Age and Gender Prediction Using Deep Learning Algorithms on Human Face Images

Noor Kamal Al-Qazzaz[a,*], Amr aloula Ridha[a], Umniah Amjad[a]

[a]*Department of Biomedical Engineering, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad 47146, Iraq.* [*]Corresponding Author: noorbme@kecbu.uobaghdad.edu.iq

https://doi.org/10.31972/iceit2024.025

Abstract

Ageing is a difficult issue for facial recognition systems nowadays. Biological changes that occur with ageing provide a special problem since they can cause noticeable differences in face features between photos of the same individual taken at different ages. The extraction of robust facial characteristics for age-invariant face recognition is becoming increasingly important, especially when there are huge age differences between photographs of the same person's face, because the face is the most affected area of the body by ageing. This study aims to create a deep learning algorithm that accurately predicts facial image-based age with low error, firstly, to build a robust and generalizable deep learning model that can properly identify gender from facial photos, secondly, and to advance age-related research and gender prediction, which may be used for demographic analysis, targeted advertising, personalized marketing, age-specific suggestions, and customized user experiences. Convolutional Neural Networks (CNN) was investigated to illustrate the effectiveness of deep learning-based methods on the UTKFace dataset. Additionally, the MobileNetV2 model consistently has the best mean accuracy rate when feature extraction is performed using the MobileNetV2 model, suggesting that it could be the most promising option for age-invariant face recognition.

Keywords: Age, gender, deep learning, CNN, classification

1. Introduction

Research on age-related face recognition has recently been intense due to the importance of this factor in many practical face recognition system applications [1]. There might be a significant age gap between the query image and the ones in the database if that happens, and it might be hard to add the most current face images of the subject to the database [2].

In order to establish a person's identity, biometric identification and verification technologies rely on facial features. Age has the greatest impact on this trait, which in turn has a major impact on how well facial recognition systems work [3].

Aging poses limitations for multiple reasons: initially, there is no way to prevent the impacts of aging; erasing aging variation from facial image capture is simply not feasible. Ethnicity, lifestyle, environment, and other factors may all play a role in how aging impacts individuals. It has been demonstrated that biological factors such as gender, ancestry, genetics, and illnesses can hasten the aging process and exacerbate its affects on the face [4]. According to Lanitis et al. there are several environmental variables that can hasten the aging process. These include smoking, heavy alcohol use, being exposed to harsh weather, emotional stress, and experiencing significant weight fluctuations [4].

Therefore, it may be possible to avoid updating huge facial databases with more recent pictures and improve accuracy when identifying people by developing age-invariant face recognition algorithms [5].

Sawant et al. surveyed age-invariant face recognition systems, looked at how aging affects these systems' performance, and compiled a few facial databases [6].

Deep learning and CNN have made it possible to learn visual representations directly from pixels, making them the most used tool today. Methods for extracting features from faces using deep learning have shown great promise, especially when it comes to tracking changes in age [7, 8].

The convolutional neural network (CNN) architectures are the mainstay of deep learning-based feature extraction methods [9], which in turn produce more accurate and age-invariant face representations. Reusing pretrained deep-CNN models is an alternative that could be very useful since this method utilizes a huge number of labeled databases for training, which is usually impractical in real-world applications. By testing the discriminatory power and invariance of pre-trained deep-CNN models in the setting of face recognition throughout age progression, we aim to enhance our understanding of feature extraction through deep-learning methodologies in this study. This study takes into account five well-known deep-CNN models for face feature extraction: AlexNet [10], GoogleNet [11], Inception V3 [12], SqueezeNet [13] and ResNet50 [14]. This group of pretrained CNN-models was hand-picked because of their widespread use in academic studies.

In contrast, Mehdipour Ghazi et al. [15] demonstrated that a pose and illumination normalization step must precede the CNN-based feature extraction stage to improve performance under varying conditions and increase accuracy rates [16].

This study aims to develop a deep learning algorithm that accurately predicts the age and gender of individuals based on their facial images with high precision and minimize errors in age estimation. Moreover, to build a graphical user interface (GUI) system for the visual representation of age and gender prediction of individuals based on their facial images. Furthermore, to contribute to the advancements in age-related research and reliable gender prediction, which can have various applications, including demographic analysis, targeted advertising, personalized marketing, age-specific recommendations and customized user experiences.

