Bridging hybrid deep learning detection and lightweight handcrafted features for robust single sample face recognition

Faulinda Ely Nastiti[1], Sopingi[1,2], Dedy Hariyadi[3], Sri Sumarlinda[1,4]
[1]Department of Information System, Faculty of Computer Science, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia
[2]Department of Informatics Engineering, Faculty of Engineering, Universitas Hasanuddin, Makassar, Indonesia
[3]Department of Information Technology, Faculty of Engineering and Information Technology, Universitas Jenderal Achmad Yani, Sleman, Indonesia
[4]Department of Information Technology, Universiti Kuala Lumpur, Kuala Lumpur, Malaysia

Article Info

Article history:

Received Jun 16, 2025
Revised Dec 31, 2025
Accepted Jan 22, 2026

Keywords:

Anchor box
Biometric
Local binary patterns
Non-maximum suppression
Single sample face recognition

ABSTRACT

Single sample face recognition (SSFR) remains a challenging task due to the limitation of having only one reference image per identity, which reduces embedding diversity and decreases robustness under variations of pose, expression, and illumination. This study proposed a hybrid framework that integrates deep learning-based detection through anchor box optimization and non-maximum suppression (NMS) with lightweight handcrafted feature extraction using local binary pattern (LBP). The detection stage leverages deep learning to ensure robust face localisation, while LBP maintains computational efficiency under limited-sample conditions. The training process showed accuracy improvement from 47.5% at the initial epoch to 98.0% at epoch 72, while testing accuracy stabilized at 85-88% with the best value of 87.9%. Evaluation on 48 new facial images achieved 89.6% accuracy, 95.3% precision, 91.1% recall, 93.1% F1-score, and 0.94 area under the receiver operating characteristic curve (AUC ROC). Real-world implementation on Android and iOS-based attendance applications further validated the model, reaching 88.46% accuracy across 52 tests under 50-400 lux illumination. The findings proved that the proposed hybrid design provides improved accuracy and stability compared with previous approaches.

Corresponding Author:

Faulinda Ely Nastiti
Department of Information System, Faculty of Computer Science, Universitas Duta Bangsa Surakarta
Km. 20 Ki Mangun Sarkoro Road, Surakarta, Central Java, Indonesia
Email: faulinda_ely@udb.ac.id

1. INTRODUCTION

Face recognition has emerged as one of the critical technologies in digital security and authentication systems [1], with various applications in public surveillance [2], [3], access control [4], and identity verification [5], [6]. Its advantages in providing efficient security and user convenience have made it a preferred choice across multiple sectors, including government, business, and the internet of things (IoT) [7]–[9]. Increasingly, these systems are adopted in embedded devices, such as mobile devices [10], [11], which require real-time identification with high accuracy [12], [13]. However, implementing facial recognition in embedded devices presents significant challenges, particularly in the context of single sample face recognition (SSFR), where only a single facial image is available for training purposes [14]–[16]. This condition commonly arises in security systems permitting only a one-time registration, such as electronic identity cards (e-ID), passports, or user verification in mobile banking applications.

Although deep learning methods have demonstrated superior performance in facial recognition tasks [17]–[19], their effectiveness decreases significantly in onboard applications under SSFR conditions. In such cases, the models suffer from poor generalization across variations in lighting, pose, and expression making SSFR one of the most challenging scenarios in mobile-based face recognition [20], [21]. SSFR was regarded as an extreme case of few-shot learning, as each identity was represented by only a single facial image [22]. As a challenge within the low-data regime, SSFR required specialized strategies to enable models to generalize effectively despite the scarcity of data [23], [24]. Approaches such as metric learning, the use of local features, attention mechanisms, and even meta-learning or generative augmentation have frequently been explored. However, this study proposed a lighter alternative by focusing on the optimization of detection and feature extraction processes.

A plethora of studies have previously examined the efficacy of diverse facial recognition models in addressing the SSFR challenge. However, the attained accuracies have been deemed inadequate. Comparative studies evaluating models for SSFR have reported accuracies of principal component analysis (PCA) (10%), linear discriminant analysis (LDA) (27.69%), kernel principal component analysis (KPCA) (14.62%), kernel fisher analysis (KFA) (25.38%), regularized supervised LDA (RSLDA) (57.46%), locally robust pattern propagation with global regularization reduction (LRPP-GRR) (57.43%), deep k-nearest neighbors (KNN) (37.69%), and Dlib face recognition library (DLIB) deep learning (63.28%) [25]. Even with the introduction of more advanced models such as, ArcFace attained 52% accuracy [26], conditional generative adversarial network (CGAN) achieved 76% [27], MobileFaceNet [28], GhostFaceNet [29], [30], both approaches were not originally designed for SSFR, thereby limiting their contributions to multi-sample scenarios. Rendering them inadequate for single-sample facial recognition tasks. Therefore, strengthening the stages of face detection and extraction became essential prior to the embedding process.

To enhance SSFR performance, precise face detection was considered essential prior to feature extraction. In modern facial recognition pipelines, anchor box-based approaches were adopted to enable multi-scale and multi-aspect ratio predictions, while non-maximum suppression (NMS) was applied to eliminate overlapping predictions, retaining only the most probable candidates [31]. The outputs from detection remained as localization results and required further processing through feature extraction to represent facial data numerically. At this stage, local binary patterns (LBP) were widely employed due to their simplicity, speed, and robustness against illumination variations [32]. Studies reported that LBP achieved accuracy rates of 76% with elliptical masks, 84.1% with rectangular masks, and 78% in standard implementations [33]. However, LBP was still limited under extreme pose variations, indicating that its integration with modern anchor-NMS-based detection was viewed as a promising strategy to improve SSFR accuracy.

Previous studies had primarily focused on feature embedding or data augmentation, while the optimization of anchor-NMS-based detection under SSFR scenarios had not been extensively explored. Furthermore, although LBP had been proven efficient and robust against illumination changes, its application had rarely been strategically integrated with modern detection techniques. Evaluations under real mobile device conditions also remained limited, despite the high demand for biometric applications on such platforms.

