

# Automated Deep Learning based Age and Gender Classification Model using Facial Features for Video Surveillance

<sup>1</sup>S. Sivachandiran, <sup>2</sup>Dr. K. Jagan Mohan, <sup>3</sup>Dr. G. Mohammed Nazer

<sup>1</sup>Research Scholar, Department of Computer Science & Engineering, Annamalai University, Tamil Nadu, India.

<sup>2</sup>Associate Professor, Department of Information Technology, Annamalai University, Tamil Nadu, India.

<sup>3</sup>Professor, Department of Computer Science, RAAK College of Arts and Science, Puducherry. [sivachandiran.s@gmail.com](mailto:sivachandiran.s@gmail.com), [aucsejagan@gmail.com](mailto:aucsejagan@gmail.com), [kgmohammednazer@gmail.com](mailto:kgmohammednazer@gmail.com)

---

## ABSTRACT

Video surveillance plays a vital role in ensuring security in public and private places over the globe. Facial image analysis using computer vision (CV) and artificial intelligence (AI) based tools become essential. Automated age and gender classification in facial image analysis is a primary process that finds useful in several real time applications such as target advertisements, forensics, human computer, etc. But age and gender classification remains a challenging process due to differences in visual angle, facial expression, background, and facial image appearance. It is more challenging in the unconstrained imaging conditions. In this aspect, this article introduces an automated deep convolutional neural network based age and gender classification (ADCNN-AGC) model using facial images. The proposed ADCNN-AGC model aims for determining the age and gender of the persons who exist in the facial images. To accomplish this, the ADCNN-AGC model follows a two stage process namely face recognition and age/gender classification. Primarily, Multi-task Cascaded Convolutional Networks (MTCNN) model is utilized for the detection of faces in the input images. Besides, the Efficient Net model is applied for the proper extraction of feature vectors which are then passed into the One-Dimensional Convolutional Neural Network (1-DCNN) for classification procedure. The performance validation of the ADCNN-AGC model has been tested using benchmark datasets and the outcomes are observed in many aspects. The experimental outcomes reported the enhanced performance of the ADCNN-AGC model over recent state of art approaches.

**Keywords:** Video surveillance, Facial images, Age classification, Deep learning, Gender classification, Face recognition.

---

## 1. Introduction

In recent times, the usage of video surveillance for security reasons has been increased [1]. Recognize people with automated face detection system that have great significance these days; however, the lower quality of the videos make it harder and remain a challenge that several authors are trying to resolve. In the past few years, Facial analysis has received considerable attention in the computer vision field [2]. The human face comprises features that describe age, identity, emotions, ethnicity, and the gender of person [3]. Amongst these features, gender and age classification could be supportive in many real-time applications involving electronic customer relationship management, electronic vending machines, biometrics, entertainment, human-computer interaction, forensic art, cosmetology, security, and video surveillance [4]. Estimation and Automatic facial recognition of age and gender using machine learning (ML) model has gained considerable interest over a year [5] and become more relevant because of the large amount of face images on the web, and particularly on social networking media. But various problems in gender and age classification remain a challenge. Age and gender classification of unfiltered real-time faces are still meeting the requirement of real-time and commercial applications despite the progress CV keeps on making with the constant development of the novel technologies [6].

In recent times, several techniques were introduced to resolve classification problems. Most of the techniques are hand-engineered that unsatisfactorily performed on the gender and age prediction of unrestrained in-the-wild images [7]. This handcrafted method depends on the difference in dimensional of face descriptor and facial feature [8] that cannot deal with the varying degree of variation observed in the challenging unconstrained imaging condition. The image in those classes

has some variation in noise, appearance, lighting, and pose that might affect the capacity of manually developed CV method for precisely classifying the gender and age of the images. Newly, deep learning (DL)-based method shows outstanding performance, especially in the gender and age classifications of unfiltered face images [9, 10].

This article introduces an automated deep convolutional neural network based age and gender classification (ADCNN-AGC) model using facial images. The proposed ADCNN-AGC model aims to determine the age and gender of the persons exist in the facial images. Primarily, Multi-task Cascaded Convolutional Networks (MTCNN) model was utilized for the detection of faces in the input images. Besides, Efficient Net model is applied for the proper extraction of feature vectors which are then passed into the One-Dimensional Convolutional Neural Network (1-DCNN) for classification process. The performance validation of the ADCNN-AGC model has been tested using benchmark datasets and the outcomes are examined in numerous aspects. In short, the paper contributions are summarized as follows.

- To develop a new ADCNN-AGC model for age and gender classification using facial images.
- To employ MTCNN model for effective recognition of faces in input images.
- To introduce an EfficientNet based feature extractor to derive feature vectors.
- To exploit 1D-CNN model for age and gender classification on surveillance videos.
- To validate the performance of the ADCNN-AGC model on benchmark datasets.

## 2. Literature Review

Lu et al. [11] presented a scheme named EAGR for recognizing gender, emotion, and age that perceive user gender, emotion, and age according to the face recognition. First, The EAGR scheme employs normalized facial cropping (NFC) as a pre-processing technique to train information beforehand data augmentation, later employs CNN as three training methods to recognize seven sentiments, two genders, and four age groups. Rouhsedaghat et al. [12] proposed FaceHop that provides an interpretable non-parametric ML solution. It has desired features namely a low training complexity, small model size, low-resolution input images, and a small training data amount.

Fayyaz et al. [13] developed an architecture that considers the integration of deep CNN and traditional methods for classifying gender. For realizing it, HOG- and LOMO-enabled lower-level feature is extracted for handling illumination, rotation, and viewpoint variances in the image. At the same time, VGG19- and ResNet101-based typical deep CNN architecture is applied for acquiring the deep feature that is stronger against pose variation. Hsu et al. [14] developed a data augmentation technique by changing the training image that resembles real-time images to enhance the efficiency of the network by offering different varieties to the training sample.

