Comparative analysis of gender classification methods using convolutional neural networks

Panca Dewi Pamungkasari[1], Ilhan Alim Asfandima[2], Achmad Pratama Rifai[3], Nguyen Huu Tho[4]
[1]Information Systems Study Program, Faculty of Computer Information Technology, Universitas Nasional, Jakarta, Indonesia
[2]Informatics Study Program, Faculty of Computer Information Technology, Universitas Nasional, Jakarta, Indonesia
[3]Department of Mechanical and Industrial Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia
[4]Faculty of Engineering and Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Oct 16, 2024
Revised Jun 19, 2025
Accepted Jul 10, 2025

Keywords:

Convolutional neural networks
Face recognition
Gender classification
Image processing
Preprocessing method

ABSTRACT

Gender classification has become an important application in the fields of system automation and artificial intelligence, having important implications across various fields. The main challenge in this classification task is variation in illumination that affects the quality of facial images. This study presents method for identifying genders with convolutional neural networks (CNNs). To address this issue, various preprocessing methods are applied, including self quotient image (SQI), locally tuned inverse sine nonlinear (LTISN), histogram equalization (HE), difference of gaussian (DoG), and gamma intensity correction (GIC), to stabilize the effects of illumination variations before the images are processed by CNN. The CNN architecture used consists of 5 convolutional blocks and 2 fully connected blocks, which have proven effective in image recognition. The results of study show that model trained with DoG method achieved accuracy of 91.07%, making it the best preprocessing technique compared to other methods such as SQI and HE, which achieved accuracy of 90.39% and 88.76%, respectively. These findings demonstrate that application of SQI in CNN can improve accuracy of gender classification on facial images, providing better performance than previous methods. These findings are expected to serve as foundation for further developments in facial image classification and its applications in various fields.

Corresponding Author:

Achmad Pratama Rifai
Department of Mechanical and Industrial Engineering, Faculty of Engineering, Universitas Gadjah Mada
St. Grafika No. 2, Yogyakarta 55284, Indonesia
Email: achmad.p.rifai@ugm.ac.id

1. INTRODUCTION

The human face is the first indicator used to identify a person's gender in everyday life. Gender recognition through facial analysis has various important applications in the field of security, such as verifying gender during registration or security surveillance in sensitive areas [1]. In marketing, data about the gender of customers can also be used to tailor marketing or commercial [2], software [3], or services to more precisely target the right audience. Additionally, it can be utilized in various types of data analysis to identify useful trends in fields such as sociology, psychology [4], or economics. The technology used for this task must be able to capture and analyze facial features, which can then be used to classify whether a person is male or female. Gender classification methods employ distinct characteristics, such as pattern recognition methods, to differentiate between masculine and feminine attributes [5]. Meanwhile, methods for identifying gender can be broadly categorized into feature-based [6] and appearance-based approaches [7], [8].

Classifying objects based on image data can be achieved using advanced machine technologies. The technology behind the ability of machines to recognize a person's face is machine learning. Machine learning is a branch of artificial intelligence, specifically focused on how computers can learn from data to improve their intelligence [9]. Machine learning plays a crucial role in classifying and detecting objects. One of the evolving fields is deep learning (DL). DL is a branch of machine learning closely related to artificial neural networks (ANN). Unlike traditional ANNs, DL involves deeper architecture composed of multiple layers, enabling the model to learn hierarchical representations of data through successive convolutions [10].

Various algorithms enable machines to classify image data, with the convolutional neural network (CNN) being one of the most widely used methods [11], [12]. CNN is another effective technique in the feature extraction process. CNN is a type of DL algorithm that processes input images by applying filters or kernels to identify and extract important features [13]. This technique can effectively extract hierarchical features from images, enabling more accurate and complex pattern recognition. The success of CNN in tasks such as object recognition and classification motivates this study to use CNN for gender classification on facial images. There are many factors that influence the process of face recognition, as mentioned in [14], which states that differences in illumination intensity-higher or lower in some areas-will affect the face recognition process, where these methods did not provide good results. As mentioned in [15], which states that because the illumination intensity of the environment differs, the face recognition rate is often low. In gender classification on facial images, image processing is closely related. There are many methods that can be used to enhance images. Raghavan and Ahmadi [16] proved that adding preprocessing techniques improves accuracy in face recognition. The best-performing technique in the study of Raghavan and Ahmadi [16] was SQI, which improved accuracy by 2.6% compared to results before using SQI.

The background of this study begins with the understanding that gender classification on facial images is an important topic in various fields, as it plays a crucial role in enhancing the accuracy of biometric identification, improving user-centric services, and enabling advanced analytics in sectors such as surveillance, healthcare, and digital marketing. Mustafa and Meehan [17] successfully applied CNN with an accuracy of 85%. In another study, Islam *et al.* [18] developed a face recognition model using transfer learning via Pareto frontier CNN. Benkaddour [19] also proposed CNN models for gender classification and age estimation. Before CNN, several studies proposed various methods for these problems, such as local binary patterns [20], bio-inspired features [21], support vector machine [22], and hierarchical classifier [23]. However, such methods are relatively out of favor in recent years due to the superiority of CNN in image classification, as indicated by [24], [25]. Therefore, this study opts to develop CNN models with various preprocessing methods for the gender classification problem.

The contribution of applying various preprocessing methods—self quotient image (SQI), locally tuned inverse sine nonlinear (LTISN), histogram equalization (HE), difference of Gaussian (DoG), and gamma intensity correction (GIC) is to enhance the quality and consistency of facial images, which is crucial for improving gender classification accuracy. Each method addresses specific challenges in image preprocessing: SQI reduces the effects of varying lighting conditions, HE enhances image contrast, LTISN minimizes non-uniform illumination, GIC adjusts intensity for better feature visibility, and DoG emphasizes edges and textures by reducing noise while retaining essential details. By systematically evaluating the performance of these methods, this study aims to identify the most effective approach or combination of techniques that can significantly improve the robustness and reliability of CNN-based gender classification models.

2. PREPROCESSING METHOD

Preprocessing techniques are used to enhance the quality of images so that the information contained in the images is easier to extract and analyze. These techniques include resizing, normalization, noise reduction, and contrast adjustment, which aim to standardize image data and minimize variations caused by lighting conditions, background clutter, or image resolution. Effective preprocessing is crucial for improving the accuracy and robustness of image classification and recognition systems. Preprocessing focuses on reducing noise, illumination variations, and low contrast in images.