This study proposed a hybrid framework that combined anchor-NMS for detection preprocessing with LBP for feature extraction. Anchor-NMS was employed to ensure precise face localization by suppressing duplicates, while LBP was utilized to provide lightweight texture representation resistant to illumination variations. This integration was expected to yield a more accurate, robust, and efficient SSFR system suitable for mobile device implementation. Additionally, the study aimed to evaluate the system's robustness against variations in lighting, pose, and facial attributes, while also demonstrating its implementation feasibility through cosine similarity-based evaluation and real-world mobile application scenarios.

2. RESEARCH METHOD

2.1. Materials

The use of a SSFR for each subject has been identified as a fundamental limitation, as it not only increases the likelihood of false positives but also constrains the model's ability to achieve reliable generalization. To overcome this condition, an experimental design was formulated as shown in Figure 1. So that the effectiveness of the algorithm could be tested directly in a limited data situation, with results measured through accuracy percentage, detection error rates, and computational efficiency.

To ensure that the model has a high level of accuracy in accordance with the SSFR scenario, face detection experiments were conducted in a Python-based virtual laboratory environment. The research dataset was collected directly by researchers at Universitas Duta Bangsa Surakarta. The dataset consists of 239 subjects allocated for data training 191 subjects and data testing 48 subjects, with only one face image per subject in accordance with the SSFR principle. Each image was processed through a pre-processing stage and detected using you only look once version 5 (YOLOv5), which had been optimised with anchor-NMS.

In addition, 48 new subjects from Universitas Duta Bangsa Surakarta students were also collected and used exclusively in the cosine similarity-based evaluation stage. With this approach, the performance of the optimised model was not only tested for detection capabilities but also for face verification and identification processes. This strategy was designed to represent a realistic scenario of SSFR use on mobile devices, where training data is limited, but the model is still required to be able to generalise identities that have not been involved before.

All procedures in this study were conducted in compliance with recognized ethical research standards. The facial images employed for experimentation were collected with informed consent from the participants and were used exclusively for academic and scientific purposes. Throughout the process of data collection and analysis, confidentiality was maintained, potential risks were minimized, and the rights of participants were fully respected in line with ethical research principles.

2.2. Methods

The process begins as shown in Figure 1 with facial images that are processed through pre-processing and detection stages using YOLOv5, which is optimised using anchor-NMS. The detection results are then extracted using RGB color mapping combine with LBP and projected into vector space, which is subsequently used in the testing process with both training and test data. Next, an evaluation based on cosine similarity and onboard mobile is performed.

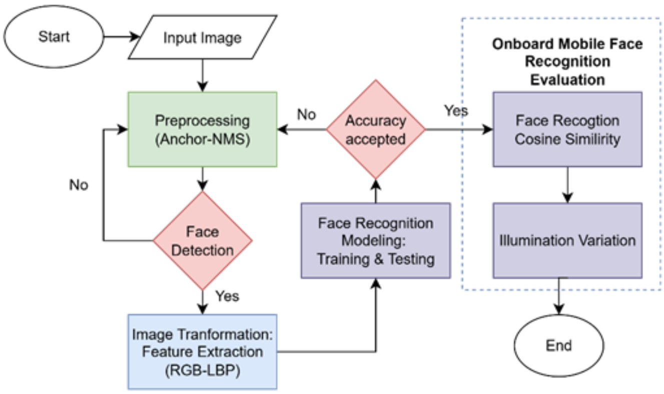


Figure 1. The pipeline for face recognition modeling

Phase one aimed to minimize bias and enhance generalization, all images were collected through a randomized sampling process covering three facial conditions, as shown in Figure 2. Figure 2(a) frontal angle eyewear, Figure 2(b) hijab slight angle, and Figure 2(c) frontal angle angle without additional attributes. The random selection ensured that each subject had an equal probability of representation across different pose variations, preventing overfitting toward specific facial orientations or attributes. This approach also simulates realistic deployment scenarios of SSFR, where system users may appear with diverse visual conditions that cannot be predetermined during model training.

(a) (b) (c)

Figure 2. Dataset face condition sample of (a) frontal angle eyewear, (b) hijab slight angle, and (c) frontal angle angle without additional attributes

Phase two, focused on preprocessing optimization to ensure accurate face detection and consistent extraction. As illustrated in Figure 3, the process began with the application of anchor boxes to generate multiple bounding box proposals with varying sizes and aspect ratios according to the shape of the face. These proposals were subsequently filtered using NMS, which removed overlapping or low-confidence bounding boxes, leaving only the most representative detection for each face. After optimization, the detected facial regions were cropped from the original images and prepared for the following steps, such as normalisation and feature extraction. This refinement ensured that only the relevant face regions were processed, reducing background noise and improving recognition accuracy.

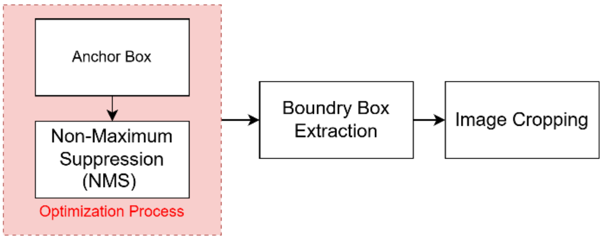


Figure 3. Image preprocessing optimization steps

To carry out this detection process, YOLOv5 was employed exclusively as the face detection module. While Figure 3 illustrates the sequential preprocessing steps, Figure 4 presents the YOLOv5 architecture that enabled this stage. The architecture consisted of three sequential components: a backbone for hierarchical feature extraction, a neck (path aggregation network (PANet)) for aggregating multi-scale features, and a head for generating the final bounding box predictions. In this study, YOLOv5 was optimised with anchor-NMS to achieve robust face localisation without extending its functionality to the recognition stage.