Khattak et al. [15] developed an effective deep learning method with a CNN system for emotion classification from facial image and identification of gender and age from the facial expression effectively. Ramya et al. [16] developed a CNN-based technique that doesn't employ raw facial images for learning the feature. Firstly, featured image is built from raw image to characterize a set of important characteristics. Next, they are extracted by utilizing CNN-based network where the featured image is provided as input. Later, they are collectively combined and provided as input to the classification, SVM, for recognizing the gender. In [17], developed a hierarchical system that creates higher age estimation precision. It comprises a collection of pre-trained 2-classes CNN (Google-Net) with the finest age gap that could establish the face images in the age groups

## 3. The Proposed Model

In this article, a new ADCNN-AGC model was developed to determine the age and gender of the persons who exist in the facial images. The ADCNN-AGC model utilized an MTCNN model for the detection of faces in the input images. Besides, Efficient Net model is applied for the proper extraction of feature vectors that are then passed into the 1-DCNN for classification process. Fig. 1 depicts the block diagram of ADCNN-AGC technique.

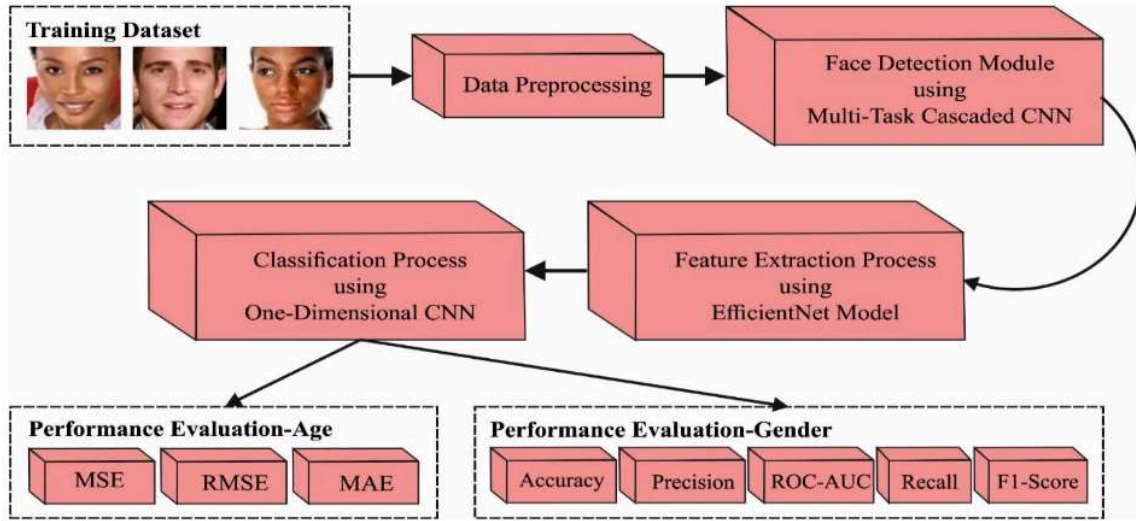


Fig. 1. Block diagram of ADCNN-AGC model

### 3.1. Face Detection using MTCNN Model

At the initial stage, the MTCNN model is utilized for the recognition of faces in the input images [18]. MTCNN is a CNN-based face detection model which comprises network *P*-Net that is a fully convolutional network, *R*-Net and *O*-Net are normal CNN. The size of input image of MTCNN could be some size. Assumed an image, resize them into distinct scales to construct an image pyramid as input of the three-phase cascaded architecture. Three tasks should be performed for training the network that is facial landmark localization, face classification and bounding box regression.  $y_i^{box}$  indicates the bounding box result attained from the network and  $y_i^{box}$  denotes the nearby ground truth. The loss for face classification is the cross-entropy loss, in which  $p_i$  denotes the possibility of the face created by network, and  $y_i^{det} \in \{0,1\}$  denotes the ground truth label.

$$L_i^{det} = -\left(y_i^{det} \log(p_i) + (1 - y_i^{det})(1 - \log(p_i))\right) \quad (1)$$

$$L_i^{box} = \|y_i^{box} - y_i^{box}\|_2^2 \quad (2)$$

$$L_i^{landmark} = \|y_i^{landma:k} - y_i^{landma:k}\|_2^2 \quad (3)$$

Fixed a threshold  $t_1$ , and the region where probability of humans face higher than  $t_1$  are the input of NMS. Afterward the NMS method, the residual box is measured to the size of  $48 \times 48$ . Afterward the NMS method, the residual box is the output of the MTCNN.

### 3.2. EfficientNet based Feature Extraction

For gender and age classification, the initial stage is to extract the feature vectors using the EfficientNet model [19]. The EfficientNet was established by Google in 2019. Compared with other typical CNN, it has maximum accuracy and minimum parameters. The typical network of EfficientNet was projected by a multi-objective neural infrastructure search, afterward, the typical network was scaled dependent upon the resolution, depth, and width to obtain a balance amongst them. The compound scaling approaches were defined in the subsequent formula as:

$$depth = \alpha^\phi, \quad (4)$$

$$width = \beta^\phi, \quad (5)$$

$$resolution = \gamma^\phi, \quad (6)$$

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2, \quad (7)$$

$$\begin{aligned} \alpha &> 1, \\ \beta &> 1, \\ \gamma &> 1, \end{aligned} \quad (8)$$

where  $\alpha, \beta$ , &  $\gamma$  is measured utilizing a lesser grid search. Primarily, an EfficientNetB0 acts a  $3 \times 3$  convolutional function on input image, then the subsequent 16 mobile inverted bottleneck convolutional were employed to moreover extract the image features. Eventually,  $1 \times 1$  convolutional and global average pooling function, the classifier outcome was reached from the FC layer. Afterward, for every convolution operation from the networks, batch normalization was executed. The loss function was commonly utilized to evaluate the effect of elements. With an enormous amount of data, the machine discovered the law utilizing autonomous learning and making forecasts. The loss function was executed to evaluate the degree of deviation amongst the outcome of the actual and forecasted value. During the network trained approach, the function was continually upgraded still an optimum appropriate outcomes are introduced for decreasing the error.