2.1. Self quotient image

SQI is an illumination-invariant algorithm designed to address variations in lighting and shadows. It is computed by taking the ratio between the original image intensity and a smoothed version of the same image, as shown in (1).

$$Q(x,y) = \frac{I(x,y)}{S(x,y)} = \frac{I(x,y)}{F(x,y)*I(x,y)} \quad (1)$$

Where $I(x, y)$ is the face image and $S(x, y)$ is the smoothed version of the image, and * is the convolution operation. F is the kernel for smoothing the image, which in this case is a Gaussian filter, and Q is the result of the SQI calculation [16].

2.2. Histogram equalization

An image histogram is a graphical representation that shows how pixel intensity values are distributed across an entire image or a selected region of it. HE is an image enhancement method where the pixel histogram of the image becomes more spread out and uniform. Since the histogram represents the probability of pixels with certain gray levels, the formula for calculating HE is used in (2).

$$S_k = (L-1)\sum_{i=0}^{k} P_r(r_j) \tag{2}$$

Where the gray level kkk is normalized against the highest gray level (L-1). The value $r_j = 0$ represents black, and $r_j = 1$ represents white on a defined grayscale [26].

2.3. Locally tuned inverse sine non-linear

LTISN is a nonlinear approach that operates on each pixel. The corrected intensity value is calculated by applying an inverse sine function. This function uses adjustable parameters based on the surrounding pixel values, as given in (3) [16].

$$I_{enh}(x, y) = \frac{2}{\pi} sin^{-1}(I_n(x, y)^{\frac{q}{2}}) \tag{3}$$

2.4. Gamma intensity correction

Gray intensity correction (GIC) is a nonlinear gray-level transformation technique that adjusts the image's gray levels by replacing each original gray value with a corresponding GIC-adjusted gray level, as described in (4).

$$GIC(x, y) = I(x, y)^{1/\gamma} \tag{4}$$

For a gamma value less than 1.0, the image will become darker, and for a gamma value greater than 1.0, the image will become brighter. When the gamma value is 1.0, no effect is produced [16].

2.5. Difference of Gaussian

DoG is a grayscale image enhancement algorithm that involves subtracting the smoothed version of the original image from another version of the original image that is not as smoothly filtered. The smoothed images are obtained by convolving the grayscale image with a Gaussian filter kernel with different standard deviations, as given in (5) [16].

$$DOG(x, y) = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\pi\sigma_1^2}} - \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\pi\sigma_2^2}} \tag{5}$$

where σ_1 and σ_2 these are the widths of the Gaussian filter kernel.

3. SYSTEM DESIGN

This system consists of several stages to classify gender using the CNN method: collecting datasets from reliable sources, preprocessing the datasets to improve image quality and consistency, and training the data using the CNN model to recognize and predict gender with high accuracy.

3.1. Dataset

This study uses a dataset consisting of facial images of both men and women. The research utilizes two types of data: secondary data obtained from Kaggle and primary data, which was personally requested from the sources with permission. The dataset will be divided into three parts: Kaggle training data for model learning, Kaggle test data for initial performance evaluation, and primary test data for final validation to assess the model's accuracy and robustness on external data.

The dataset consists of both training and test data sourced from Kaggle and primary data specifically collected for this research that as seen in Figure 1. The training dataset includes 10,582 images-5,291 male and 5,291 female-featuring facial images that are not influenced by illumination conditions. The test dataset contains 2,646 images-1,323 male and 1,323 female-with facial images affected by illumination from various angles. Whereas the primary dataset also includes 95 images-45 male and 50 female-depicting facial images under different illumination conditions from multiple directions.

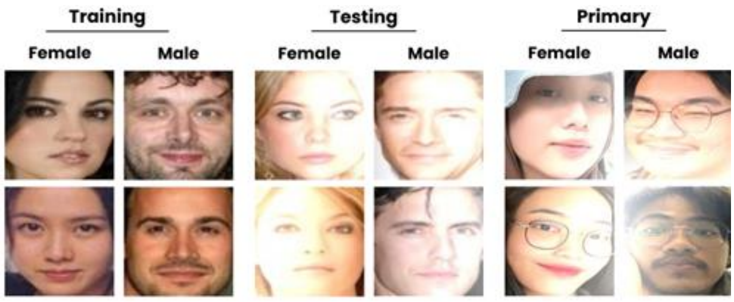

Figure 1. Example of training data

The test data features varying illumination conditions with the goal of evaluating the performance of the preprocessing methods. It allows us to determine which method is most effective in handling different illumination conditions. The distribution of facial images in the dataset across different categories is divided into three parts: training data, test data, and primary data, as seen in Table 1. The test data features varying illumination conditions with the goal of evaluating the performance of the preprocessing methods. It allows us to determine which method is most effective in handling different illumination conditions. The distribution of facial images in the dataset across different categories is divided into three parts: 5,291 images of the training data, 1,323 images of the test data, and the primary data consists of 50 images and 45 images of the females primary and males, respectively. All models are trained and evaluated using the same training, testing, and primary data to ensure fair comparison of the models' performance. In addition, here, the testing and primary data for evaluation are drawn from different subjects to assess the generalization of the trained models in recognizing new data.

Afterward, data augmentation is performed to artificially increase the size and diversity of a dataset by applying various transformations to the existing data. Data augmentation is especially valuable when there is limited data available. Generating augmented samples helps create a larger and more diverse dataset, which is critical for training DL models that typically require large datasets. Data augmentation that implemented geometric transformations (rotation, flipping, scaling, cropping), and photometric adjustments (brightness, contrast, noise addition). These techniques enhance variation, reducing overfitting and improving generalization. Data distribution is carefully maintained to prevent class imbalance; typically, each class receives an equal proportion of augmented samples. Ensuring a diverse yet balanced dataset allows the model to learn invariant facial features, enhancing recognition accuracy across different lighting conditions, angles, and occlusions in real-world applications. The number of images after data augmentation is presented in Table 2.