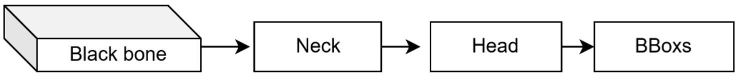


Figure 4. YOLOv5 steps

Phase three involved the transformation of cropped RGB face images into discriminative numerical representations. To achieve this, LBP were applied to the grayscale-converted images to extract texture-based descriptors. This method compares the intensity of each central pixel with that of its neighboring pixels, producing a robust local texture signature. The binary codes generated from this process were converted into numerical arrays and subsequently normalised using the L2 method to form unit-length feature vectors. The decision to use LBP was based on its computational efficiency and suitability for mobile deployment. Unlike convolutional neural network (CNN)-based feature extractors, which typically demand high processing power and memory, LBP produced lightweight descriptors with low complexity, thereby reducing inference time while preserving discriminative capability. This made LBP particularly relevant in the context of SSFR, where both efficiency and generalisation were critical under resource-constrained environments. The normalised feature vectors then served as the foundation for the training and testing stages.

Phase four outlined the training and testing procedure of the recognition model using LBP feature vectors. From 239 subjects, 191 were allocated for training and 48 for testing, with one image per subject to follow the SSFR principle. Training was conducted in a Python-based environment for up to 72 epochs. Performance was monitored using four metrics: training loss, validation loss, training accuracy, and testing accuracy. An early stopping mechanism was applied once validation performance stabilised to prevent overfitting. Logs and plots were generated throughout the process to provide data for analysis in the results and discussion section.

In addition, to validate the effectiveness of the proposed optimization, an ablation study was conducted with three configurations: i) oriented features from accelerated segment test and rotated binary robust independent elementary features (ORB) as a baseline non-face descriptor, ii) MobileFaceNet with

ArcFace loss, and iii) the hybrid anchor-NMS and LBP. This allowed the contribution of each component to be examined separately. A comparative analysis against prior approaches was also included to highlight performance gains under the SSFR scenario.

Phase five involved evaluating the recognition model using a cosine similarity approach. In this stage, 48 additional subjects, not included in the training or testing sets, were collected to assess the system's verification and identification capabilities under SSFR conditions. Cosine similarity was applied to compare the feature vectors of query images with those stored in the reference database. A similarity score close to 1 indicated a match, while lower values indicated dissimilarity. This procedure enabled the system to be tested on unseen identities, providing a realistic evaluation of robustness, generalisation, and feasibility for mobile-based face recognition.

Phase six, as the final evaluation, was conducted in an integrated manner through a mobile application developed for the Android and iOS platforms. The testing scenario followed SSFR, whereby one image per identity was used as an enrolment template. Since SSFR does not provide additional images to adapt to changes in illumination, lighting variations were identified as the main confounding factor that needed to be tested for resilience. The test data consisted of 48 images of students from Universitas Duta Bangsa Surakarta who were not involved in the previous model training or testing process. Each image was evaluated in two illumination ranges: i) normal ≈350-400 lux and ii) dim ≈50-100 lux. Illumination values were measured using a lux meter on the face area to ensure range suitability. The accuracy evaluation procedure for the onboard mobile face recognition system, as defined in (1), is simply the ratio of correct predictions to the total number of samples under each illumination condition.

$$Accuracy = \frac{T}{N} \times 100\% \tag{1}$$

Where *T* represents the number of correct predictions and *N* represents the number of samples in a given test condition. In the identification scheme (1:N), a prediction is considered correct if the face scanned by the mobile application matches the correct label; in the verification scheme (1:1).

3. RESULTS AND DISCUSSION

3.1. Image preprocesing optimization

The anchor box initialization process is an important initial step in the SSFR based facial detection system. The model must ascertain the starting position and dimensions of the bounding box prior to additional regression and filtering processes. The bounding box regression formula used (2) to (5) [34].

$$b_x = \sigma(t_x) + C_x \tag{2}$$

$$b_y = \sigma(t_y) + C_y \tag{3}$$

$$b_w = p_w e^{t_w} \tag{4}$$

$$b_h = p_h e^{t_h} \tag{5}$$

In this context, b_x and b_y indicates the coordinates of the predicted center, while b_w and b_h specifies the dimensions of the predicted bounding box. Additionally, t_x, t_y, t_w, and t_h are the parameters predicted by the model. The symbol σ represents the sigmoid activation function, which ensures that the coordinates stay within a valid range.

This research employs a scale of 1 to ensure that the initial bounding box size remains proportional to the feature map. This approach was selected because the facial images in the dataset are of uniform size, which allows the model to operate more stably and efficiently without requiring significant adjustments to the bounding box size. The anchor box used in this experiment maintains a 1:1 ratio, with an initial size of 60×60 pixels on the feature map. After initialization, the model adjusts the position and size of the bounding box through coordinate transformation-based regression. This transformation utilizes sigmoid and exponential functions to ensure that the bounding box values stay within a valid range. The results indicate that the bounding box, following regression, has center coordinates of (101.05, 100.98). Its dimensions have changed to 146.57×108.58 pixels, which still closely approximates the initial 1:1 ratio.

Figure 5 shows an experiment result of three rectangular anchor boxes that are intentionally centered right on the subject's face to illustrate the concept of anchor placement before the inference process begins.

The boxes are of different scales: small, medium, and large with a 1:1 ratio, so that each scale can represent the face at different camera distances. The smallest anchor (turquoise) covers the narrowest area of the face, the medium anchor (yellow) overlaps with the ground-truth, while the largest anchor (red) slightly extends beyond the contour of the face.

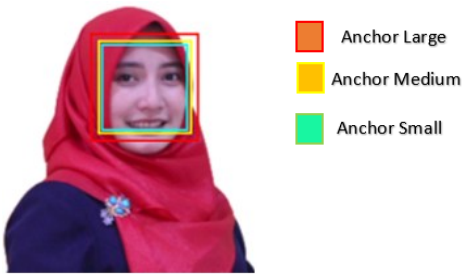


Figure 5. Anchor box

The 3 anchor boxes generated are then processed with the NMS method. In NMS, only one box with the highest confidence score and intersection over union (IoU) in (6) [35] is selected after the network regresses each anchor and calculates the confidence score. This method also removes overlapping boxes above a certain level. Therefore, even though there are three initial anchors, the detector ultimately produces only one correct face box.

$$IoU = \frac{Aoverlap}{Aunion} \tag{6}$$

After the NMS process is complete, the next step involves extracting the bounding box coordinates and performing image cropping to isolate the detected facial regions. The results of this process, as shown in Figure 6, illustrate the optimized outputs of the preprocessing stage. Each cropped image represents a face that has been consistently localized, regardless of variations in orientation or visual attributes. Specifically, Figure 6(a) demonstrates the successful localization of a face with eyewear, while Figure 6(b) shows the result on a subject wearing a hijab. Figure 6(c) serves as a clear example of the optimized output for a frontal angle without additional attributes.

