 ISSN 2579-7298

Performance of Deep Face Recognition Models under Adaptive Margin Loss: A Real-Time Evaluation

Kevin Muhammad Tegar Aditama [a,1], Anan Nugroho [b,2,*], Subiyanto [b,3], Arthur Gregorius Pongoh [b,4]

[a] Computer Engineering, Faculty of Engineering, Universitas Negeri Semarang, Indonesia
[b] Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Semarang, Indonesia
[b] Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Semarang, Indonesia
[b] Computer Engineering, Faculty of Engineering, Universitas Negeri Semarang, Indonesia
[1] kevinadittama@students.unnes.ac.id ; [2] anannugroho@mail.unnes.ac.id [*] ; [3] subiyanto@mail.unnes.ac.id : [4] arthurgregorius333@students.unnes.ac.id
* corresponding author

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Face recognition
Deep learning
Biometrics
Backbone networks
Open-Set Recognition

ABSTRACT

Real-time face recognition systems encounter a critical trade-off between high-security demands and computational efficiency, particularly when deployed in unconstrained open-set environments. This study presents a comprehensive benchmarking of four distinct deep learning backbones ResNet100, GhostFaceNets, LAFS, and TransFace specifically trained using the Adaptive Margin Loss (AdaFace) function to handle image quality variations. The primary objective is to identify the optimal architecture for secure attendance systems operating on standard hardware with limited training data. The evaluation protocol employs a rigorous real-world open-set test to quantify performance using False Acceptance Rate (FAR) and False Rejection Rate (FRR). The experimental results demonstrate that ResNet100 establishes the highest security standard, achieving a 0.00% FAR at strict thresholds. Meanwhile, GhostFaceNets emerges as the most balanced solution for resource-constrained deployments, delivering competitive accuracy above 93% with significantly lower computational complexity. Conversely, the Vision Transformer (TransFace) fails to generalize in this low-data regime, resulting in unacceptable false acceptance rates. These findings definitively recommend GhostFaceNets for efficient edge-based implementations, while ResNet100 remains the superior choice for mission-critical security applications.

Introduction

Face recognition technology has witnessed substantial progress in recent years, catalyzed by breakthroughs in artificial intelligence (AI) and deep learning [1], [2], [3]. Convolutional Neural Networks (CNNs), particularly ResNet architectures [4], have established themselves as a strong baseline for feature extraction due to their ability to learn robust representations from large-scale datasets [4], [5], [6]. However, a significant disparity remains between laboratory performance and real-world deployment [7], [8], [9]. In practical applications such as real-time attendance systems, the model must operate in an unconstrained open-set environment [10], [11], where it encounters unregistered individuals (impostors) and varying image qualities due to uncontrolled lighting or motion blur [5], [7], [10], [12]. Moreover, deploying deep models on standard computing units (CPUs) for real-time inference poses an additional challenge in balancing recognition accuracy with computational efficiency [13], [14], [15].

To improve robustness in unconstrained environments, margin-based loss functions such as ArcFace [16] have been widely adopted to strengthen discriminative face embeddings by increasing

intra-class compactness and inter-class separation. However, ArcFace applies a fixed margin uniformly across samples, which can be suboptimal when input quality varies in real deployments (e.g., low illumination, blur, or partial occlusion), as commonly observed in real-time attendance systems. To address this limitation, AdaFace [17] introduces a quality-adaptive margin mechanism that modulates the margin according to quality proxies derived from the feature representation, aiming to improve recognition reliability under heterogeneous capture conditions. Consequently, evaluation should not rely solely on closed-set accuracy; instead, deployment-oriented open-set behavior must be assessed through security-related error trade-offs (e.g., FAR/FRR) at specific decision thresholds [18].

In addition to loss design, the selection of backbone architecture is critical for real-time deployment. Deep CNN backbones such as ResNet variants remain strong baselines for face representation learning [4], yet their computational cost can be challenging for CPU-based real-time applications and surveillance-style verification settings [8], [10]. To address efficiency constraints, lightweight architectures such as GhostFaceNets leverage cheap operations to reduce inference complexity while maintaining competitive recognition performance [13], and subsequent variants further improve efficiency-accuracy trade-offs through architectural refinements [15], [19]. Beyond CNN efficiency, landmark-aware approaches such as LAFS incorporate geometric cues and attention to improve alignment and robustness under pose variations [20]. More recently, transformer-based models introduce a global feature learning paradigm through self-attention [21], supported by broader vision transformer surveys [22] and data efficient training strategies [23]. In face recognition specifically, TransFace adapts transformer training from a data-centric perspective to improve stability and performance [24], while recent comparative studies highlight the practical trade-offs between ViT-based and CNN-based face recognition under different training regimes [25]. However, despite the rapid growth of these heterogeneous backbone paradigms, a deployment-oriented benchmark that systematically compares them under a unified quality-adaptive margin loss and evaluates open-set security behavior (e.g., FAR/FRR sensitivity to operating thresholds), particularly when trained from scratch on limited real-world datasets, remains under-explored [6], [9], [11], [18].