Table 1. Data distribution

Category	Training data	Test data	Primary data
Female	5,291	1,323	50
Male	5,291	1,323	45
Total	10,582	2,646	90

Table 2. Data distribution after augmentation

Category	Training data	Test data	Primary data
Female	8,142	1,985	835
Male	8,142	1,985	825
Total	16,284	3,970	1,660

3.2. Apply preprocessing methods to images

In the preprocessing stage, images are resized to 96×96 pixels to ensure uniform input dimensions, which is crucial for consistent model performance. After resizing, normalization is applied to scale the pixel values to a [0, 1] range, standardizing the data and enhancing the learning process. Illumination conditions can significantly impact image quality, so handling illumination effectively during preprocessing is vital. Various methods, including grayscale conversion, SQI, HE, LTSIN, GIC, and DoG, are employed to address these variations. The effects of these illumination adjustment techniques on preprocessing are illustrated in Figure 2, where different methods demonstrate their effectiveness in creating more consistent and reliable data for model training. The SQI and DoG methods effectively reduce illumination conditions. However, the DoG method primarily sharpens edges, which can lead to a loss of information. HE and GIC also help in minimizing illumination variations, but both methods tend to darken areas that are not affected by light.

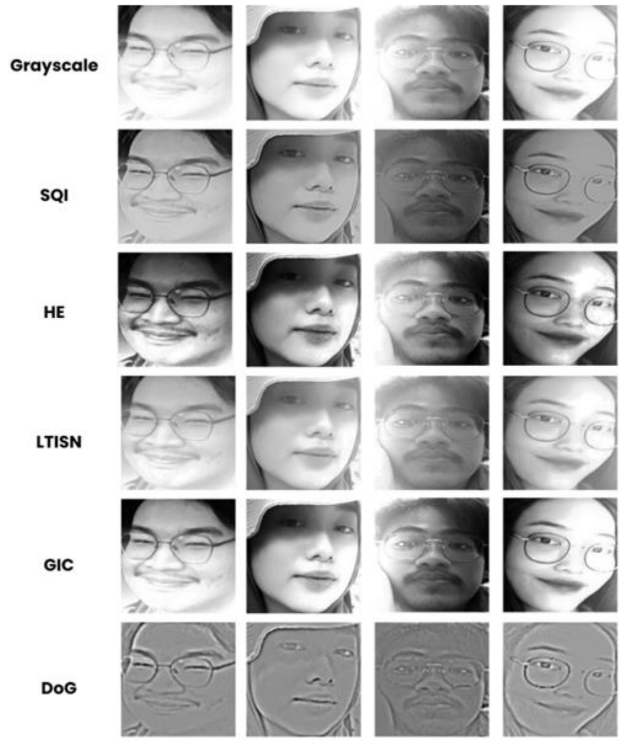

Figure 2. Comparison of preprocessing methods

3.3. Convolutional neural networks

A CNN is a type of ANN designed for image recognition. CNNs automatically learn and extract features from input images and integrate these features with a classification mechanism. One of the key advantages of CNN classifiers is their relatively simple architecture compared to other methods, with a clear sequence of layers that transform input data into output predictions [27]. Typically, CNNs operate in two main stages: feature extraction and classification. The feature extraction phase derives meaningful features from the original dataset, which helps reduce computational resources while preserving critical information [28]. In the classification phase, fully connected layers act as the classifier, assigning probabilities to predict the object present in the image based on the extracted features [29].

The CNN model architecture, consists of 5 convolutional blocks and 2 fully connected layers. The input layer has a size of 96×96 pixels with a color channel of 1, resulting in an input layer composition of (96, 96, 1). The image then undergoes feature learning through convolutional operations performed by the 5 convolutional blocks. Subsequently, the image will be flattened, converting the results of feature learning into a vector. The final stage involves classification using the fully connected layers. Figure 3 illustrates the model architecture used in this study.

As shown in Figure 3, there are 5 convolutional blocks and 2 fully connected blocks. The convolutional layer is the earliest layer in a CNN. The parameters in the convolutional layer are determined by the number of kernels used in the convolutional operation. The operation is performed on the input to produce the neurons output [to enhance -30]. This is followed by pooling to extract features from the input image sequentially. The pooling layer, also known as subsampling or downsampling, reduces the spatial dimensions of each feature map while preserving the most important information [30]–[33]. The purpose of using pooling is to reduce the number of features from the convolutional layer's output or feature map, thereby speeding up the model's training process. After passing through all the layers, the data is forwarded to the fully connected/dense layer. In this layer, each neuron is linked to every activation from the preceding layer. When these fully connected layers are combined with a SoftMax function, they form a multi-layer perceptron (MLP), which serves as the classifier in the network [14].

