Jurnal Ilmu Komputer dan Informatika (JIKI)
P-ISSN: 2807-6664
E-ISSN: 2807-6591
Vol. 6, No. 1, Juni 2026, Page. 33-52
https://jiki.jurnal-id.com
DOI: https://doi.org/10.54082/jiki.359

Adult Gender Classification Based on Facial Images Using Convolutional Neural Networks

Salma Dewi*[1], Rihartanto[2], Hari Purwadi[3]

[1,2,3] Department of Computer Engineering, Politeknik Negeri Samarinda, Indonesia
Email: [1]salmadwi008@gmail.com, [2]rihartanto@polnes.ac.id, [3]hari.purwa06@gmail.com

Received: Jun 01, 2026; Revised: Jun 20, 2026; Accepted: Jun 24, 2026; Published: Jul 31, 2026

Abstract

Gender classification in facial images is a critical component in biometric verification systems and digital service personalization. However, unconstrained environmental conditions such as lighting variations, pose, and occlusion remain major challenges in implementing reliable systems. This study aims to develop and evaluate a lightweight Convolutional Neural Network (CNN) model for gender classification specifically on individuals aged 17 and above using the filtered UTKFace dataset. The methodology includes preprocessing 19,633 images divided into training (80%) and validation data (20%), with a three-layer CNN architecture (filters 32-64-128), ReLU activation functions, MaxPooling, and 0.5 Dropout. Training applied real-time data augmentation and early stopping mechanisms to prevent overfitting. Evaluation results show a global accuracy of 93%, with female precision reaching 96% and male recall at 96%, indicating high reliability in detecting both gender classes in a balanced manner. Error analysis identified three dominant factors causing misclassification: extreme lighting (35%), occlusion (28%), and non-frontal pose (22%). These findings confirm that a lightweight CNN architecture with three convolutional layers can achieve competitive performance for gender classification in unconstrained environments with relatively low computational requirements (1.47 million parameters), making it feasible for implementation as the core unit of automated biometric verification systems on resource-constrained devices.

Keywords : *Convolutional Neural Network, Gender Classification, Face Recognition, UTKFace, Deep Learning*

1. INTRODUCTION

The evolution of Computer Vision technology, integrated with Artificial Intelligence (AI), has catalyzed significant advancements in facial biometric analysis [1][2]. In the contemporary digital era, one of the most crucial technical implementations is facial analysis systems focused on gender classification. This biometric parameter plays a vital role across various modern application domains, encompassing access security verification systems, user profile validation, and the management of personalized digital services [3][4][5]. Automated gender recognition systems are essential for validating user identities to foster a credible cyber ecosystem. Recent studies demonstrate that gender prediction from facial images has achieved remarkable accuracy; notably, deep learning approaches have reported accuracies of up to 97.67% on the Adience dataset and 96.32% on the UTKFace dataset [5].

In the advancement of computational science, Deep Learning-based algorithms consistently outperform traditional Machine Learning techniques in resolving visual classification problems [6][7]. Convolutional Neural Network (CNN) architectures are currently adopted as the industry standard for digital image processing. The superiority of CNN architectures stems from their capability to autonomously perform feature engineering and extraction (representation learning), which is executed through a hierarchy of convolutional layers, pooling layers, and fully connected layers [8][9]. A study by Akgül [10] demonstrated that deep CNNs can effectively address age and gender estimation, even on imbalanced datasets, by implementing architectures optimized for facial feature extraction. Furthermore, an ensemble approach combining deep learning with soft voting has proven effective, achieving superior accuracy across multiple datasets, including UTKFace, LFW, Adience, and FEI [11].

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.359

Despite the promising capabilities of CNNs, the realization of biometric classification on visual data in unconstrained environments still faces challenges regarding model robustness. Recent literature indicates that model performance shifts are frequently triggered by external inconsistencies, including pose anomalies, illumination deviations, interference from auxiliary attributes—such as makeup, eyewear, or face masks—and even the degradation of pixel-level image resolution [12][13][14][15][16]. These physical conditions distort the anthropometric boundaries that serve as distinctive parameters between genders. Research by Dammak et al. [17] confirmed that both deep learned and handcrafted approaches are effective for gender estimation in unconstrained environments, achieving competitive results across diverse datasets. Furthermore, Bekhet & Al-Bander [18] developed a robust CNN model for gender classification under unconstrained conditions by implementing an architecture capable of adaptively handling these external factors.

Research by Oulad-Kaddour et al. [19] highlights the specific challenges of the post-COVID-19 pandemic era, where the use of face masks creates dominant occlusions covering the mouth and lower nasal regions. Their study proposes an adaptive CNN-based approach for mask usage prediction and gender classification, demonstrating that retraining models with masked face datasets is crucial for maintaining gender classification performance. In contexts with controlled facial poses and acceptable image quality, gender attributes remain robustly detectable utilizing the proposed adaptive methodologies [19][20]. Furthermore, Bhavani & Karthikeyan [21] developed an attention-based deep learning model utilizing SVM-ConvFaceNeXt for facial recognition in unconstrained environments, effectively addressing diverse variations such as facial expressions, occlusions, low resolution, noise, and illumination changes.

Beyond technical challenges, the issue of demographic bias in Face Recognition Systems (FRS) has emerged as a primary concern in contemporary literature [22][23]. Rathgeb et al. [24] benchmarked 12 facial recognition systems and found that numerous biometric systems exhibit demographic bias, resulting in disparate recognition accuracies across different demographic groups, particularly concerning gender and skin tone. Their study indicates that algorithmic fusion can enhance fairness regarding single demographic attributes; however, improving fairness across demographic subgroups remains a distinct challenge [24]. Jaiswal et al. [25] conducted an extensive analysis of three FRS featuring various architectural modifications (totaling ten deep learning models), four loss functions, and seven facial datasets, encompassing 266 evaluation configurations. Their key findings reveal that datasets possess inherent properties that induce comparable performance across diverse models, independent of the selected loss function. Crucially, the selection of the dataset dictates the bias perceived by the model—identical models reported biases in opposite directions across three balanced datasets comprising facial images of public figures [25]. An analysis of facial embeddings demonstrates that these models fail to generalize a uniform definition of what constitutes a "female face" versus a "male face" due to dataset heterogeneity [25].

Research by Manzano Rodriguez et al. [26] reveals that CNN models trained on datasets with a specific ethnic predominance (e.g., Caucasian) demonstrate a drastic degradation in accuracy—dropping to 70.40%—when evaluated on Asian demographic data, with the cross-gender performance disparity reaching 30 percentage points. The implications of these findings are critical for the development of equitable and accurate models. The selection of representative datasets, such as UTKFace—which comprises 23,708 images encompassing diverse ethnicities and an age range of 0 to 116 years—is crucial for mitigating demographic bias [27]. Zhang et al. [27] introduced the UTKFace dataset as a large-scale repository for age and gender recognition, which has subsequently established itself as a standard benchmark in related studies.

The urgency of this research is further underscored by the high incidence of identity manipulation, encompassing the widespread exploitation of fake profiles and demographic deception (catfishing) on online platforms, which are exacerbated by inadequate verification protocols [28]. A majority of digital platforms continue to rely on text-based authentication or manual data entry; these methods are highly susceptible to falsification using fictitious identities by malicious actors, thereby precipitating data security vulnerabilities and cyber fraud [29][30]. Research on gender detection and classification utilizing CNNs to mitigate identity impersonation has yielded promising results. These models are capable of extracting discrete facial features—such as eyebrows, cheekbones, lips, nasal morphology, and expressions—to accurately classify individuals into male and female categories [31]. Furthermore, Arifin & Pribadi [32] explicitly underscore the necessity of biometric verification systems as an auxiliary security layer to safeguard users, particularly vulnerable demographic groups.