In this paper, we delve into the topic of discriminative approaches, particularly regarding the feature extraction stage. This is a crucial part of face recognition as it enables the extraction of discriminant visual representations from face images, which in turn fully describe ageing face characteristics. Conventional methods rely on manually calculated features derived from statistical representations and low-level qualities in order to extract features and generate visual representations. Nowadays, nevertheless, the most popular tool is learning visual representations directly from pixels, thanks to advancements in deep learning and convolutional neural networks (CNN).

2. Materials and Methods

This section illustrates an overview of the main stages applied in this study. Starting with the facial image, preprocessing stage, CNNs and ending with the Classification stage (Figure 1.

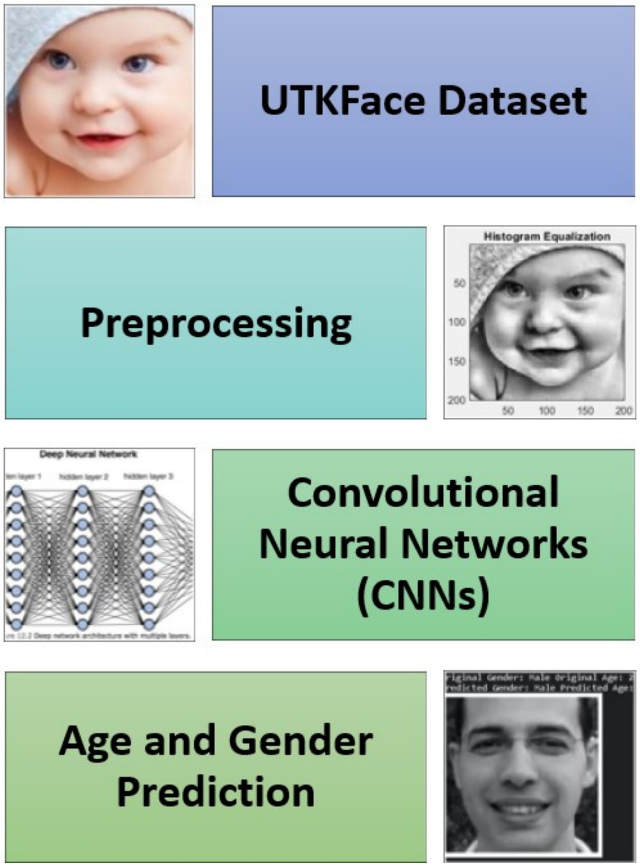

UTKFace Dataset
Preprocessing
Histogram Equalization
50
100
150
200
50
100
150
200
Deep Neural Network
hidden layer 2
hidden layer 3
Convolutional Neural Networks (CNNs)
Age and Gender Prediction
Original Gender: Male Original Age:
Predicted Gender: Male Predicted Age

Figure 1: Schematic diagram of this study on ASD detection through EEG signals.

2.1. UTKFace Dataset

UTKFace dataset is a large-scale face dataset with a long age span (ranging from 0 to 116 years old). The dataset consists of about 23,000 face images with annotations of age, gender, and ethnicity. The images cover large variations in pose, facial expression, illumination, occlusion, resolution, etc.

The UTKFace public database of age and gender images used in this work was downloaded from: https://www.kaggle.com/datasets/jangedoo/utkfacenew. The labels of each face image are embedded in the file name, formatted like this: [*age*] [*gender*] [*race*] _[*dateandtime*].*jpg* [age] is an integer from 0 to 116, indicating the age [gender] is either 0 (male) or 1 (female) [race] is an integer from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern). [date and time] is in the format of yyyymmddHHMMSSFFF, showing the date and time an image was collected to the UTKFace dataset. In this study, two folders were created in which the ages and the genders were separated from the dataset.

2.2. Preprocessing

Preprocessing face data is an important step in preparing the data for use in neural network models and the data will be preprocessed.

