Smart Techno (Smart Technology, Informatic, and Technopreneurship)
Volume 08, Number 01, 2026 pp. 74-82
E-ISSN : 2541-0679
Open Access: https://lppm.primakara.ac.id/jurnal/index.php/smart-techno/

Implementation of DeepFace for Gender Prediction Based on Facial Images

Aditya Wijaya*, Sadam Dwi Langit[2], Abdurrozzaq Musyaffa[3]
[1,2,3] Informatics, Faculty of Engineering and Informatics, Bina Sarana Informatika University, Jakarta, Indonesia

ARTICLE INFO
Article history:
Initial submission
03-12-2025
Received in revised form
07-01-2026
Accepted 13-01-2026
Available online 27-02-2026

Keywords:
DeepFace, gender classification, facial images, UTKFace, deep learning, pretrained model

DOI:
https://doi.org/10.59356/smart-techno.v7i1

ABSTRACT

This study evaluates the performance of a pretrained DeepFace model for gender classification based on facial images using the UTKFace dataset. A total of 100 facial images were employed as test data, consisting of 50 male and 50 female samples selected through controlled random sampling to maintain class balance. Image preprocessing was conducted automatically using the DeepFace.analyze() function, which includes face detection, alignment, size normalization, and facial cropping. The study did not involve model retraining and relied solely on the inference capability of the pretrained DeepFace model. The experimental results show that the model correctly classified 45 male and 44 female images, achieving accuracies of 90% and 88% for the male and female classes, respectively, with an overall accuracy of 89%. Confusion matrix analysis indicates that misclassifications were primarily influenced by image quality factors such as lighting variations, camera angles, and facial expressions. Overall, the findings demonstrate that DeepFace is effective for gender classification without retraining; however, further improvements in preprocessing techniques and dataset diversity may enhance classification performance in future research.

1. INTRODUCTION

Estimating a person's age solely from facial appearance is a challenging task due to the unique and complex characteristics of facial images. Age-related changes in human facial features such as wrinkles, skin texture, and structural transformations vary significantly among individuals. The human face is a critical source of visual information that can be used to determine age, gender, race, and other personal attributes. Human development is commonly classified into five age categories: newborns, children, adolescents, adults, and the elderly. As individuals age, fundamental facial changes occur, including increased wrinkling, alterations in cheekbone structure, and shifts in the relative distances between the mouth, nose, and eyes, which are among the most significant facial characteristics (Informatika et al., 2024).

In recent years, access to adult content has increased substantially. According to a report by Jawapos.com, Indonesia ranked second globally in terms of access to pornographic content, after India, during 2015 and 2016. Approximately 74% of users were young individuals, while the remaining proportion consisted of older generations (Melangi, 2020). This phenomenon highlights the growing importance of reliable age estimation systems to support content regulation and digital safety.

The human face represents a multifaceted visual expression of identity and emotion. From facial appearance, it is possible to infer an individual's gender and estimate their age (Opencv et al., 2023). Age is defined as a measure of time that quantifies the existence of an object or organism, whether living or non-living, according to the Indonesian Ministry of Health. Facial recognition studies demonstrate a strong relationship between age and identity recognition. Facial recognition refers to the capability of computers or other devices to identify human faces from digital images (Nabila, 2024).

* *Corresponding Author: A Wijaya: adityawjaya1710@gmail.com*

In Latin, the term imago refers to an image or illustration that represents, resembles, or imitates an entity or object. Images can be categorized into visible and invisible forms. Visible images encountered in daily life include paintings, sketches, and photographs, while invisible images include mathematical function representations and image data stored digitally in files (Ilmiah & Pendidikan, 2024). Image processing techniques for object detection have been widely developed in previous studies. These studies indicate that color features within images can be utilized as references for object identification using cameras and other image acquisition technologies (Andrekha & Huda, 2021).

Deep learning has been extensively applied across various domains in recent years, including social media applications such as face filters, immigration systems, and facial recognition features on smartphones (Sachi et al., 2023). One prominent deep learning algorithm is the Convolutional Neural Network (CNN), which consists of three-dimensional neuron layers characterized by width, height, and depth. The width and height represent the spatial dimensions of neuron groups, while depth indicates the number of feature maps within a layer. Convolution serves as the core operation in CNNs, utilizing localized regions of an image to compute convolution operations between the input and filters in order to extract specific features. CNNs perform these operations through multiple interconnected neural layers (Algoritme et al., 2023).

