Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 1, February 2026, pp. 438~449
ISSN: 2302-9285, DOI: 10.11591/eei.v15i1.9936

Comparative evaluation of AlexNet, SqueezeNet, VGG16, and ResNet50 for gender and hijab detection

Aji Supriyanto[1], Theresia Dwiati Wismarini[2], Herny Februariyanti[3], Arief Jananto[3], Fitri Damaryanti[1], Hilmy Nurakmal Satria[4]
[1]Department of Information Technology, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia
[2]Department of Informatics Engineering, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia
[3]Department of Information Systems, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia
[4]Department of Computer Engineering, Faculty of Engineering, Diponegoro University, Semarang, Indonesia

Article Info

Article history:

Received Jan 23, 2025
Revised Sep 3, 2025
Accepted Sep 27, 2025

Keywords:

AlexNet
Facial images
Gender detection
Hijab features
ResNet50
SqueezeNet
VGG16

ABSTRACT

This study aims to detect gender based on facial images with and without hijab features, with the expected outcome of distinguishing gender from these facial features. The method involves comparing the performance of four convolutional neural network (CNN) architectures: AlexNet, SqueezeNet, VGG16, and ResNet50. A total of 170 facial images were directly collected using smartphone cameras. The dataset consists of two classes: 68 male faces and 102 female faces, among which 78 images of females feature hijabs, while 24 do not. The validation stage with 40 images (15 males and 25 females) showed that the AlexNet architecture achieved the highest validation accuracy at 100%, followed by ResNet50 with 97.50%, VGG16 with 95%, and SqueezeNet with 92.50%. The testing stage with 40 images (20 males and 20 females, including 10 females with hijabs and 10 without) showed that ResNet50 classified 38 images correctly, achieving 95% accuracy. AlexNet classified 37 images correctly with 92.50% accuracy, SqueezeNet classified 36 images correctly with 90% accuracy, and VGG16 classified 34 images correctly with 85% accuracy. The contribution of this research shows that AlexNet achieves the highest validation accuracy, while ResNet50 provides the best accuracy in facial image detection for determining gender and hijab features.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Aji Supriyanto
Department of Information Technology, Faculty of Information Technology and Industry
Universitas Stikubank
St. Tri Lomba Juang No.1 Semarang, Indonesia
Email: ajisup@edu.unisbank.ac.id

1. INTRODUCTION

Various convolutional neural network (CNN) architectures can be employed to detect human faces. Human face detection is one of the most extensively researched areas in computer vision [1]. However, selecting the wrong architecture can lead to various issues, primarily inadequate face recognition accuracy [2]. Many current architectures are ineffective in handling partially covered facial objects such as masks, hijabs, or niqabs, and gender [1], [3], [4]. In some cases, facial changes due to aging [1], [5], [6], the use of hats and scarves [7], and other issues like lighting, occlusion, pose variations, and facial expressions pose additional challenges [2], [8], [9].

CNN-based facial recognition systems allow models to automatically extract facial features from images, ranging from low-level features like edges and corners to high-level features like facial shapes and textures [10]. Conversely, challenges arise in complex environments, including irregular poses, poor lighting, and especially occlusion when parts of the face are obscured by objects such as masks or hijabs [11]. These challenges become even more critical to address, considering that hijabs are cultural elements widely used in many countries, particularly in the middle east and Southeast Asia, influencing the success of automated facial recognition [12], [13]. Wearing a hijab can reduce the performance of facial recognition models reliant on full facial features, such as the jawline and mouth [14]. In this research, facial recognition in hijab conditions is considered an occlusion case. Occlusion refers to a situation where parts of the object to be recognized are not visible due to being covered by other objects [8], [15], [16].

Another issue in facial recognition involves gender classification, i.e., distinguishing between male and female genders. Sometimes, it is difficult to determine gender [1], [4]. Even faces that are mostly covered by masks or niqabs present significant challenges in detecting landmarks (such as eyes, nose, and mouth corners) [8], [9]. Accurate detection of gender and hijab status requires models that can adapt to the various facial expressions found in datasets [17]. A CNN model was also developed to classify skin lesion images into classes and subclasses to improve accuracy up to 96.2% [18]. However, particularly occlusion can affect the accuracy and performance of facial recognition systems [14], CNN architectures can be utilized to overcome these issues [3], [19], [20] including modified CNN architectures [21]. The selection of a CNN architecture that has a high level of accuracy is expected to be able to overcome the grafting problem.

Various CNN architectures have been extensively utilized, such as LeNet, AlexNet, ZfNet, VGG, GoogleNet, ResNet, SqueezeNet, DenseNet, MobileNet, U-Net, EfficientNet, Faster R-CNN, and NasNet [22]. Each CNN architecture has its own strengths and weaknesses [1], [3], [23]. Therefore, this study conducts an evaluation comparing several CNN architectures: AlexNet, SqueezeNet, VGG16, and ResNet50.

Based on research by Thaher *et al*. [4], which stated that the deep learning method is superior to traditional methods in detecting occluded faces, especially region-based CNN (R-CNN and Faster R-CNN) and optimized single-shot detectors (you only look once (YOLO) and single shot multibox detector (SSD)). Then, research by Hassanat *et al*. [19], which examined face, gender, and expression recognition on fully veiled faces with a focus on the eye area to overcome occlusion. However, that study did not compare ResNet50 and VGG16, where ResNet's deeper architecture can address vanishing gradient issues [24]. Although SqueezeNet is efficient in terms of parameter usage, it often has limitations in handling facial conditions with hijabs [11]. AlexNet faces challenges in dealing with complex data, such as faces with occlusions or poor lighting [24]. FaceNet is capable of deeply learning facial representations and efficiently measuring the distances between faces using triplet loss [25]. ResNet exhibits higher accuracy in handling facial detection under challenging conditions, including variations in lighting, masks, and facial expressions, although with heavier computational requirements [3]. Conversely, ResNet50 has proven to be highly efficient in dealing with irregular facial poses and severe occlusions [13], [26].