In order to classify the UTKFace images using CNN, all the images have been resized into 224 × 224 dimensions to meet the MobileNet specifications, firstly and 128 × 128 to meet the proposed 2D-CNN specifications, secondly.

2.3. Convolutional Neural Networks (CNNs)

Building a deep learning model for age and gender prediction involves designing the architecture of the neural network and defining the training procedure. The architecture of the neural network typically consists of multiple convolutional layers followed by fully connected layers. The convolutional layers extract features from the input images, while the fully connected layers classify the extracted features into age and gender categories.

2.3.1. MobileNetV2

MobileNetV2 is considered as a CNN network that has considerably low complexity and size owing to the use of depth-wise Separable Convolution, which makes it suitable to run on devices with low computational power. MobileNetV2 expanded the feature extraction and introduced an inverted residual structure.

The model architecture consists of a convolutional layer followed by a series of residual bottleneck layers. Kernel size for all spatial convolution operations is taken as ReLU6 is used as the non-linearity along with batch normalization and dropout during the training phase. Each bottleneck block consists of 3 layers, starting with a (1 × 1) convolutional layer followed by the aforementioned (3 × 3) depth-wise convolution layer and finally another (1 × 1) convolutional layer without ReLU6 activation. MobileNetV2 has wide applications at the current time due to its excellent feature extraction capabilities and small size [17].

MobilenetV2 is a pre-trained model for image classification. Pre-trained models are deep neural networks that are trained using a large image dataset [17]. Using the pre-trained models, the developers need not build or train the neural network from scratch, thereby saving time for development (Figure 2).

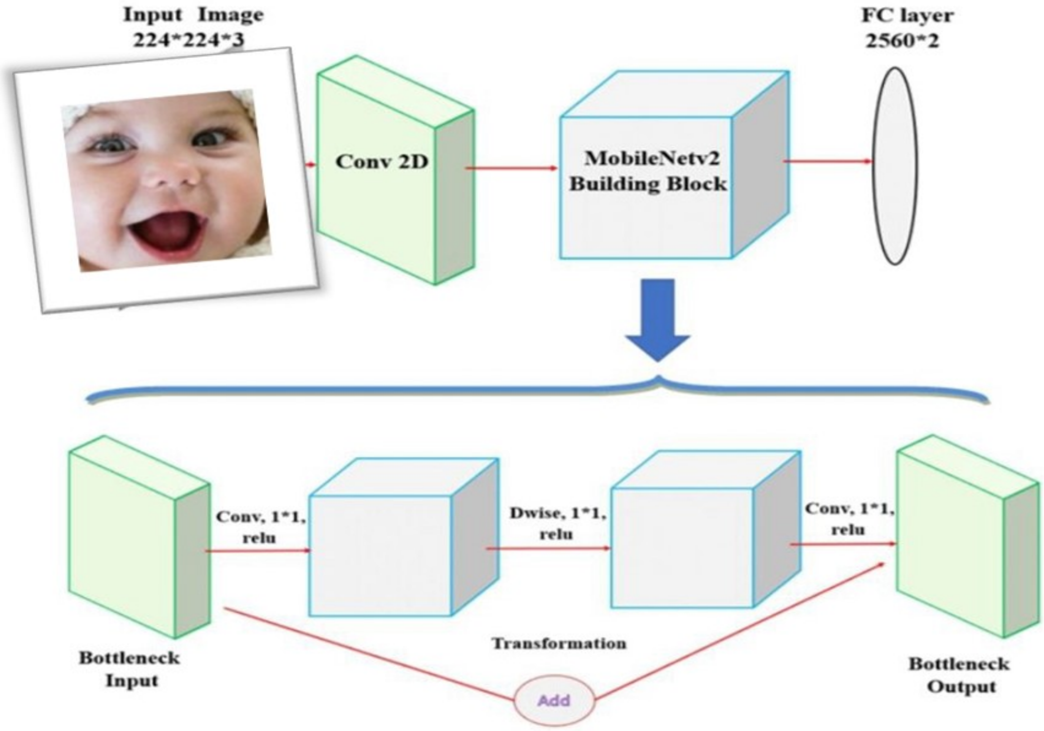


Figure 2: MobileNetV2 architecture.