This study establishes a deployment-oriented benchmark for real-time face recognition by evaluating four representative backbone architectures: ResNet100, GhostFaceNets, LAFS, and TransFace [4], [13], [20], [21] trained under a unified quality-adaptive margin loss (AdaFace) [17]. In contrast to evaluations that emphasize closed-set accuracy alone, we adopt an open-set perspective that reflects practical access-control and attendance requirements by analyzing security-related error trade-offs, including False Acceptance Rate (FAR) and False Rejection Rate (FRR), across operating decision thresholds [11], [18]. In addition, considering the computational constraints commonly encountered in real-time deployments, we relate recognition performance to efficiency considerations relevant to CPU/edge settings [10], [19], [26].

The novelty of this work lies in providing a deployment-oriented benchmark that jointly analyzes quality-adaptive margin loss training (AdaFace) and heterogeneous backbone architectures using open-set security metrics (FAR/FRR) under real-time CPU/edge constraints.

The main contributions of this paper are as follows:

- A unified benchmarking protocol that compares heterogeneous backbone paradigms under the same AdaFace objective, enabling fairer performance attribution across architectures[17].
- An open-set evaluation that quantifies security-usability trade-offs using FAR/FRR sensitivity to operating thresholds, which is critical for real-world attendance and access-control deployment [11], [18].
- Deployment-relevant insights by discussing the performance-efficiency trade-off of each backbone for real-time operation on resource-constrained platforms [10], [19], [26].

Method

Figure 1 summarizes the end-to-end workflow adopted in this study for deployment-oriented benchmarking of deep face recognition in a real-time open-set setting. The process starts with data acquisition and dataset construction, where images are organized into a gallery of registered identities and a probe set representing real-time queries, which may include both genuine users and impostors.

Next, preprocessing and standardization are applied via face detection/alignment followed by resizing and normalization to ensure consistent inputs across all models. Each backbone is then trained under a unified configuration to learn discriminative embeddings, which are subsequently used for similarity-based verification. Performance is evaluated through a threshold-based accept/reject rule across three operating points (Strict/Moderate/Loose) to analyze security usability trade-offs, quantified using open-set metrics such as FAR and FRR [11], [18]. Finally, results are analyzed comparatively across backbones and operating thresholds; the training–evaluation loop is repeated for each backbone architecture to ensure fair and reproducible comparison. No questionnaire-based assessment is involved, and the term "decision" refers strictly to the automated threshold-based verification outcome.

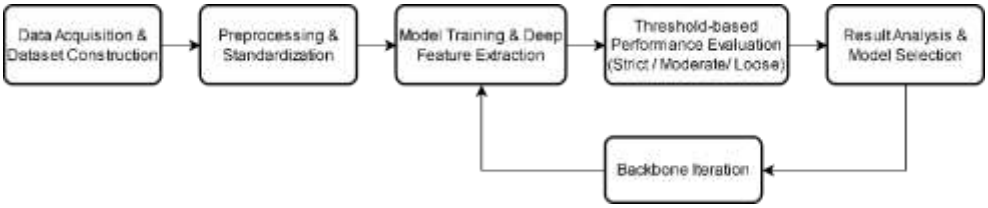

Figure 1. Structural workflow of the comparative benchmarking methodology.

2.1. Data Acquisition and Dataset Construction

The experimental data were collected from a real-world biometric attendance environment to approximate practical deployment conditions. Image acquisition was performed using a standard webcam positioned at an approximate distance of 20–40 cm from the subject, reflecting a typical check-in scenario [8], [10]. The dataset comprises 46 unique identities and includes natural variations in head pose (yaw and pitch), facial expression, and illumination (natural and artificial) to represent unconstrained capture conditions [7].

To support open-set verification, the dataset was partitioned into a Gallery and a Probe subset. The Gallery serves as the enrollment/reference database and contains curated images of 36 registered identities. The Probe subset contains real-time captures used for evaluation and is divided into (i) Genuine samples, i.e., new images of the registered identities, and (ii) Impostor samples from 10 unregistered identities not present in the Gallery. This configuration enables measuring security-related errors in open-set operation, particularly false acceptances (FAR) when impostors are incorrectly accepted [11], [18]. The dataset distribution is summarized in Table 1.