The cross entropy loss function is named as Softmax function is mostly employed to measure the gap amongst the actual and forecasted values but the CNN approach controls classifiers problem. The Softmax function is defined in the subsequent formula:

$$L_{softmax\ loss} = -\frac{1}{N} \sum_{i=1}^N \log \left( \frac{e^{a_k}}{\sum_j e^{a_j}} \right), \quad (9)$$

In which  $N$  stands for the neuron count in the resultant layer.  $a_k$  defines the input signal. Next, it makes sure of bias correction:

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (10)$$

$$\widehat{v}_t = \frac{\gamma_t}{1 - \beta_2^t}. \quad (11)$$

The final formula to weighted upgrade is

$$\omega_{t+1} = \omega_t - lr \cdot \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t + \epsilon}}, \quad (12)$$

In which  $lr$  demonstrated the rate of learning and  $\epsilon$  signifies the hyper-parameter (fixed to  $1e - 3$ ).

### 3.3. Age and Gender Classification using 1D-CNN Model

In the classification procedure, the feature extracting vectors are passed as to 1D-CNN technique to classify age and gender [20]. A 1D-CNN is similar to a typical NN however, it generally has raw dataset as input rather than hand-engineered feature. The input dataset can be processed by trainable convolution layer to learn a suitable representation of the input. Based on the “local connectivity” concept, the neuron in a layer is interconnected with a smaller region of the preceding layer.

$$T = F(X|\theta) = f_L(\dots f_2(f_1(X|\theta_1)|\theta_2)|\theta_L) \quad (13)$$

Whereas  $L$  indicates the amount of hidden layers. For the convolution layer, the function of  $l$ -th layer is formulated by:

$$T_l = f_l(X_l|\theta_l) = h(W \otimes X_l + b), \theta_l = [W, b] \quad (14)$$

Which  $\otimes$  represents the convolutional process,  $X_l$  denotes a 2D input matrix of  $N$  feature maps,  $W$  denotes a group of  $N$  1D kernel utilized to extract a novel group of features under the input array,  $b$  denotes the bias vector, and  $h$  indicates the



activation function. Also, pooling layer is employed between the convolution layer to increase the area covered by the following receptive field.

$$T_l = f_l(X_l|\theta_l) = h(WX_l + b), \theta_l = [W, b] \quad (15)$$

In the trained method, the parameter of the network is attuned based on the backpropagated classifier error and the parameter of the network is enhanced for minimizing a suitable loss function. Fig. 2 demonstrates the structure of 1D-CNN method.

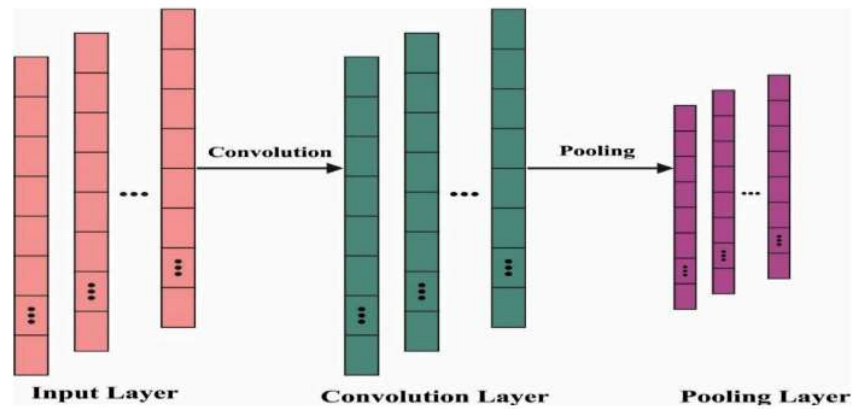


Fig. 2. Structure of 1D-CNN model

#### 4. Performance Validation

In this study, the experimental validation of the ADCNN-AGC model is tested using the UTKFace dataset [21]. It can be large scale facial dataset with age span of 0-116 years. Fig. 3 shows sample test images. The dataset contains facial images with annotations of age, gender, and ethnicity. The images cover large differences in pose, resolution, illumination, occlusion, facial expression, etc.