MobileNetV2 expects images of 224 × 224 pixels with three colour channels. In other words, it expects an input of shape (224 × 224 × 3) [17].

The output of MobileNetV2 is a probability distribution over a set of pre-defined classes. The number of output classes depends on the specific task for which the network was trained. For example, if the network were trained to classify images into 1000 different categories, the output would be a vector of 1000 values, each representing the probability of the input image belonging to a specific category. The predicted class is usually the category with the highest probability value [17].

Finally, classification is a process of categorization based on the knowledge acquired in a dataset that contains observations for which the category is known. In this case, classification consisted of categorizing the UTKFace images using a mobileNetV2 with transfer learning by changing the layers of mobileNetV2 from 1000 to 2 layers depending on the acquired dataset.

2.3.2. Two-dimensional convolutional neural network (2D-CNN)

The proposed 2D-CNN model included necessary layers like the input layer, two 2D convolutional layers, two max-pooling layers, a flattened layer, a fully connected layer, two dense layers with ReLU activation function, two dropout layers to drop out some of the neurons to prevent overfitting and two output layers to display the general prediction results (Figure 3).

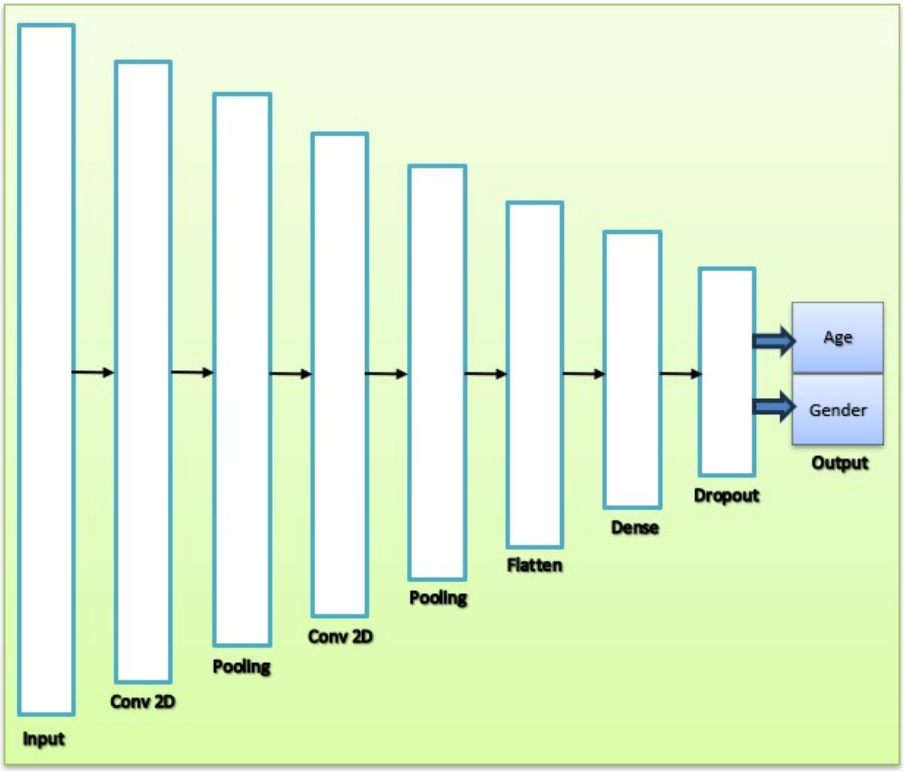


Figure 3: The proposed two-dimensional convolutional neural network (2D-CNN) deep learning architecture.

2.4. Age and Gender Prediction

Training a deep learning model for age and gender prediction can performed using classification and illustrated the obtained results visually using a graphical user interface (GUI) framework.

Age and Gender Prediction are computationally intensive tasks, especially when dealing with large datasets. It requires powerful hardware such as graphics processing units (GPUs) to accelerate the training process.

2.4.1. Networks Classifications

The classification procedure involves feeding the labeled images into the network, computing the prediction error, and updating the network's weights using an optimization algorithm such as stochastic gradient descent. This process is repeated for multiple epochs until the model achieves satisfactory performance on the training dataset. For the training and classification test procedures of this 2D-CNN, training and test sets with 80% and 20% of data have been used, respectively.

