randomly reflecting 227 X 227 pixels from the input image in each forward and backward training cycle. This is similar to the several crop and mirror versions that were used.

The training itself is done using a fifty-image batch size and stochastic gradient descent. The initial learning rate is e^{-3}, however after 10K iterations, it drops to e^{-4}.

Prediction. The following two techniques are applied to forecast the age and gender of new faces:

1) **Center Crop**: Cropped image of a face fed to the network with a 227 x 227 cropped around its centre
2) **Over-sampling**: From the edges of the facial image, we take out four 227 X 227-pixel crop sections and an additional 227-pixel crop region that covers the centre. Each of the five images, as well as their horizontal reflections, are shown to the network. In order to come at its final forecast, it uses the average of all of these variations as a foundation.

The quality of our findings could be significantly impacted by even a tiny portion of misalignment in the Adience images due to occlusions, motion blur, or other issues. As a second example, the network is given numerous translations of the same face in order to adjust for minor alignment problems, rather than enhancing alignment quality.

VIII. EXPERIMENTAL RESULTS

The proposed work is implemented in Python with Anaconda Platform.

A. Dataset

We have used the face pictures from IMDB and Adience dataset from https://www.kaggle.com/ttungl/adience-benchmark-gender-and-age-classification which is a benchmark dataset for face images which has various real-world imaging constraints. It has a total of 26,580 photos of 2,284 candidates under various age groups. For age and gender prediction, pre-trained age and gender models have been used.

For age and gender, the age_deploy and gender_depoly prototxt files are used for training, which describe the network configuration and the age_net and gender_net caffemodel files define the internal states of the parameters of the layers.

XI. PREDICTED RESULTS

The following figure shows the predicted age and gender estimations for the input images using CNN-LSTM classifier.

Musik in Bayern
ISSN: 0937-583x Volume 89, Issue 8 (Aug -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*256*

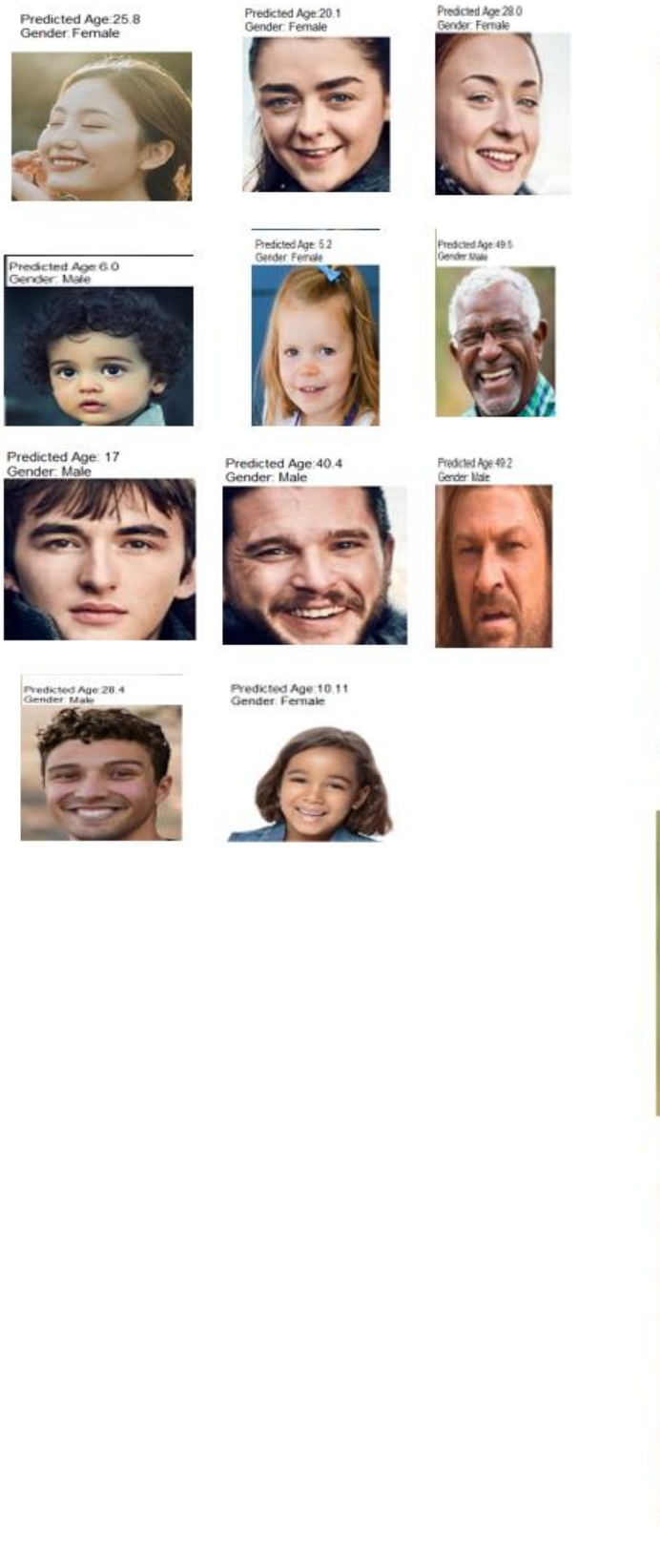

Fig.1. Input images of Actual Age

Musik in Bayern
ISSN: 0937-583x Volume 89, Issue 8 (Aug -2024)
https://musikinbayern.com
DOI https://doi.org/10.15463/gfbm-mib-2024-*256*

Fig. 2. Predicted age and gender for the input images

Figure 3 shows the loss function against the Epoch values for training and validation phases. The iteration with the minimum validation loss will have the optimum weights.