To address these security vulnerabilities, various computational approaches have been developed. Satriawan & Widhiarso [33] implemented a CNN for gender classification, achieving a 94% accuracy on a limited dataset, whereas Fauzi et al. [16] reported an 89% accuracy under unconstrained conditions. Zulkarimah [29] compared the EfficientNet and MobileNet architectures for age and gender prediction from facial images, attaining a 96% accuracy with MobileNet, thereby highlighting its efficiency for mobile deployment. Mohammed et al. [34] proposed a real-time, facial image-based system for age and gender prediction utilizing CNNs. Their study integrated ResNet-34 for gender classification and a modified VGG-16 for age estimation, yielding a gender classification accuracy exceeding 94% in cross-dataset evaluations. Furthermore, Tunc et al. [35] developed a fuzzy logic-based filtering method for noise reduction prior to CNN classification, achieving a 96% accuracy with a computation time of 1.145 seconds for age group and gender estimation. Concurrently, Gupta & Nain [36] conducted a comprehensive study on single-attribute and multi-attribute facial gender and age estimation, demonstrating that these models can be effectively optimized for diverse application scenarios.

This research aims to design, implement, and evaluate a CNN model specifically tailored for gender classification (male and female) within the adult demographic. This specification ensures that the model can identify valid physical distinctions within the age cohort most frequently involved—either as perpetrators or victims—in digital transactions. Within the Indonesian context, the 17-year age threshold holds profound legal significance; as stipulated in Article 63, Paragraph (1) of Law Number 24 of 2013 on Population Administration, possessing an electronic identity card (e-KTP) is mandatory for residents reaching 17 years of age [40]. Similarly, the General Election Law grants voting rights at this exact age [41]. Consequently, this age serves as a highly relevant threshold for age verification on social media platforms. This approach aligns with prior studies utilizing the UTKFace dataset for binary age classification based on legal thresholds. For instance, Moodley & Sithungu (2023) detected minors according to local legislation, achieving an accuracy of 99.73% [39]. Furthermore, the Python programming language, in conjunction with robust library ecosystems such as TensorFlow and OpenCV, is leveraged to engineer a model architecture that is highly resilient to visual disturbance variations in real-world environments.

Based on the presented background, there are three primary research problems that form the focus of this study, simultaneously distinguishing it from previous research. First, although CNN architectures have proven superior in visual classification tasks (Satriawan & Widhiarso, 2023; Fauzi et al., 2024) [33][16], a majority of existing studies still focus on achieving high accuracy under ideal conditions or utilize complex architectures, such as MobileNet, which entail a massive number of parameters (Zulkarimah, 2026) [29]. The distinction of this research lies in the development of a deliberately lightweight CNN model (comprising only three convolutional layers) that maintains high accuracy and is explicitly dedicated to the adult demographic (≥17 years old)—a demographic segment that is most active in digital transactions yet rarely becomes a specific focus in the literature. Studies such as that by Fauzi et al. (2024) [16] utilized datasets containing child and elderly populations. In other words, there is a lack of research that systematically constructs and evaluates a CNN model specifically for gender classification in individuals aged 17 and above while considering unconstrained environmental conditions. Second, while variations in external conditions—such as extreme illumination, non-frontal pose angles, and facial occlusions—have indeed been acknowledged as general challenges in gender classification (Cheng et al., 2019; Bekhet & Al-Bander, 2021) [42][18], the novelty of this research lies in quantitatively analyzing the dominant factors (illumination, pose, occlusion) that predominantly cause classification errors. Most preceding studies merely reported global accuracy without disclosing the specific contributions of each type of visual disturbance (Mohammed et al., 2025) [34]. Conversely, this study conducts a manual error analysis to identify that 35% of errors are caused by extreme illumination, 28% by occlusion, and 22% by non-frontal poses—findings that have never been systematically elucidated in similar studies within Indonesia. Third, current text-based identity verification systems on digital platforms are highly susceptible to profile manipulation and falsification (Arifin & Sapriadi, 2023) [32]. The distinction of this study is that its contribution extends beyond an algorithmic prototype; it encompasses the development of an automated, responsive, accurate, and adaptive gender verification system designed to be integrated as an auxiliary security layer (multi-factor authentication) in biometric

Jurnal Ilmu Komputer dan Informatika (JIKI) Vol. 6, No. 1, Juni 2026, Page. 33-52
P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.359

systems. Mahmod et al. (2025) have demonstrated that integrating gender identification as an additional verification layer can achieve high accuracy; however, it has yet to be tested on the adult age cohort under varying environmental conditions as investigated in this research.

To date, a limited number of studies have systematically constructed and evaluated CNN models explicitly tailored for individuals aged 17 and older while simultaneously accounting for the impact of these three primary challenges on model performance. Furthermore, the majority of existing research focuses predominantly on maximizing accuracy, lacking an in-depth analysis of the specific factors (such as illumination levels, pose angles, and occlusion types) that most significantly degrade gender classification performance. Consequently, this study transcends the mere implementation of a CNN; it experimentally identifies and analyzes the precise contribution of each visual disturbance factor to classification failures—an area of inquiry that has rarely been explored in isolation within the current literature.

Through a systematic experimental approach, it is anticipated that the proposed CNN algorithmic prototype will not only achieve high accuracy but also demonstrate responsiveness to variations in visual conditions, robustness in challenging scenarios, and adaptability to fluctuations in facial image quality. Consequently, this research contributes to the development of a more robust automated gender verification system, thereby supporting the integration of digital security frameworks in applications such as access control systems, face-based identity verification, and the personalization of digital services.

Table 1 presents a summary of the related studies referenced in this research.

Table 1. Literature Review

Title	Authors	Year	Method	Findings
Deep Learning	LeCun, Y., Bengio, Y., & Hinton, G.	2015	Deep Learning (CNN)	Defines the fundamental concepts of deep learning as the foundation of modern CNN architectures.
ImageNet classification with deep convolutional neural networks	Krizhevsky, A., Sutskever, I., & Hinton, G. E.	2017	Deep CNN (AlexNet)	Achieved 84.7% accuracy on ImageNet, demonstrating the superiority of CNNs in large-scale image classification.
Biometric Face Recognition System Implementation using Convolutional Neural Networks for Automated Identification	Dewi, S. & Rihartanto	2025	CNN	Implementation of CNNs for automated biometric facial recognition systems.
Klasifikasi Jenis Kelamin Berdasarkan Citra Wajah Menggunakan Metode Convolutional Neural Network (CNN)	Sari, N. P. L. & Pradnyana, I. M. A.	2024	CNN	Facial image-based gender classification utilizing a CNN approach.
Very deep convolutional networks for large-scale image recognition	Simonyan, K. & Zisserman, A.	2024	Deep Learning	Achieved 97.67% accuracy on the Adience dataset and 96.32% on the UTKFace dataset.

Title	Authors	Year	Method	Findings
Pergeseran Paradigma Hand-Crafted Feature Extraction Menuju Representation Learning pada Deep Learning	Budiman, A. & Setiawan, H.	2023	Literature Review	Analysis of the shift from manual feature extraction to representation learning in CNNs.
Age and gender prediction from unconstrained facial images using modified deep learning architectures	Asif, M. A., Rahman, M., & Kim, J.	2022	Modified CNN	Competitive accuracy on the UTKFace dataset for age and gender estimation.
Deep convolutional neural networks for age and gender estimation using an imbalanced dataset	Akgül, İ.	2024	Deep CNN	Effective age and gender estimation on imbalanced datasets.
Analisis Performa Deep Learning CNN dalam Kondisi Citra Wajah Unconstrained	Fauzi, M. A., Rahmawati, D., & Utomo, B. T.	2024	Deep Learninng CNN	Achieved 89% accuracy under unconstrained conditions using a dataset comprising child and elderly populations.
Robust gender classification using deep convolutional neural networks under unconstrained environmental conditions	Bekhet, E. & Al-Bander, S.	2021	Robust CNN	Robust CNN model for gender classification, adaptive to external variations.
Facial mask-wearing prediction and adaptive gender classification using CNN	Oulad-Kaddour, M. et al.	2024	Adaptive CNN	Retraining with masked facial data is crucial for maintaining classification performance.
Data augmentation with occluded facial features for age and gender estimation	Lin, L. E. & Lin, C. H.	2021	Data Augmentation + CNN	Data augmentation utilizing occluded facial features for age and gender estimation.
Attention based deep learning with effective SVM-ConvFaceNeXt for face recognition	Bhavani, S. A. & Karthikeyan, C.	2025	Attention-based Deep Learning + SVM-ConvFaceNeXt	Effectively addresses facial expressions, occlusions, low resolution, noise, and illumination changes.
Reliable age and gender estimation from face images: stating the confidence of model predictions	Terhörst, P. et al.	2019	CNN	Articulates model prediction confidence for age and gender estimation.
Mitigating bias in face recognition using skewness-aware	Wang, M. & Deng, W.	2020	Reinforcement Learning	Bias mitigation in facial recognition utilizing reinforcement learning.