To improve the MobileNetV2 model's performance, techniques such as transfer learning and fine-tuning can be applied. Transfer learning involves using a pre-trained model as a starting point and fine-tuning it on the specific age and gender prediction task. This approach can save significant training time and improve the model's accuracy.

During the fine-tuning process, the weights of the pre-trained model are adjusted to better fit the target task. This is done by training the model on the labeled face images specific to age and gender prediction. Fine-tuning helps the model adapt to the specific characteristics of the age and gender prediction task and improves its overall performance.

The validation curve helps evaluate the model's performance on unseen data during training. It measures the validation accuracy and loss, computed on a separate dataset called the validation set. The validation set contains examples the model has not seen during training, allowing you to estimate its ability to generalize to new, unseen data. The validation accuracy represents the model's performance on the validation set, while the validation loss reflects the discrepancy between predicted and actual values for the validation data. Monitoring the validation curve is crucial to identifying overfitting or under-fitting The proposed network is trained with 8 epochs using mobileNetV2 and 9840 iterations with 1185 iterations per epoch, the frequency of validation is 3 iterations. Whereas for the proposed 2D-CNN, the network is trained with 50 epochs and 32 batch sizes. The performance of the proposed system was evaluated using average classification accuracy.

2.4.2. Graphical User Interface Framework

In this study, a graphical user interface (GUI) framework was developed to illustrate our suggested methods visually. GUI system was built for discriminating persons from the visual representation of age and gender with the prediction of individuals based on their face photos.

3. Results and Discussions

In recent years, deep learning frameworks have shown tremendous potential in medical image analysis, particularly in detecting age and gender from face images.

The initial step involved creating two folders: the age folder, which included labels, and the gender folder, which included labels from each image caption.

Additionally, the age label is at the zero index, as the first component is the age and the second component is the gender, and both were converted into integers because both of these are expressed in numbers. There are only two genders: male and female. For males, the number is zero, and for females, it is one. Ultimately, the process involves sequentially appending the age, corresponding gender, and path of each image. After that, a simple dictionary was created just for our convenience, where zero means male and one means female. Next, we converted each of the created lists into a separate column. This was done to ensure the creation of a clear and concise data frame. For each image path, we have a corresponding age and gender.

3.1. Results of Preprocessing

Preprocessing also involves the resizing of all the input images into the size of resizing that image into (224 × 224) and (128 × 128), to be used by the MobileNetV2 and the proposed 2D-CNN, respectively.

For the MobileNetV2, transfer learning was performed by sharing weights or knowledge extracted from one problem domain to solve other related problems. The CNN MobileNetV2 pre-trained model is assigned with their weights and then output layers are removed to customize the output layer. A softmax layer with three neurons is employed to produce the final classification. The pre-trained model obtained from the above process needs training with face image data. Thus, the models are trained with face image datasets for the detection of age and gender.

3.2. Results of Convolutional Neural Networks Classifications

Once the preprocessing of the data analysis is complete, the CNNs have performed feature extractions. For that, a custom function for extracting image features has been created, in which the list of image parts has been passed. So this image is actually a list that would contain image parts. Then an empty list called features is created, traversing through all the image parts, and loading that image as a number array. The process involved converting the images into a number array and then appending them to the features list.

After that, all the feature lists were converted into a number array and then reshaped into the number of images, the dimension of each image, and the channel, therefore, the number of the extracted features was 9882. The next step involves normalizing all the features, which are arrays, for each image. Naturally, the task at hand involves predicting the gender and estimating the age based on an image provided as input.

Accuracy or performance metrics are crucial measures for evaluating the effectiveness and reliability of a trained neural network, Performance metrics provide quantitative measures that quantify how well a neural network performs on specific tasks such as classification and object detection. These metrics serve as benchmarks for evaluating the quality of predictions made by the network and help in comparing different models or algorithms. One of the most fundamental performance metrics is accuracy, which measures the proportion of correct predictions made by the neural network over a given dataset.