Table 1. The distribution of the private dataset for open-set evaluation

Dataset subset	Category	Number of Subjects	Description	Role in Experiment
Gallery	Registered	36	Curated images of authorized personnel	Reference Database (Enrollment)
Probe	Genuine	36	Real-time captures of registered users	Validating True Positives (TP)
	Impostor	10	Unregistered individuals (Unknown)	Evaluating False Acceptances (FP)
Total		46	Total unique identities	Open-Set Benchmark

The private dataset used in this study contains 46 identities and is intended to reflect low-data attendance deployments rather than large-scale public benchmarks. Accordingly, the results should be interpreted as controlled comparative evidence across backbones and operating thresholds, not as

population-level performance estimates [3], [6]. Because FAR/FRR are computed from a finite number of genuine and impostor trials, rate uncertainty increases when the number of trials is limited; therefore, we complement threshold-specific results with ROC/AUC and DET-style analyses and recommend site-specific threshold recalibration prior to deployment [11], [18], [27], [28].

2.2. Preprocessing and Standardization

To ensure a fair comparison across backbone architectures, all raw images were processed using the same standardized preprocessing pipeline consisting of face detection, alignment, cropping, resizing, and normalization [5], [6], [7]. Face regions were detected using Multi-task Cascaded Convolutional Networks (MTCNN) due to its robustness under pose and illumination variations [29]. Using the facial landmarks provided by MTCNN (e.g., eyes, nose, and mouth corners), each detected face was aligned via a similarity transformation to reduce geometric discrepancies caused by head pose changes. The aligned faces were then cropped and resized to 112 × 112 pixels to match the input specification of the evaluated models. Finally, pixel intensities were normalized to the range [-1, 1] to stabilize optimization and improve convergence during training and inference. This uniform preprocessing protocol ensures that performance differences observed in subsequent experiments can be attributed primarily to the backbone architecture and training objective rather than inconsistencies in input preparation.

2.3. Proposed Framework (Deep Learning Architectures and Loss Functions)

This study establishes a comparative benchmarking framework by evaluating four backbone architectures that represent different design paradigms in deep face recognition: ResNet100 [4], GhostFaceNets [13], LAFS [20], and TransFace [24]. ResNet100 is employed as a strong baseline due to its deep residual learning formulation [4]. GhostFaceNets represent lightweight CNNs designed for efficiency through cheap operations [13], while LAFS integrates landmark-aware attention mechanisms to emphasize geometrically informative facial regions [20]. TransFace represents a transformer-based face recognition approach that leverages self-attention to capture global contextual information [21], [24].

To address the challenges of face recognition in unconstrained environments, we employ the AdaFace (Adaptive Margin Loss) function [17]. Unlike standard margin-based losses such as ArcFace [16] which apply a fixed penalty (m) to all samples regardless of their quality, AdaFace adapts the margin based on image quality. Image quality is approximated using the feature norm $\|z_i\|$. AdaFace emphasizes hard samples for high-quality images while ignoring unlearnable samples for low-quality images to prevent overfitting. The loss function is formulated as:

$$L_{Ada} = -\frac{1}{n}\sum_{I=1}^{N} log \frac{e^{s(\cos(\theta_{yi}+g_{angle}))-g_{add}}}{e^{s(\cos(\theta_{yi}+g_{angle}))-g_{add}} + \sum_{j\neq yi} e^{s\,cos\theta_j}} \quad (1)$$

where θ_{yi} is the angle between the feature and the class center, s is the scaling parameter, and N is the batch size. The terms g_{angle} and g_{add} are adaptive margin functions controlled by the feature norm $\|z_i\|$. Specifically, if the image quality is high, the margin concentrates on angular component (g_{angle}), whereas for low-quality images, it shifts towards additive margin (g_{add}) to stabilize the gradient [17].

For fair benchmarking, all backbone models were trained from scratch with randomly initialized weights (i.e., without external pretrained weights) under the same experimental protocol, input resolution (112 × 112), embedding dimension (512), and AdaFace objective [17]. During training, each input image is mapped to a 512-dimensional embedding vector by the backbone network, and the AdaFace classification head produces logits for supervised optimization. After training, only the backbone is retained for feature extraction in the verification stage.

For similarity-based verification, cosine similarity is computed between a probe embedding and the gallery embeddings, and the maximum similarity score is used as the matching score. The final accept/reject decision is obtained by comparing the matching score against the operating threshold.

To reflect deployment requirements, decision behavior is later analyzed across multiple operating points (Strict/Moderate/Loose) using open-set security metrics such as FAR and FRR [11], [18].

2.4. Experimental Setup

The hardware specifications include an Intel Core i9-13900K processor and 32 GB DDR5 RAM running on Windows. Although a discrete GPU was available, the training and inference processes were explicitly executed on the CPU to simulate resource-constrained environments typical of low-cost attendance systems. The implementation was built using the PyTorch framework [30], with NumPy [31] and scikit-learn [32] utilized for matrix operations and performance metric calculations.