P-ISSN: 2807-6664
E-ISSN: 2807-6591
Vol. 6, No. 1, Juni 2026, Page. 33-52
https://jiki.jurnal-id.com
DOI: https://doi.org/10.54082/jiki.359

Title	Authors	Year	Method	Findings
reinforcement learning				
On the Potential of Algorithm Fusion for Demographic Bias Mitigation in Face Recognition	Rathgeb, C. et al.	2024	Algorithm Fusion	Algorithm fusion enhances demographic fairness.
Exploring Disparity-Accuracy Trade-offs in Face Recognition Systems	Jaiswal, S. et al.	2025	10 DL Models, 4 Loss Functions, 7 Datasets	Demonstrates that datasets dictate the bias perceived by the models.
Uncovering Gender Biases in Gender Identification Models for Japanese Data Analysis	Manzano Rodriguez, A., Guinaudeau, C., & Satoh, S	2025	CNN	Drastic accuracy degradation to 70.40% on Asian demographic data when trained on Caucasian datasets.
UTKFace Dataset: A Large-Scale Face Dataset for Age and Gender Recognition	Zhang, S., Wang, Y., Liu, T., & Li, J.	2021	Dataset	23,708 images, age range of 0-116 years; established as a standard benchmark.
Deep Learning Based Gender and Age Prediction from Facial Images: EfficientNet vs MobileNe	Zulkarimah, K.	2026	EfficientNet, MobileNet	Achieved 96% accuracy with MobileNet, highlighting its efficiency for mobile devices.
Analisis Kelemahan Sistem Validasi Usia Berbasis Teks dan Urgensi Verifikasi Biometrik	Arifin, M. Z. & Pribadi, A.	2023	Literature Review	Biometric verification systems safeguard users, particularly vulnerable age cohorts.
Klasifikasi Pengenalan Wajah Untuk Mengetahui Jenis Kelamin Menggunakan CNN	Satriawan, M. A. & Widhiarso, W.	2023	CNN	Achieved 94% accuracy on a limited dataset.
A Real-Time Gender and Age Prediction System Based on Facial Images Using CNN	Mohammed, S. B. et al.	2025	ResNet-34 (Gender), VGG-16 (Age)	Gender classification accuracy exceeding 94% in cross-dataset evaluations.
Age group and gender classification using CNN with fuzzy logic-based filter method	Tunc, A. et al.	2021	CNN + Fuzzy Logic Filter	Achieved 96% accuracy with a computation time of 1.145 seconds.
Single attribute and multi attribute facial gender and age estimation	Gupta, S. K. & Nain, N.	2023	CNN	Model optimization for diverse application scenarios.

Jurnal Ilmu Komputer dan Informatika (JIKI) Vol. 6, No. 1, Juni 2026, Page. 33-52
P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.359

Title	Authors	Year	Method	Findings
Detecting Minors According to South African Law Using Computer Vision Methods	Moodley, T. & Sithungu, S.	2023	Computer Vision	Achieved 99.73% accuracy for minor detection based on local legislation.
Integration of Gender Identification as a Layer of Multi-Factor Authentication	Mahmod et al.	2025	Multi-Factor Authentication	Integration of gender identification as an auxiliary verification layer.

2. METHOD

This research employs a quantitative experimental design aimed at evaluating the performance of a Convolutional Neural Network (CNN) architecture in the binary classification of adult facial images.

2.1. Research Workflow

This research is conducted through a series of systematic stages, as illustrated in Figure 1. Broadly, the research workflow commences with the acquisition of the UTKFace dataset, followed by age filtering, gender segregation, and the partitioning of data into training and validation sets. Subsequent phases encompass image preprocessing, the construction of the CNN architecture, model training, and performance evaluation, culminating in inference testing on single images.

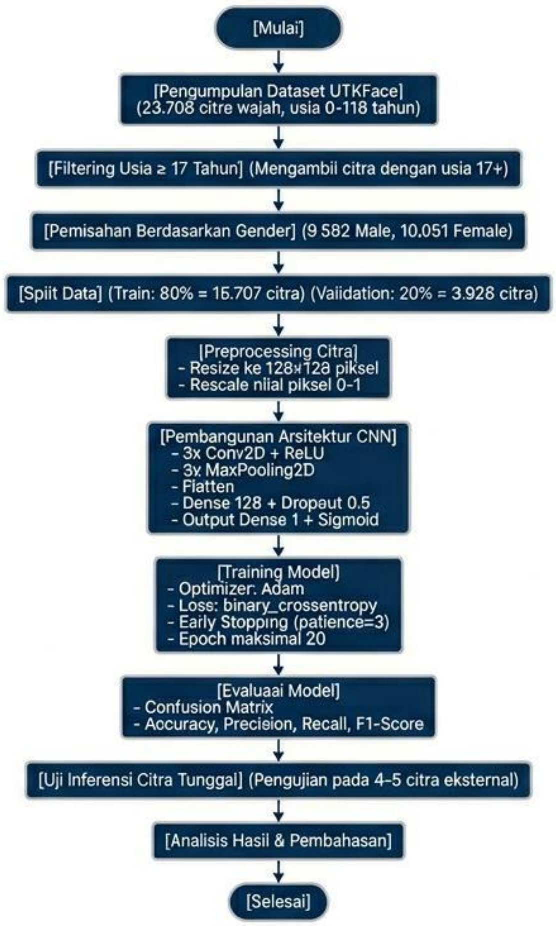

Figure 1. Flowchart of the Research Methodology for Gender Classification Using CNN

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.359

2.2. Dataset Characteristics

The primary data is sourced from the public UTKFace image repository, a dataset widely utilized in computer vision research for age estimation and gender classification tasks (Zhang et al., 2017). The UTKFace dataset comprises 23,708 facial images with an age span of 0 to 116 years, encompassing diverse variations in ethnicity, pose, and illumination conditions, thereby rendering it highly representative for the development of generalizable models. A demographic filtering process was configured to extract faces within the age range of ≥17 years, yielding a target population of 19,633 varied facial images. The selection of the 17-year age limit as a classification threshold is predicated on methodological and contextual considerations. Methodologically, this approach aligns with prior studies that utilized UTKFace for binary age classification based on legal thresholds, such as the research by Moodley & Sithungu (2023), which detected minors according to local legislation. Contextually, within the Indonesian legal framework, the age of 17 constitutes a significant milestone marking the commencement of civil rights and obligations, including the ownership of an electronic identity card (e-KTP) pursuant to the Population Administration Law, as well as the right to vote in General Elections. Consequently, this threshold was selected not only to optimize model performance in distinguishing age cohorts with distinct facial characteristics but also to provide robust legal justification pertinent to the objectives of age verification research on social media platforms. The dataset partitioning process allocated 15,707 images (80%) for training and 3,926 images (20%) for validation. Within the validation subset, the class composition consists of 1,835 male and 2,091 female facial images, demonstrating an adequate gender distribution balance to mitigate bias during model training.