3.2.1. Results of MobileNetV2

This diverse dataset enables the deep learning model to learn and differentiate between different ages and genders cases. The MobileNetV2 architecture is a well-known and widely used deep learning model, specifically designed for mobile and resource-constrained environments. Its efficient design allows for accurate and fast inference on mobile devices, making it an ideal choice for the development of a mobile application aimed at age and gender detection from face images. Although the quantitative results are quite promising and show the effectiveness of the proposed approach still, the qualitative results must be investigated properly and calculated using the average classification accuracy.

A notable deep learning framework, MobileNetV2, has been employed to develop a system with an average classification accuracy of 91.14% for gender detection (Figure 4). The achieved average classification accuracy is an impressive result, indicating the effectiveness of the CNN MobileNetV2 framework in distinguishing between males and females. However, MobileNetV2 was unable to predict the ages with high accuracy.

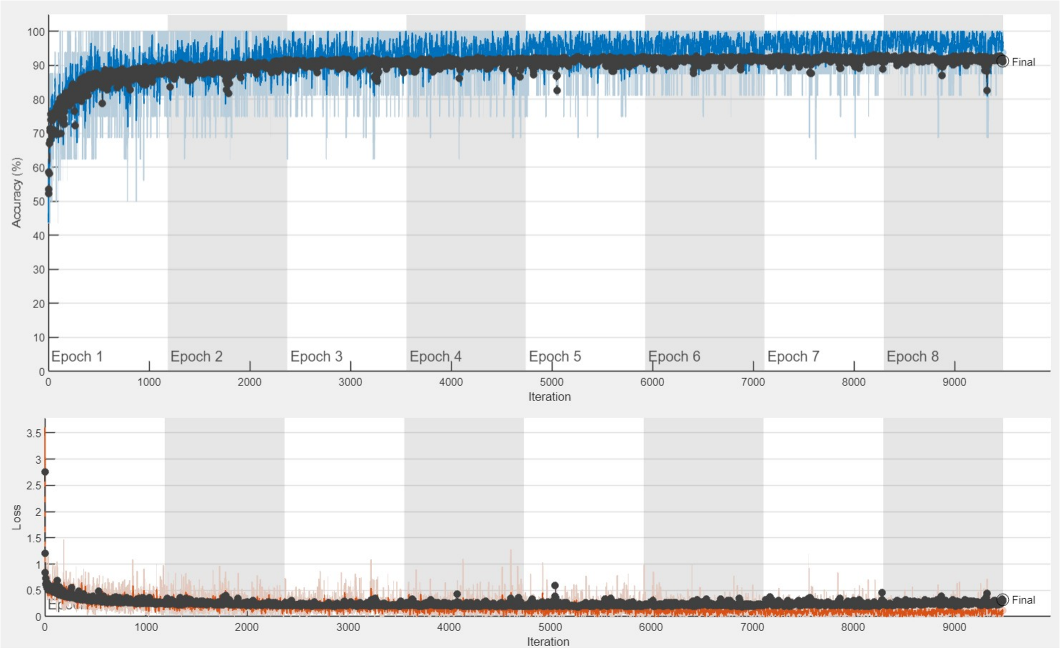


Figure 4: The Process of training the MobileNetV2.

This accuracy demonstrates the potential of deep learning during the training phase of a neural network, the model learns from a labeled dataset by adjusting its parameters through an iterative optimization process. Moreover, Figure 4 illustrates the training curves of the proposed 2D-CNN and the model's progress as it improves its performance over successive iterations or epochs to predict both age and gender. Typically, training curves depict the training loss, which measures the discrepancy between the predicted outputs of the network and the true labels in the training data. The goal is to minimize this loss, indicating that the network is learning and adjusting its weights to make more accurate predictions.

3.2.2. Results of 2D-CNN

It can be noted that there is a significant increase in loss values at the beginning of the training, which decreases substantially in the later stage of the training. The main reason for this sharp increase and decrease is attributed to the number of data in the age and gender class, these rapid ups and downs are slowly reduced in the later part of the training. The multi-class classification performance of the model has been evaluated and the average classification of 77% performance of the model is obtained (Figure 5).