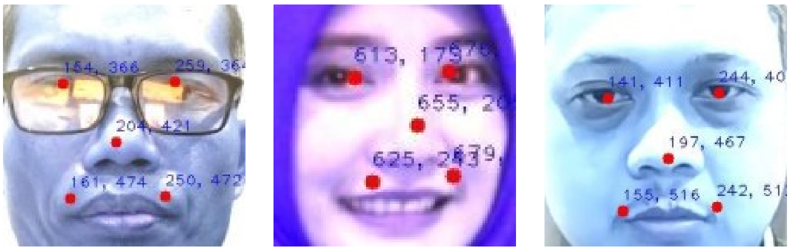
(a) (b) (c)

Figure 6. Example of image preprocessing optimization results of (a) frontal angle eyewear, (b) hijab slight angle, and (c) frontal angle angle without additional attributes

3.2. Image transformation

The results of the image transformation process following the face detection and cropping stages are illustrated in Figure 7, which displays RGB color mapping and the five-landmark point format. As demonstrated in Figures 7(a)-7(c), the image transformation process successfully forms a proper bounding box around the face, regardless of variations such as eyewear Figure 7(a) shows the frontal angle eyewear; Figure 7(b) shows the hijab slight angle; and Figure 7(c) shows the frontal angle without additional attributes. Furthermore, the system accurately detects five main landmarks (both eyes, the tip of the nose, and both corners of the mouth). These landmarks are visualized as red dots with numeric coordinates, serving as essential references for spatial normalization and the precise arrangement of facial features.

The blue and purple colors that appear in the image are the result of the transformation process to intensify and emphasize important feature areas. The main purpose of this conversion is to reduce the dimension of the data (eliminate irrelevant color information) and sharpen the local texture pattern that will be used in the feature extraction process. After facial landmarks are detected and the intensity conversion is completed from the original RGB image, facial features are extracted using the LBP method. This process produces a fixed-dimensional numeric feature array, which describes the micro-texture pattern of the face in the form of data that can be processed mathematically. This array is then input for the normalization and classification stages in the next process. Visually, it can be observed that the system works consistently on various facial conditions, including subjects wearing glasses, hijabs, and frontal poses. This shows that the transformation stage has succeeded in forming an initial numeric representation that is ready to be used in the facial recognition process.



(a) (b) (c)

Figure 7. RGB color mapping and five landmark points formats of (a) frontal angle eyewear, (b) hijab slight angle, and (c) frontal angle without additional attributes

3.3. Face recognition prediction modeling

The need for SSFR on mobile devices requires a solution that can maintain detection speed and accuracy in limited resource conditions. In this study, anchor-NMS combined with LBP are used as a modern approach that balances computational efficiency with face detection accuracy. This strategy is designed to ensure that the model remains lightweight, accurate, and resistant to overfitting. To test the effectiveness of this design, the training and testing results are presented through a comparison of epochs and loss accuracy up to 72 epochs from 191 subjects data training and data testing 48 subjects. Table 1 summarises the training loss, validation loss, training accuracy, and testing accuracy values at each epoch.

Table 1. Epoch vs. loss training-testing log

Epoch	Train_loss	Test_loss	Train_acc	Test_acc
1	1.7468	1.9235	0.475	0.4134
15	1.019	1.1507	0.6937	0.6221
30	0.6107	0.6182	0.8458	0.7533
50	0.2666	0.2719	0.9191	0.831
51	0.2977	0.3346	0.936	0.8561
52	0.2733	0.2444	0.9408	0.8336
53	0.2576	0.2988	0.9307	0.8435
54	0.2739	0.3421	0.9436	0.85
55	0.2731	0.2424	0.945	0.8624
56	0.2622	0.2515	0.9352	0.8412
57	0.2183	0.2682	0.9524	0.8638
58	0.2206	0.2473	0.9565	0.8562
59	0.2255	0.2136	0.9638	0.85
60	0.2307	0.2602	0.9654	0.8631
…	…	…	…	…
70	0.1349	0.2971	0.9768	0.8564
71	0.1498	0.2113	0.9683	0.8718
72	0.1683	0.2669	0.9802	0.8786

Table 1 shows the development of loss and accuracy in training and testing data up to 72 epochs. In the early stages (epoch 1), the train loss value was still high at 1.7468 with a training accuracy of 0.475, while the validation loss was recorded at 1.9235 with a testing accuracy of 0.4134. This is normal because the model is still in the initial adaptation phase to the data. As the number of epochs increases, the loss value decreases consistently and the accuracy increases. For example, at epoch 30, the train loss has dropped to 0.6107 with an accuracy of 0.8458, while the test accuracy has increased to 0.7533.

In the middle stage (epochs 50-60), it can be seen that the training loss is getting smaller (0.2666-0.2307) with a training accuracy above 0.92. The testing accuracy also increases significantly to around 0.83-0.86, indicating that the model not only learns from the training data but is also able to generalise to the test data. Early stopping was applied at epoch 58, where the training accuracy reached 0.9565 and the testing accuracy reached 0.8562. This decision was important to avoid overfitting, even though the accuracy value continued to increase in the following epochs.

Although early stopping was applied at epoch 58 to prevent overfitting, subsequent evaluation up to epoch 71 demonstrated that model performance remained stable, with a marginal accuracy improvement to 0.8718. This indicates that the model had reached convergence while maintaining generalization without signs of overfitting. The numerical data in Table 1 shows a positive trend in the training process, but the convergence pattern and indications of overfitting will be more clearly visible in the graphs. Therefore, Figures 8 and 9 are presented to show the dynamics of accuracy and loss over 72 epochs.