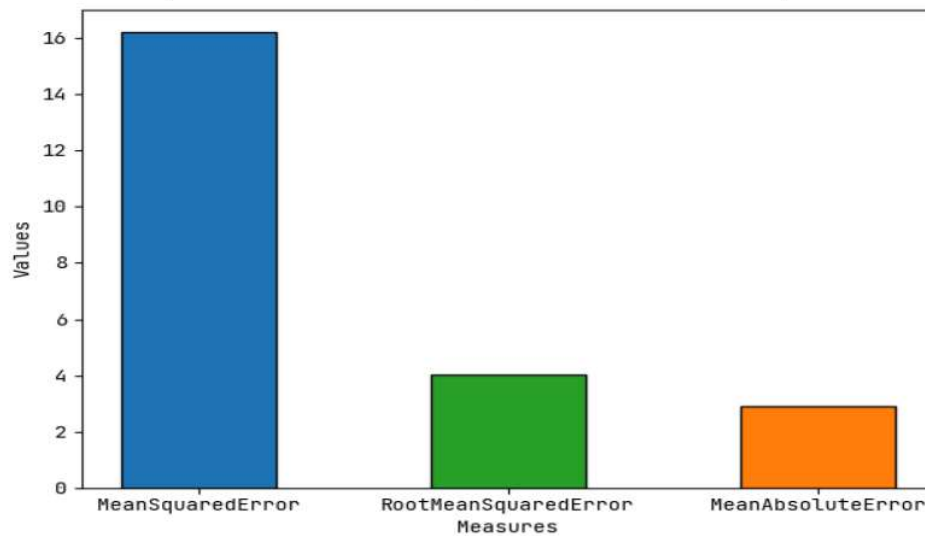


Fig. 3. Sample Test Sequences

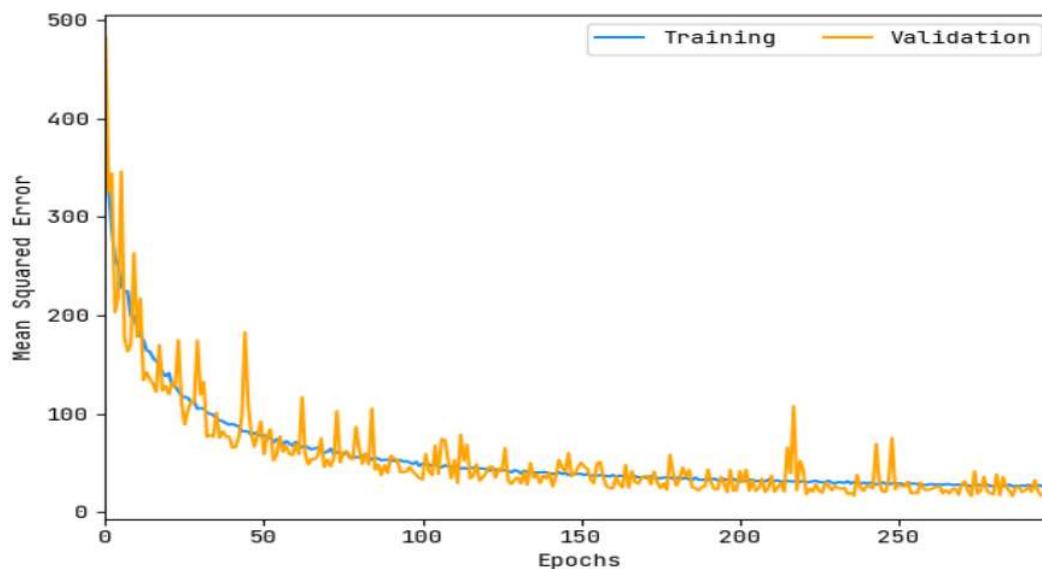
Table 1 and Fig. 4 provides the age prediction results offered by the ADCNN-AGC technique. The experimental results indicated that the ADCNN-AGC approach has accomplished effective results with MSE, RMSE, and MAE of 16.1974, 4.0246, and 2.8983 respectively. Fig. 5 shows the MSE offered by the proposed model on the training and validation processes. The results indicated that the proposed model has resulted to least MSE values.

**Table 1** Age prediction results of ADCNN-AGC model

Measures	Values
MSE	16.1974
RMSE	4.0246
MAE	2.8983



**Fig. 4.** Age prediction results of ADCNN-AGC model



**Fig. 5.** MSE analysis of ADCNN-AGC model for Age Prediction

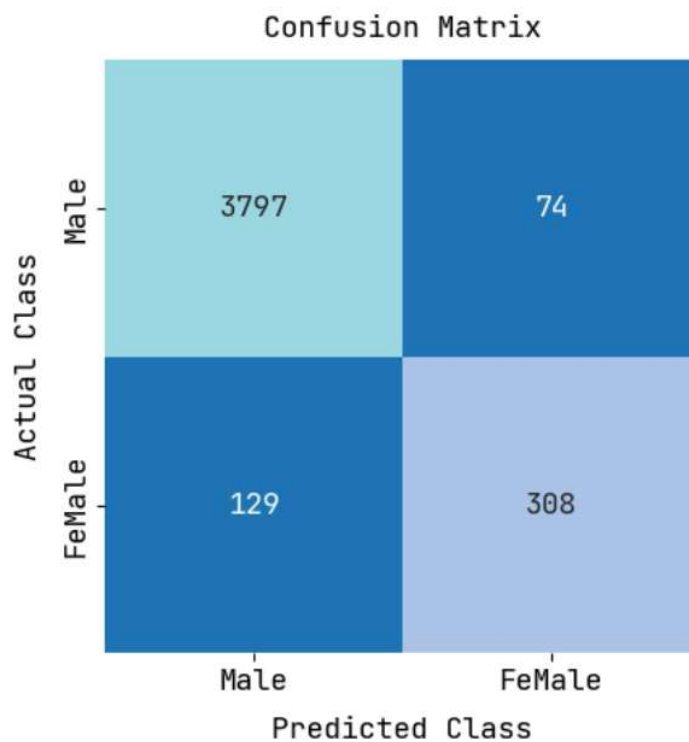
**Fig. 6.** Confusion matrix of ADCNN-AGC model

Fig. 6 depicts the confusion matrix offered by the ADCNN-AGC model on gender classification process. The figure indicated that the ADCNN-AGC model has identified a total of 3797 samples under male class and 308 samples under female class.

Table 2 and Fig. 7 report the age classification outcomes of the ADCNN-AGC model on the test data. The results implied that the ADCNN-AGC model has accomplished effective age classification results with  $accu_y$  of 95.29%,  $prec_n$  of 88.67%,  $reca_l$  of 84.28%,  $F1_{score}$  of 86.31%, and  $ROC_{score}$  of 95.35%.