Various comparative studies on CNN architectures have been conducted, such as by Krichen [27], which concluded that ResNet is the best CNN architecture based on its ability to train deeper networks and effectively address issues faced by other architectures like visual geometry group (VGG), AlexNet, LeNet, and Inception Net. ResNet offers a superior solution for deeper and more complex networks. Similarly, Shah *et al*. [28] asserted that ResNet50 is the most effective model for early disease detection in rice compared to models like Inception V3, VGG16, and VGG19. In the study by Madkour *et al*. [29], automatic facial segmentation using CNN was applied to women wearing hijabs. The images were automatically segmented into three classes: skin, hijab, and background. The dataset consisted of 250 images, divided into 150 training and 100 testing. The FCN method with 91 layers was employed. Validation results included global accuracy: 92%, mean accuracy: 92.69%, and mean IoU: 84.4%. Evaluation results showed skin accuracy: 95.43%, hijab accuracy: 90.61%, and background accuracy: 92.05%.

Kocacinar *et al*. [20] developed a lightweight, fine-tuned CNN for mobile applications to detect masked faces. The architectures used included Mobile Net, VGG16, and ResNet, with a dataset comprising 1,849 facial samples from 12 individuals. Mobile Net achieved the highest validation accuracy of 90.40% for identifying correctly or incorrectly worn masks. Fine-tuning with VGG16 and ResNet resulted in validation accuracies of 87.60% and 51.74%, respectively. Ali *et al*. [5] used CNN for age and gender prediction, achieving an accuracy of 95% for gender and 92% for age. The dataset contained approximately 26,000 labeled facial images, covering variations in lighting, facial expressions, poses, and image quality.

Hassanat *et al*. [19] research shows that an eye-focused approach using a lightweight CNN and data augmentation can achieve ±92% accuracy for fully veiled face identification, ±90% for gender classification, and ~78% for expression recognition, with performance comparable to large models but faster for real-time applications. Shah *et al*. [28] automated the diagnosis of blast disease in rice plants using Inception V3, VGG16, VGG19, and ResNet50. The dataset consisted of 2,000 images divided into two classes (1,200 infected images and 800 healthy leaf images). ResNet50 delivered the best performance, achieving an accuracy of 99.75%, a loss rate of 0.33, a validation accuracy of 99.69%, a precision of 99.50%, an F1-score of 99.70, and an AUC of 99.83%.

Izdihar *et al*. [30] compared two CNN architectures (VGG16 and ResNet50) for detecting pneumonia in chest x-ray (CXR) images. Results indicated that ResNet50 outperformed VGG16 in performance. Nugraha *et al*. [31] compared the performance of GoogleNet, AlexNet, VGG-16, LeNet-5, and ResNet-50 in recognizing Arabic handwritten patterns. The dataset consisted of 8,400 handwritten Arabic images from various individuals. Training data is 80%, and testing is 20%. ResNet-50 and GoogleNet demonstrated the best accuracy and training speed. Although AlexNet and VGG-16 yielded lower accuracies, their results were acceptable, whereas LeNet-5 had low accuracy and was not recommended. Akhand *et al*. [32] developed facial emotion recognition (FER) using transfer learning (TL) with a DCNN model. DenseNet-161 achieved the highest accuracy: 96.51% on the KDEF dataset and 99.52% on the JAFFE dataset using 10-fold cross-validation.

Naseer *et al*. [33] compared intrinsic CNN architectures LeNet, AlexNet, VGG16, ResNet-50, and Inception-V1 for detecting lung cancer, with the LUNA16 dataset. AlexNet, optimized with SGD, achieved the highest validation accuracy for lung cancer detection using CT images, with an accuracy of 97.42%, a classification error rate of 2.58%, sensitivity of 97.58%, specificity of 97.25%, positive predictive value of 97.58%, negative predictive value of 97.25%, false omission rate of 2.75%, and an F1-score of 97.58%. The performance of CNN architectures is significantly influenced by the selection of appropriate hyperparameters, impacting accuracy, training speed, and generalization ability [34], [35]. This principle applies to facial recognition, including masked faces [14], [26], [36] and hijabs [4], [29]. The aim of this research is to identify the most effective CNN architecture for facial recognition based on gender and hijab status. In addition, the aim of this research is also to understand the strengths and weaknesses of each architecture in handling various types of occlusions. The contribution of this research is to identify the best performance of CNN architectures in handling the simultaneous detection of gender and hijab features.

2. METHOD

This research began by identifying the problem of facial image detection to determine gender and hijab status using CNN architectures. Based on the findings, several CNN architectures were selected for comparison in facial image detection tasks. The next step involved collecting facial image data under various poses, genders, and with or without hijabs. The collected data was labeled and cleaned to ensure usability for further analysis. The data was then classified based on labels and analyzed to determine its suitability for testing. Once the dataset was deemed appropriate, testing was implemented using four CNN architectures: AlexNet, SqueezeNet, VGG16, and ResNet50. Training and validation were performed for each architecture, evaluating metrics such as accuracy, precision, recall, and F1-score. All validation results were then assessed, culminating in conclusions drawn from the evaluation. The research stages are illustrated in Figure 1.