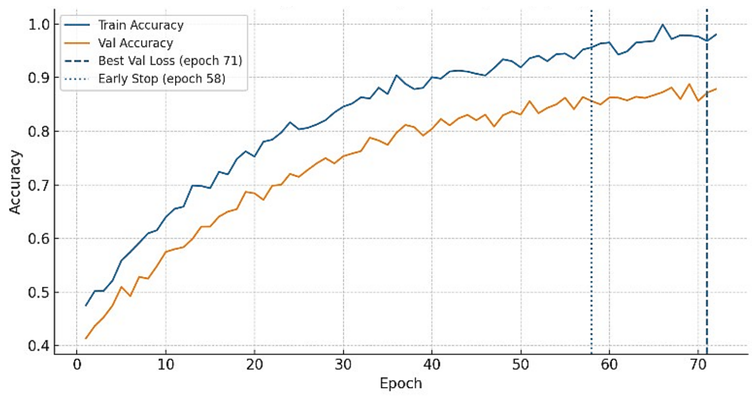


Figure 8. Accuracy vs. epoch

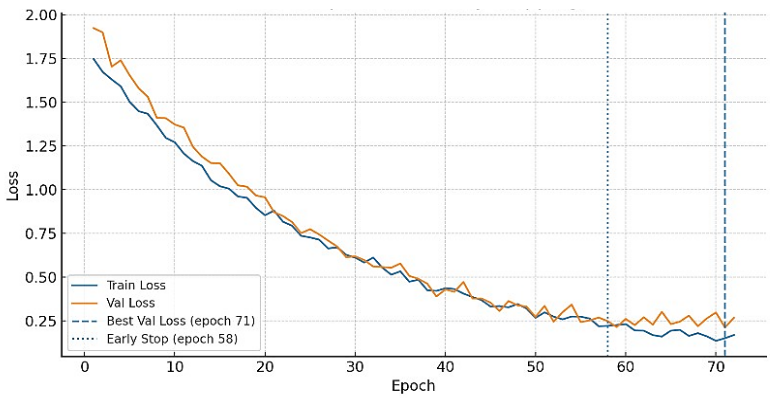


Figure 9. Loss vs. epoch

The results of model training and testing are shown in Figures 8 and 9. The accuracy graph in Figure 8 shows that training accuracy increased consistently to nearly 1.0, while testing accuracy also showed a steady upward trend to around 0.9. The difference between the two curves indicates overfitting, although it is still within acceptable limits. The early stopping mechanism applied at epoch 58 was able to prevent overfitting, although the best validation accuracy was recorded at epoch 71.

The loss graph in Figure 9 shows a similar pattern, where both training loss and validation loss decreased consistently throughout the training process. After epoch 50, a small gap began to appear between the two curves, indicating the early signs of overfitting. However, the loss value at the end of training was relatively stable, indicating that the model had reached a converged state. This convergence indicates that the model parameters have adapted optimally to the training data without losing their generalisation ability on the test data.

To evaluate the robustness and generalization capability of the proposed SSFR framework, a comparative experiment was conducted against two baseline models, ORB and MobileFaceNet. The same 48 subject test set used for model evaluation was employed to ensure fair benchmarking. As shown in Table 2, ORB, utilized as a baseline non-face descriptor, achieved only 45.2% accuracy, reflecting its limitations in handling pose variations and occlusions. MobileFaceNet, optimized through data augmentation and adaptive threshold tuning, achieved 82.4% accuracy. Despite being trained on diverse facial data, its performance decreased in SSFR settings due to the lack of multiple reference samples per identity, which constrained embedding diversity.

In contrast, the proposed hybrid model integrating anchor-based detection, NMS, and LBP achieved a converged testing accuracy of 85.6%, which was selected as the official evaluation value for comparative analysis. The use of this converged value, rather than the peak accuracy of 87.18% achieved at epoch 71, was intended to ensure the validity and reproducibility of results while preventing overfitting bias. The convergence-based evaluation reflects a balance between accuracy and model stability, providing a reliable measure of performance in SSFR. Accordingly, the 3.2% improvement over MobileFaceNet indicates that the proposed model remains competitive and can outperform conventional deep learning architectures under SSFR, where feature diversity and generalization are typically limited.

Table 2. Comparative experimental models

Model testing	Accuracy (%)
ORB	45.2
MobileFaceNet	82.4
Proposed model (anchor-NMS-LBP)	85.6

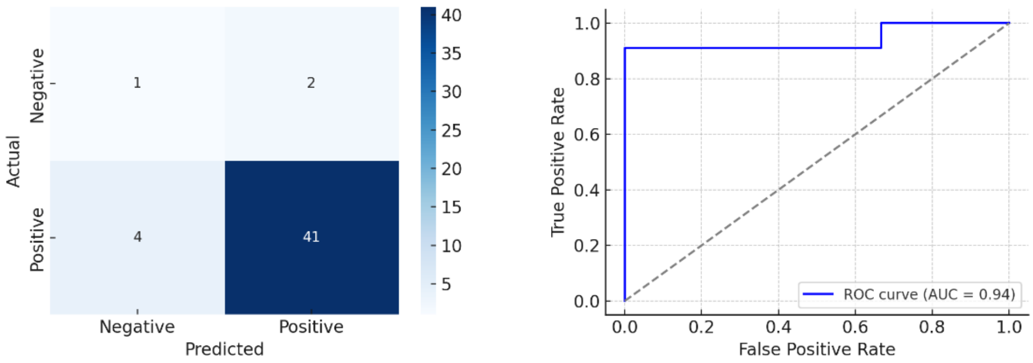
3.4. Face recognition cosine similarity evaluation

The evaluation of the face recognition model in this study was carried out using 48 new facial images that included variations in lighting, angles, and expressions. To evaluate the model performance, cosine similarity was used as the main metric, with a minimum similarity level of 30% as the acceptance threshold. The threshold of 30% was chosen based on research results showing that this level of similarity is still sufficient to uniquely distinguish individuals while still allowing for natural variations in facial expression and lighting. In addition, this value has also been tested in various face recognition scenarios, where the model performance remains optimal without significantly increasing the number of false positives.