**Table 2** Gender classification results of ADCNN-AGC model

Metrics	Values
Accuracy	95.29
Precision	88.67
Recall	84.28
F1-Score	86.31
ROC Score	95.35

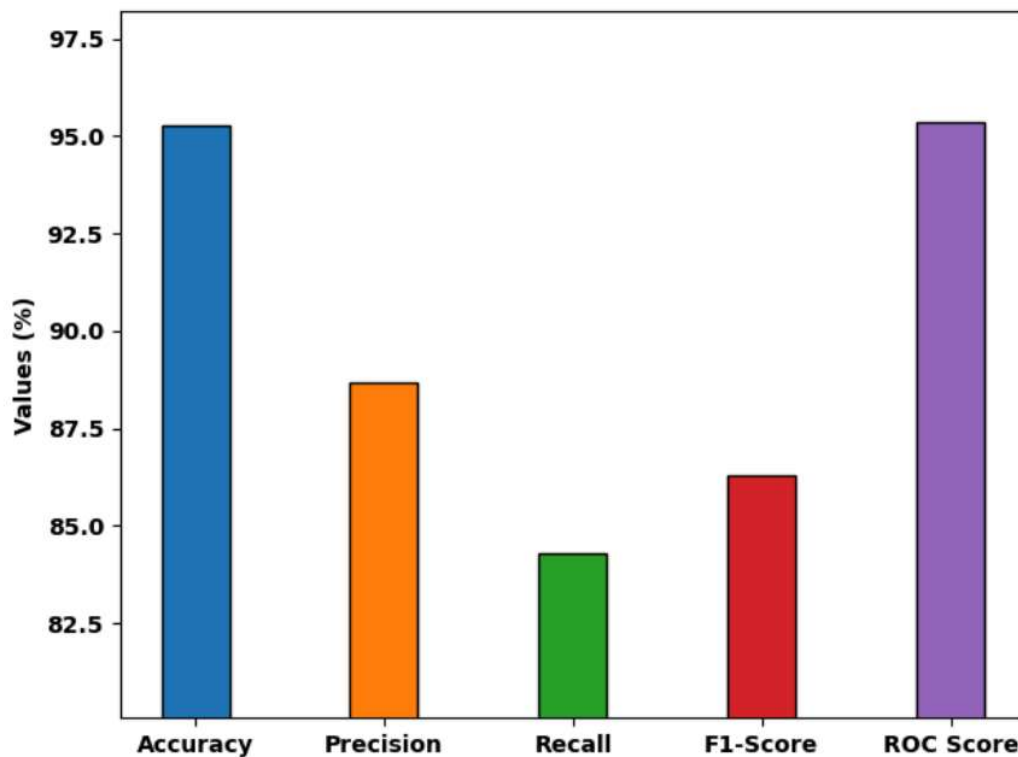


Fig. 7. Overall Gender classification results of ADCNN-AGC model

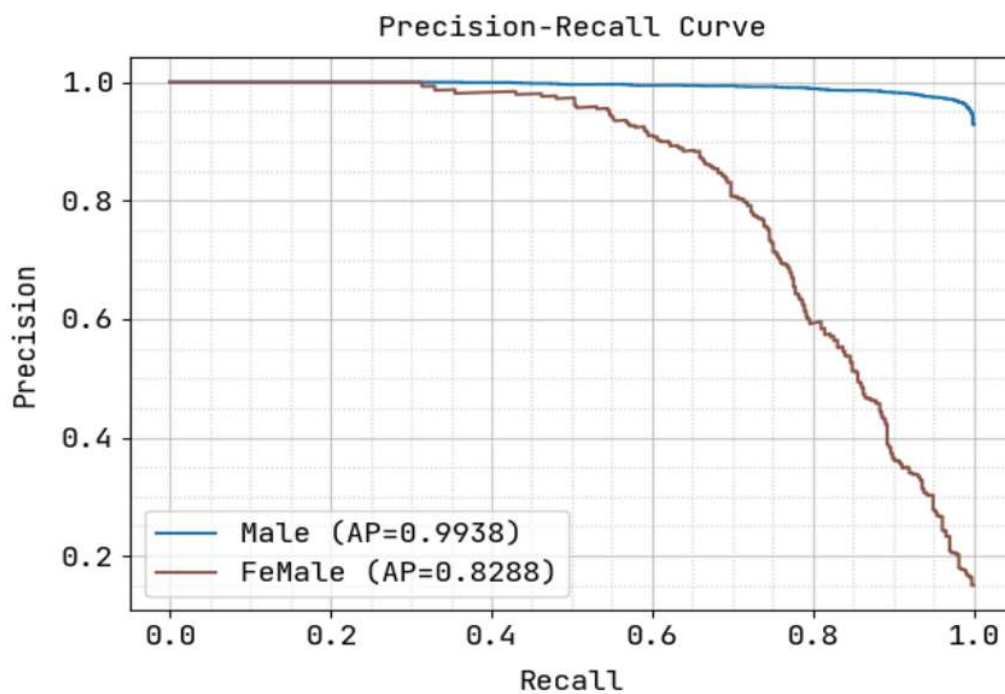
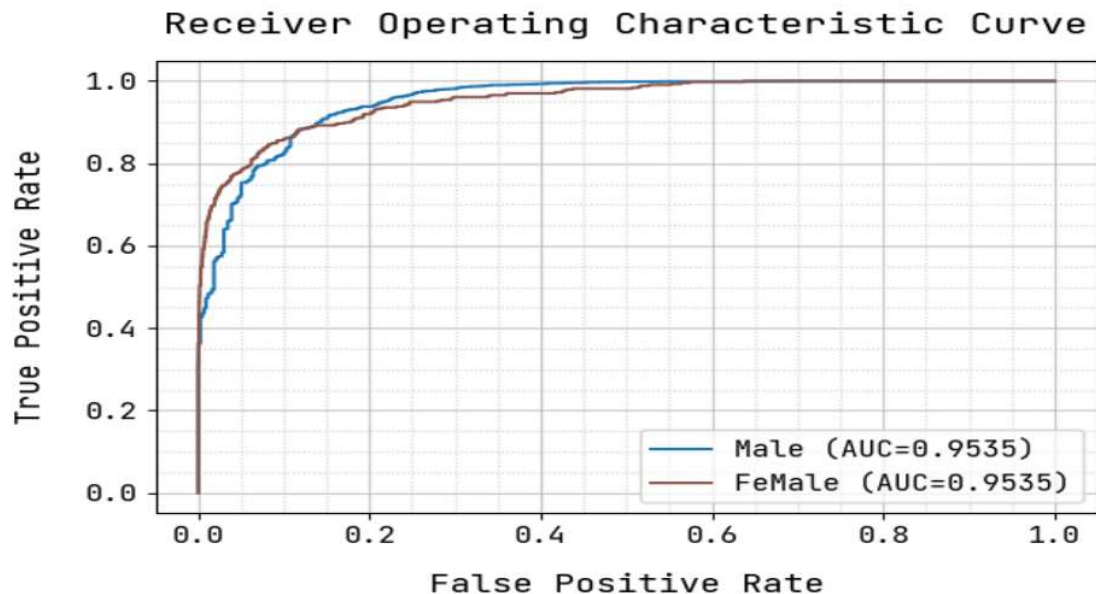