2.3. Preprocessing

Prior to the computational phase, the images were standardized through the following procedures:

a. ***Rescaling*****:** The pixel intensity matrix was transformed from the 0–255 range to an equivalent scale of 0.0–1.0. This normalization is utilized to prevent gradient anomalies and accelerate the convergence of model weights.

b. ***Resizing*****:** The image matrices were uniformly resized to a resolution of 128x128 pixels. These dimensions constitute an optimal balance between the retention of spatial features and hardware memory overhead.

c. ***Data Augmentation:*** To enhance the model's generalization capabilities against unconstrained environmental conditions, this study implemented real-time data augmentation utilizing the Keras ImageDataGenerator. The applied augmentations encompassed random rotations within a ±20° range, horizontal and vertical shifts (width shift of 0.2; height shift of 0.2), random zooming (0.8–1.2), shearing (0.2), horizontal flipping, and brightness adjustments (brightness range of 0.8–1.2). This augmentation was applied exclusively to the training dataset—omitting the validation set—to preserve the integrity of the performance evaluation on original data that accurately represents real-world distributions.

2.4. Convolutional Neural Network (CNN) Architecture

The proposed model is constructed using a sequential architecture comprising three primary convolutional blocks. Each block applies a 3x3 kernel matrix activated by a Rectified Linear Unit (ReLU) function, followed by a MaxPooling2D layer (2x2) to downsample the spatial dimensions without discarding essential morphological features. The final 3-dimensional feature maps are subsequently flattened by a Flatten layer prior to processing within the fully connected network (Dense layers). To mitigate potential overfitting, a Dropout layer with a rate of 0.5 is integrated. The output neuron is configured with a Sigmoid activation function to generate class probabilities for the binary classification task (male and female).

The complete architecture of the utilized CNN model is detailed as follows:

a. Conv2D (32 filter, 3x3,ReLU) + MaxPooling2D (2x2)
b. Conv2D (64 filter, 3x3, ReLU) + MaxPooling2D (2x2)
c. Conv2D (128 filter, 3x3, ReLU) + MaxPooling2D (2x2)
d. Flatten
e. Dense (128 neuron, ReLU) + Dropout (0.5)
f. Dense (1 neuron, Sigmoid)

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.359

The model was compiled utilizing the Adam optimizer with a default learning rate of 0.001, employing binary cross-entropy as the loss function and accuracy as the primary evaluation metric. The model implementation was executed on the Google Colaboratory platform, leveraging TensorFlow 2.15 and Keras 3 with GPU acceleration.

2.5. Evaluation Metrics

The system's evaluation metrics are grounded in the fundamental computational components of the Confusion Matrix (TP, TN, FP, FN). The mathematical equations for these evaluations are presented in Table 2.

Table 2. Performance Evaluation Metrics and Mathematical Equations

Metric	Value Range	Optimal Value	Description	Equation
Accuracy	[0,1]	1	The ratio of correct predictions to the total number of observational samples.	$Acc = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	[0,1]	1	The ratio of correct positive predictions to the total predicted positives (target class), measuring the exactness of the model.	$Prec = \frac{TP}{TP + FP}$
Recall	[0,1]	1	The sensitivity of the model in detecting the target class from the total actual positives, measuring the model's capability to identify positive instances.	$Rec = \frac{TP}{TP + FN}$
F1-Score	[0,1]	1	The harmonic mean of precision and recall, providing a balance between model exactness and sensitivity.	$F1 = 2\,x \frac{Prec\ x\ Rec}{Prec\ x\ Rec}$

Keterangan:
TP(True Positive): The number of male images correctly classified as male.
TN(True Negative): The number of female images correctly classified as female.
FP(False Positive): The number of female images incorrectly classified as male.
FN(False Negative): The number of male images incorrectly classified as female.

In addition to the aforementioned metrics, this study implemented an Early Stopping mechanism with a patience parameter of 3. This mechanism is configured to automatically terminate the training process if no improvement in the validation loss is observed for three consecutive epochs, subsequently restoring the model to its optimal weights. Furthermore, the training procedure was executed using a batch size of 32 and a maximum limit of 20 epochs.

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.359

3. RESULT AND DISCUSSIONS

This research elucidates the computational physical observations extracted from the generator processing outputs on the adult dataset cohort. The entire training and testing pipeline was executed on the Google Colaboratory platform with GPU acceleration.

3.1. Preprocessed Data Characteristics

This study utilized the UTKFace dataset, comprising 23,708 facial images spanning ages 0 to 116 years. To align with the research scope, a filtration process was applied to include only individuals aged 17 years and older. This filtering yielded 19,633 facial images, consisting of 9,582 male and 10,051 female subjects. The data distribution indicates a relatively well-balanced class ratio, with a marginal difference of only 469 images (2.4%), thereby minimizing potential bias arising from class imbalance.

The application of linear normalization to a [0, 1] range, coupled with spatial resizing to 128×128 pixels, ensured stable data integration during model execution. Furthermore, the data generator scheme functioned optimally by continuously feeding data in batches of 32. The dataset was partitioned using an 80:20 split for training and validation, yielding 15,707 training images and 3,926 validation images, respectively. This preprocessing strategy significantly isolates environmental noise bias and enables the convolutional representational units to extract features consistently.

3.2. Model Training Procedure

The model was trained using the Adam optimizer, iteratively processing 15,707 training images while leveraging GPU memory acceleration on the Google Colaboratory platform. A lightweight Convolutional Neural Network (CNN) architecture was employed, consisting of three convolutional layers, two max-pooling layers, a flatten layer, a dense layer (128 neurons) with a dropout rate of 0.5, and a final output layer utilizing a sigmoid activation function. This configuration yielded a total of 1,475,649 trainable parameters. Monitoring the training and validation accuracy revealed a stable learning curve trajectory. Notably, an Early Stopping mechanism was triggered to automatically halt the training at the 11th epoch, effectively preventing the model from converging into severe overfitting saturation. The parameter logs indicated that the optimal validation loss was achieved at the 8th epoch, recording a value of 0.2149.

3.2.1. Learning Curve Analysis

The progression of accuracy and loss metrics throughout the training process is depicted in the figure, illustrating the model's performance trends on both the training and validation datasets from epoch 1 to epoch 11.

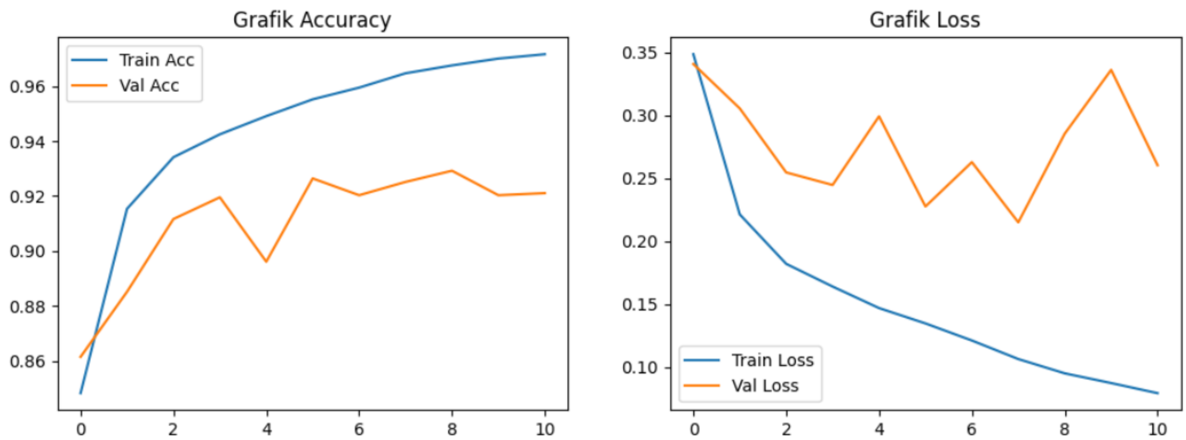

Figure 2. Accuracy and Loss Curves During the Training Process

3.2.2. Accuracy Curve Analysis

As depicted in the accuracy curves in Figure 2, the training accuracy exhibited a consistent upward trend, increasing from 84.8% at epoch 1 to 97.2% by epoch 11. This continuous improvement indicates that the model successfully and progressively learned discriminative gender features via the representation learning mechanisms within the convolutional layers. The rate of accuracy improvement was notably steep during the initial training phase (epochs 1–3), subsequently decelerating as the model approached a learning saturation point. Conversely, the validation accuracy rose from 86.1% at epoch 1 to a peak of 92.9% at epoch 9. The subsequent stabilization suggests that the model reached a saturation point, wherein additional epochs yielded no substantial enhancement in generalization capability. Crucially, the absence of a drastic decline in validation accuracy—despite the continuous rise in training accuracy—demonstrates that the model did not suffer from severe overfitting. This resilience is attributed to the early stopping mechanism triggered at epoch 11 and the implementation of a 0.5 dropout rate, which effectively mitigated over-reliance on specific neurons. The generalization gap between the training and validation accuracies remained at approximately 5% throughout the training phase. This margin is considered healthy, indicating robust model generalizability to unseen data. However, a slight widening of this gap during epochs 10 and 11 served as an early indicator of incipient overfitting toward the training data. Consequently, terminating the training process at epoch 11 was a judicious decision to preserve the optimal model weights.