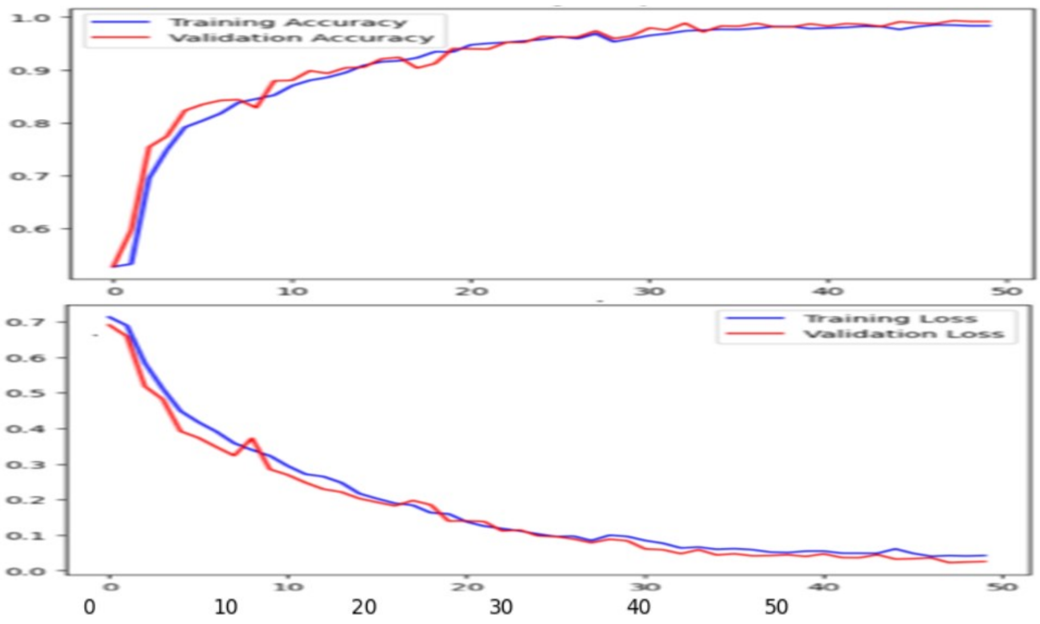
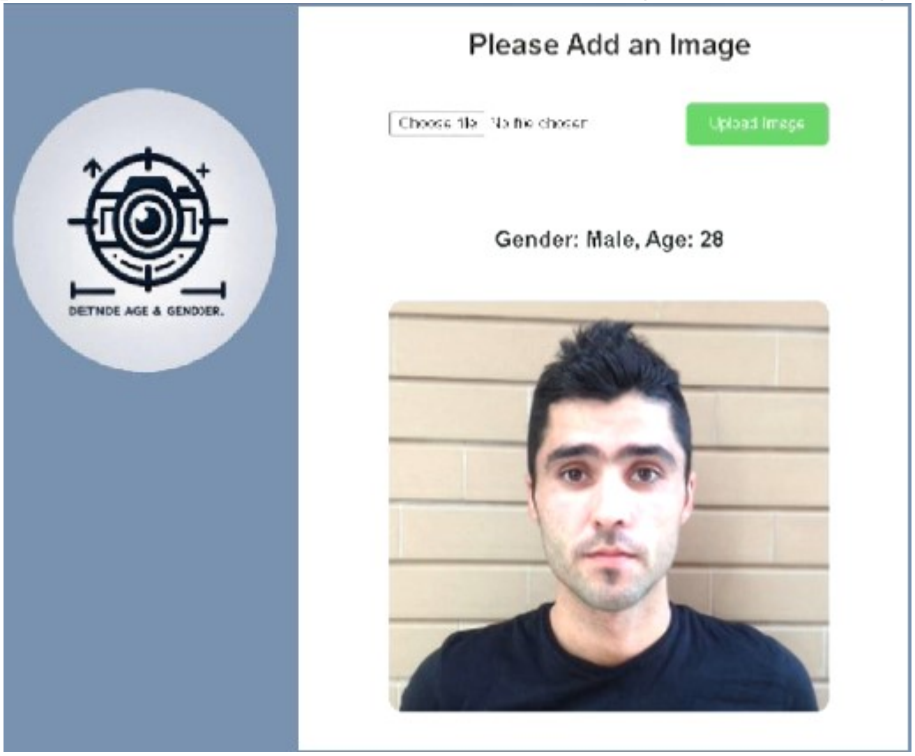


Figure 5: The Process of training the MobileNetV2.