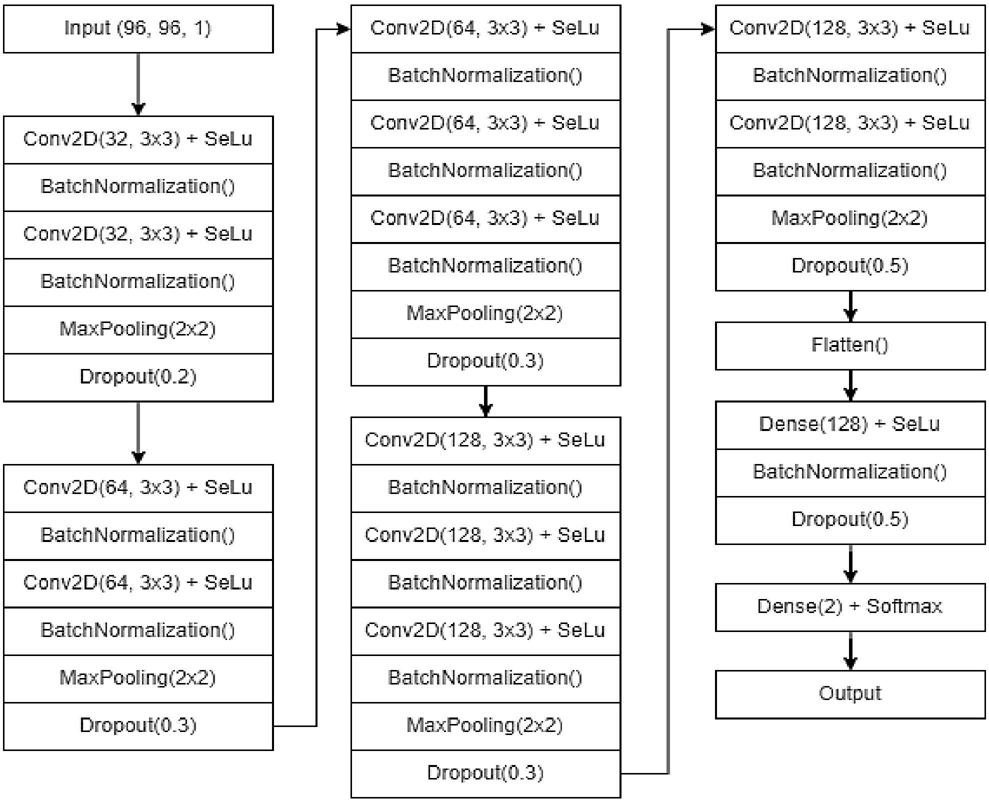

Figure 3. CNN model architecture

In the CNN architecture, each layer includes the scaled exponential linear unit (SeLU) activation function to enable high-level abstract representations. SeLU has the property of self-normalization, which automatically drives the output toward a mean of zero and varying units [34]–[36]. This helps maintain gradient stability during training and accelerates convergence. The SeLU equation can be seen in (6).

$$SeLU(x) = \lambda \begin{cases} x & for\ x \geq 0 \\ \alpha(e^x - 1) & for\ x < 0 \end{cases} \quad (6)$$

Additionally, each convolutional layer incorporates batch normalization and dropout. Batch normalization helps minimize internal covariate shift by normalizing the outputs of the previous activation layer—this is done by subtracting the batch mean and dividing by the batch standard deviation, which improves the network's stability and speeds up training [37]. The dropout technique is introduced to mitigate overfitting by temporarily deactivating a subset of neurons during training, based on a specified probability $p\ (where\ 0 < p < 1)$ [38].

The experimental conditions are outlined in Table 3. The table shows that the batch normalization size is set to 64, ensuring stable training. Optimization is performed using the Adam optimizer with a learning rate of 0.001, providing efficient convergence. The output layer employs the SoftMax function, suitable for classification tasks, while the hidden layers use SeLU for self-normalization. Additionally, 2×2 pooling is applied to reduce spatial dimensions, and input data is rescaled by 1/255, normalizing pixel values to a [0, 1] range for consistent and effective training.

Table 3. Model specifications

Specification	Value
Optimizer	Adam
Learning rate	0.0001
Batch size	64
Epoch	40
Activation function (hidden)	SeLU
Activation function (output)	SoftMax
Pooling	2×2
Rescale	1/255

4. RESULTS AND DISCUSSION

At this stage, the training of the model will be conducted using secondary data, with the training data 10,582 samples and the test data 2,098 samples with preprocessing method. The model will be trained according to the specifications model. During training, the model will evaluate the test data and produce accuracy metrics at each epoch. Once all models are trained, the evaluation will be performed on primary data to determine the best preprocessing method for gender classification.

4.1. Training results

The CNN models are trained in 40 epochs. The obtained accuracy and loss of each epoch during training progress of the models are shown in Figure 4. For the grayscale model, the accuracy graph in Figure 4(a) shows a consistent upward trend with minor ripples, indicating periods of slower gains. The loss graph similarly reflects a general decline with occasional fluctuations, suggesting the model's ongoing adjustments in minimizing errors. The SQI test data results in a smoother curve compared to the grayscale model as visualized in Figure 4(b), indicating that the model was trained adequately to handle variations in illumination.

Meanwhile, the results from the HE models in Figure 4(c) reveal a wavy and uneven curve, suggesting that the model was not effectively trained to handle variations in illumination. This irregularity indicates that the model struggles with generalizing across different illumination conditions. Despite these issues, the model was evaluated with primary data and achieved an accuracy of 88.76%. While this accuracy is relatively strong, the observed curve characteristics highlight limitations in the model's training, pointing to potential improvements needed in handling diverse illumination scenarios to enhance overall performance and consistency. Figure 4(d) illustrates the LTISN model, which displays a relatively stable accuracy trend, though minor fluctuations are present in later epochs. This indicates that the model was moderately successful in adapting to variations in illumination, showing better consistency compared to the HE model but still leaving room for improvement.

Similar to the HE models, the test data curve of GIC model produces a jagged and uneven curve, as shown in Figure 4(e), with accuracy exhibiting an unstable decline after epoch 25. This suggests that the model has not been adequately trained to manage variations in illumination. Finally, the test data curve of DoG model shown in Figure 4(f) produces a relatively smooth curve, with a minor accuracy drop observed after epoch 34. This suggests that the model is adequately trained to manage variations in illumination. Evaluated with primary data, the model achieved an accuracy of 91.07%. This performance indicates that the model is effectively handling the challenges posed by different illumination conditions, demonstrating its robustness and reliability in gender classification tasks. The smooth curve and high accuracy underscore the model's proficiency in adapting to variations and ensuring consistent results.

4.2. Test results

The test results are further presented in the confusion matrix in Figure 5. The grayscale model for gender classification demonstrates an overall accuracy of 87%. As shown in Figure 5(a), the model shows high precision for females (97%), but a lower recall (76%), indicating a tendency to misclassify females as males. In contrast, male classifications exhibit a recall of 97% but a lower precision (80%), suggesting occasional misclassification of males as females. The F1-scores for females and males are 0.85 and 0.88, respectively, reflecting a good balance between precision and recall. While the model performs well overall, improving recall for female classifications could enhance its effectiveness.

The SQI model achieves an accuracy of 90% for gender classification with the confusion matrix shown in Figure 5(b). Precision and recall for both female and male classes are balanced, with values of 0.91 and 0.89 for females, and 0.90 and 0.91 for males, respectively. The F1-scores for both classes are 0.90. These results indicate that the SQI preprocessing technique effectively enhances classification performance under varying illumination conditions. Similarly, the HE model achieves an accuracy of 89% for gender classification. The precision and recall for the female class are 0.88 and 0.90, while for the male class they are 0.90 and 0.87, respectively. Both classes have an F1-score of 0.89. These results indicate that the HE preprocessing technique also provides balanced performance as visualized in Figure 5(c), though slightly lower than the SQI method, with effective handling of illumination variations.