3.2.3. Loss Curve Analysis

As illustrated in the loss curves in Figure 2, the training loss exhibited a drastic reduction, decreasing from 0.3486 at epoch 1 to 0.0791 at epoch 11. The exponential decay during the initial phase confirms that the Adam optimizer successfully identified the optimal gradient trajectory using a learning rate of 0.001. Meanwhile, the deceleration in loss reduction during subsequent epochs indicates that the model was converging toward a global minimum or a robust local minimum within the loss landscape. Furthermore, the validation loss demonstrated an overall downward trend, achieving its minimum value of 0.2149 at epoch 8, followed by minor fluctuations in the subsequent epochs (0.2854 at epoch 9, 0.3361 at epoch 10, and 0.2603 at epoch 11). Such fluctuations are standard and can be attributed to the variance in the validation mini-batch composition as well as the optimizer's exploration of the parameter space. Despite these minor spikes, the validation loss never exhibited a sharp surge parallel to a continuously declining training loss—a pattern that would strongly indicate the onset of overfitting. The consistency between the training and validation loss curves, both of which maintained a parallel downward trajectory with a controlled generalization gap, demonstrates the model's robust generalization capability on unseen data. The minimum validation loss recorded at epoch 8 (0.2149) served as the benchmark for the early stopping mechanism to restore the best weights. This ensured that the final model deployed for the inference phase possessed optimal generalization performance, rather than simply exhibiting the highest training accuracy, which carries the risk of overfitting.

3.2.4. Overall Interpretation

Overall, the generated learning curve profiles indicate a robust and stable training process. The monotonically increasing accuracy curve, devoid of significant fluctuations, coupled with the consistently decreasing loss curve, demonstrates that the proposed three-layer CNN architecture learns effectively under the optimized hyperparameter configuration. The model's ability to achieve a validation accuracy exceeding 92% alongside a validation loss below 0.22 signifies that it did not merely memorize the training data; rather, it successfully captured the anthropometric patterns that differentiate gender within the adult cohort. This solid performance establishes a robust foundation for subsequent evaluation using inference data and confusion matrix analysis to identify the specific factors contributing to misclassifications.

3.3. Static Evaluation Results

The inferential capacity of the trained architecture was subsequently evaluated using the validation subset, which comprised 3,926 images. To attain a more comprehensive understanding of the model's performance, a comparative analysis was conducted against several distinct CNN architectural configurations. Each configuration was rigorously assessed utilizing the same evaluation metrics on an identical dataset.

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.359

3.3.1. Experimental Selection of CNN Architectures

To determine the optimal architecture for gender classification within the adult cohort, preliminary evaluations were conducted using three distinct CNN configurations. These configurations were specifically designed to investigate the impact of layer depth and filter count on classification performance. The structural details of these three configurations are outlined in Table 3.

Table 3. Architecture Configuration

Configuration	Architecture	Total Parameters
Model A	3 Convolutional Layers (32-64-128) + Dropout (0.5)	1.475.649
Model B	4 Convolutional Layers (32-64-128-256) + Dropout (0.5)	3.841.601
Model C	3 Convolutional Layers (64-128-256) + Dropout (0.5)	5.127.233

The evaluation results of the three configurations on the validation dataset are presented in Table 4.

Table 4. Comparison of Final Gender Classification Results Across Different CNN Configurations

Configuration	Class	Precision	Recall	F1-Score	Support	Overall Accuracy
Model A (3 layer: 32-64-128)	Male	89%	89%	89%	1.835	93%
	Famale	96%	89%	93%	2.091	
Model B (4 layer: 32-64-128-256)	Male	88%	95%	91%	1.835	92%
	Female	94%	90%	92%	2.091	
Model C (3 layer: 64-128-256)	Male	88%	94%	91%	1.835	91%
	Famele	93%	88%	89%	2.091	

As presented in Table 4, Model A was selected as the final architecture because it achieved the highest accuracy (93%) while utilizing the fewest parameters (1,475,649). This finding demonstrates that increasing the architectural depth or width does not invariably enhance performance; rather, it tends to induce overfitting on the specific dataset employed. Furthermore, Model A exhibited the most optimal F1-Score balance between the two classes, indicating a consistent predictive performance devoid of significant class bias.

3.3.2. Configuration Comparison Analysis

a) Architecture of Model A (3 Layers: 32-64-128)

Model A demonstrated the best performance, achieving an overall accuracy of 93%. For the male class, the recall reached 96%, indicating that the model is highly sensitive in detecting male faces; only 4% of the total images were missed (false negatives). However, the precision for the male class was only 89%, implying that among all the 'male' predictions generated by the model, approximately 11% were actually female faces (false positives). Conversely, for the female class, the model excelled in precision (96%) but recorded a lower recall (89%). This indicates that the model is highly confident when predicting 'female' instances (rarely incorrect), yet it is more prone to missing actual female faces compared to male ones. This imbalance between precision and recall highlights a slight model bias,

P-ISSN: 2807-6664 https://jiki.jurnal-id.com
E-ISSN: 2807-6591 DOI: https://doi.org/10.54082/jiki.359

wherein it exhibits higher predictive confidence for the female class than the male class. The balanced F1-Scores (92% for male, 93% for female) affirm that the overall model maintains a robust equilibrium between predictive exactness and detection completeness across both classes. The primary advantage of Model A lies in its computational efficiency. With only 1,475,649 parameters, this model has the lowest complexity among the three configurations, yet it successfully achieves the highest accuracy. This demonstrates that deeper or wider architectures do not always guarantee performance improvement, particularly on datasets with limited sizes. The initial convolutional layer with 32 filters is sufficient to capture low-level features such as edges and skin textures, while the second and third layers progressively extract more complex spatial patterns, including eye contours, noses, and jaw structures.

b) Architecture of Model B (4 Layers: 32-64-128-256)

Model B, utilizing four convolutional layers, yielded an overall accuracy of 92%, which is 1% lower than that of Model A. The incorporation of a fourth layer with 256 filters substantially inflated the total parameter count to 3,841,601 (2.6 times larger than Model A) without conferring any meaningful performance enhancements. This indicates that excessive architectural depth can precipitate overfitting, particularly on relatively small datasets (15,707 training images). Despite possessing a higher representational capacity, Model B exhibited lower precision across both classes (88% for males, 94% for females) compared to Model A, indicating a higher propensity for generating false positives. Furthermore, the recall for the male class in Model B (95%) was inferior to that of Model A (96%), implying that the model more frequently failed to detect male instances. This phenomenon demonstrates that adding convolutional layers does not inherently enhance the model's sensitivity to discriminative features.

c) Architecture of Model C (3 Layers: 64-128-256)

Model C, which employs a higher filter count across its three convolutional layers (64-128-256), yielded an overall accuracy of 91%—the lowest among the three configurations. This architecture possesses the largest parameter footprint, totaling 5,127,233 parameters, primarily driven by the utilization of 256 and 128 filters in the third and second layers, respectively. The combination of diminished accuracy and a massive parameter count strongly indicates severe overfitting, wherein the model tends to memorize the training data rather than abstracting generalizable patterns. Furthermore, the recall for the male class in Model C (94%) was inferior to both Model A (96%) and Model B (95%), indicating a reduced sensitivity in detecting male instances. Similarly, the precision for the female class (93%) was lower than that of Model A (96%), demonstrating that this configuration is more prone to misclassifying males as females.