3.3. Results using Graphical User Interface Framwork

The GUI systems were sufficiently and successfully developed for visual representation of age and gender prediction of individuals based on their facial images as shown in Figure 6.


Please Add an Image

Gender: Male, Age: 28

Figure 6: GUI system for age and gender prediction of individuals based on their face images.

4. Discussion

In this chapter of our project on age and gender detection from face images using CNN deep learning. we focused on acquiring and preprocessing the dataset. The dataset consists of male, female and age cases. The dataset is crucial in training our deep learning model to classify face images accurately. To begin, the UTKFace images were downloaded from Kaggle, the largest data science community with powerful tools and resources to achieve data https://www.kaggle.com/datasets/jangedoo/utkface-new.

After acquiring the dataset, we performed preprocessing steps to ensure compatibility with our chosen deep learning model, CNN MobileNetV2. In this chapter of our project on age and gender detection from UTKFace dataset using MoileNetV2 and 2D-CNN deep learning frameworks, we focused on preprocessing the dataset to group them into age folders and gender folders that were labeled with the respective classes. Then we resizing the images to ensure consistency in the dimensions of the images to meet the MoileNetV2 and 2D-CNN networks.

After preprocessing the dataset, we proceeded to train our deep learning models. Training and evaluating several deep learning models, we found that MobileNetV2 exhibited the highest accuracy of 91.14% gives us confidence in its ability to accurately classify gender into two classes: males and females. This accuracy is a crucial aspect of gender classification however it was unable to classify ages as ages includes, therefore, we proposed 2D-CNN which detects both age and gender with 77% classification accuracy.

References

[1] H. El Khiyari and H. Wechsler, "Face recognition across time lapse using convolutional neural networks," *Journal of Information Security*, vol. 7, no. 03, p. 141, 2016.

[2] A. Jain and S. Li, *Handbook of Face Recognition*. Springer, 2011, vol. 1.

[3] H. El Khiyari and H. Wechsler, "Age invariant face recognition using convolutional neural networks and set distances," *Journal of Information Security*, vol. 8, no. 03, p. 174, 2017.

[4] A. Lanitis, "Facial biometric templates and aging: Problems and challenges for artificial intelligence," in *AIAI Workshops*, 2009, pp. 142–149.

[5] N. Ramanathan, R. Chellappa, and S. Biswas, "Age progression in human faces: A survey," *Journal of Visual Languages and Computing*, vol. 15, pp. 3349–3361, 2009.

[6] M. Sawant and K. Bhurchandi, "Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging," *Artificial Intelligence Review*, pp. 1–28, 2018.

[7] Y. Wang, D. Gong, Z. Zhou, X. Ji, H. Wang, Z. Li, W. Liu, and T. Zhang, "Orthogonal deep features decomposition for age-invariant face recognition," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 738–753.

[8] M. Sajid, T. Shafique, S. Manzoor, F. Iqbal, H. Talal, U. Samad Qureshi, and I. Riaz, "Demographic-assisted age-invariant face recognition and retrieval," *Symmetry*, vol. 10, no. 5, p. 148, 2018.

[9] Y. Li, G. Wang, L. Lin, and H. Chang, "A deep joint learning approach for age invariant face verification," in *CCF Chinese Conference on Computer Vision*. Springer, 2015, pp. 296–305.

[10] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[13] F. Iandola, S. Han, M. Moskewicz, K. Ashraf, W. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[15] M. Mehdipour Ghazi and H. Kemal Ekenel, "A comprehensive analysis of deep learning based representation for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 34–41.

[16] K. Grm, V. Struc, A. Artiges, M. Caron, and H. Ekenel, "Strengths andˇ weaknesses of deep learning models for face recognition against image degradations," *Iet Biometrics*, vol. 7, no. 1, pp. 81–89, 2017.

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.