The LTISN model achieves an accuracy of 82%. Figure 5(d) shows high precision for females (0.97) but lower recall (0.66), while males have a lower precision (0.74) but a high recall (0.98). The F1-scores are 0.78 for females and 0.84 for males. These results indicate an imbalance in the model's performance, particularly in misclassifying a significant number of females as males, which affects overall accuracy. The GIC model achieves an accuracy of 86%. The confusion matrix of GIC model in Figure 5(e) shows strong precision for females (0.92) but lower recall (0.80), while males have a lower precision (0.82) but a higher recall (0.93). The F1-scores are 0.85 for females and 0.87 for males. These results indicate balanced overall performance, with a slightly higher ability to correctly classify males.

The DoG model achieves an accuracy of 91%. Figure 5(f) shows that DoG model has high precision (0.87) and recall (0.96) for females, while males class demonstrate higher precision (0.96) but lower recall (0.86). The F1-scores are 0.92 for females and 0.91 for males. These results indicate strong overall performance, with the model showing better recall for females and higher precision for males, leading to balanced classification outcomes.

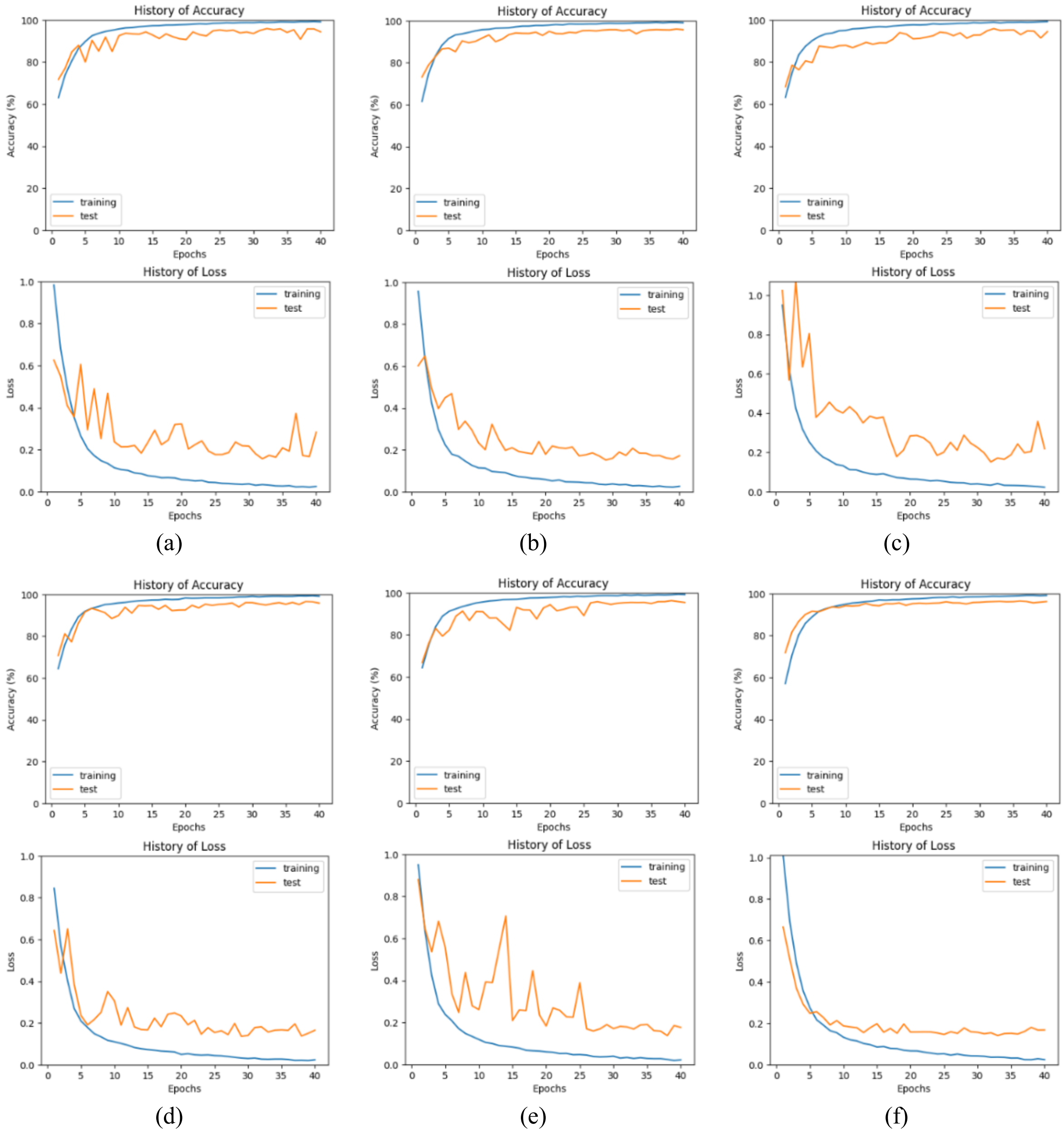

Figure 4. Training progress of (a) grayscale model, (b) SQI model, (c) HE model, (d) LTISN model, (e) GIC model, and (f) DoG model