Input images were resized to 112 × 112 pixels and normalized to the range [-1, 1]. To enhance model generalization, random horizontal flip augmentation was applied during training. The training process was conducted for 125 epochs with a batch size of 16. Distinct optimization strategies were employed tailored to architecture type: stochastic gradient descent (SGD) with momentum 0.9 and weight decay 5×10^{-4} was used for CNN-based models (ResNet, GhostFaceNets, and LAFS), while AdamW with weight decay 0.05 was utilized for the transformer-based model (TransFace) to ensure stable convergence [24]. A MultiStepLR scheduler was applied for CNNs, whereas a CosineAnnealingLR scheduler was adopted for TransFace. Furthermore, gradient clipping (max norm = 5.0) was implemented as a stabilization measure, particularly for lightweight models, to prevent gradient explosion and ensure consistent convergence.

2.5. Evaluation Metrics

To quantitatively assess the robustness of the proposed system, we employ an evaluation protocol based on the confusion matrix adapted to an open-set face verification scenario. In this setting, the system must distinguish between genuine users (registered in the gallery) and impostors (unregistered identities). Given a probe image, the model produces a 512-dimensional embedding and computes the cosine similarity against the gallery embeddings. The highest similarity score is compared with a decision threshold to generate an automated accept/reject outcome. No questionnaire based or human subjective assessment is used; the term “decision” strictly refers to the automated accept/reject output produced by the similarity-threshold rule .Based on this threshold rule, prediction outcomes are categorized into four cases as summarized in Table 2: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN).

Table 2. Confusion matrix for open-set face verification (threshold-based decision)

Predicted: Accepted	True Positive (TP)	Actual: Unregistered (Impostor)
(Score > Threshold)	Correctly accepted registered user.	False Positive (FP)Incorrectly accepted impostor (Security Breach).
Predicted: Rejected	False Negative (FN)	True Negative (TN)
(Score < Threshold)	Incorrectly rejected registered user (Inconvenience).	Correctly rejected impostor.
	Actual: Registered (Genuine)	Actual: Unregistered (Impostor)

Using these outcomes, we compute security- and usability-related metrics. The False Acceptance Rate (FAR) measures the proportion of impostor attempts that are incorrectly accepted, while the False Rejection Rate (FRR) measures the proportion of genuine attempts that are incorrectly rejected. The corresponding formulations are defined as:

$$\text{FAR}\frac{FP}{FP+TN},\ FRR = \frac{FN}{FN+TP} \quad (2)$$

In addition, overall accuracy is reported for completeness:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Threshold selection rationale. In operational face verification, the decision threshold τ defines the system operating point and directly controls the security usability trade off: a stricter τ reduces false acceptances (lower FAR) at the cost of potentially increasing false rejections (higher FRR), while a looser τ tends to increase FAR but improves user convenience. In this study, the three operating points were instantiated using cosine-similarity thresholds $\tau \in \{0.70, 0.50, 0.30\}$ for Strict, Moderate, and Loose settings, respectively The values were selected through score calibration on development/validation trials by examining the separation between genuine and impostor score distributions, consistent with common biometric evaluation and deployment practice.[27], [28]

Results and Discussion

3.1. Internal Benchmark Analysis (Closed-Set Performance)

The first phase of evaluation assesses the fundamental learning capacity of each backbone architecture under the AdaFace loss function using an internal closed-set protocol. In this scenario, the probe images contain only registered (enrolled) identities, so the evaluation focuses on false rejection behavior (FRR) and its complementary genuine acceptance accuracy, while FAR is not applicable in this phase. Performance across three operating thresholds (Strict, Moderate, and Loose) is summarized in Table 3.

Table 3. Comparison of verification accuracy and FRR on internal closed-set benchmark using AdaFace

Backbone Architecture	Threshold	Accuracy (%)	FRR (%)
GhostFaceNets	Strict	95.92	4.08
	Moderate	98.51	1.49
	Loose	98.71	1.29
LAFS (Landmark-Aware)	Strict	95.43	4.57
	Moderate	98.51	1.49
	Loose	99.20	0.80
ResNet100	Strict	96.62	3.38
	Moderate	98.71	1.29
	Loose	99.40	0.60
TransFace (ViT)	Strict	5.47	94.53
	Moderate	5.47	94.53
	Loose	5.47	94.53

To improve comparability across backbones, Figure 2 visualizes the performance trends across threshold levels for (a) Accuracy and (b) FRR. For the CNN-based backbones (ResNet100, LAFS, and GhostFaceNets), loosening the threshold generally increases accuracy and reduces FRR, indicating a more permissive acceptance of genuine matches. Among all models, ResNet100 consistently demonstrates the most stable performance, achieving the highest accuracy and the lowest FRR across all threshold levels (e.g., reaching 99.40% accuracy with 0.60% FRR under the Loose setting). LAFS follows closely, showing strong performance and a notable reduction in FRR as the

threshold becomes looser, while GhostFaceNets remains competitive with slightly higher FRR under the Strict setting.