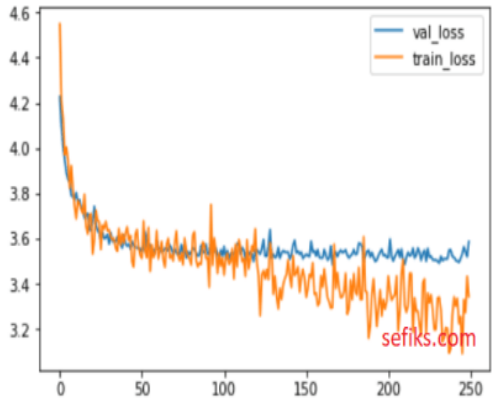

Fig. 3. Loss Curves for trainig and validation

IX. COMPARISON RESULTS

In this section, we compare the performance of the proposed CNN-LSTM model with DCNN, SVM and K-Means clustering algorithms. The performance is evaluated in terms of the metrics accuracy, precision, recall and F1-Score.

Table 1 and Figure 7 show the comparison results of the metrics for these algorithms.

TABLE I
COMPARISON RESULTS FOR VARIOUS ALGORITHMS

Metrics	**CNN-LSTM**	**DCNN**	**SVM**	**K-Means**
Accuracy	0.9814	0.9537	0.9246	0.9345
Precision	0.9641	0.9425	0.9011	0.9132
Recall	0.9547	0.9207	0.9057	0.9131
F1-Score	0.9514	0.9277	0.9019	0.9081
Accuracy	0.9814	0.9537	0.9246	0.9345

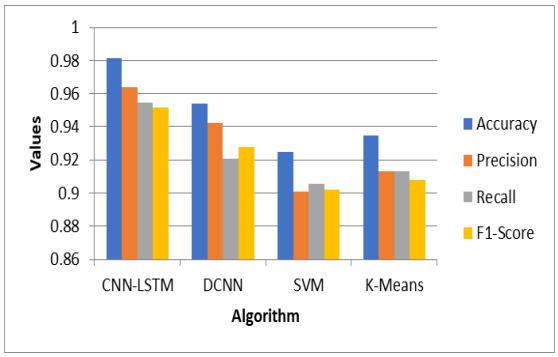

Fig. 3. Performance Comparison of Various Algorithm

As seen from Figure 4, the proposed CNN-LSTM shows the superior performance over the other algorithms by attaining accuracy around 98%, precision around 96%, recall around 95% and F1-score around 75%.

DOI https://doi.org/10.15463/gfbm-mib-2024-*256*

X. CONCLUSION

A hybrid CNN-LSTM model has been developed for age and gender estimation from face images. Initially the dataset has been preprocessed by applying data cleaning and data augmentation operations. Then CNN-LSTM model is applied for predicted the gender and age from human face images. We have used the face pictures from IMDB and Adience dataset which is a benchmark dataset for face images. The performance of the CNN-LSTM model has been compared with DCNN, SVM and K-Means clustering algorithms. The performance is evaluated in terms of the metrics accuracy, precision, recall and F1-Score. From the results, it has been seen that the proposed CNN-LSTM shows the superior performance over the other algorithms by attaining accuracy around 98%, precision around 96%, recall around 95% and F1-score around 75%.

REFERENCES

[1] Meghana, A. S., Sudhakar, S., Arumugam, G., Srinivasan, P., & Prakash, K. B. (2020). "Age and Gender prediction using convolution, resnet50 and inception resnetv2". *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(2), 1328-1334.

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

[3] Sheoran, V., Joshi, S., & Bhayani, T. R. (2020, December). "Age and Gender Prediction using Deep CNNs and Transfer Learning". In *International Conference on Computer Vision and Image Processing* (pp. 293-304). Springer, Singapore. B. R. Jackson and T. Pitman, U.S. Patent No. 6,345,224 (8 July 2004)

[4] Hassan, M., Wang, Y., Wang, D., Li, D., Liang, Y., Zhou, Y., & Xu, D. (2021). "Deep learning analysis and age prediction from shoeprints". *Forensic Science International*, *327*, 110987.

[5] 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*. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[6] Islam, M., & Baek, J. H. (2021). "Deep Learning Based Real Age and Gender Estimation from Unconstrained Face Image towards Smart Store Customer Relationship Management". *Applied Sciences*, *11*(10), 4549.

[7] Bhat, S. F., & Dar, T. A. (2019, September). "Gender prediction from images using deep learning techniques". In *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)* (pp. 1-6). IEEE.

[8] Rajendra, G., Sumanth, K., Anjali, C., Pardhasai, G., & Supraja, M. (2021, August). "Gender Prediction using Deep Learning Algorithms and Model on Images of an Individual". In *Journal of Physics: Conference Series* (Vol. 1998, No. 1, p. 012014). IOP Publishing.

[9] Verma, I., Marhatta, U., Sharma, S., & Kumar, V. "Age Prediction using Image Dataset using Machine Learning". *International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN*, 2278-3075.

[10] 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*.

[11] Sheoran, V., Joshi, S., & Bhayani, T. R. (2020, December). "Age and Gender Prediction using Deep CNNs and Transfer Learning". In *International Conference on Computer Vision and Image Processing* (pp. 293-304). Springer, Singapore.

[12] Duan, M., Li, K., Yang, C., & Li, K. (2018). "A hybrid deep learning CNN–ELM for age and gender classification". *Neurocomputing*, *275*, 448-461.