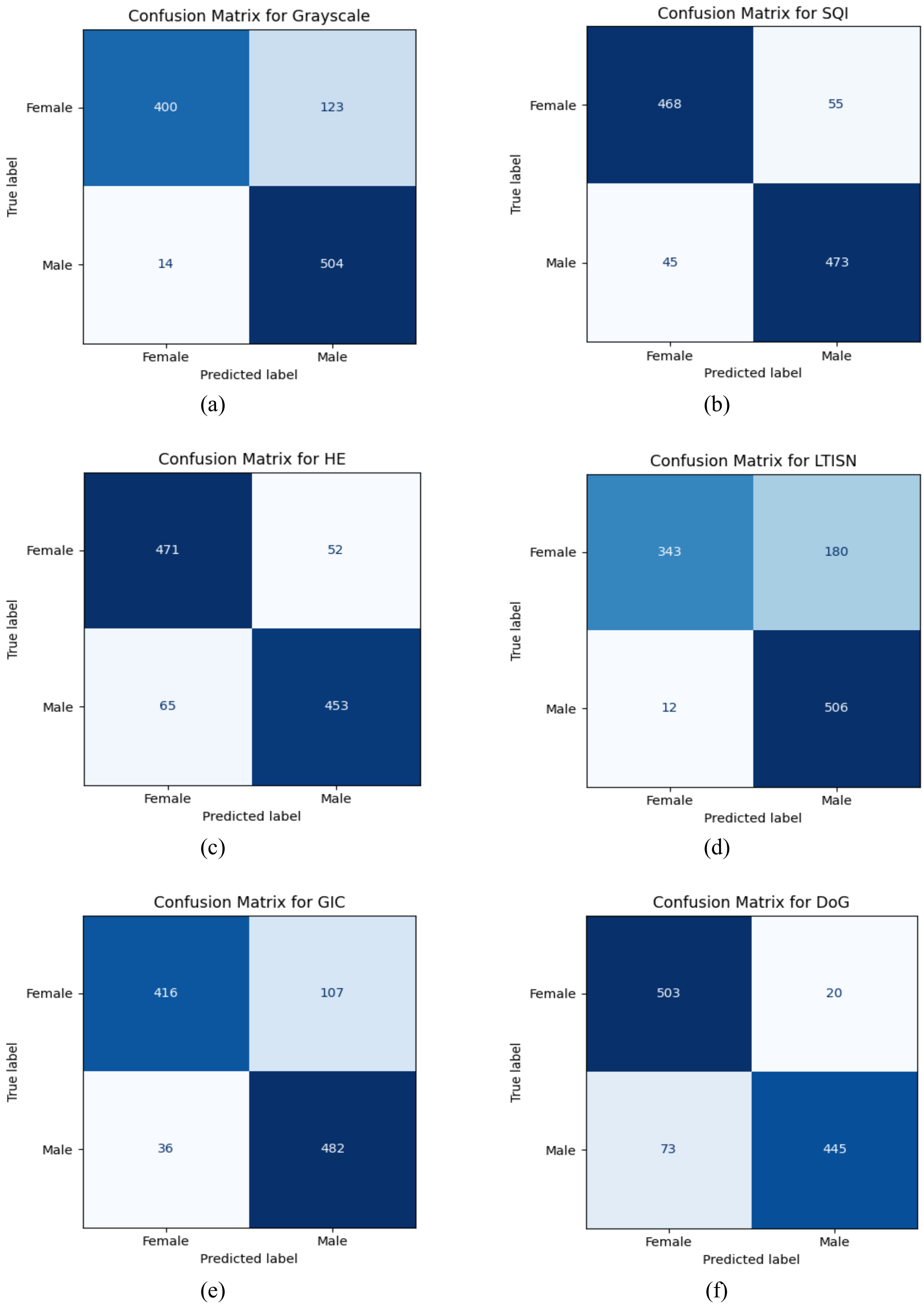

Figure 5. Confusion matrix of all models: (a) grayscale model, (b) SQI model, (c) HE model, (d) LTISN model, (e) GIC model, and (f) DoG model

4.3. Model comparison

Table 4 is provided to simplify the comparison of performance metrics from various preprocessing method applied, allowing a clearer evaluation of their impact on the model's accuracy and efficiency. This comparison helps identify which method delivers the most optimal result for the given dataset. The comparison is based on key performance metrics such as accuracy, precision, recall, and F1-score for both female and male classifications. Accuracy is a critical metric, reflecting the overall correctness of the model. The DoG model achieved the highest accuracy of 91%, closely followed by SQI with 90%, and HE with 89%. These results suggest that the DoG and SQI methods are more effective in improving the model's ability to correctly classify both genders across various illumination conditions. On the lower end, LTISN produced an accuracy of 82%, indicating less reliable overall performance, especially in correctly identifying female images.

Table 4. Model performance metrics

Model	Accuracy (%)	Loss	Precision (Female)	Recall (Female)	F1 score (Female)	Precision (Male)	Recall (Male)	F1 score (Male)
Grayscale	86.84	0.3088	0.97	0.76	0.85	0.80	0.97	0.88
SQI	90.39	0.1085	0.91	0.89	0.90	0.90	0.91	0.90
HE	88.76	0.3330	0.88	0.90	0.89	0.90	0.87	0.89
LTISN	81.56	0.1937	0.97	0.66	0.78	0.74	0.98	0.84
GIC	86.26	0.3830	0.92	0.80	0.85	0.82	0.93	0.87
DoG	91.07	0.2504	0.87	0.96	0.92	0.96	0.86	0.91

Precision measures the percentage of correctly identified instances out of all instances predicted as a specific gender. The LTISN model demonstrated the highest precision for the female class (0.97), but this is offset by lower recall values, indicating that while it is highly accurate in the predictions it makes, it fails to detect many true positives. The DoG model, while having a slightly lower precision for females (0.87), maintains a much more balanced performance with high precision for males (0.96), indicating fewer false positives across both classes. SQI and HE also show balanced precision across both classes, with values around 0.90, demonstrating reliable performance in correctly identifying both genders.

Recall, which reflects the ability of the model to correctly identify all instances of a particular class, is where the DoG model stands out, especially for females, with a recall of 0.96. This high recall for females means the DoG model is highly effective in identifying almost all female images, with minimal false negatives. In contrast, LTISN, despite its high precision for females, has a significantly lower recall (0.66), meaning it fails to correctly classify a substantial portion of female images. The GIC model exhibits a more balanced recall between both genders, with values of 0.80 for females and 0.93 for males, reflecting a reliable ability to capture most instances of each gender, though slightly biased toward male classifications.

The F1-score balances precision and recall, offering a more comprehensive view of model performance. DoG and SQI again demonstrate superior results, with F1-scores of 0.92 for females and 0.91 for males in the DoG model and balanced 0.90 scores for both genders in the SQI model. These scores highlight their ability to maintain a strong trade-off between precision and recall, making them the most balanced models in this comparison. In contrast, LTISN, despite its high precision for females, suffers from a lower F1-score (0.78 for females) due to its weak recall, limiting its overall effectiveness.

Each model presents a different set of strengths and weaknesses. DoG emerges as the best overall model, achieving high accuracy, balanced precision, and recall, along with strong F1-scores for both genders. This balance suggests that DoG is highly effective in real-world scenarios where illumination conditions can vary significantly. SQI also performs well across all metrics, making it another strong candidate for reliable gender classification. LTISN, while excelling in precision for females, struggles with recall, leading to many misclassifications in this class. Its high precision but low recall for females suggests that while it makes accurate predictions for the images it classifies as female, it fails to identify a significant portion of actual female images. This imbalance makes it less suitable for applications where it is critical to correctly classify all female instances. HE and GIC offer more balanced performance than LTISN, but their accuracy and F1-scores, while competitive, fall slightly short of the top-performing models, DoG and SQI. These models might be considered reliable but less robust when dealing with highly varied or challenging image datasets.