The objective of this study is to develop a method for determining an individual's age and gender based on facial photographs. Gender recognition research is particularly important because it can enhance the performance of face detection and face verification systems. Security systems can leverage facial recognition software for gender identification purposes (Irawan & Humaini, 2016). This approach mimics the facial recognition mechanisms of the human visual cortex by employing CNN-based architectures (Zein, 2020; Arifandi, 2022).

2. METHOD

The research methodology refers to a systematic set of procedures employed by researchers to collect and analyze data in order to answer research questions or evaluate research hypotheses. As illustrated in Figure 1, this study follows a structured workflow consisting of several sequential phases, beginning with dataset collection and continuing through preprocessing, modeling, evaluation, and testing stages (Fajar & Huda, 2025).

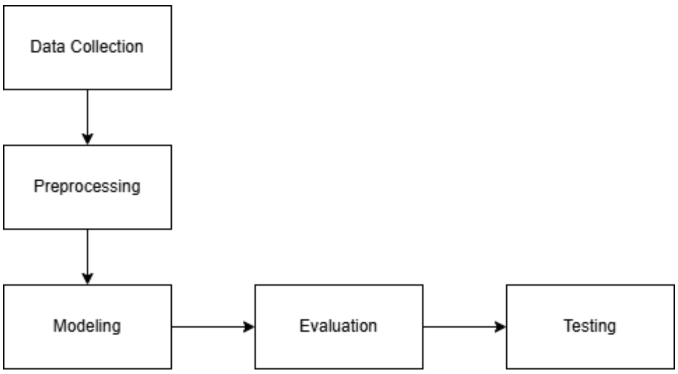

Figure 1. Research Methodology

1. Dataset Collection

   In this study, the UTKFace dataset obtained from the Kaggle repository was used. UTKFace is a large-scale facial image dataset containing more than 20,000 face images annotated with age, gender, and race information. This dataset has been widely used in facial recognition research due to its high data diversity, including a wide age range, variations in facial expressions, and diverse image backgrounds. Such diversity makes UTKFace an appropriate data source for objectively evaluating the performance of the DeepFace model in gender prediction based on facial images.

In this research, the UTKFace dataset was used exclusively as testing data to evaluate the model's performance. From a total of 23,708 images in the UTKFace dataset, 100 images were selected as test samples, consisting of 50 male and 50 female facial images. Accordingly, the data composition in this study can be summarized as follows. The results of the testing process were then analyzed to measure the prediction accuracy of the model in classifying gender based on facial images from the UTKFace dataset.

Table 1. UTKFace dataset

Data Type	Number of Images	Description
Training Data	-	Not used (pretrained model)
Validation Data	-	Not used
Testing Data	100	UTKFace dataset (random sampling)

Because the UTKFace dataset contains a very large number of images, this study does not process the entire dataset. Instead, a sampling procedure is applied using Simple Random Sampling (SRS) without replacement. This technique is selected because it provides an equal probability for each unit in the population to be chosen as a sample and is commonly employed in large-population studies. In addition, SRS is widely recommended in modern statistical research due to its unbiased nature and ease of implementation.

The determination of the minimum sample size is conducted using Slovin's formula, which is frequently applied in studies involving large populations with unknown variance. Slovin's formula is expressed as follows:

$$N = \frac{N}{1+N \cdot e^2} \quad (1)$$

where:

- nnn = sample size
- NNN = population size (23,708 UTKFace images)
- eee = margin of error

In this study, a margin of error of 10% (0.10) is applied, which is considered appropriate for exploratory research aimed at obtaining a general overview of the classification model's performance without requiring high precision.

- Substitution of values:

$$N = \frac{23708}{1 + 23708 \cdot (0.10^2)}$$

$$N = \frac{23708}{1 + 23708 \cdot (0.01)}$$

$$N = \frac{23708}{1 + 237.08}$$

$$N = \frac{23708}{238.08} = 96.6$$

The calculation results indicate that the minimum required sample size is 100 image samples. Accordingly, this study utilizes 100 images from the UTKFace dataset as testing data for evaluating the performance of the DeepFace model.

2. Preprocessing

The preprocessing stage in this study was conducted automatically within a Streamlit-based application utilizing the DeepFace.analyze() function. Each input image, whether uploaded as a file or captured via a camera, was processed prior to gender prediction analysis.

The process begins by reading the image in array format using the PIL module and converting it into a NumPy array. The image is then temporarily stored using the tempfile and cv2 modules. This step is necessary because the DeepFace.analyze() function requires a file path as input, rather than a direct array. Subsequently, DeepFace performs internal preprocessing, including face detection, face cropping, image size normalization, and alignment adjustment to match the model's input format.