The evaluation was conducted using the cosine similarity approach and produced a performance distribution as shown in Figure 10. Figure 10(a) shows the confusion matrix. Out of the 48 images, 41 images were correctly recognised (true positive (TP)), while 2 images were misidentified as other identities (false positive (FP)), and 4 images were not recognised (false negative (FN)). One other image was recorded as true negative (TN). These results provide an overall accuracy of 89.6%, which shows the consistency of the model in dealing with variations in test conditions.

Furthermore, discrimination evaluation was performed using the receiver operating characteristic (ROC) curve as shown in Figure 10(b), with an area under curve (AUC) value of 0.94. This value indicates that the model has a fairly strong ability to distinguish between different identities, although there is still a reasonable margin of error. In addition, the evaluation results can be summarised into performance metrics. The precision rate was 95.3%, calculated from the ratio of correctly recognised faces (TP) to all positive predictions (TP+FP). The recall value was 91.1%, derived from the ratio of correctly recognised faces to all actual face data (TP+FN). Meanwhile, the F1-score value reached 93.1% as a result of the harmonisation between precision and recall. These three metrics show that the model is not only accurate in recognising faces correctly, but also capable of maintaining a balance between prediction accuracy and detection completeness.

The evaluation results demonstrate that the proposed model is robust and capable of maintaining a balanced trade-off between precision and recall. Compared with prior studies, such as ArcFace (52%) [26], and CGAN (76%) [27] the proposed optimization proved to be more accurate, efficient, and stable. The performance also surpassed the study by Ding *et al.* [36], which reported 87.3% accuracy of SSFR using uniform generic representation and deep feature on labeled faces in the wild (LFW). These findings confirm that the integration of anchor box, NMS, and LBP is effective and robust in the context of SSFR, in contrast to most previous works that relied on multi-sample configurations.



(a) (b)

Figure 10. Cosine similarity evaluation of (a) confusion matrix and (b) ROC curve

3.5. Illumination variation evaluation

The system evaluation was conducted in an integrated manner through a mobile application developed for the Android and iOS platforms. Since facial recognition was applied in a SSFR scenario, lighting variations emerged as the main obstacle; the system did not have additional images to adapt to changes in illumination. Therefore, the evaluation was conducted by measuring the model's resilience to these obstacles. The application was tested on 48 images of students from Universitas Duta Bangsa Surakarta, which were not used in the training or testing process. The evaluation results Table 3 show that under normal lighting conditions of around 350-400 lux, the model successfully recognised 43 out of 48 faces with an accuracy of 89.58%. Meanwhile, under low lighting conditions of 50-100 lux, the system was still able to identify 41 faces with an accuracy of 85.42%. The limited difference between these two values indicates that the proposed approach is quite tolerant to illumination variations, even though it is equipped with only one sample per identity. With inference latency in the range of tens of milliseconds on mobile devices, this system not only meets SSFR performance requirements but is also suitable for implementation on mobile platforms with limited power and computing resources.

Table 3. Mobile face recognition lux evaluation

Lux	Number of facial tested	Correct prediction	Accuracy %
350-400	48	43	89.58
50-100	48	41	85.42

4. CONCLUSION

The hybrid model for SSFR integrating anchor, NMS, and LBP improved training accuracy from 47.5-98.0% at epoch 72, with testing accuracy remaining stable in the range of 85-88% (best at 87.9%). Evaluation on 48 new facial images achieved 89.6% accuracy, 95.3% precision, 91.1% recall, 93.1% F1-score, and 0.94 AUC ROC. Practical deployment via Android and iOS-based attendance applications confirmed the transferability of the model with levels between 50 lux and 400 lux, it produces an accuracy of 88.46%. These findings confirm that the proposed framework is not only accurate but also more robust than classical baselines and previous deep learning approaches, while remaining computationally efficient and suitable for real-time biometric authentication in public sector, education, and IoT ecosystems. Future studies should expand dataset diversity and include broader demographic validation to mitigate potential bias and fairness issues in SSFR. Incorporating face alignment, adaptive thresholding, and privacy-preserving techniques such as on-device inference or federated learning may further enhance robustness and security. Additionally, large-scale real-time multi-user evaluations are recommended to assess scalability in mobile and IoT environments.

ACKNOWLEDGMENTS

Gratitude is first expressed to the Ministry of Higher Education, Science, and Technology, Republic of Indonesia, as this paper represents one of the outputs of the Fundamental Research Grant. The authors also extend appreciation to Universitas Duta Bangsa Surakarta, for providing administrative support and facilitation. Special thanks are conveyed to students Lia Ana Pratama and Stanley Chen Ho for their valuable contributions to data collection. Furthermore, the authors acknowledge the use of generative AI (ChatGPT, OpenAI, QuillBoot) to assist in improving the clarity and readability of the manuscript; all generated outputs were carefully reviewed and edited to ensure accuracy and compliance with ethical publication standards.

FUNDING INFORMATION

This research was funded by the Program Hibah Penelitian Fundamental organized by the Ministry of Higher Education, Science, and Technology, Republic of Indonesia, with grant number 127/C3/DT.05.00/PL/2025.

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
Faulinda Ely Nastiti	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Sopingi		✓				✓		✓	✓	✓	✓	✓		
Dedy Hariyadi	✓		✓	✓			✓			✓	✓		✓	✓
Sri Sumarlinda					✓			✓	✓					

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

The authors state that no conflicts of interests exist in this study.

INFORMED CONSENT

We have secured written informed consent from all workers of Universitas Duta Bangsa Surakarta participating in this study. The consent encompasses authorisation to utilise facial data for processing, analysis, and dissemination of research findings in a scientific format. The complete facial recognition data processing is conducted in a regulated and secure manner on the Universitas Duta Bangsa Surakarta Data Science server. Data management is conducted in alignment with the principles of confidentiality and personal data protection.

ETHICAL APPROVAL

Research involving human data in this study complied with all relevant national regulations (Law No. 27 of 2022 on Personal Data Protection) and the institutional policies of Universitas Duta Bangsa Surakarta. All facial data collection procedures were conducted with full respect for subject privacy and were strictly limited to the development of the attendance system and scientific research purposes.