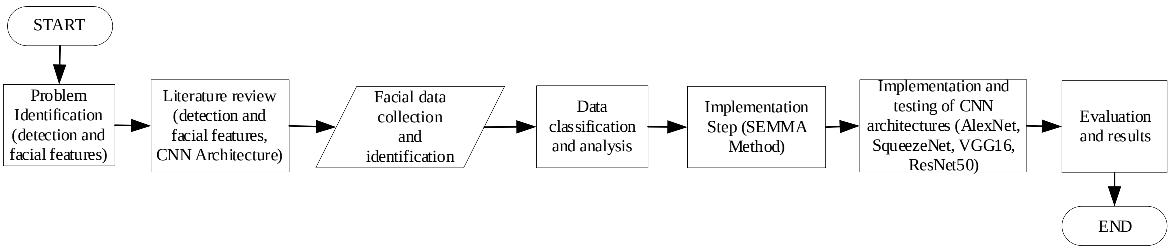


Figure 1. Research methodology stages

2.1. Problem identification

Based on the background discussed in the previous subsection, identifying gender and hijab status from facial images requires performance evaluation of various CNN architectures. This study aimed to compare four major CNN architectures: AlexNet, SqueezeNet, VGG16, and ResNet50, in the task of gender and hijab status detection based on facial images. CNN was utilized to create a model capable of classifying gender as male or female. Subsequently, for female classifications, facial images were evaluated to determine whether hijab features were present. The research tested the accuracy of these CNN architectures and evaluated the best-performing model based on classification results.

2.2. Literature analysis

Literature searches were conducted on electronic publication platforms such as Google Scholar, IEEE Xplore, MDPI, SpringerOpen, ProQuest, and Academia.edu. Over 90% of references used were indexed by Scopus, with more than 75% ranked as Q1 and Q2. The primary themes included CNN architecture usage for facial detection, gender classification, and hijab differentiation.

2.3. Data collection

The research used human facial image data consisting of two classifications: male and female. A total of 170 images were used, comprising 68 male faces and 102 female faces. The images were collected directly using smartphone cameras with average quality. Subjects ranged from 18 to 48 years old, photographed in relaxed poses from distances of 0.8 to 3 meters. Cameras operated in auto mode without specific lighting or settings. Photos were taken both indoors and outdoors to achieve natural results, reflecting real-world applications. The dataset of 170 images was divided into male and female classes. Male images accounted for 40% (68 images), while female images comprised 60% (102 images), further divided into 78 hijab-wearing images (45.9%), and 24 non-hijab images (14.1%). For female-only data (102 images), 76.5% were hijab-wearing, and 23.5% were non-hijab.

2.4. Convolutional neural network architecture implementation

Following data collection, selection, and cleaning, the dataset was divided into training, validation, and testing datasets. Data resizing was performed to standardize image sizes for each CNN architecture. Training and validation were conducted for AlexNet, SqueezeNet, VGG16, and ResNet50 architectures. This research employed the Sample, Explore, Modify, Model, and Assess (SEMMA) framework to preprocess data, apply discriminant models using four machine learning algorithms, evaluate their performance, and test algorithms with the best discriminant verification. Figure 2 illustrates the dataset processing flowchart with CNN architectures.

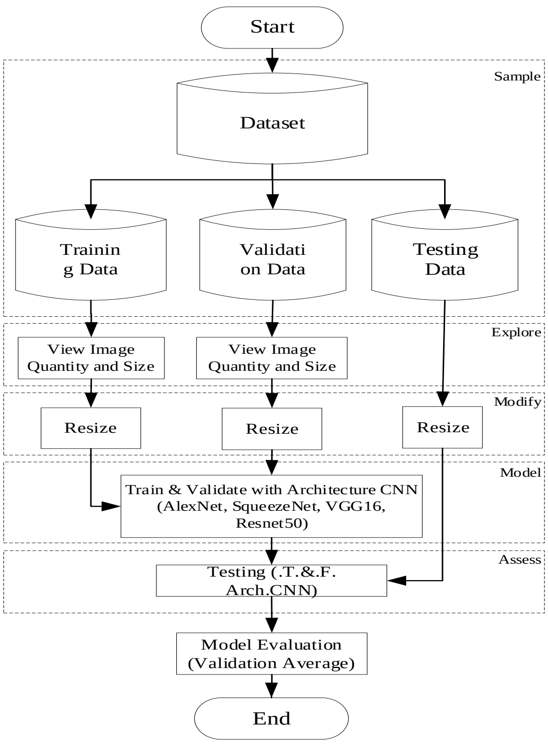


Figure 2. Dataset processing flowchart with CNN architectures

a. Sample stage

The sample stage involves collecting and selecting data to be used in the analysis. The data in this study consists of human facial images, divided into two classes: male faces and female faces. The dataset is divided into three parts: 90 data points for training, 40 data points for validation, and 40 data points for testing. This division is made to optimize the learning process and evaluate the developed model. Table 1 explains the data distribution for training, validation, and testing purposes.

Table 1. Data distribution

No	Data type	Male	Female	Agregat
1	Training	33	57	90
2	Validasi	15	25	40
3	Testing	20	20	40
	Total	68	102	170

b. Explore stage

The explore stage is the initial step in data analysis to identify patterns, trends, and characteristics of the collected data. In this study, the explore stage involves examining the resolution and quantity of the collected image data.

c. Modify stage

The modify stage aims to prepare the data for modeling. The modify process involves resizing images to ensure uniformity and reduce size. For AlexNet, SqueezeNet, and VGG16 architectures, images are resized to 227×227×3. For the ResNet50 architecture, images are resized to 224×224×3. This resizing is necessary to ensure that the image data can be appropriately processed by the input layer of each CNN architecture (AlexNet, SqueezeNet, VGG16, and ResNet50).

d. Modeling stage

The modeling stage involves building models using CNN with four different architectures. CNN processes data layer by layer, with each layer responsible for extracting increasingly complex features from the data. By applying these architectures, models are developed with varying capabilities in understanding patterns and visually representing the data used.

e. Assess stage

The assess stage in data analysis is the evaluation step for the built models. During this stage, testing is conducted using testing data to examine the performance of the developed models. By using testing data, the models are evaluated to determine how well they classify gender with the expected accuracy. The results of the assess stage will aid in evaluating the effectiveness and reliability of the developed models.