Fig. 8. Precision-recall graph analysis of ADCNN-AGC technique



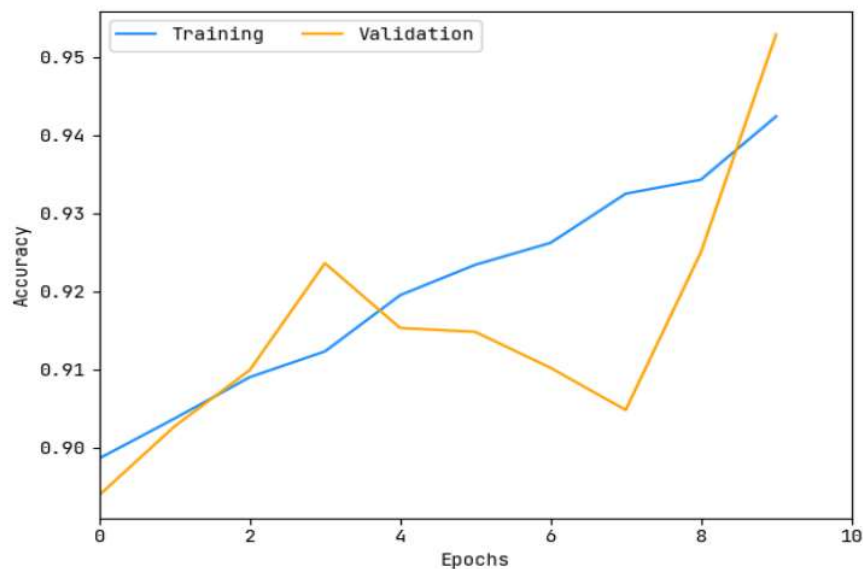
A brief precision-recall examination of the ADCNN-AGC model on test dataset is portrayed in Fig. 8. By observing the figure, it is noticed that the ADCNN-AGC model has accomplished maximum precision-recall performance under test dataset.



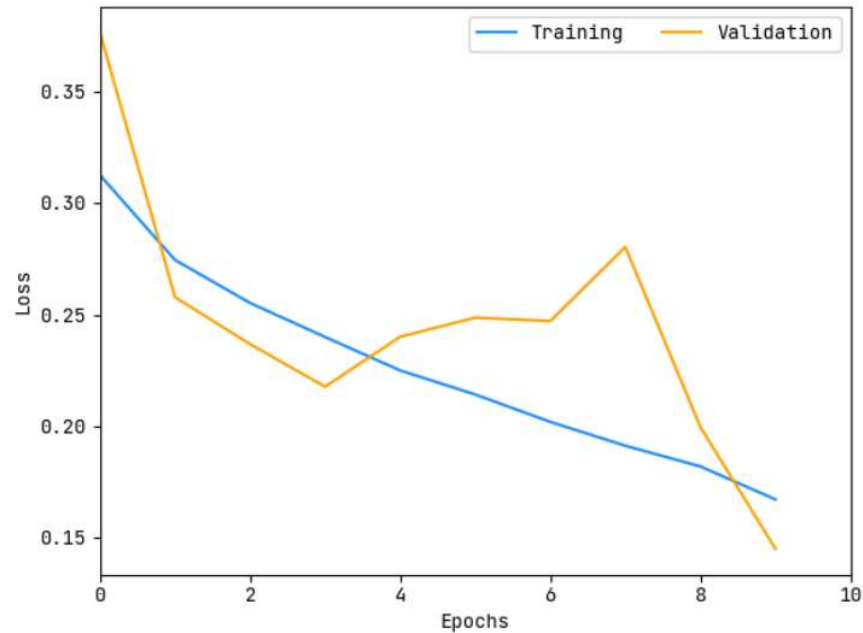
**Fig. 9.** ROC curve analysis of ADCNN-AGC technique

Fig. 9 portrays the ROC examination of the ADCNN-AGC approach on test dataset. The figures indicated that the ADCNN-AGC system has resulted in maximum ROC values on the categorization of male and female classes.

Fig. 10 illustrates the training and validation accuracy inspection of the ADCNN-AGC method on test dataset. The figure conveyed that the ADCNN-AGC system has offered maximum training/validation accuracy on classification process.