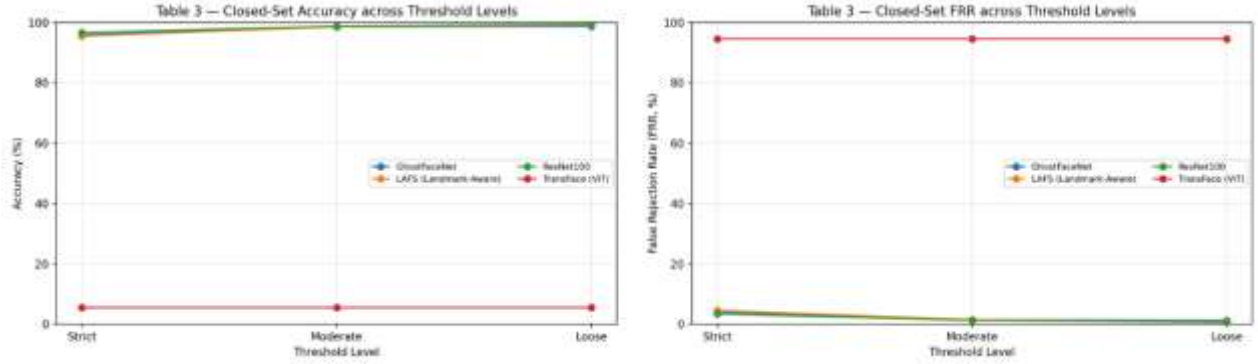

Figure 2. Comparative closed-set performance across threshold levels: (a) Accuracy and (b) FRR (side-by-side).

In contrast, TransFace exhibits a severe performance degradation in the internal benchmark, with extremely low accuracy and consistently high FRR across all thresholds. This behavior suggests a failure to generalize under the current training/data regime, thus TransFace is not recommended for subsequent deployment-oriented evaluation in this study. Overall, the closed-set results indicate that ResNet100 provides the most reliable identification performance, while LAFS and GhostFaceNets offer viable alternatives depending on resource constraints and tolerance to FRR.

3.2. Real-World Robustness Evaluation (Open-Set Analysis)

This section evaluates deployment-oriented robustness in an open-set face verification setting, where the system must accept enrolled identities while rejecting non-enrolled individuals. In contrast to closed-set benchmarking, performance reflects a security–usability trade-off captured by FAR and FRR. We evaluate three operating thresholds (Strict, Moderate, and Loose) to assess operating-point sensitivity. Table 4 summarizes results across backbones and thresholds.

Table 4. Comprehensive security performance (Accuracy, FAR, FRR) across three threshold levels under AdaFace

Backbone	Threshold Level	Accuracy (%)	FAR (%) ↓	FRR (%) ↓
ResNet100	Strict	97.00	0.00	4.00
	Moderate	89.00	37.00	2.33
	Loose	73.50	99.00	2.33
LAFS	Strict	96.50	1.00	4.33
	Moderate	92.25	23.00	2.67
	Loose	74.25	98.00	1.67
GhostFaceNets	Strict	93.75	10.00	5.00
	Moderate	84.50	52.00	3.33
	Loose	72.50	100.00	3.33
TransFace	Strict	6.50	100.00	91.33
	Moderate	6.50	100.00	91.33
	Loose	6.50	100.00	91.33

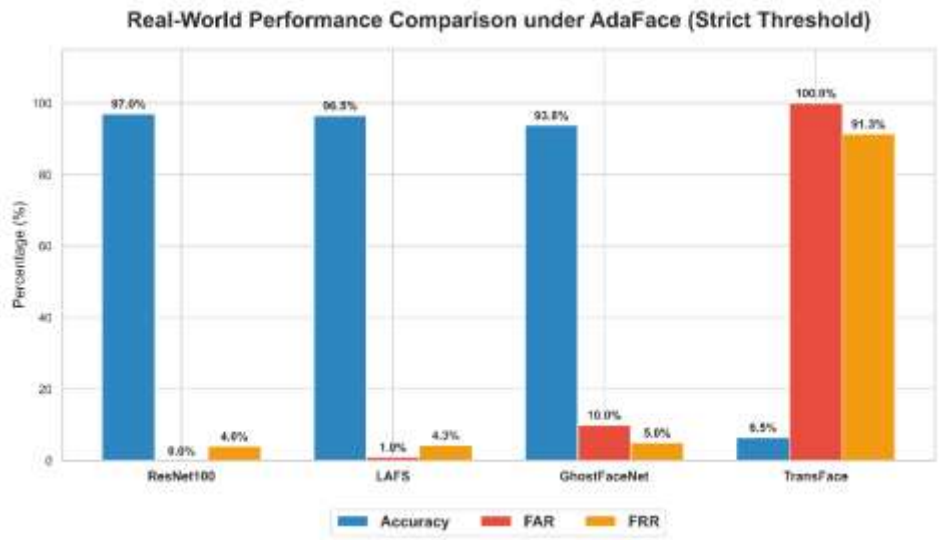

Figure 3. Real-World Performance Comparison (Strict Threshold)