2.5. Model evaluation

Evaluation aimed to measure the success and quality of the developed models. This process used a confusion matrix as recommended by Krichen [27], to obtain metrics such as accuracy, precision, recall, and F1-score for the CNN architectures tested. The formulas for these metrics are shown in (1)-(4):

$$Accuracy = \frac{(\mathrm{TP+TN})}{\mathrm{TP+FP+FN+TN}} \times 100\% \tag{1}$$

$$Precission = \frac{(\mathrm{TP})}{\mathrm{TP+FN}} \tag{2}$$

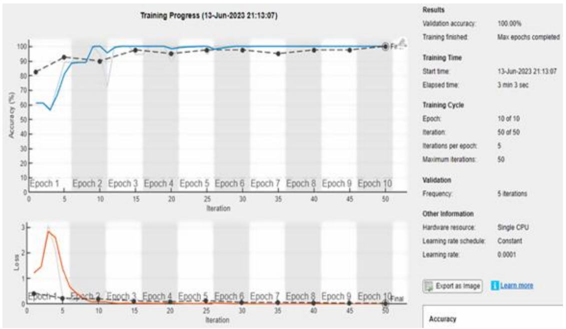
$$Recall = \frac{(\mathrm{TP})}{\mathrm{FP+TP}} \tag{3}$$

$$F1\ score = 2\ \times\ \frac{(recall\ \mathrm{x}\ precission)}{(recall + \mathrm{pecission})} \tag{4}$$

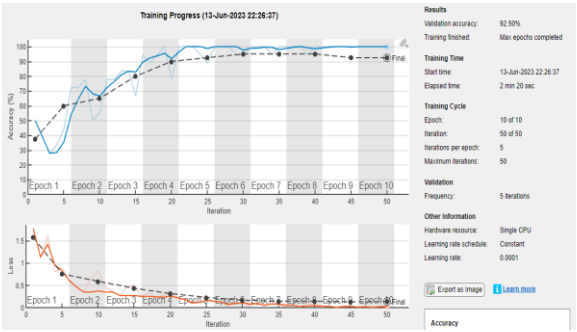
3. RESULTS AND DISCUSSION

Based on the stages outlined in the previous subsections, the training phase employed four CNN architectures: AlexNet, SqueezeNet, VGG16, and ResNet50. The objective was to compare these architectures to identify the best-performing model in a facial recognition dataset designed to determine gender and hijab or non-hijab status. The training parameters included the Adam optimizer method, a learning rate of 0.0001, a maximum of 10 epochs, and a batch size of 18. During training, the models were trained and validated every five iterations. Training results are illustrated in Figure 3, each depicting the training and validation processes for AlexNet, SqueezeNet, VGG16, and ResNet50 architectures. The red and blue curves in Figures 3(a)-(d) represent model performance on training and validation data, respectively. The red curve shows the model's performance improvement over iterations or epochs during training, while the blue curve reflects the model's ability to generalize patterns learned from training data to validation data.

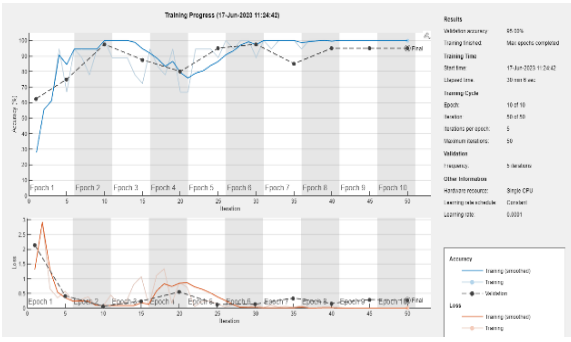
Based on the four figures, the curves can be analyzed to determine whether the model is experiencing overfitting (overtrained), underfitting (undertrained), or if it is capable of good generalization on new data. A red curve that increases rapidly on training data indicates that the model is effectively adapting to the training data. However, if the blue curve for validation data is significantly below the red curve, this may suggest that the model is overfitting and unable to generalize well on new data. Conversely, if both curves are close to each other and achieve good performance on validation data, the model demonstrates good generalization ability and can recognize patterns in previously unseen data.



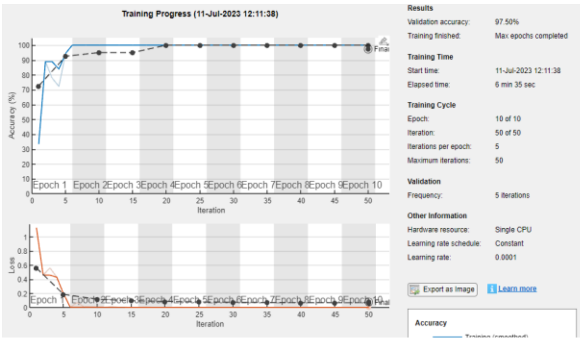
(a)



(b)



(c)



(d)

Figure 3. CNN architecture training and validation; (a) Alexnet, (b) SqueezeNet, (c) VGG16, and (d) ResNet50

3.1. Validation results

The evaluation was conducted for the four CNN architectures: AlexNet, SqueezeNet, VGG16, and ResNet50. This was carried out to identify images based on gender classification, distinguishing between male and female. Figure 4 illustrates the evaluation using the confusion matrix from the validation results of the models developed with these four architectures.

From the AlexNet confusion matrix in Figure 4(a), out of 40 data used for model validation, the model correctly classified all (100%) data according to their respective classes. Table 2 shows the calculated accuracy, precision, recall, and F1-score based on the AlexNet confusion matrix. Based on Table 2, the validation results using the AlexNet model achieved an average accuracy of 100%, a precision of 1.0000, a recall of 1.0000, and an F1-score of 1.0000.