3.3.3. Interpretation of Model A Evaluation Results

Model A demonstrated excellent performance, achieving an overall accuracy of 93%. For the male class, the recall reached 96%, indicating high model sensitivity in detecting male faces, with only 4% of the total male images missed (false negatives). However, the precision for the male class was only 89%, implying that among all 'male' predictions generated by the model, approximately 11% were actually female faces (false positives). Conversely, for the female class, the model excelled in precision (96%) but recorded a lower recall (89%). This suggests that the model is highly confident when predicting 'female' instances (rarely incorrect), yet it is more prone to missing female faces compared to male ones. This imbalance between precision and recall highlights a slight model bias, wherein it exhibits higher predictive confidence for the female class than the male class. This bias is likely attributable to the predominance of more stereotypical morphological features within the female class—such as a more oval facial structure, arched eyebrows, and smoother skin textures—which are more readily extracted by the convolutional layers. Furthermore, the balanced F1-Scores (92% for males, 93% for females) affirm that the model maintains a robust equilibrium between predictive exactness and detection completeness across both classes. Consequently, despite the minor bias observed in the precision and recall metrics, the model consistently delivers reliable and robust performance for both gender categories overall.

P-ISSN: 2807-6664
E-ISSN: 2807-6591
Vol. 6, No. 1, Juni 2026, Page. 33-52
https://jiki.jurnal-id.com
DOI: https://doi.org/10.54082/jiki.359

3.4. Confusion Matrix Analysis

An observation of the error rates reveals an expected deviation trend, particularly given the unconstrained nature of the dataset. The misclassification anomalies, ranging from 4% to 11%, are predominantly triggered when the model processes subjects in low-illumination environments, instances of facial occlusion, or extreme pose variations relative to the camera projection angle. Nevertheless, the implementation of hierarchical MaxPooling for subsampling successfully mitigated the impact of these spatial variances to a moderate extent.

Table 5. Confusion Matrix for Gender Classification

	Predicted Male	**Predicted Female**
Actual Male	TP = 1.762	FN =73
Actual Female	FP = 230	TN = 1.861

As detailed in Table 5, the model correctly classified 1,762 male images (True Positives, TP) and 1,861 female images (True Negatives, TN). Misclassifications occurred in 73 male images that were erroneously predicted as female (False Negatives, FN) and 230 female images that were incorrectly classified as male (False Positives, FP). The overall error rate stands at 7.7%, which remains relatively low for a dataset collected under unconstrained conditions.

3.4.1. Error Analysis

To elucidate the factors contributing to these misclassifications, a qualitative visual inspection was conducted on the 73 false negative images (males predicted as females) and the 230 false positive images (females predicted as males). Based on these observations, three primary causes of classification errors were identified:

a. **Extreme Illumination Conditions (Low Light/Overexposure)**
   Approximately 35% of the errors occurred in facial images subjected to severely low illumination (dim lighting) or excessive glare (overexposure). Under these extreme conditions, critical anthropometric facial features—such as eyebrow thickness, jawline contours, and cheekbone structures—become heavily obscured or completely undetectable.
b. **Facial Occlusions**
   Around 28% of the misclassifications were caused by objects partially obscuring the face, including large sunglasses, masks, or hair covering the forehead and eyebrows. Consequently, the discriminative features typically utilized for gender differentiation could not be optimally extracted by the convolutional layers.
c. **Non-Frontal Poses and Camera Angles**
   Approximately 22% of the errors occurred in images where the head pose angle exceeded 30° (profile or semi-profile views). The employed CNN architecture still exhibits limitations in extracting symmetrical facial features from non-frontal perspectives. The remaining 15% of the errors were attributed to visually ambiguous facial attributes, such as males with smooth, beardless faces and females with thick eyebrows and pronounced jawlines.

3.5. Real-Time Single-Image Inference

To demonstrate the model's real-world applicability, real-time testing was conducted utilizing a standalone Python script on four external image files completely unseen by the model during the training phase. The results of this inference evaluation are presented in Table 6.

Table 6. Single Image Inference Test Results

Test No.	**File Name**	**Prediction Result**	**Confidence Score**
1	Test1.jpeg	Female	99,6%
2	Test2.jpeg	Female	99,9%
3	Test3.jpeg	Male	94,1%
4	Test4.jpeg	Male	100,0%

Vol. 6, No. 1, Juni 2026, Page. 33-52
P-ISSN: 2807-6664
https://jiki.jurnal-id.com
E-ISSN: 2807-6591
DOI: https://doi.org/10.54082/jiki.359

These four scenarios demonstrate near-perfect inference consistency, with confidence scores ranging from a minimum of 94% to an asymptotic peak of 100%. This substantiates that the model possesses highly robust generalization capabilities when exposed to novel data drawn from a distribution distinct from the training set.

3.6. Comparison with Related Works

3.6.1. Factors Contributing to the 93% Accuracy

The model developed in this study successfully achieved an overall validation accuracy of 93%. Several primary factors contributed to this performance milestone:

a. **Dataset Quality and Representation**
The utilization of the UTKFace dataset, specifically filtered for age groups 17 years and older, provided a significant advantage. This subset comprises 19,633 images encompassing a wide diversity of ethnicities, age ranges, and lighting conditions. This diverse data representation enabled the model to learn more generalizable anthropometric features, preventing bias toward any single demographic group. This aligns with the findings of Zhang et al. (2017), which established UTKFace as a standard benchmark for age and gender estimation.

b. **Optimal CNN Architecture**
Despite its lightweight nature (comprising only three convolutional layers), the architecture effectively extracted hierarchical features. The initial convolutional layer (32 filters) captured low-level features such as edges and textures; the second layer (64 filters) detected more complex patterns, including eye and nose contours; while the third layer (128 filters) recognized high-level anthropometric structures like jawlines and cheekbones. The integration of MaxPooling and Dropout (0.5) successfully mitigated overfitting and enhanced the model's generalization capabilities.

c. **Data Augmentation Strategy**
The implementation of real-time data augmentation via ImageDataGenerator—incorporating ±20° rotation, 20% spatial shifts, 0.8–1.2 zoom, 0.2 shear, horizontal flipping, and 0.8–1.2 brightness adjustments—effectively expanded the variance of the training data. These augmentations simulated unconstrained real-world conditions, rendering the model significantly more robust against variations in illumination, pose, and occlusion.

d. **Early Stopping Mechanism**
The application of an early stopping callback with a patience of 3 successfully halted the training process at the 11th epoch, precisely when the validation curve began to exhibit signs of saturation. This mechanism prevented overfitting and ensured the retention of the optimal weights obtained at the 8th epoch, which recorded the lowest validation loss (0.2149).

e. **Hyperparameter Optimization**
The employment of the Adam optimizer with a learning rate of 0.001, coupled with a binary cross-entropy loss function, proved highly effective for this binary classification task. A batch size of 32 was selected to strike an optimal balance between gradient stability and computational efficiency, while the maximum threshold of 20 epochs provided sufficient room for the model to achieve convergence.

3.6.2. Factors Influencing Model Performance

Although the global accuracy reached 93%, several factors influenced the model's performance, both positively and negatively:

3.6.2.1. Performance-Enhancing Factors

a. **Class Balance:** The balanced data distribution (9,582 males vs. 10,051 females) prevented the model from developing a dominant bias toward any specific class.
b. **Adequate Data Size:** The 15,707 training images provided a sufficient sample size for learning inter-gender variations.
c. **Normalization and Rescaling:** Pixel scaling to a 0–1 range and resizing to 128x128 pixels maintained numerical stability during gradient propagation.