**Fig. 10.** Accuracy graph analysis of ADCNN-AGC technique



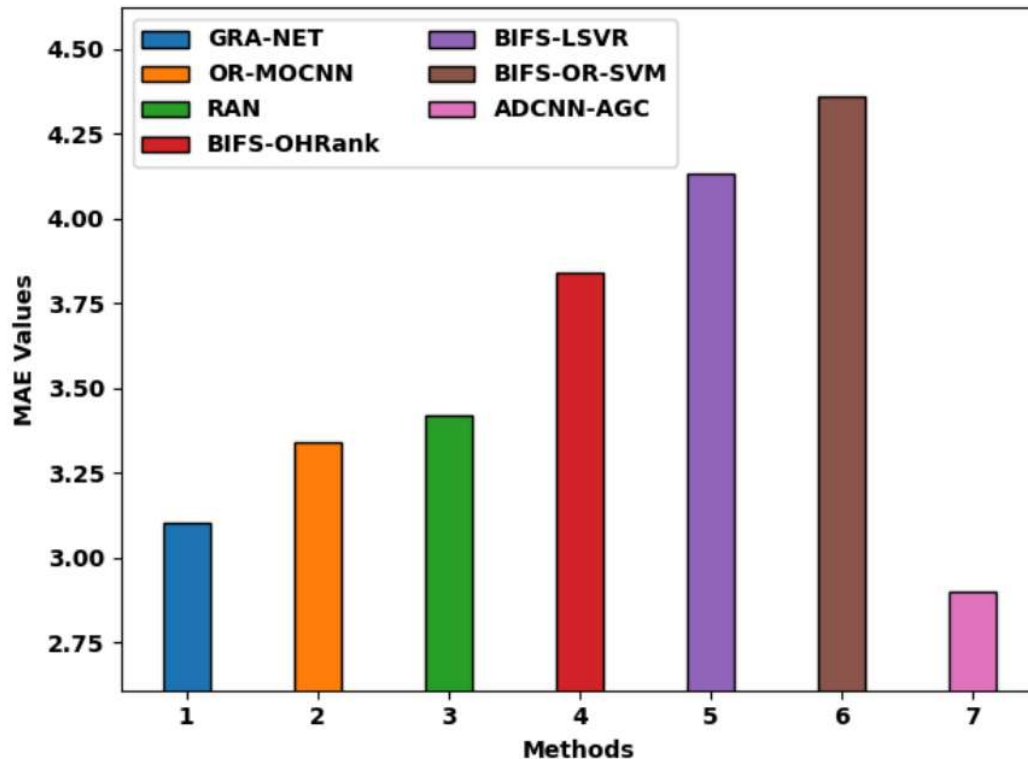
**Fig. 11.** Loss graph analysis of ADCNN-AGC technique

Next, Fig. 11 exemplifies the training and validation loss inspection of the ADCNN-AGC technique on test dataset. The figure reported that the ADCNN-AGC methodology has offered reduced training/accuracy loss on the classification process of test data.

For ensuring the enhanced outcomes of the ADCNN-AGC system on age prediction, a comparison study with recent models is made in Table 3 and Fig. 12. The experimental values indicated that the BIFS-LSVR and BIFS-OR-SVM systems have obtained ineffectual outcomes with MAE values of 4.1300 and 4.3600 respectively. Along with that, the BIFS-OHRank approach has gained somewhat higher performance with MAE of 3.8400. Followed by, the GRA-NET, OR-MOCNN, and RAN models have accomplished closer MAE of 3.1000, 3.3400, and 3.4200 respectively. However, the ADCNN-AGC model has accomplished effectual outcomes with MAE of 2.8983.

**Table 3** Comparative MAE analysis of ADCNN-AGC model

Methods	MAE
GRA-NET	3.1000
OR-MOCNN	3.3400
RAN	3.4200
BIFS-OHRank	3.8400
BIFS-LSVR	4.1300
BIFS-OR-SVM	4.3600
ADCNN-AGC	2.8983



**Fig. 12.** Comparative MAE analysis of ADCNN-AGC technique with recent methods

Finally, a detailed accuracy examination of the ADCNN-AGC model with recent models is made in Table 4 and Fig. 13 [22]. The experimental values indicated that the LBP and FPLBP techniques have gained poor performance with  $accu_y$  of 73.40% and 72.60% respectively. In line with, the GRA-NET, RAN, and FBLBP-PCA models have resulted in moderately closer  $accu_y$  values of 81.40%, 77.30%, and 76.10% respectively. Though the Facenet model has resulted in reasonable  $accu_y$  of 81.40%, the presented ADCNN-AGC model has accomplished superior results with maximum  $accu_y$  of 95.29%.

**Table 4** Accuracy analysis of ADCNN-AGC approach with recent methods

Methods	Accuracy (%)
Facenet	91.20
GRA-NET	81.40
RAN	77.20
LBP	73.40
FPLBP	72.60
FBLBP-PCA	76.10
ADCNN-AGC	95.29

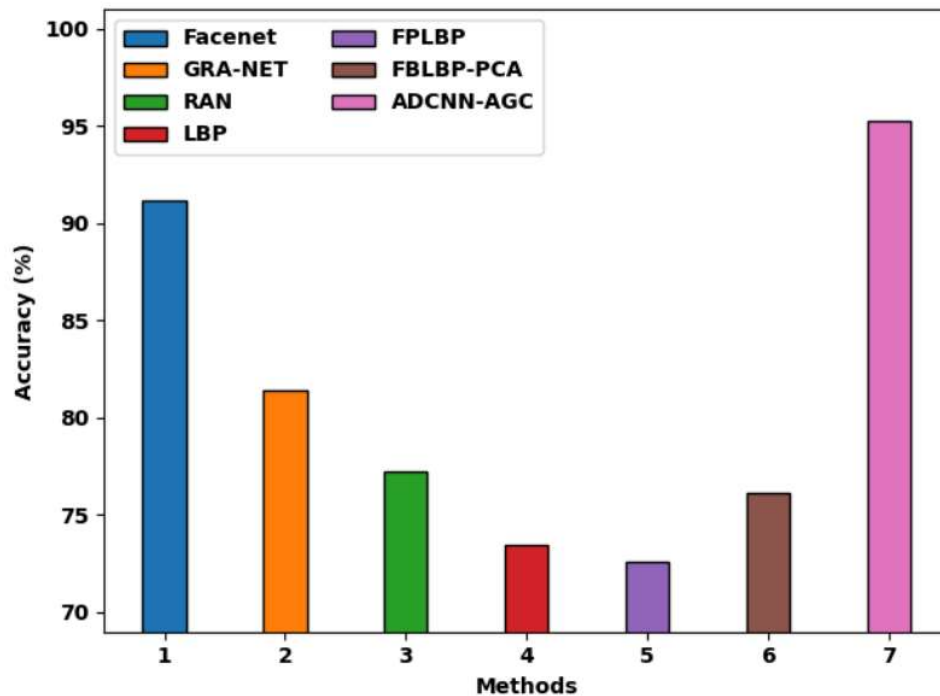


Fig. 13. Accuracy analysis of ADCNN-AGC technique with recent methods

After investigative the aforementioned tables and figures, it is apparent that the ADCNN-AGC model has accomplished superior performance with maximum age and gender classification outcomes.