DATA AVAILABILITY

The dataset used in this study cannot be shared publicly as it contains facial data of university employees, which are classified as sensitive personal information and protected under institutional privacy policies. However, the implementation of the face recognition model developed in this research can be accessed through the UDB Dosen mobile application, available at:

- Google Play Store: https://play.google.com/store/apps/details?id=id.ac.udb.dosen.
- Apple App Store: https://apps.apple.com/id/app/udb-dosen/id6443435938?l=id.

REFERENCES

[1] 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.

[2] S. Park, H. Na, and D. Choi, "Verifiable facial de-identification in video surveillance," *IEEE Access*, vol. 12, pp. 67758–67771, 2024, doi: 10.1109/ACCESS.2024.3399230.

[3] P. Zhang, F. Huang, D. Wu, B. Yang, Z. Yang, and L. Tan, "Device-edge-cloud collaborative acceleration method towards occluded face recognition in high-traffic areas," *IEEE Transactions on Multimedia*, vol. 25, pp. 1513–1520, 2023, doi: 10.1109/TMM.2023.3240884.

[4] D. Mouafo and U. Biaou, "Face recognition system for control access to restrictive domain," *International Journal of Safety and Security Engineering*, vol. 12, no. 2, pp. 251–257, Apr. 2022, doi: 10.18280/ijsse.120214.

[5] A. Baobaid, M. Meribout, V. K. Tiwari, and J. P. Pena, "Hardware accelerators for real-time face recognition: a survey," *IEEE Access*, vol. 10, pp. 83723–83739, 2022, doi: 10.1109/ACCESS.2022.3194915.

[6] W. Gao, J. Yu, R. Hao, F. Kong, and X. Liu, "Privacy-preserving face recognition with multi-edge assistance for intelligent security systems," *IEEE Internet of Things Journal*, vol. 10, no. 12, pp. 10948–10958, Jun. 2023, doi: 10.1109/JIOT.2023.3240166.

[7] H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas, and F. Flammini, "A comprehensive review of face recognition techniques, trends, and challenges," *IEEE Access*, vol. 12, pp. 107903–107926, 2024, doi: 10.1109/ACCESS.2024.3424933.

[8] M. Eman, T. M. Mahmoud, M. M. Ibrahim, and T. A. El-Hafeez, "Innovative hybrid approach for masked face recognition using pretrained mask detection and segmentation, robust PCA, and KNN classifier," *Sensors*, vol. 23, no. 15, Jul. 2023, doi: 10.3390/s23156727.

[9] X. Wang, Y. C. Wu, M. Zhou, and H. Fu, "Beyond surveillance: privacy, ethics, and regulations in face recognition technology," *Frontiers in Big Data*, vol. 7, Jul. 2024, doi: 10.3389/fdata.2024.1337465.

[10] T. Jindanil, L. Xu, R. C. Fontenele, M. C. de L. Perula, and R. Jacobs, "Smartphone applications for facial scanning: a technical and scoping review," *Orthodontics & Craniofacial Research*, vol. 27, no. S2, pp. 65–87, Dec. 2024, doi: 10.1111/ocr.12821.

[11] G. Srivastava and S. Bag, "Modern-day marketing concepts based on face recognition and neuro-marketing: a review and future research directions," *Benchmarking: An International Journal*, vol. 31, no. 2, pp. 410–438, Feb. 2024, doi: 10.1108/BIJ-09-2022-0588.

[12] T. J. Hartmann, J. B. J. Hartmann, U. F.-Hoffmann, C. Lato, W. Janni, and K. Lato, "Novel method for three-dimensional facial expression recognition using self-normalizing neural networks and mobile devices," *Geburtshilfe und Frauenheilkunde*, vol. 82, no. 9, pp. 955–969, Sep. 2022, doi: 10.1055/a-1866-2943.

[13] N. Yashvardhini, D. K. Jha, and S. Bhattacharya, "Identification and characterization of mutations in the SARS-CoV-2 RNA-dependent RNA polymerase as a promising antiviral therapeutic target," *Archives of Microbiology*, vol. 203, no. 9, pp. 5463–5473, Nov. 2021, doi: 10.1007/s00203-021-02527-9.

[14] T. Tsai and P. Chi, "A single-stage face detection and face recognition deep neural network based on feature pyramid and triplet loss," *IET Image Processing*, vol. 16, no. 8, pp. 2148–2156, Jun. 2022, doi: 10.1049/ipr2.12479.

[15] M. S. E. Saadabadi, S. R. Malakshan, A. Zafari, M. Mostofa, and N. M. Nasrabadi, "A quality aware sample-to-sample comparison for face recognition," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2023, pp. 6118–6127. doi: 10.1109/WACV56688.2023.00607.

[16] F. Li, T. Yuan, Y. Zhang, and W. Liu, "Face recognition in single sample per person fusing multi-scale features extraction and virtual sample generation methods," *Frontiers in Applied Mathematics and Statistics*, vol. 8, Apr. 2022, doi: 10.3389/fams.2022.869830.

[17] A. Farkhod, A. B. Abdusalomov, M. Mukhiddinov, and Y.-I. Cho, "Development of real-time landmark-based emotion recognition CNN for masked faces," *Sensors*, vol. 22, no. 22, Nov. 2022, doi: 10.3390/s22228704.

[18] M. Kowalski, A. Grudzień, and K. Mierzejewski, "Thermal–visible face recognition based on CNN features and triple triplet configuration for on-the-move identity verification," *Sensors*, vol. 22, no. 13, Jul. 2022, doi: 10.3390/s22135012.

[19] L. He, L. He, and L. Peng, "CFormerFaceNet: efficient lightweight network merging a CNN and transformer for face recognition," *Applied Sciences*, vol. 13, no. 11, May 2023, doi: 10.3390/app13116506.

[20] P. Alirezazadeh, F. Dornaika, and A. Moujahid, "A deep learning loss based on additive cosine margin: application to fashion style and face recognition," *Applied Soft Computing*, vol. 131, Dec. 2022, doi: 10.1016/j.asoc.2022.109776.