Table 2. AlexNet validation results

No	Class	Accuracy (%)	Precision	Recall	F1-score
1	Male	100	1.0000	1.0000	1.0000
2	Female	100	1.0000	1.0000	1.0000
Average		100	1.0000	1.0000	1.0000

From the SqueezeNet confusion matrix in Figure 4(b), out of 40 validation data points, 37 images were correctly classified, and 3 images were misclassified. Specifically, 14 male images were correctly classified, while 1 male image was misclassified as female. Furthermore, 23 female images were correctly classified, while 2 female images were misclassified as male. Table 3 presents the calculated accuracy, precision, recall, and F1-score based on Figure 4(b). Based on Table 3, the validation results using the SqueezeNet model achieved an average accuracy of 92.50%, precision of 0.9167, recall of 0.9265, and F1-score of 0.921.

Table 3. SqueezeNet validation results

No	Class	Accuracy (%)	Precision	Recall	F1-score
1	Male	92.50	0.875	0.933	0.9031
2	Female	92.50	0.9583	0.92	0.9388
Average		92.50	0.9167	0.9265	0.921

From the VGG16 confusion matrix in Figure 4(c), out of 40 validation data points, 38 images were correctly classified, and 2 images were misclassified. Specifically, 15 male images were correctly classified, while 23 female images were correctly classified. However, 2 female images were misclassified as male. Table 4 presents the calculated accuracy, precision, recall, and F1-score based on Figure 4(c). Based on Table 4, the validation results using the VGG16 model achieved an average accuracy of 95%, precision of 0.9412, recall of 0.9600, and F1-score of 0.9479.

Table 4. VGG16 validation results

No	Class	Accuracy (%)	Precision	Recall	F1-score
1	Male	95	0.8824	1.0000	0.9375
2	Female	95	1.0000	0.92	0.9583
Average		95	0.9412	0.96	0.9479

From the ResNet50 confusion matrix in Figure 4(d), out of 40 validation data points, 39 images were correctly classified, and 1 image was misclassified. Specifically, 14 male images were correctly classified, while 25 female images were correctly classified. However, 1 male image was misclassified as female. Table 5 presents the calculated accuracy, precision, recall, and F1-score based on Figure 4(d). Based on Table 5, the validation results using the ResNet50 model achieved an average accuracy of 97.50%, precision of 0.9808, recall of 0.9667, and F1-score of 0.9730. These results are better than the findings of [5], [20], [29].

Table 5. ResNet50 validation results

No	Class	Accuracy (%)	Precision	Recall	F1-score
1	Male	97.50	1.0000	0.9333	0.9655
2	Female	97.50	0.9615	1.0000	0.9804
Average		97.50	0.98077	0.96667	0.97295

(a) (b) (c) (d)

Figure 4. Confusion matrices for validation; (a) AlexNet, (b) SqueezeNet, (c) VGG16, and (d) ResNet50

3.2. Testing results

Testing involved evaluating the models using unseen data to assess their generalization performance. The testing dataset consisted of 40 facial images, with 20 male and 20 female images, including 10 hijab-wearing female images and 10 non-hijab female images. Table 6 summarizes the testing results for AlexNet, SqueezeNet, VGG16, and ResNet50 architectures.

From Table 6, the ResNet50 architecture correctly classified 38 images and misclassified 2, achieving a testing accuracy of 95%. Next, the AlexNet architecture correctly classified 37 images and misclassified 3, achieving a testing accuracy of 92.5%. Furthermore, the SqueezeNet architecture correctly classified 36 images and misclassified 4, achieving a testing accuracy of 90%. Last, the VGG16 architecture correctly classified 34 images and misclassified 6, achieving a testing accuracy of 85%. The results of the hijab and gender face detection evaluation are shown in Figure 5.

Table 6. Testing results for AlexNet, SqueezeNet, VGG16, and ResNet50

No	Architecture	Testing result	
		True	False
1	AlexNet	37	3
2	SqueezeNet	36	4
3	VGG16	34	6
4	ResNet50	38	2