Under the Strict setting, ResNet100 achieves the lowest FAR while maintaining high accuracy and low FRR, indicating the most favorable security profile in this study's deployment-oriented setting. LAFS remains competitive with similarly low FAR and strong accuracy, suggesting that landmark-aware attention can preserve discriminative features under tighter decision boundaries. GhostFaceNets shows a higher FAR even under Strict conditions, indicating greater susceptibility to false acceptances. TransFace exhibits extremely high error rates across metrics, suggesting poor generalization under the current training/data regime.

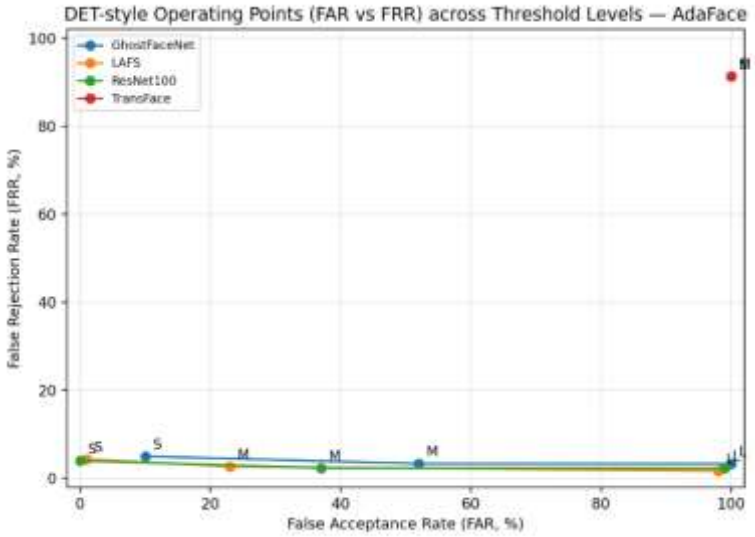

Figure 4. DET-style operating-point plot (FAR vs FRR) for each backbone across threshold levels (S=Strict, M=Moderate, L=Loose).

Figure 4 presents DET-style operating points by plotting FAR (x-axis) against FRR (y-axis) for each backbone, with points corresponding to Strict (S), Moderate (M), and Loose (L). This view summarizes the security usability trade-off across operating thresholds: relaxing the threshold typically reduces FRR but can sharply increase FAR, which is undesirable for security-sensitive attendance systems. Therefore, backbones whose operating points remain in the low-FAR region while keeping FRR acceptable are preferable.

Overall, ResNet100 and LAFS exhibit the most stable balance across thresholds, maintaining low FAR under stricter settings while keeping FRR manageable. GhostFaceNets shows higher FAR growth as thresholds loosen, indicating that stricter calibration is required for deployment. TransFace

remains unsuitable under the evaluated low-data, from-scratch training regime due to consistently poor operating behavior.

3.3. ROC Curve and Stability Analysis

While FAR and FRR provide performance snapshots at specific thresholds, the Receiver Operating Characteristic (ROC) curve offers a holistic view of a model's discriminative ability across all possible operating thresholds in the open-set verification setting. The Area Under the Curve (AUC) serves as a scalar measure of stability, where values closer to 1.0 indicate better separability between genuine and impostor distributions. Figure 5 illustrates the ROC curves for the four backbone architectures trained under AdaFace.