[21] S. Bajpai, G. Mishra, R. Jain, D. K. Jain, D. Saini, and A. Hussain, "RI-L_1 approx: a novel Resnet-Inception-based Fast L_1-approximation method for face recognition," *Neurocomputing*, vol. 613, Jan. 2025, doi: 10.1016/j.neucom.2024.128708.

[22] H. Chen, H. Li, Y. Li, and C. Chen, "Sparse spatial transformers for few-shot learning," *Science China Information Sciences*, vol. 66, no. 11, Nov. 2023, doi: 10.1007/s11432-022-3700-8.

[23] Y. Song, T. Wang, P. Cai, S. K. Mondal, and J. P. Sahoo, "A comprehensive survey of few-shot learning: evolution, applications, challenges, and opportunities," *ACM Computing Surveys*, vol. 55, no. 13s, pp. 1–40, Dec. 2023, doi: 10.1145/3582688.

[24] W. Jiang, K. Huang, J. Geng, and X. Deng, "Multi-scale metric learning for few-shot learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 1091–1102, Mar. 2021, doi: 10.1109/TCSVT.2020.2995754.

[25] A. R. Butt, Z. U. Rahman, A. U. Haq, B. Ahmed, and S. Manzoor, "Unconstrained face recognition using infrared images," *International Journal of Image and Graphics*, vol. 25, no. 6, Nov. 2025, doi: 10.1142/S0219467825500561.

[26] A. Firmansyah, T. F. Kusumasari, and E. N. Alam, "Comparison of face recognition accuracy of ArcFace, Facenet and Facenet512 models on deepface framework," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, Feb. 2023, pp. 535–539. doi: 10.1109/ICCoSITE57641.2023.10127799.

[27] M. A. Iqbal, W. Jadoon, and S. K. Kim, "Synthetic image generation using conditional GAN-provided single-sample face image," *Applied Sciences*, vol. 14, no. 12, Jun. 2024, doi: 10.3390/app14125049.

[28] F. P.-Montes, J. O.-Mercado, G. S.-Perez, G. B.-Garcia, L. P.-Tixteco, and O. L.-Garcia, "Analysis of real-time face-verification methods for surveillance applications," *Journal of Imaging*, vol. 9, no. 2, Jan. 2023, doi: 10.3390/jimaging9020021.

[29] K. Han *et al.*, "GhostNets on heterogeneous devices via cheap operations," *International Journal of Computer Vision*, vol. 130, no. 4, pp. 1050–1069, Apr. 2022, doi: 10.1007/s11263-022-01575-y.

[30] M. Alansari, O. A. Hay, S. Javed, A. Shoufan, Y. Zweiri, and N. Werghi, "GhostFaceNets: lightweight face recognition model from cheap operations," *IEEE Access*, vol. 11, pp. 35429–35446, 2023, doi: 10.1109/ACCESS.2023.3266068.

[31] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local gradient hexa pattern: a descriptor for face recognition and retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 171–180, Jan. 2018, doi: 10.1109/TCSVT.2016.2603535.

[32] V. N. T. Le, S. Ahderom, B. Apopei, and K. Alameh, "A novel method for detecting morphologically similar crops and weeds based on the combination of contour masks and filtered local binary pattern operators," *GigaScience*, vol. 9, no. 3, Mar. 2020, doi: 10.1093/gigascience/giaa017.
[33] J. Raghavan and M. Ahmadi, "Performance evaluation of entropy based LBP for face recognition," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2021, pp. 241–245. doi: 10.1109/MWSCAS47672.2021.9531690.
[34] X. Wei, L. Yu, S. Tian, P. Feng, and X. Ning, "Underwater target detection with an attention mechanism and improved scale," *Multimedia Tools and Applications*, vol. 80, no. 25, pp. 33747–33761, Oct. 2021, doi: 10.1007/s11042-021-11230-2.
[35] I. H. Barnhart, S. Lancaster, D. Goodin, J. Spotanski, and J. A. Dille, "Use of open-source object detection algorithms to detect Palmer amaranth *(Amaranthus palmeri)* in soybean," *Weed Science*, vol. 70, no. 6, pp. 648–662, Nov. 2022, doi: 10.1017/wsc.2022.53.
[36] Y. Ding, F. Liu, Z. Tang, and T. Zhang, "Uniform generic representation for single sample face recognition," *IEEE Access*, vol. 8, pp. 158281–158292, 2020, doi: 10.1109/ACCESS.2020.3017479.

BIOGRAPHIES OF AUTHORS

Faulinda Ely Nastiti received the Ph.D. degree in Information Technology. she was a certified data center professional and big data certified associate. She is a senior lecturer and researcher at Universitas Duta Bangsa Surakarta, Indonesia. She has authored more than 70 research papers published in journals and conferences. Her areas of interest include machine learning, deep learning, and intelligence systems. She can be contacted at email: faulinda_ely@udb.ac.id.

Sopingi is a senior lecturer and researcher at Universitas Duta Bangsa Surakarta, Indonesia. He earned his master's degree from Universitas Amikom Yogyakarta, specializing in Web Services. He holds a professional competency certification in Programming. His research interests include information system development and web services. He has authored 49 research publications in journals and conference proceedings. He is currently a doctoral student from Hasanudin University, South Sulawesi. He can be contacted at email: sopingi@gmail.com.

Dedy Hariyadi is currently pursuing his doctoral degree at Universitas Gadjah Mada, Yogyakarta, Indonesia. He is a researcher in the field of cybersecurity and holds the cybersecurity awareness professional certification. His research interests include privacy, data security, and the modeling of machine learning in cybersecurity. He has authored 65 publications in academic journals and conferences. He can be contacted at email: dedy@unjaya.ac.id.

Sri Sumarlinda is currently pursuing her doctoral studies at Universiti Kuala Lumpur, Malaysia. She holds a professional competency certification from the Data Science Association. Her research focuses on the development of machine learning models. She is an active researcher with 63 publications in international journals and conferences. She can be contacted at email: sumarlinda.sri@s.unikl.edu.my.