All preprocessing steps are executed without manual intervention, as the parameter enforce_detection=False allows DeepFace to proceed even if face detection is not perfect. Therefore, all images tested in this study underwent standard preprocessing according to the DeepFace pipeline, ensuring consistency and uniformity of the data before being analyzed for gender classification.

3. Modeling

This study does not perform model training from scratch, as the DeepFace.analyze() function provides a pretrained model that can be used directly. The built-in model has been trained on large-scale datasets encompassing diverse conditions, including variations in pose, lighting, facial expressions, and demographic attributes. In this study, the model is employed to perform inference on the UTKFace dataset without any re-training. This approach allows the researchers to evaluate the generalization capability of the DeepFace model on new images that were not part of its original training data. The model extracts facial features from the input images and predicts gender based on the recognized feature patterns.

4. Evaluation

The primary evaluation was conducted using a Confusion Matrix, which is a table that compares the model's predictions with the true labels. The Confusion Matrix helps to understand the types of errors that occur, including false positives and false negatives. The key components of the Confusion Matrix are:

- True Positive (TP): The number of images correctly classified as the target class. For example, male images correctly predicted as male.
- False Positive (FP): The number of images incorrectly classified as the target class when they actually belong to a different class. For example, female images incorrectly predicted as male.
- False Negative (FN): The number of images that fail to be recognized as the target class. For example, male images incorrectly classified as female.
- True Negative (TN): The number of images correctly classified as not belonging to the target class. For example, female images correctly classified as not male in the context of evaluating the male class.

Accuracy is calculated using the following formula:

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$ (2)

Male Accuracy $= \frac{TP}{TP+FN} \times 100\%$(3)

Female Accuracy $= \frac{TN}{TN+FP} \times 100\%$(4)

In addition to accuracy, this study includes other evaluation metrics to provide a more comprehensive assessment of the model's performance, namely precision, recall, and F1-score. These metrics were calculated based on the Confusion Matrix obtained from testing on 100 facial images.

Based on the classification results, the evaluation metric values are presented in the following table.

Table 2. Evaluation Metric Values

Class	Precision	Recall	F1-Score

Male	0.88	0.90	0.89
Female	0.90	0.88	0.89
Average (Macro Avg)	0.89	0.89	0.89

The precision for the male class of 0.88 indicates that the majority of the male predictions made by the model are correct, although there are still some misclassifications of female images as male. A recall of 0.90 demonstrates the model's strong capability in correctly identifying male images overall. For the female class, a precision of 0.90 reflects a high accuracy in predictions, while a recall of 0.88 indicates that a small portion of female images was still misclassified as male.

5. Testing

The testing phase was conducted by submitting each of the 100 facial images to the DeepFace.analyze() function.

Figure 2. Testing Results

Each image was processed independently using the DeepFace.analyze() function with default parameters, including enforce_detection=False to allow predictions even if face detection is not perfect, detector_backend='opencv' for fast face detection, actions=['gender'] to restrict the analysis to gender prediction, and align=True to standardize the positions of the eyes, nose, and mouth according to the model's input requirements. The model outputs the predicted gender (Dominant Gender) along with confidence scores (Prob. Male % and Prob. Female %) for each class. Each image was analyzed independently, and the model produced outputs consisting of the gender prediction and associated confidence values. The testing process was performed consistently with default parameters across all images to ensure objective measurement. The predicted results were then compared with the ground-truth labels to determine whether the model produced correct or incorrect predictions. This evaluation was conducted entirely without human intervention, ensuring that all outcomes reflect the pure inference capability of the pretrained DeepFace model.

3. RESULT AND DISCUSSION

The testing was conducted on 100 images from the UTKFace dataset, consisting of 50 male and 50 female images. Each image was analyzed using the DeepFace.analyze() function without retraining, as DeepFace utilizes a pretrained deep learning model. The results showed that the model correctly recognized 45 male images while misclassifying 5 male images as female. For the female image group, the system accurately classified 44 images and produced errors on 6 images. Based on per-class accuracy, the model achieved **90%** accuracy for male images and **88%** accuracy for female images.