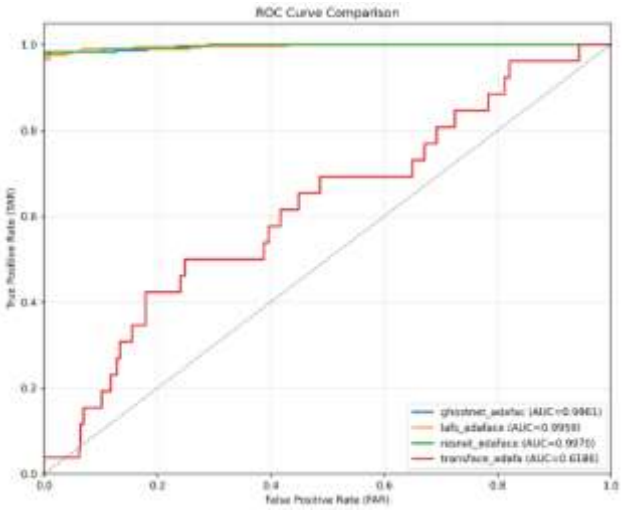

Figure 5. ROC Curve for AdaFace

As depicted in Figure 5, the ROC curves for ResNet100, GhostFaceNets, and LAFS rise sharply toward the top-left corner, indicating high sensitivity (True Positive Rate) even at low False Positive Rates. Quantitatively, ResNet100 achieves a near-perfect AUC of 0.9970, indicating strong and stable discriminative performance under a high-security backbone. Remarkably, the lightweight GhostFaceNets achieves a competitive AUC of 0.9961, suggesting that well-designed lightweight CNNs can learn highly separable manifolds suitable for real-time verification, even if their strict-threshold FAR is higher than ResNet100. LAFS obtains the highest AUC of 0.9959, supporting the effectiveness of landmark-aware attention in preserving discriminative facial cues.

Table 5. Comparison of Area Under Curve (AUC) scores

Backbone Architecture	AUC Score
ResNet100	0.9970
GhostFaceNets	0.9961
LAFS	0.9959
TransFace (ViT)	0.6186

Table 5 summarizes the comparison of AUC scores across models. Conversely, the TransFace (ViT) curve remains close to the diagonal line, resulting in a poor AUC of 0.6186. This visualization indicates that, under the current training/data regime, TransFace approaches random performance and fails to establish a clear margin between genuine and impostor distributions.

Conclusion

This study presented a deployment-oriented benchmark of four deep face recognition backbones (ResNet100, GhostFaceNets, LAFS, and TransFace) trained under a unified AdaFace objective for

real-time open-set verification. Across both closed-set and open-set evaluations, several conclusions can be drawn.

First, ResNet100 demonstrates the strongest security profile. Under the Strict operating point, it achieves 0.00% FAR while maintaining high accuracy (97.00%), indicating robust impostor rejection under unconstrained conditions. Second, GhostFaceNets provides a practical accuracy efficiency trade-off for resource-constrained deployments. Although lightweight, it remains competitive under Strict settings (93.75% accuracy), but its FAR (10.00%) suggests that stricter calibration is required when security constraints are stringent. Third, TransFace exhibits poor generalization under the low-data, from-scratch training regime, leading to consistently high error rates (including 100% FAR across thresholds), and is therefore unsuitable under the evaluated configuration.

Overall, ResNet100 is recommended for security-critical access-control scenarios where minimizing false acceptances is paramount. For CPU/edge-based attendance systems where efficiency is prioritized, GhostFaceNets is a viable alternative provided that thresholds are tuned to control FAR. Given the limited dataset scale in this study, these findings should be interpreted as comparative evidence under the same protocol; future work will investigate transfer learning or pretraining to improve transformer convergence and explore hybrid designs to further optimize the security efficiency trade-off.

Acknowledgment

The authors gratefully acknowledge the AIRIoT Laboratory (Artificial Intelligence, Robotics, and Internet of Things) at the Department of Electrical Engineering, Universitas Negeri Semarang, for providing the high-performance computational resources and facilities required to conduct this study.

Declarations

Author contribution. Author contributions are reported in accordance with the Contributor Roles Taxonomy (CRediT) guidelines.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kevin Muhammad Tegar Aditama	✓	✓	✓		✓	✓		✓	✓	✓	✓			
Anan Nugroho	✓	✓		✓			✓			✓		✓	✓	
Subiyanto	✓	✓		✓						✓		✓	✓	
Arthur Gregorius Pongoh	✓	✓	✓	✓	✓									

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

Funding statement. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

Data and Software Availability Statements

The source code, training scripts, and benchmark protocols used in this study are openly available at https://github.com/KevinAdittama/skripsi to facilitate reproducibility. The private face dataset generated during the current study involves sensitive biometric information and is not publicly available to ensure the privacy of the participants. However, the data supporting the findings of this study are available from the corresponding author upon reasonable request

References

[1] M. Hassaballah and S. Aly, “Face recognition: Challenges, achievements and future directions,” Aug. 01, 2015, *Institution of Engineering and Technology*. doi: 10.1049/iet-cvi.2014.0084.

[2] Y. Xu *et al.*, "Artificial intelligence: A powerful paradigm for scientific research," Nov. 28, 2021, *Cell Press*. doi: 10.1016/j.xinn.2021.100179.

[3] A. K. . Jain, P. J. . Flynn, and A. A. . Ross, *Handbook of biometrics*. Springer, 2008.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385

[5] B. Prihasto *et al.*, "A survey of deep face recognition in the wild," in *2016 International Conference on Orange Technologies (ICOT)*, 2016, pp. 76–79. doi: 10.1109/ICOT.2016.8278983.