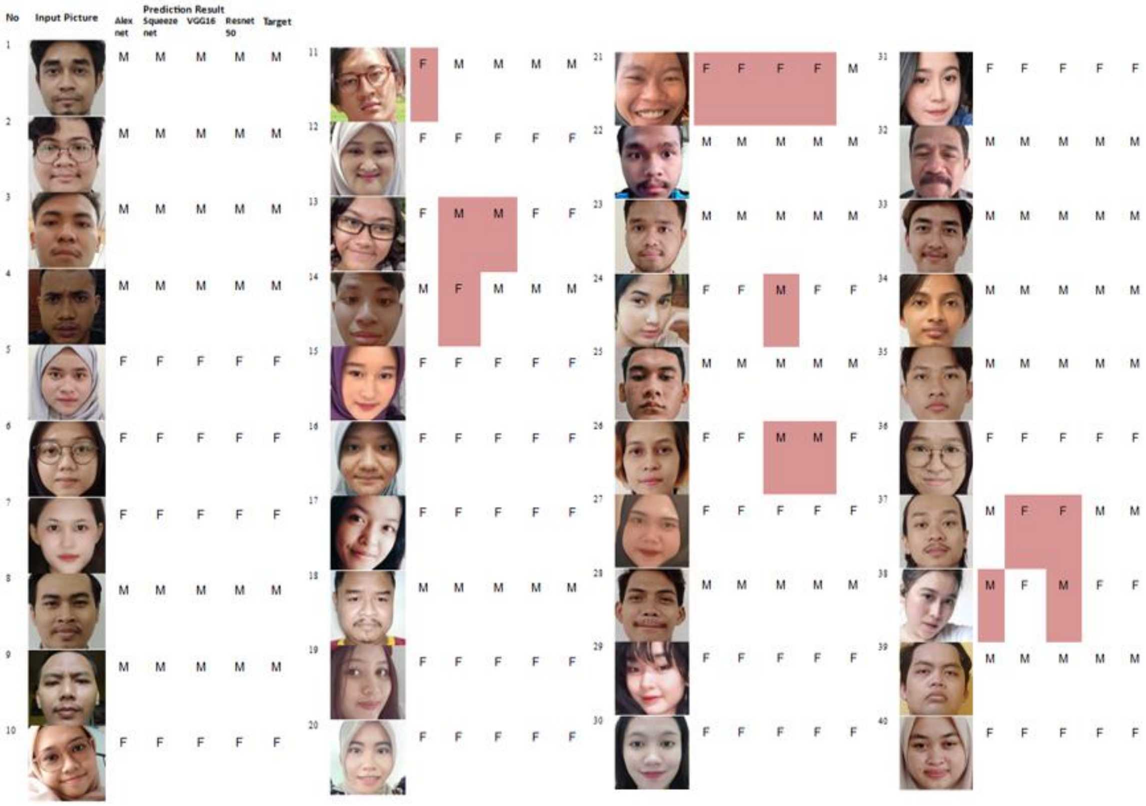


Figure 5. Results of the evaluation of hijab face detection and gender

3.3. Evaluation summary

Based on the validation and testing results and Figure 5, a summary of the average performance metrics for each architecture is provided in Table 7.

Table 7. Evaluation summary for AlexNet, SqueezeNet, VGG16, and ResNet50

No	Architecture	Average validation				Testing	
		Accuracy (%)	Precision	Recall	F1-score	True	False
1	AlexNet	100	1	1	1	37	3
2	SqueezeNet	92.5	0.9167	0.9265	0.921	36	4
3	VGG16	95	0.9412	0.96	0.9479	34	6
4	ResNet50	97.5	0.98077	0.96667	0.97295	38	2
Amount of data						145	15

From the evaluation:

- ResNet50 demonstrated the highest testing accuracy (95%) and maintained high validation accuracy (97.5%), making it the most robust model in this study.
- AlexNet achieved perfect validation accuracy (100%) but performed slightly lower during testing, classifying 37 images correctly (92.5% testing accuracy).
- SqueezeNet showed competitive performance with a testing accuracy of 90%, emphasizing computational efficiency.

– VGG16 had high validation accuracy (95%) but lower testing accuracy (85%), indicating potential overfitting or limitations in generalization.

Validation accuracy: AlexNet (100%), ResNet50 (97.5%). Test accuracy: ResNet50 (95%), AlexNet (92.5%). The superior performance of ResNet50 and its implications for real-world applications.

This study has several limitations: i) the size and diversity of the dataset; ii) non-uniform image resolutions; and iii) evaluation procedures that require further development. Recommendations for future research are: i) increasing the dataset size and diversity; ii) standardizing image resolutions; and iii) adopting more complex evaluation methods, including consequential advancements in self-supervised learning (SSL) within deep learning contexts.

4. CONCLUSION

This study successfully tested the effectiveness of four CNN architectures: AlexNet, SqueezeNet, VGG16, and ResNet50 in detecting gender from facial images, both with and without hijab features. These CNN models were trained using the Adam optimization method with a learning rate of 0.0001, over 10 epochs, and a batch size of 18. The findings of this research are that the AlexNet model achieved the highest validation accuracy at 100%. Meanwhile, ResNet50, SqueezeNet, and VGG16 achieved validation accuracies of 97.50%, 92.50%, and 95%, respectively. This indicates that all four architectures perform well in classifying gender based on facial images. In the testing stage with new, unseen data, the ResNet50 model achieved the highest accuracy of 95%, correctly classifying 38 images. Despite variations in accuracy among the four models, this study demonstrates the potential of CNN architectures in detecting gender from facial images with appropriate training parameters. The findings of this research indicate that AlexNet achieved the highest validation accuracy, while ResNet50 provided the best accuracy in facial image detection for determining gender and hijab features.

This study has several limitations that need to be addressed in future research. The dataset used in this study is relatively small. The performance of models heavily depends on the quality and diversity of the dataset used for training and evaluation. Future research should include a larger dataset with diverse poses, genders, ages, and hijab models. Comparative analysis with other CNN architectures is necessary to provide a more comprehensive comparison and identify more accurate alternatives. Both the dataset and CNN architectures should be further developed for real-time testing scenarios, such as video processing or CCTV monitoring, to assess their performance in practical, everyday applications. Real-time implementation or testing on more complex datasets, such as datasets with varying lighting conditions or cultural clothing variations.

ACKNOWLEDGMENTS

We would like to thank the facial data collection team and Unisbank for their assistance with funding.

FUNDING INFORMATION

Thank you to Stikubank University (UNISBANK) for funding this research.

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
Aji Supriyanto	✓	✓		✓					✓			✓		
Theresia Dwiati Wismarini					✓					✓				✓
Herny Februariyanti	✓						✓		✓					
Arief Jananto		✓			✓			✓		✓				
Fitri Damaryanti						✓			✓				✓	
Hilmy Nurakmal Satria			✓			✓				✓	✓			

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

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author [AS] on request.

REFERENCES

[1] Z. S. A. Ghrban and N. K. EL Abbadi, "Gender and Age Estimation from Human Faces Based on Deep Learning Techniques: A Review," *International Journal of Computing and Digital Systems*, vol. 14, no. 1, pp. 201–220, 2023, doi: 10.12785/ijcds/140118.

[2] H. Benradi, A. Chater, and A. Lasfar, "A hybrid approach for face recognition using a convolutional neural network combined with feature extraction techniques," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, pp. 627–640, 2023, doi: 10.11591/ijai.v12.i2.pp627-640.

[3] K. M. Hosny, N. A. Ibrahim, E. R. Mohamed, and H. M. Hamza, "Artificial intelligence-based masked face detection: A survey," *Intelligent Systems with Applications*, vol. 22, May 2024, doi: 10.1016/j.iswa.2024.200391.