Table 3. Gender Classification Testing Results on 100 Facial Images

No	Nama File	Gender Dominan	Prob. Laki-laki (%)	Prob. Perempuan (%)
0	25_1_0_20170103180311751.jpg.chip.jpg	Woman	0.03	99.97
1	25_1_0_20170109214731480.jpg.chip.jpg	Woman	2.93	97.07
2	25_1_0_20170103180606727.jpg.chip.jpg	Woman	0.00	100.00
3	25_1_0_20170105000803577.jpg.chip.jpg	Woman	0.00	100.00
4	25_1_0_20170105183634552.jpg.chip.jpg	Woman	0.13	99.87
...	...	...	...	...
98	25_0_0_20170113210319573.jpg.chip.jpg	Man	98.00	2.00
99	25_0_0_20170116205006744.jpg.chip.jpg	Man	99.81	0.19

This table presents the results of gender classification on 100 facial images from the UTKFace dataset using the DeepFace model. Each row corresponds to a single image and its prediction outcome. The columns are described as follows:

1. No: Sequential number of the image in the dataset (0-99), used for identification and reference.
2. File Name: The name of the image file as stored in the UTKFace dataset. File names typically include age, gender, and date information, but here they serve as unique identifiers for each tested image.
3. Dominant Gender: The predicted gender by the DeepFace model for each image. The label indicates the class with the highest predicted probability: “Man” for male and “Woman” for female.
4. Prob. Male (%): The probability (in percentage) assigned by the model that the image belongs to the male class. Values closer to 100% indicate higher model confidence in predicting male.
5. Prob. Female (%): The probability (in percentage) assigned by the model that the image belongs to the female class. Values closer to 100% indicate higher model confidence in predicting female.

This table allows researchers to examine the model’s performance on an image-by-image basis, identify misclassifications, and analyze cases where predicted probabilities are near the decision threshold. By reviewing the table, overall accuracy, error patterns, and the distribution of model confidence across classes can be assessed.

Table 4. Summary Table of Gender Classification Testing Results on the UTKFace Dataset

Kategori	Jumlah	Persentase
Laki-laki terdeteksi benar	45	90%
Laki-laki salah deteksi	5	10%
Perempuan terdeteksi benar	44	88%
Perempuan salah deteksi	6	12%
Akurasi keseluruhan	**89/100**	**89%**

This table summarizes the results of gender classification on the UTKFace dataset, showing the number of images correctly and incorrectly classified by the DeepFace model.

1. Correctly Detected Male (45 images / 90%): Indicates that out of 50 male images tested, the model correctly classified 45 images as male.
2. Incorrectly Detected Male (5 images / 10%): Indicates that 5 male images were misclassified as female by the model.
3. Correctly Detected Female (44 images / 88%): Indicates that out of 50 female images, the model correctly classified 44 images as female.
4. Incorrectly Detected Female (6 images / 12%): Indicates that 6 female images were misclassified as male by the model.
5. Overall Accuracy (89/100 = 89%): Represents the model's success rate in classifying all 100 images. This value is calculated from the total number of correct predictions (45 males + 44 females = 89) divided by the total number of images (100).

This table provides a clear overview of the model's performance, showing which class is more easily recognized and the extent of prediction errors. It also aids in evaluating the model's effectiveness and identifying potential areas for improvement. The overall system accuracy is calculated using the formula:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

Based on the testing dataset:

$$\text{Accuracy} = +\frac{45+44}{45+44+5+6} \times 100\% = \frac{89}{100} \times 100\% = 89\%$$

In addition to the overall accuracy, per-class accuracy can also be calculated:

- Male Accuracy

  $$\text{Male Accuracy} = +\frac{TP}{TP+FN} \times 100\% = \frac{45}{45+5} \times 100\% = 90\%$$

- Female Accuracy

  $$\text{Female Accuracy} = +\frac{TN}{TN+FP} \times 100\% = \frac{44}{44+6} \times 100\% = 88\%$$

Based on the testing of 100 facial images from the UTKFace dataset, the overall accuracy of the DeepFace model in gender classification was 89%, obtained from correctly identifying 45 male and 44 female images out of the total 100 images.

For the male class, out of 50 images tested, the model correctly recognized 45 images (90%), while 5 images (10%) were misclassified as female. For the female class, out of 50 images, the model correctly identified 44 images (88%), with 6 images (12%) misclassified as male.

These results indicate that DeepFace can achieve fairly accurate predictions even without retraining, as the model leverages knowledge from its pretraining. A bar chart illustrates the distribution of predicted results, showing that male images tend to be recognized more accurately than female images. Meanwhile, a pie chart visualizes the overall proportion of correct and incorrect classifications, with the largest portion corresponding to correctly predicted images.

Overall, the findings suggest that DeepFace is effective in automatically classifying gender based on facial images. However, misclassifications were still observed, which may be influenced by image conditions such as camera angle, lighting, or facial expression. Therefore, improving the dataset quality or incorporating additional preprocessing steps could enhance system performance in future studies.

The test results show that DeepFace can classify gender from facial images with relatively high accuracy, even without retraining. Accuracy for male images was slightly higher (90%) compared to female images (88%), which may be attributed to differences in facial feature characteristics between the two classes, as well as variations in expressions, camera angles, or lighting conditions.