[6] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021, doi: 10.1016/j.neucom.2020.10.081.

[7] A. Zhalgas, B. Amirgaliyev, and A. Sovet, "Robust Face Recognition Under Challenging Conditions: A Comprehensive Review of Deep Learning Methods and Challenges," Sep. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/app15179390.

[8] H. Lee, S. H. Park, J. H. Yoo, S. H. Jung, and J. H. Huh, "Face recognition at a distance for a stand-alone access control system," *Sensors (Switzerland)*, vol. 20, no. 3, Feb. 2020, doi: 10.3390/s20030785.

[9] P. Chitrapu, M. Kumar Morampudi, and H. Kumar Kalluri, "Robust Face Recognition Using Deep Learning and Ensemble Classification," *IEEE Access*, vol. 13, pp. 99957–99969, 2025, doi: 10.1109/ACCESS.2025.3575192.

[10] F. Perez-Montes, J. Olivares-Mercado, G. Sanchez-Perez, G. Benitez-Garcia, L. Prudente-Tixteco, and O. Lopez-Garcia, "Analysis of Real-Time Face-Verification Methods for Surveillance Applications," *J Imaging*, vol. 9, no. 2, Feb. 2023, doi: 10.3390/jimaging9020021.

[11] A. Mahdavi and M. Carvalho, "A Survey on Open Set Recognition," Aug. 2021, doi: 10.1109/AIKE52691.2021.00013.

[12] J. Zhao *et al.*, "Towards Pose Invariant Face Recognition in the Wild." [Online]. Available: https://zhaoj9014.github.io/.

[13] 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.

[14] S. Sunardi, A. Yudhana, and S. A. Wijaya, "Face Detection Analysis of Digital Photos Using Mean Filtering Method," *International Journal of Artificial Intelligence Research*, vol. 6, no. 2, Jul. 2022, doi: 10.29099/ijair.v6i2.307.

[15] R. Nachet, J. Garrigós, and T. B. Stambouli, "GhostFaceNet++: boosting efficiency and accuracy via CSP bottlenecks and Channel Attention," *J Real Time Image Process*, vol. 22, no. 6, Dec. 2025, doi: 10.1007/s11554-025-01768-x.

[16] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4685–4694. doi: 10.1109/CVPR.2019.00482.

[17] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality Adaptive Margin for Face Recognition," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 18729–18738. doi: 10.1109/CVPR52688.2022.01819.

[18] M. Günther, S. Cruz, E. M. Rudd, and T. E. Boult, "Toward Open-Set Face Recognition."

[19] J. N. Kolf *et al.*, "EFaR 2023: Efficient Face Recognition Competition," in *2023 IEEE International Joint Conference on Biometrics (IJCB)*, 2023, pp. 1–12. doi: 10.1109/IJCB57857.2023.10448917.

[20] Z. Sun, C. Feng, I. Patras, and G. Tzimiropoulos, "LAFS: Landmark-Based Facial Self-Supervised Learning for Face Recognition," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 1639–1649. doi: 10.1109/CVPR52733.2024.00162.

[21] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2010.11929

[22] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," Jan. 2022, doi: 10.1145/3505244.

[23] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," 2021.

[24] J. Dan *et al.*, "TransFace: Calibrating Transformer Training for Face Recognition from a Data-Centric Perspective," Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.10133

[25] M. Rodrigo, C. Cuevas, and N. García, "Comprehensive comparison between vision transformers and convolutional neural networks for face recognition tasks," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-72254-w.

[26] A. George, C. Ecabert, H. O. Shahreza, K. Kotwal, and S. Marcel, "EdgeFace: Efficient Face Recognition Model for Edge Devices," *IEEE Trans Biom Behav Identity Sci*, vol. 6, no. 2, pp. 158–168, 2024, doi: 10.1109/TBIOM.2024.3352164.

[27] "FIDO Biometrics Requirements." [Online]. Available: https://github.com/fido-alliance/biometrics-requirements/pull/9.

[28] P. Grother, M. Ngan, and K. Hanaoka, "NISTIR XXXX Draft Ongoing Face Recognition Vendor Test (FRVT) Part 1: Verifcation." [Online]. Available: https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing

[29] J. Xiang and G. Zhu, "Joint Face Detection and Facial Expression Recognition with MTCNN," in *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, 2017, pp. 424–427. doi: 10.1109/ICISCE.2017.95.

[30] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," Dec. 2019, [Online]. Available: http://arxiv.org/abs/1912.01703

[31] C. R. Harris *et al.*, "Array programming with NumPy," Sep. 17, 2020, *Nature Research*. doi: 10.1038/s41586-020-2649-2.

[32] F. Pedregosa FABIANPEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," 2011. [Online]. Available: http://scikit-learn.sourceforge.net.