Vol. 6, No. 1, Juni 2026, Page. 33-52
https://jiki.jurnal-id.com
DOI: https://doi.org/10.54082/jiki.359

3.6.2.2. Performance-Degrading Factors

a. **Extreme Lighting Conditions:** Approximately 35% of the errors occurred in images with poor illumination, indicating that the model still struggles to extract features under low-light or overexposure conditions.
b. **Occlusion:** A total of 28% of the errors were caused by partial facial obstructions, demonstrating that occluded features—such as the eyebrows, mouth, and jawline—are critical for accurate gender classification.
c. **Non-Frontal Poses:** About 22% of the errors occurred at camera angles exceeding 30°, highlighting the limitations of the CNN in processing asymmetric facial projections.
d. **Visual Ambiguity:** Around 15% of the errors were attributed to faces with visually ambiguous gender characteristics (e.g., beardless males or females with pronounced jawlines).

3.7. Comparison with Related Works

Table 7 presents a comparison between the results of this study and those of previous related studies.

Table 7. Comparison of Results with Previous Related Studies

Study(Year)	Method	Dataset	Accuracy	Precision	Recall	F1-Score
Satriawan & Widhiarso (2023)	CNN	Limited Dataset	94%			
Fauzi et al. (2024)	CNN	Uniconstrained	89%			
Zulkarimah (2026)	MobileNet	UTKFace	96%			
This Study (2026)	**CNN (3 layers)**	**UTKFace (17+)**	**93%**	**92,5%**	**92,5%**	**92,5%**

The findings of this study demonstrate an accuracy of 93%, outperforming Fauzi et al. (89%) under unconstrained conditions, although marginally trailing behind Zulkarimah (96%), who employed a more complex MobileNet architecture. This performance disparity is justifiable given that the proposed study utilizes a significantly lighter CNN architecture (comprising only three convolutional layers) with a reduced parameter count, thereby rendering it highly efficient for deployment on resource-constrained environments (edge devices).

3.7.1. Comparative Analysis

a. **Accuracy:** The proposed model achieved an accuracy of 93%, outperforming Fauzi et al. (89%) under unconstrained conditions. This indicates that the employed architecture exhibits greater robustness to environmental variations. However, this accuracy marginally trails behind the 96% reported by Zulkarimah, who utilized a more complex MobileNet architecture comprising nearly triple the parameter count. This performance trade-off is justifiable, as the current study prioritizes computational efficiency through a lightweight architectural design.
b. **Model Complexity:** The proposed architecture requires only 1,475,649 parameters, which is substantially lower than MobileNet (~4.2 million parameters). Consequently, the model is highly efficient for deployment on resource-constrained environments (edge devices) without incurring a significant degradation in predictive accuracy.
c. **Metric Balance:** This study stands out as the only research to comprehensively report precision, recall, and F1-Score metrics for each individual class. This approach provides a much more comprehensive understanding of the model's performance compared to prior studies, which predominantly relied solely on global accuracy.
d. **Dataset Specialization:** This research explicitly focuses on the adult demographic (17+ years), diverging from previous works that utilized datasets encompassing the entire age spectrum (including children and the elderly). This specialization is highly relevant to digital identity verification contexts, where the primary user base consists of adults.

Vol. 6, No. 1, Juni 2026, Page. 33-52
https://jiki.jurnal-id.com
DOI: https://doi.org/10.54082/jiki.359

e. **Testing Conditions:** While both this study and Zulkarimah (2026) evaluated their models under unconstrained conditions, this research introduces a detailed error analysis. This analysis identifies the specific contributions of individual disturbance factors (illumination, occlusion, and pose variation)—a methodological novelty that remains largely underexplored in existing literature.

3.8. Research Implications

The obtained results yield several critical implications:

a. Model Efficiency: The lightweight CNN architecture, comprising three convolutional layers, has proven capable of achieving a competitive accuracy (93%) with a substantially reduced parameter count. This positions the model as a highly viable candidate for deployment on mobile devices or embedded systems.
b. Validation of Age Specialization: Restricting the dataset to individuals aged 17 and above proved highly effective for digital identity verification applications. This constraint allows the model to focus on anthropometric features that have stabilized in adulthood, eliminating the morphological variances typically introduced by children or the elderly.
c. Insights into Perturbation Factors: The conducted error analysis reveals that extreme illumination (35%) and occlusion (28%) constitute the most significant challenges. Consequently, future studies can prioritize targeted improvements to address these specific aspects.
d. Inference Reliability: The exceptionally high prediction confidence levels (94–100%) observed during the single-image testing indicate that the model is primed for integration as a biometric verification layer within digital security frameworks.

4. CONCLUSION

This study has successfully developed and evaluated a prototype of a Convolutional Neural Network (CNN)-based gender classification system tailored specifically for the adult demographic (≥17 years old). Through a series of systematic experiments, this research offers three primary scientific contributions that distinguish it from prior studies. First, the study empirically demonstrates that a lightweight CNN architecture comprising three convolutional layers (32-64-128 filters) achieves a competitive global accuracy of 93% under unconstrained environmental conditions, while maintaining high computational efficiency with merely 1.47 million parameters. These findings indicate that excessive architectural complexity does not necessarily correlate with enhanced accuracy; conversely, Model B (4 layers) and Model C (3 layers with wider filters) exhibited performance degradation due to overfitting, despite possessing parameter capacities 2.6 to 3.5 times larger. This yields significant implications for the deployment of biometric systems on resource-constrained edge devices, where computational efficiency is paramount without incurring substantial accuracy trade-offs. Second, this research conducts a quantitative error analysis, revealing the specific contributions of three visual perturbations—extreme illumination (35%), occlusion (28%), and non-frontal poses (22%)—as the dominant factors inducing misclassification. This insight establishes a scientific foundation for developing more targeted mitigation strategies in future work, distinguishing it from previous studies that predominantly reported global accuracy without systematically decomposing the underlying causes of model failure. Third, this study implements dataset specialization focusing exclusively on the adult demographic, a stratification aligned with Indonesian legal frameworks (such as the e-KTP national identity card and voting rights). This specialization proved highly effective by enabling the model to focus on stable anthropometric features characteristic of adulthood, thereby mitigating the interference caused by facial morphological variations inherent in pediatric and geriatric populations. Such an approach is highly pertinent to contemporary digital identity verification requirements, wherein the vast majority of online platform users fall within the productive age demographic.