The choice of a 25-year age threshold in this study was made methodologically to reduce variability in facial data that could affect gender classification results. In the age range below 25, particularly for children and adolescents, human faces undergo significant biological growth, such as changes in facial proportions, jaw development, and secondary feature formation. This variability may lead to notable differences in facial feature patterns even within the same gender class.

Prediction errors were largely influenced by external factors, such as:

- Camera angle: images with non-frontal face poses may reduce detection accuracy.
- Lighting: uneven illumination or shadows can affect the recognition of facial features.
- Facial expression: non-neutral expressions (e.g., smiling, partially closed eyes) may make facial features more difficult for the model to recognize.

Nevertheless, the distribution of test results, with the majority being correct, indicates that the DeepFace model is effective in extracting facial features that distinguish between male and female subjects. The analyzed bar charts and pie charts show a higher proportion of correct predictions compared to incorrect ones, confirming the model's consistency.For future research, several recommendations could be made to further improve the system's performance, including:

1. Expanding the diversity of the test dataset, including various ages and ethnicities, to provide a more comprehensive evaluation of the model's generalization.
2. Advanced preprocessing, such as image augmentation (rotation, flipping, lighting adjustments), to stabilize prediction results.
3. Image quality: ensuring sharp images, even lighting, and consistent face angles can reduce prediction errors

By implementing these improvements, the DeepFace model has the potential to achieve higher and more stable performance under diverse real-world image conditions, making it suitable for practical applications such as security systems, surveillance, and gender-based content verification.

4. CONCLUSION

Based on the testing results on the UTKFace dataset, it can be concluded that the pretrained DeepFace model is capable of performing gender classification on facial images with a high level of accuracy without requiring retraining. Out of 100 tested images, the model correctly identified 45 male images and 44 female images, with class-specific accuracies of 90% and 88%, respectively, and an overall accuracy of 89%. Prediction errors were mostly influenced by image quality factors, such as face angle, lighting conditions, and facial expressions. The automatic preprocessing stage performed by DeepFace, including face detection, face cropping, image size normalization, and orientation adjustment, proved effective in preparing images for the classification process. Thus, DeepFace demonstrates strong capability in automatically classifying gender from facial images, although improvements in image quality or the addition of advanced preprocessing steps could further enhance the system's performance in future studies.

5. ACKNOWLEDGEMENT

To successfully complete this research, the author would like to express sincere gratitude to the Informatics Study Program, Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, for providing the facilities and a supportive academic environment. The author also wishes to thank their supervisor, colleagues, and all parties who have assisted in this research, both directly and indirectly.

6. REFERENCES

Algoritme, J., Satriawan, M. A., Widhiarso, W., Multi, U., Palembang, D., Learning, D., & Kelamin, J. (2023). Klasifikasi Pengenalan Wajah Untuk Mengetahui Jenis Kelamin Menggunakan Metode Convolutional Neural Network 1,2. 4(1), 43–52. https://doi.org/10.35957/algoritme.xxxx

Andrekha, M. Z., & Huda, Y. (2021). Deteksi Warna Manggis Menggunakan Pengolahan Citra dengan Opencv Python P - ISSN : 2302-3295. 9(4).

Arifandi, A. (2022). Menggunakan Algoritma Convolutional Neural Network (CNN). 4(2), 89–96.

Fajar, Y. I., & Huda, U. N. (2025). Penerapan tensorflow dalam prediksi jenis kelamin dengan menggunakan algoritma cnn. 2(1), 122–129.

Ilmiah, J., & Pendidikan, W. (2024). No Title. 10(20), 168–177.

Informatika, P. S., Sain, F., Kristen, U., & Yogyakarta, I. (2024). Klasifikasi Usia Berdasarkan Citra Wajah Menggunakan Algoritma You Only Look Once Age Classification Based on Face Image Using You Only Look Once Algorithm. 13(105), 996–1003.

Melangi, S. (2020). Klasifikasi Usia Berdasarkan Citra Wajah Menggunakan Algoritma Artificial Neural Network dan Gabor Filter. 2, 60–67.

Nabila, A. (2024). DEEP LEARNING MENGIDENTIFIKASI UMUR MANUSIA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN). 12(3), 1836–1843.

Opencv, I., Recognition, F., Deteksi, P. R., & Dan, U. (2023). Implementation of Opencv Face Recognition in Real-Time Age and Gender Detection Using Python with Classification Method Pendahuluan. 1(September).

Sachi, H., Wijayanti, L., & Octaviani, S. (2023). Klasifikasi Gender Berdasarkan Gambar Menggunakan Metode Deep Learning Pada MATLAB. 16(2).