In conclusion, the DoG and SQI models demonstrate superior performance across key metrics, particularly in balancing precision, recall, and F1-scores for both genders. DoG slightly edges out SQI with its higher recall for females and overall accuracy. These models are thus well-suited for applications that require high accuracy in gender classification under varying illumination conditions. While LTISN and GIC exhibit strong performance in specific areas, they are less reliable due to imbalances between precision and recall, making them less effective for general use. Future improvements might focus on refining models like LTISN to improve recall, particularly for the female class, without sacrificing the high precision it achieves.

5. CONCLUSION

The implementation of preprocessing methods such as grayscale, SQI, HE, LTSIN, GIC, and DoG is carried out. These preprocessing methods are applied to handle illumination variations in facial image classification, and a comparison is made to identify the best preprocessing methods for handling illumination variations in images. The best preprocessing method is DoG, achieving an accuracy of 91.07%, followed by the SQI model with an accuracy of 90.39% and the HE model with an accuracy of 88.76%. The grayscale model achieved an accuracy of 86.84%, the GIC model achieved an accuracy of 86.26%, and the LTISN model achieved an accuracy of 81.56%. The results of this training demonstrate that the DoG preprocessing technique has a significant advantage in handling various illumination conditions. The preprocessing techniques explored in this study have not fully optimized their parameters. For instance, SQI and DoG might see improvements by testing various Gaussian kernels, while LTSIN could be enhanced by tuning the parameters of the inverse sinus function. Additionally, GIC could be refined by experimenting with different gamma values. Such adjustments would enable each preprocessing method to achieve better parameter settings for gender classification in facial image systems handling varying illumination conditions.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Panca Dewi Pamungkasari	✓	✓		✓	✓	✓	✓		✓	✓		✓		
Ilhan Alim Asfandima		✓	✓		✓	✓		✓	✓		✓			
Achmad Pratama Rifai	✓	✓		✓	✓	✓	✓		✓	✓			✓	
Nguyen Huu Tho				✓	✓	✓	✓			✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] M. Demirkus, K. Garg, and S. Guler, "Automated person categorization for video surveillance using soft biometrics," *Biometric Technology for Human Identification VII*, vol. 7667, 2010, doi: 10.1117/12.851424.

[2] R. Maldonado, P. Tansuhaj, and D. D. Muehling, "The impact of gender on ad processing: a social identity perspective," *Academy of Marketing Science Review*, vol. 2003, no. 3, pp. 1–15, 2003.

[3] L. Beckwith, M. Burnett, S. Wiedenbeck, and V. Grigoreanu, "Gender HCI: results on date regarding issues in problem-solving software," in *AVI 2006 Gender and Interaction: Real and Virtual Women in a Male World Workshop paper*, 2006, pp. 1–4.

[4] K. Delac and M. Grgic, "A survey of biometric recognition methods," in *Proceedings Elmar - International Symposium Electronics in Marine*, 2004, pp. 184–193, doi: 10.1109/ELMAR.2004.1356372.

[5] O. Singh, G. Bommagani, V. K. Gunjan, and S. R. Ravula, "Pattern based gender classification," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 10, pp. 888–895, 2013.

[6] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004, doi: 10.1109/TCSVT.2003.818349.

[7] Y. Saatci and C. Town, "Cascaded classification of gender and facial expression using active appearance models," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006, pp. 393–400, doi: 10.1109/FGR.2006.29.

[8] E. Mäkinen and R. Raisamo, "An experimental comparison of gender classification methods," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1544–1556, 2008, doi: 10.1016/j.patrec.2008.03.016.

[9] T. Wahyono, *Python for machine learning*. Yogyakarta, Indonesia: Gava Media, 2021.

[10] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: a survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018, doi: 10.1016/j.compag.2018.02.016.

[11] M. Hussain, J. J. Bird, and D. R. Faria, "A study on CNN transfer learning for image classification," *Advances in Intelligent Systems and Computing*, vol. 840, pp. 191–202, 2018, doi: 10.1007/978-3-319-97982-3_16.

[12] M. Coskun, A. Ucar, O. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in *2017 International Conference on Modern Electrical and Energy Systems (MEES)*, 2017, pp. 376–379, doi: 10.1109/MEES.2017.8248937.

[13] R. Chauhan, K. K. Ghanshala, and R. Joshi, "Convolutional neural network (CNN) for image detection and recognition," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 2018, pp. 278–282, doi: 10.1109/ICSCCC.2018.8703316.

[14] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep convolutional neural networks architectures," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 2, pp. 1015–1027, 2020, doi: 10.11591/ijeecs.v18.i2.pp1015-1027.

[15] Y. Yang and X. Song, "Research on face intelligent perception technology impact deep learning under different illumination intensities," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 1, pp. 32–36, 2022, doi: 10.47852/bonviewJCCE19919.

[16] J. Raghavan and M. Ahmadi, "Preprocessing techniques to improve CNN based face recognition system," in *Computer Science & Information Technology (CS & IT)*, 2021, pp. 1–20, doi: 10.5121/csit.2021.110101.

[17] A. Mustafa and K. Meehan, "Gender classification and age prediction using CNN and ResNet in real-time," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, 2020, pp. 1–6, doi: 10.1109/ICDABI51230.2020.9325696.

[18] M. M. Islam, N. Tasnim, and J.-H. Baek, "Human gender classification using transfer learning via Pareto frontier CNN networks," *Inventions*, vol. 5, no. 2, 2020, doi: 10.3390/inventions5020016.

[19] M. K. Benkaddour, "CNN based feature extraction for age estimation and gender classification," *Informatica*, vol. 45, no. 5, 2021, doi: 10.31449/inf.v45i5.3262.

[20] A. Gunay and V. V. Nabiyev, "Automatic age classification with LBP," in *2008 23rd International Symposium on Computer and Information Sciences*, 2008, pp. 1–4, doi: 10.1109/ISCIS.2008.4717926.

[21] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 112–119, doi: 10.1109/CVPR.2009.5206681.

[22] M. A. B. -Nia and Z. Mousavi, "A new classification method based on pairwise support vector machine (SVM) for facial age estimation," *Journal of Industrial and Systems Engineering*, vol. 10, no. 1, pp. 91–107, 2017.