## 5. Conclusion

In this article, a novel ADCNN-AGC model was developed for determining the age and gender of the persons who exist in the facial images. The ADCNN-AGC model utilized a MTCNN model for the detection of faces in the input images. Besides, Efficient Net model is applied for the proper extraction of feature vectors that are then passed into the 1-DCNN for classification procedure. The performance validation of the ADCNN-AGC model has been tested using benchmark datasets and the outcomes are studied in numerous aspects. The experimental outcomes reported the improved performance of the ADCNN-AGC model on state of art approaches. Thus, the ADCNN-AGC technique was establish that a sutiable tool for face recognition and classification. In future, hybrid DL models with metaheuristic hyperparameter optimizers can be employed for enhanced age and gender classification using facial images.

## References

- [1] Kosinski, M., 2021. Facial recognition technology can expose political orientation from naturalistic facial images. *Scientific reports*, 11(1), pp.1-7.
- [2] Karkkainen, K. and Joo, J., 2021. Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 1548-1558).
- [3] Agbo-Ajala, O. and Viriri, S., 2021. Deep learning approach for facial age classification: a survey of the state-of-the-art. *Artificial Intelligence Review*, 54(1), pp.179-213.
- [4] Bekhet, S. and Alahmer, H., 2021. A robust deep learning approach for glasses detection in non-standard facial images. *IET Biometrics*, 10(1), pp.74-86.
- [5] Savchenko, A.V., 2021, September. Facial expression and attributes recognition based on multi-task learning of lightweight neural networks. In *2021 IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY)* (pp. 119-124). IEEE.



- [6] Lin, Y.H., Tang, C.H., Chen, Z.T., Hsu, G.S.J., Shopon, M. and Gavrilova, M., 2021, September. Age-Style and Alignment Augmentation for Facial Age Estimation. In *International Conference on Computer Analysis of Images and Patterns* (pp. 297-307). Springer, Cham.
- [7] Greco, A., Saggese, A., Vento, M. and Vigilante, V., 2021. Effective training of convolutional neural networks for age estimation based on knowledge distillation. *Neural Computing and Applications*, pp.1-16.
- [8] Dammak, S., Mliki, H. and Fendri, E., 2021. Gender effect on age classification in an unconstrained environment. *Multimedia Tools and Applications*, 80(18), pp.28001-28014.
- [9] Al-Humaidan, N.A. and Prince, M., 2021. A classification of arab ethnicity based on face image using deep learning approach. *IEEE Access*, 9, pp.50755-50766.
- [10] Saxena, A., Singh, P. and Singh, S.N., 2021, January. Gender and Age detection using Deep Learning. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 719-724). IEEE.
- [11] Lu, T.T., Yeh, S.C., Wang, C.H. and Wei, M.R., 2021. Cost-effective real-time recognition for human emotion-age-gender using deep learning with normalized facial cropping preprocess. *Multimedia Tools and Applications*, 80(13), pp.19845-19866.
- [12] Rouhsedaghat, M., Wang, Y., Ge, X., Hu, S., You, S. and Kuo, C.C.J., 2021, January. Facehop: A light-weight low-resolution face gender classification method. In *International Conference on Pattern Recognition* (pp. 169-183). Springer, Cham.
- [13] Fayyaz, M., Yasmin, M., Sharif, M. and Raza, M., 2021. J-LDFR: joint low-level and deep neural network feature representations for pedestrian gender classification. *Neural Computing and Applications*, 33(1), pp.361-391.
- [14] Hsu, C.Y., Lin, L.E. and Lin, C.H., 2021. Age and gender recognition with random occluded data augmentation on facial images. *Multimedia Tools and Applications*, 80(8), pp.11631-11653.
- [15] Khattak, A., Asghar, M.Z., Ali, M. and Batool, U., 2021. An efficient deep learning technique for facial emotion recognition. *Multimedia Tools and Applications*, pp.1-35.
- [16] Ramya, R., Anandh, A., Muthulakshmi, K. and Venkatesh, S., 2022. Gender recognition from facial images using multichannel deep learning framework. In *Machine Learning for Biometrics* (pp. 105-128). Academic Press.
- [17] Dagher, I. and Barbara, D., 2021. Facial age estimation using pre-trained CNN and transfer learning. *Multimedia Tools and Applications*, 80(13), pp.20369-20380.
- [18] Ma, L.H., Fan, H.Y., Lu, Z.M. and Tian, D., 2020. Acceleration of multi-task cascaded convolutional networks. *IET Image Processing*, 14(11), pp.2435-2441.
- [19] Marques, G., Agarwal, D. and de la Torre Díez, I., 2020. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Applied soft computing*, 96, p.106691.
- [20] Abdoli, S., Cardinal, P. and Koerich, A.L., 2019. End-to-end environmental sound classification using a 1D convolutional neural network. *Expert Systems with Applications*, 136, pp.252-263.
- [21] <https://www.kaggle.com/jangedoo/utkface-new>
- [22] Garain, A., Ray, B., Singh, P.K., Ahmadian, A., Senu, N. and Sarkar, R., 2021. GRA\_net: A deep learning model for classification of age and gender from facial images. *IEEE Access*, 9, pp.85672-85689.