[4] T. Thaher, M. Mafarja, M. Saffarini, A. H. H. M. Mohamed, and A. A. El-Saleh, "A Comprehensive Review of Face Detection Techniques for Occluded Faces: Methods, Datasets, and Open Challenges," *CMES - Computer Modeling in Engineering and Sciences*, vol. 143, no. 3, pp. 2615–2673, 2025, doi: 10.32604/cmes.2025.064857.

[5] D. Ali, F. Epifania, N. Ahmed, B. Khan, L. Marconi, and R. Matamoros, "A Modified CNN for Age and Gender Prediction," *CEUR Workshop Proceedings*, vol. 3463, pp. 0–2, 2022.

[6] R. Singh, S. Saurav, T. Kumar, R. Saini, A. Vohra, and S. Singh, "Facial expression recognition in videos using hybrid CNN & ConvLSTM," *International Journal of Information Technology (Singapore)*, vol. 15, no. 4, pp. 1819–1830, 2023, doi: 10.1007/s41870-023-01183-0.

[7] T. Sikandar, W. N. A. W. Samsudin, M. F. Rabbi, and K. H. Ghazali, "An Efficient Method for Detecting Covered Face Scenarios in ATM Surveillance Camera," *SN Computer Science*, vol. 1, no. 3, 2020, doi: 10.1007/s42979-020-00163-6.

[8] Z. R. Alqahtani, M. S. Sunar, and A. M. Artoli, "Comparative Analysis of Pre-trained Deep Learning Models for Facial Landmark Localization on Enhanced Dataset of Heavily Occluded Face Images," *Journal of Advances in Information Technology*, vol. 15, no. 11, pp. 1252–1263, 2024, doi: 10.12720/jait.15.11.1252-1263.

[9] M. Abdul-Al, G. K. Kyeremeh, R. Qahwaji, N. T. Ali, and R. A. Abd-Alhameed, "The Evolution of Biometric Authentication: A Deep Dive into Multi-Modal Facial Recognition: A Review Case Study," *IEEE Access*, vol. 12, pp. 179010-179038, 2024, doi: 10.1109/ACCESS.2024.3486552.

[10] T. Gwyn, K. Roy, and M. Atay, "Face recognition using popular deep net architectures: A brief comparative study," *Future Internet*, vol. 13, no. 7, pp. 1–15, 2021, doi: 10.3390/fi13070164.

[11] M. O. Oloyede, G. P. Hancke, and H. C. Myburgh, "Improving face recognition systems using a new image enhancement technique, hybrid features and the convolutional neural network," *IEEE Access*, vol. 6, pp. 75181–75191, 2018, doi: 10.1109/ACCESS.2018.2883748.

[12] J. O. Pinzón-Arenas, R. Jiménez-Moreno, and J. E. M. Baquero, "Comparison of convolutional neural network models for user's facial recognition," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 1, pp. 192–198, 2024, doi: 10.11591/ijece.v14i1.pp192-198.

[13] S. Gao, H. Liang, D. Hu, X. Hu, E. Lin, and H. Huang, "SAM-ResNet50: A Deep Learning Model for the Identification and Classification of Drought Stress in the Seedling Stage of Betula luminifera," *Remote Sensing*, vol. 16, no. 22, 2024, doi: 10.3390/rs16224141.

[14] R. K. Shukla and A. K. Tiwari, "Masked Face Recognition Using MobileNet V2 with Transfer Learning," *Computer Systems Science and Engineering*, vol. 45, no. 1, pp. 293–309, 2023, doi: 10.32604/csse.2023.027986.

[15] R. Alrabiah, M. Hussain, and H. A. Aboalsamh, "Unconstrained Gender Recognition from Periocular Region Using Multiscale Deep Features," *Intelligent Automation and Soft Computing*, vol. 35, no. 3, pp. 2941–2962, 2023, doi: 10.32604/iasc.2023.030036.

[16] A. A. Alashbi, M. S. Sunar, and Z. Alqahtani, "Deep-Learning-CNN for Detecting Covered Faces with Niqab," *Journal of Information Technology Management*, vol. 14, pp. 114–123, 2022, doi: 10.22059/JITM.2022.84888.

[17] T. Baltrusaitis, C. Ahuja, and L. P. Morency, “Multimodal Machine Learning: A Survey and Taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019, doi: 10.1109/TPAMI.2018.2798607.

[18] M. U. Ali, M. Khalid, H. Alshanbari, A. Zafar, and S. W. Lee, “Enhancing Skin Lesion Detection: A Multistage Multiclass Convolutional Neural Network-Based Framework,” *Bioengineering*, vol. 10, no. 11, 2023, doi: 10.3390/bioengineering10121430.

[19] A. B. Hassanat *et al.*, “Deep learning for identification and face, gender, expression recognition under constraints,” *arXiv*, 2021, doi: 10.48550/arXiv.2111.01930.

[20] B. Kocacinar, B. Tas, F. P. Akbulut, C. Catal, and D. Mishra, “A Real-Time CNN-Based Lightweight Mobile Masked Face Recognition System,” *IEEE Access*, vol. 10, pp. 63496–63507, 2022, doi: 10.1109/ACCESS.2022.3182055.

[21] A. H. Jassim and A. Altaie, “Improved CNN System for Face Mask Recognition,” *Research Square*, pp. 1–25, Apr. 19, 2024, doi: 10.21203/rs.3.rs-4251321/v1.

[22] K. C. Santosh, D. GhoshRoy, and S. Nakarmi, “A Systematic Review on Deep Structured Learning for COVID-19 Screening Using Chest CT from 2020 to 2022,” *Healthcare*, vol. 11, no. 17, pp. 1–42, 2023, doi: 10.3390/healthcare11172388.

[23] D. Mamieva, A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, “Improved Face Detection Method via Learning Small Faces on Hard Images Based on a Deep Learning Approach,” *Sensors*, vol. 23, no. 1, pp. 1–16, 2023, doi: 10.3390/s23010502.