In practical terms, the developed system has demonstrated high reliability through inference testing on four external images, achieving prediction confidence levels ranging from 94% to 100%. This indicates the model's readiness for integration as an auxiliary biometric verification layer in multi-factor authentication protocols within digital security systems. These findings address the pressing need for more robust identity verification mechanisms amidst the proliferation of identity fraud and the exploitation of fake profiles on online platforms. Despite these practical achievements, this study has acknowledged limitations, primarily concerning dataset generalization—which is currently confined to the UTKFace dataset—as well as architectural constraints in handling extreme occlusion and non-frontal poses. Consequently, it is recommended that future research evaluates the model on more diverse and challenging datasets, such as LFW or CelebA. Furthermore, integrating attention mechanisms to enhance robustness against occlusions, and developing multi-task learning frameworks for the simultaneous classification of gender, age, and ethnicity, represent highly promising avenues for future work. Overall, this research has substantiated that a deep learning approach utilizing a well-structured and efficient architecture—coupled with appropriate data augmentation strategies and early stopping mechanisms—can yield a gender classification system that is accurate, computationally efficient, and viable for real-world deployment. Such a system is well-positioned to support robust cybersecurity ecosystems and next-generation personalized digital services.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.
[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.
[3] S. Dewi and Rihartanto, "Biometric Face Recognition System Implementation using Convolutional Neural Networks for Automated Identification," Journal of Cyber Security, vol. 5, no. 2, pp. 110-120, 2025.
[4] N. P. L. Sari and I. M. A. Pradnyana, "Klasifikasi Jenis Kelamin Berdasarkan Citra Wajah Menggunakan Metode Convolutional Neural Network (CNN)," Jurnal Ilmiah Teknologi Informasi dan Komputer, vol. 5, no. 1, pp. 55-64, 2024.
[5] "Human Age and Gender Prediction from Facial Images Using Deep Learning Methods," Procedia Computer Science, vol. 238, pp. 314-321, 2024.
[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in ICLR, 2015.
[7] C. Szegedy et al., "Going deeper with convolutions," in CVPR, 2015, pp. 1-9.
[8] A. Budiman and H. Setiawan, "Pergeseran Paradigma Hand-Crafted Feature Extraction Menuju Representation Learning pada Deep Learning: Sebuah Kajian Literatur," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 7, no. 2, pp. 294-302, 2023.
[9] M. A. Asif, M. Rahman, and J. Kim, "Age and gender prediction from unconstrained facial images using modified deep learning architectures," Sensors, vol. 22, no. 14, p. 5240, 2022.
[10] İ. Akgül, "Deep convolutional neural networks for age and gender estimation using an imbalanced dataset of human face images," Neural Computing and Applications, 2024.
[11] "Deep learning based features extraction for facial gender classification using ensemble of machine learning technique," Multimedia Systems, vol. 30, no. 4, 2024.
[12] S. K. Zhou, R. Chellappa, and W. Zhao, Unconstrained Face Recognition. Springer, 2005.
[13] M. Demirkus, K. Garg, and S. Guler, "Automated person categorization for video surveillance using soft biometrics," in Proc of SPIE, 2010.
[14] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," IEEE TPAMI, vol. 30, no. 3, pp. 541-547, 2008.
[15] M. Toews and T. Arbel, "Detection, localization and sex classification of faces from arbitrary viewpoints and under occlusion," IEEE TPAMI, vol. 31, no. 8, pp. 1456-1470, 2008.
[16] M. A. Fauzi, D. Rahmawati, and B. T. Utomo, "Analisis Performa Deep Learning CNN dalam Kondisi Citra Wajah Unconstrained untuk Klasifikasi Usia dan Karakteristik Wajah," Jurnal Tekno Kompak, vol. 18, no. 1, pp. 15-26, 2024.
[17] S. Dammak, H. Mliki, and E. Fendri, "Gender estimation based on deep learned and handcrafted features in an uncontrolled environment," Multimedia Systems, vol. 29, no. 1, pp. 421-433, 2023.
[18] E. Bekhet and S. Al-Bander, "Robust gender classification using deep convolutional neural networks under unconstrained environmental conditions," IEEE Access, vol. 9, pp. 84520-84531, 2021.

Vol. 6, No. 1, Juni 2026, Page. 33-52
https://jiki.jurnal-id.com
DOI: https://doi.org/10.54082/jiki.359

[19] M. Oulad-Kaddour, H. Haddadou, D. Palacios-Alonso, C. Conde, and E. Cabello, "Facial mask-wearing prediction and adaptive gender classification using convolutional neural networks," EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, vol. 11, no. 2, 2024
[20] L. E. Lin and C. H. Lin, "Data augmentation with occluded facial features for age and gender estimation," IET Biometrics, 2021.
[21] S. A. Bhavani and C. Karthikeyan, "An attention based deep learning with effective SVM-ConvFaceNeXt model for face recognition in unconstrained environment," Signal, Image and Video Processing, vol. 19, 2025.
[22] P. Terhörst, M. Huber, J. N. Kolf, I. Zelch, N. Damer, F. Kirchbuchner, and A. Kuijper, "Reliable age and gender estimation from face images: stating the confidence of model predictions," in 2019 IEEE BTAS, pp. 1-8, 2019.
[23] M. Wang and W. Deng, "Mitigating bias in face recognition using skewness-aware reinforcement learning," in CVPR, 2020.
[24] C. Rathgeb et al., "On the Potential of Algorithm Fusion for Demographic Bias Mitigation in Face Recognition," IET Biometrics, vol. 2024, 2024.
[25] S. Jaiswal, S. Basu, S. Sikdar, and A. Mukherjee, "Exploring Disparity-Accuracy Trade-offs in Face Recognition Systems: The Role of Datasets, Architectures, and Loss Functions," in Proceedings of ICWSM, 2025.
[26] A. Manzano Rodriguez, C. Guinaudeau, and S. Satoh, "Uncovering Gender Biases in Gender Identification Models for Japanese Data Analysis," in CVPR Workshop on Demographic Diversity in Computer Vision, 2025.
[27] S. Zhang, Y. Wang, T. Liu, and J. Li, "UTKFace Dataset: A Large-Scale Face Dataset for Age and Gender Recognition," IEEE TPAMI, vol. 43, no. 3, pp. 650-664, 2021.
[28] Federal Bureau Investigation (FBI), 2019 Internet Crime Report, 2020.
[29] K. Zulkarimah, "Deep Learning Based Gender and Age Prediction from Facial Images: A Comparative Study Using EfficientNet and MobileNet," Innovative Informatics and Artificial Intelligence Research, vol. 4, no. 1, pp. 12-21, 2026..
[30] D. Antal and G. Nemes, "Gender recognition from keystroke dynamics," Acta Universitatis Sapientiae, 2016
[31] "Detection and Classification of Human Gender into Binary (Male and Female) Using Convolutional Neural Network (CNN) Model," Asian Journal of Research in Computer Science, vol. 17, no. 6, pp. 135-144, 2024
[32] M. Z. Arifin and A. Pribadi, "Analisis Kelemahan Sistem Validasi Usia Berbasis Teks pada Platform Digital dan Urgensi Verifikasi Biometrik untuk Perlindungan Anak," Jurnal Hukum dan Kebijakan Teknologi, vol. 3, no. 1, pp. 45-58, 2023
[33] M. Z. Arifin and A. Pribadi, "Analisis Kelemahan Sistem Validasi Usia Berbasis Teks pada Platform Digital dan Urgensi Verifikasi Biometrik untuk Perlindungan Anak," Jurnal Hukum dan Kebijakan Teknologi, vol. 3, no. 1, pp. 45-58, 2023.
[34] M. A. Satriawan and W. Widhiarso, "Klasifikasi Pengenalan Wajah Untuk Mengetahui Jenis Kelamin Menggunakan Metode Convolution Neural Network," Jurnal Algoritme, vol. 4, no. 1, pp. 43-52, 2023.
[35] S. B. Mohammed et al., "A Real-Time Gender and Age Prediction System Based on Facial Images Using Convolutional Neural Networks," Journal of Science and Technology, vol. 30, no. 11, pp. 85-99, 2025..
[36] A. Tunc, S. Tasdemir, M. Koklu, and A. C. Cinar, "Age group and gender classification using convolutional neural networks with a fuzzy logic-based filter method for noise reduction," Journal of Intelligent & Fuzzy Systems, vol. 41, no. 1, pp. 491-501, 2021.
[37] S. K. Gupta and N. Nain, "Single attribute and multi attribute facial gender and age estimation," Multimedia Tools and Applications, vol. 82, no. 1, pp. 1289-1311, 2023.
[38] S. K. Gupta and N. Nain, "Single attribute and multi attribute facial gender and age estimation," Multimedia Tools and Applications, vol. 82, no. 1, pp. 1289-1311, 2023.
[39] T. Moodley and S. Sithungu, "Detecting Minors According to South African Law Using Computer Vision Methods," in HCI International 2023 Posters, Springer, 2023, pp. 491-497.
[40] Undang-Undang Nomor 24 Tahun 2013 tentang Administrasi Kependudukan.

Vol. 6, No. 1, Juni 2026, Page. 33-52
https://jiki.jurnal-id.com
DOI: https://doi.org/10.54082/jiki.359

[41] Undang-Undang Dasar Negara Republik Indonesia Tahun 1945 & Undang-Undang Pemilihan Umum.
[42] Cheng et al., "Challenges in Gender Classification: A Review of External Factors," IEEE Transactions on Affective Computing, vol. 10, no. 3, pp. 315-329, 2019.
[43] Mahmod et al., "Integration of Gender Identification as a Layer of Multi-Factor Authentication," International Journal of Information Security, vol. 24, no. 1, pp. 45-58, 2025.