[23] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognition*, vol. 44, no. 6, pp. 1262–1281, 2011, doi: 10.1016/j.patcog.2010.12.005.

[24] B. Saiharsha, A. Lesle, B. Diwakar, R. Karthika, and M. Ganesan, "Evaluating performance of deep learning architectures for image classification," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, Jun. 2020, pp. 917–922, doi: 10.1109/ICCES48766.2020.9137884.

[25] O. Agbo-Ajala and S. Viriri, "Deeply learned classifiers for age and gender predictions of unfiltered faces," *The Scientific World Journal*, vol. 2020, pp. 1–12, 2020, doi: 10.1155/2020/1289408.

[26] H. Yeganeh, A. Ziaei, and A. Rezaie, "A novel approach for contrast enhancement based on histogram equalization," in *2008 International Conference on Computer and Communication Engineering*, 2008, pp. 256–260, doi: 10.1109/ICCCE.2008.4580607.

[27] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature extraction using convolution neural networks (CNN) and deep learning," in *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2018, pp. 2319–2323, doi: 10.1109/RTEICT42901.2018.9012507.

[28] S. S. Abdullah and A. M. Abdulazeez, "Facial expression recognition based on deep learning convolution neural network: a review," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 53–65, 2021.

[29] M. K. Benkaddour, S. Lahlali, and M. Trabelsi, "Human age and gender classification using convolutional neural network," in *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, 2021, pp. 215–220, doi: 10.1109/IHSH51661.2021.9378708.

[30] Z. Ma *et al.*, "Fine-grained vehicle classification with channel max pooling modified CNNs," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3224–3233, 2019, doi: 10.1109/TVT.2019.2899972.

[31] A. Zafar *et al.*, "A comparison of pooling methods for convolutional neural networks," *Applied Sciences*, vol. 12, no. 17, 2022, doi: 10.3390/app12178643.

[32] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," *arXiv-Computer Science*, pp. 1–16, 2020.

[33] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Artificial Neural Networks – ICANN 2010*, Berlin, Heidelberg: Springer, 2010, pp. 92–101, doi: 10.1007/978-3-642-15825-4_10.

[34] H. Hu, "vReLU activation functions for artificial neural networks," in *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2018, pp. 856–860, doi: 10.1109/FSKD.2018.8687140.

[35] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," *arXiv-Computer Science*, pp. 1–14, 2016.

[36] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," *arXiv-Computer Science*, pp. 1–10, 2016.

[37] P. R. Tupe, P. M. Vibhute, and M. A. Sayyad, "An architecture in combining convolutional neural network (CNN) with batch normalization for apparel image classification," in *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, 2020, pp. 1–6, doi: 10.1109/iSSSC50941.2020.9358882.

[38] Y. Cui, T. Pan, S. Chen, and X. Zou, "A gender classification method for Chinese mitten crab using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 79, no. 11–12, pp. 7669–7684, 2020, doi: 10.1007/s11042-019-08355-w.

BIOGRAPHIES OF AUTHORS

Panca Dewi Pamungkasari has been a lecturer at the Faculty of Technology and Computer Science, Universitas Nasional, since 2008. She earned her bachelor's degree in electrical engineering from Universitas Kristen Satya Wacana in 2005. Then, she pursued her master's degree in electrical engineering at Universitas Indonesia, and in 2017, completed her Ph.D. in Telecommunications from Keio University, Japan. After earning her doctorate, she became a cooperative researcher at the Josaphat Microwave Remote Sensing Laboratory (JMRSL) at Chiba University from 2017 to 2018. Her research at JMRSL focused on advanced remote sensing technologies. From 2018 to 2023, she worked as a researcher at Professor Shinichi Shirakawa's laboratory at Yokohama National University, concentrating on deep learning applications in image biomedical. Throughout her career, she has published extensively in international journals, contributing to the fields of telecommunications, biomedical, image processing, and deep learning. Her work continues to influence the academic and research communities in these areas. She can be contacted at email: panca.dewi79@gmail.com.

Ilhan Alim Asfandima is a fresh graduate from Universitas Nasional, where he earned his bachelor's degree in computer science (S.Kom.) in 2024. With a strong academic background in information technology, he has developed a solid understanding of data analysis and big data technologies. Currently, he works as a data quality at Solusi247 IT Solution Big Data, where he is responsible for ensuring the accuracy and integrity of data in large-scale projects. His expertise in data management and deep knowledge of big data make him a skilled and valuable asset in the IT industry. He can be contacted at email: ilhanalim2020@student.unas.ac.id.

Dr. Achmad Pratama Rifai holds a Ph.D. degree in integrated design engineering from Keio University, Tokyo, Japan, in 2019. He also received the M.Eng. degree in manufacturing engineering from the University of Malaya, Kuala Lumpur, Malaysia, in 2014, and the Bachelor of Engineering degree in industrial engineering from Universitas Gadjah Mada, Indonesia, in 2011. He is currently an assistant professor with the Department of Mechanical and Industrial Engineering, Universitas Gadjah Mada. In addition, he is serving as the Head of the Manufacturing Process and System Laboratory of Universitas Gadjah Mada (2022-presents). His current research interests include optimization, operations research, metaheuristics, artificial intelligence, computer vision, and manufacturing systems. He was involved in some research projects in Indonesia, Malaysia, and Japan in both manufacturing systems, optimization, and computer vision domains. He has published several articles and is active as a reviewer in reputable journals indexed in Web of Science and Scopus in the field of computing, operations research, and manufacturing. He can be contacted at email: achmad.p.rifai@ugm.ac.id.

Nguyen Huu Tho received his Ph.D. degree at the University of Malaya in 2016 under a JICA AUN/SEED-Net Ph.D. sandwich scholarship between University of Malaya and Keio University. He was a lecturer at Ho Chi Minh City University of Technology, Vietnam during the period of 2009-2017 and a lecturer at Ho Chi Minh City University of Food Industry from 2018-2021. Currently, he is a lecturer at Nguyen Tat Thanh University (NTTU), Ho Chi Minh City, Vietnam. His research interests include manufacturing systems and logistics engineering technology, material handling systems, MCDA, DOE, and intelligent manufacturing. He can be contacted at email: nhtho@ntt.edu.vn.