[24] H. K. Rinky, R. R. Bhuiyan, and H. T. Rahman, “Performance Comparison of CNN Architectures for Detecting Malaria Diseases,” B.Sc. thesis, Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh, pp. 1–39, 2020

[25] F. Hidayat, U. Elviani, G. B. G. Situmorang, M. Z. Ramadhan, F. A. Alunjati, and R. F. Sucipto, “Face Recognition for Automatic Border Control: A Systematic Literature Review,” *IEEE Access*, vol. 12, pp. 37288–37309, 2024, doi: 10.1109/ACCESS.2024.3373264.

[26] O. K. Sikha and B. Bharath, “VGG16-random fourier hybrid model for masked face recognition,” *Soft Computing*, vol. 26, no. 22, pp. 12795–12810, 2022, doi: 10.1007/s00500-022-07289-0.

[27] M. Krichen, “Convolutional Neural Networks: A Survey,” *Computers*, vol. 12, no. 8, pp. 1–41, 2023, doi: 10.3390/computers12080151.

[28] S. R. Shah *et al.*, “Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a Rice Disease,” *Agronomy*, vol. 13, no. 1633, pp. 1–13, 2023, doi: 10.3390/agronomy13061633.

[29] D. M. Madkour, A. Madani, and M. W. Fakhr, “Automatic face and hijab segmentation using convolutional network,” *International Journal of Integrated Engineering*, vol. 11, no. 7, pp. 61–66, 2019, doi: 10.30880/ijie.2019.11.07.008.

[30] N. Izdihar, S. B. Rahayu, and K. Venkatesan, “Comparison Analysis of CXR Images in Detecting Pneumonia Using VGG16 and ResNet50 Convolution Neural Network Model,” *International Journal on Informatics Visualization*, vol. 8, no. 1, pp. 326–332, 2024, doi: 10.62527/joiv.8.1.2258.

[31] G. S. Nugraha, M. I. Darmawan, and R. Dwiyansaputra, “Comparison of CNN’s Architecture GoogleNet, AlexNet, VGG-16, Lenet -5, Resnet-50 in Arabic Handwriting Pattern Recognition,” *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 8, no. 2, May 2023, doi: 10.22219/kinetik.v8i2.1667.

[32] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, “Facial emotion recognition using transfer learning in the deep CNN,” *Electronics*, vol. 10, no. 9, 2021, doi: 10.3390/electronics10091036.

[33] I. Naseer, S. Akram, T. Masood, A. Jaffar, M. A. Khan, and A. Mosavi, “Performance Analysis of State-of-the-Art CNN Architectures for LUNA16,” *Sensors*, vol. 22, no. 12, 2022, doi: 10.3390/s22124426.

[34] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00444-8.

[35] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, “Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification,” *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3840–3854, 2020, doi: 10.1109/TCYB.2020.2983860.

[36] Y. Mao, Y. Lv, G. Zhang, and X. Gui, “Exploring Transformer for Face Mask Detection,” *IEEE Access*, vo. 12, pp. 118377–118388, 2024, doi: 10.1109/ACCESS.2024.3449802.

BIOGRAPHIES OF AUTHORS

Aji Supriyanto is an associate professor in the Master of Information Technology Program, Faculty of Information Technology and Industry, Universitas Stikubank (UNISBANK) Semarang, Indonesia. Graduated with a Bachelor of Informatics in 1997. Continued his Masters and Doctoral studies in Computer Science at Gadjah Mada University (UGM) Yogyakarta Indonesia, and graduated with a Doctorate in 2019. His main research is in the field of decision support systems and AI-based decisions, and data security for the application of e-Gov and social fields. He can be contacted at email: ajisup@edu.unisbank.ac.id.

Theresia Dwiati Wismarini completed her master's degree in the Master of Computer Science Program at Universitas Gadjah Mada (UGM), Yogyakarta, in 2013. She has been actively engaged in research since 2008 in the fields of Data Science and Geospatial Data. Currently studying for a Doctorate in Information Technology at Satya Wacana Christian University of Indonesia. She is currently a full-time lecturer in the Informatics Engineering Program, Faculty of Information Technology and Industry, Universitas Stikubank. She can be contacted at email: thwismarini@edu.unisbank.ac.id.

Herny Februariyanti received the B.Sc. degree in Management of Informatics and Computer Engineering from Institute of Science and Technolgy "Akprind" Yogyakarta in 1998 and the M.Sc. degree in Computer Science from Gadjah Mada University, Yogyakarta, Indonesia, in 2010. Currently, she is a lecturer at Faculty of Information Technology, Stikubank University, Indonesia. Her research interests are in the areas of information retrieval and information security. She can be contacted at email: hernyfeb@edu.unisbank.ac.id.

Arief Jananto is a lecturer in the Information Systems S1 study program, Faculty of Information Technology and Industry. He is degree with Cumlaude status and Best in Information Systems from Stikubank University, Semarang, Indonesia in 2008. He then continued his S2 studies and completed his M.Cs. in Computer Science from Gadjah Mada University, Indonesia in 2010. His research interests are mainly in the fields of data mining and database design and data mining techniques, especially in predicting the timeliness of student graduation and sales transaction data associations. He can be contacted at email: ajananto09@edu.unisbank.ac.id.

Fitri Damaryanti is a master's student of information technology at Stikubank University. She works at the Meteorology, Climatology and Geophysics Agency (BMKG) Semarang Indonesia. She is currently working on a thesis in the field of archival document classification related to BMKG's administrative work. She can be contacted at email: rr.fitridamaryanti@gmail.com.

Hilmy Nurakmal Satria is a final year undergraduate student of the computer engineering study program, Faculty of Engineering, Diponegoro University. He is working on his final assignment with the theme of UI/UX for stunting applications with QRCode interfaces. Have an interest in UI/UX design and international relations. He can be contacted at email: hilmysatria@students.undip.ac.id.